Geometric Group Theory
Preliminary Version
Still under revision

Cornelia Druţu and Michael Kapovich

Preface

The goal of this book is to present several central topics in geometric group theory, primarily related to the large scale geometry of infinite groups and spaces on which such groups act, and to illustrate them with fundamental theorems such as Gromov's Theorem on groups of polynomial growth, Tits' Alternative, Mostow's Rigidity Theorem, Stallings' theorem on ends of groups, theorems of Tukia and Schwartz on quasiisometric rigidity for lattices in real-hyperbolic spaces, etc. We give essentially self-contained proofs of all the above mentioned results, and we use the opportunity to describe several powerful tools/toolkits of geometric group theory, such as coarse topology, ultralimits and quasiconformal mappings. We also discuss three classes of groups central in geometric group theory: Amenable groups, hyperbolic groups, and groups with Property T.

The key idea in geometric group theory is to study groups by endowing them with a metric and treating them as *geometric objects*. This can be done for groups that are *finitely generated*, i.e. that can be reconstructed from a finite subset, *via* multiplication and inversion. Many groups naturally appearing in topology, geometry and algebra (e.g. fundamental groups of manifolds, groups of matrices with integer coefficients) are finitely generated. Given a finite generating set S of a group G, on can define a metric on G by constructing a connected graph, the *Cayley graph* of G, with G serving as the set of vertices and the oriented edges labeled by elements in S. A Cayley graph $\mathcal{G}$, as any other connected graph, admits a natural metric invariant under automorphisms of $\mathcal{G}$: The distance between two points is the length of the shortest path in the graph joining these points (see Section 1.8.3). The restriction of this metric to the vertex set G is called the *word metric* dist_S on the group G. The first obstacle to "geometrizing" groups in this fashion is the fact that a Cayley graph depends not only on the group but also on a particular choice of finite generating set. Cayley graphs associated with different generating sets are not isometric but merely *quasiisometric*.

Another typical situation in which a group G is naturally endowed with a (pseudo)metric is when G acts on a metric space X: In this case the group G maps to X *via* the *orbit map* $g \mapsto gx$. The pull-back of the metric to G is then a pseudo-metric on G. If G acts on X isometrically, then the resulting pseudometric on G is G-invariant. If, furthermore, the space X is proper and geodesic and the action of G is *geometric* (i.e., properly discontinuous and cocompact), then the resulting (pseudo)metric is quasiisometric to word metrics on G (Theorem 5.35). For example, if a group G is the fundamental group of a closed Riemannian manifold M, the action of G on the universal cover $\widetilde{M}$ of M satisfies all these properties. The second class of examples of isometric actions (whose origin lies in functional analysis and representation theory) comes from isometric actions of a group G on Hilbert spaces. The square of the corresponding metric on G is known in the literature as a *conditionally negative semidefinite kernel*. In this case, the relation between the word metric and the metric induced from the Hilbert space is more loose than quasiisometry; nevertheless, the mere existence of such a metric (induced from an isometric action on a Hilbert space) has many interesting implications, detailed in Chapter 17.

In the setting of the geometric view of groups, the following questions become fundamental:

QUESTIONS. (A) *If G and G' are quasiisometric groups, to what extent do G and G' share the same algebraic properties?*

(B) *If a group G is quasiisometric to a metric space X, what geometric properties (or structures) on X translate to interesting algebraic properties of G?*

Addressing these questions is the primary focus of this book. Several striking results (like Gromov's Polynomial Growth Theorem) state that certain algebraic properties of a group can be reconstructed from its loose geometric features.

Closely connected to these considerations are two foundational problems which appeared in different contexts but both render the same sense of existence of a "demarcation line" dividing the class of infinite groups into "abelian-like" groups and "free-like" groups. The invariants used to draw the line are quite different (existence of a finitely-additive invariant measure in one case and behavior of the growth function in the other); nevertheless, the two problems/question and the classification results that grew out of these questions, have much in common.

The first of these questions was inspired by work investigating the existence of various types of group-invariant measures, that originally appeared in the context of Euclidean spaces. Namely, the *Banach-Tarski paradox* (see Chapter 15), while denying the existence of such measures on the Euclidean plane, inspired John von Neumann to formulate two important concepts: That of *amenable groups* and that of *paradoxical decompositions and groups* [**vN28**]. In an attempt to connect amenability to the algebraic properties of a group, von Neumann made the observation, in the same paper, that the existence of a free subgroup excludes amenability. Mahlon Day (in [**Day50**] and [**Day57**]) extended von Neumann's work, introduced the terminology *amenable groups*, defined the class of *elementary amenable groups* and proved several foundational results about amenable and elementary amenable groups. In [**Day57**, p. 520] he also noted [1]:

- It is not know whether the class of elementary amenable groups equals the class of amenable groups and whether the class of amenable groups coincides with the class of groups containing no free non-abelian subgroups.

This observation later became commonly known as the von Neumann–Day problem (or conjecture):

QUESTION (The von Neumann–Day problem). *Is non-amenability of a group equivalent to the existence of a free non-abelian subgroup?*

[1] Contrary to the common belief, Day neither formulated a conjecture about this issue nor attributed the problem to von Neumann.

The second problem appeared in the context of Riemannian geometry, in connection to attempts to relate, for a compact Riemannian manifold M, the geometric features of its universal cover $\widetilde{M}$ to the behavior of its fundamental group $G = \pi_1(M)$. Two of the most basic objects in Riemannian geometry are the *volume* and the *volume growth rate*. The notion of volume growth extends naturally to discrete metric spaces, such as finitely generated groups. The *growth function* of a finitely generated group G (with a fixed finite generating set S) is the cardinality $\mathfrak{G}(n)$ of the ball of radius n in the metric space (G, dist_S). While the function $\mathfrak{G}(n)$ depends on the choice of the finite generating set S, the *growth rate* of $\mathfrak{G}(n)$ is independent of S. In particular, one can speak of groups of linear, polynomial, exponential growth, etc. More importantly, the growth rate is preserved by quasi-isometries, which allows to establish a close connection between the Riemannian growth of a manifold $\widetilde{M}$ as above, and the growth of $G = \pi_1(M)$.

One can easily see that every abelian group has polynomial growth. It is a more difficult theorem (proven independently by Hyman Bass [**Bas72**] and Yves Guivarc'h [**Gui70, Gui73**]) that all nilpotent groups also have polynomial growth. We prove this result in Section 12.2. In this context, John Milnor [**Mil68c**] and Joe Wolf [**Wol68**] asked the following question:

QUESTION. *Is it true that the growth of each finitely generated group is either polynomial (i.e. $\mathfrak{G}(n) \leqslant Cn^d$ for some fixed C and d) or exponential (i.e. $\mathfrak{G}(n) \geqslant Ca^n$ for some fixed $a > 1$ and $C > 0$)?*

Note that Milnor stated the problem in the form of a question, not a conjecture, however, he conjectured in [**Mil68c**] that each group of polynomial growth is virtually nilpotent, i.e., contains a nilpotent subgroup of finite index.

The answer to the Question is positive for *solvable groups:* This is the *Milnor–Wolf Theorem*, which states that *solvable groups of polynomial growth are virtually nilpotent*, see Theorem 12.37 in this book (the theorem is a combination of results due to Milnor and Wolf). This theorem still holds for the larger class of *elementary amenable groups* (see Theorem 16.56); moreover, such groups with non-polynomial growth must contain a free non-abelian subsemigroup.

The proof of the Milnor–Wolf Theorem essentially consists of a careful examination of increasing/decreasing sequences of subgroups in nilpotent and solvable groups. Along the way, one discovers other features that nilpotent groups share with abelian groups, but not with solvable groups. For instance, *in a nilpotent group all finite subgroups are contained in a maximal finite subgroup, while solvable groups may contain infinite strictly increasing sequences of finite subgroups.* Furthermore, *all subgroups of a nilpotent group are finitely generated, but this is no longer true for solvable groups.* One step further into the study of a finitely generated subgroup H in a group G is to compare a word metric dist_H on the subgroup H to the restriction to H of a word metric dist_G on the ambient group G. With an appropriate choice of generating sets, the inequality $\mathrm{dist}_G \leqslant \mathrm{dist}_H$ is immediate: All the paths in H joining $h, h' \in H$ are also paths in G, but there might be some other, shorter paths in G joining h, h'. The problem is to find an upper bound on dist_H in terms of dist_G. If G is abelian, the upper bound is linear as a function of dist_G. If dist_H is bounded by a polynomial in dist_G, then the subgroup H is said to be *polynomially distorted* in G, while if dist_H is approximately $\exp(\lambda \mathrm{dist}_G)$ for some $\lambda > 0$, the subgroup H is said to be *exponentially distorted.* It turns out that *all subgroups in a nilpotent group are polynomially distorted, while some solvable groups contain finitely generated subgroups with exponential distortion.*

Both von Neumann-Day and Milnor-Wolf questions were answered in the affirmative for linear groups by Jacques Tits:

THEOREM (Tits' Alternative). *Let F be a field of zero characteristic and let Γ be a subgroup of $GL(n, F)$. Then either Γ is virtually solvable or Γ contains a free nonabelian subgroup.*

We prove Tits' Alternative in Chapter 13. Note that this alternative also holds for fields of positive characteristic, provided that Γ is finitely generated.

There are other classes of groups in which both von Neumann-Day and Milnor questions have positive answers, they include: Subgroups of Gromov–hyperbolic groups ([**Gro87**, §8.2.F], [**GdlH90**, Chapter 8]), fundamental groups of closed Riemannian manifolds of nonpositive curvature [**Bal95**], subgroups of the mapping class group [**Iva92**] and the groups of outer automorphisms of free groups [**BFH00, BFH05**].

The von Neumann-Day question in general has negative answer: The first counterexamples were given by A. Ol′shanskiĭ in [**Ol′80**]. In [**Ady82**] it was shown that the free Burnside groups $B(n, m)$ with $n \geqslant 2$ and $m \geqslant 665$, m odd, are also counterexamples. Finally, finitely presented counterexamples were constructed by A. Ol′shanskiĭ and M. Sapir in [**OS02**]. These papers have lead to the development of certain techniques of constructing "infinite finitely generated monsters". While the negation of amenability (i.e. the paradoxical behavior) is, thus, still not completely understood algebraically, several stronger properties implying non-amenability were introduced, among which are various fixed-point properties, most importantly *Kazhdan's Property* T (Chapter 17). Remarkably, amenability (hence paradoxical behavior) is a quasiisometry invariant, while Property T is not.

Milnor–Wolf question, in full generality, likewise, has negative answer: The first groups of *intermediate growth*, i.e. growth which is super-polynomial but subexponential, were constructed by Rostislav Grigorchuk. Moreover, he proved the following:

THEOREM (Grigorchuk's Subexponential Growth theorem). *Let f be an arbitrary sub-exponential function larger than $2^{\sqrt{n}}$. Then there exists a finitely generated group Γ with subexponential growth function $\mathfrak{G}(n)$ such that:*

$$f(n) \leqslant \mathfrak{G}(n)$$

for infinitely many $n \in \mathbb{N}$.

Later on, Anna Erschler [**Ers04**] adapted Grigorchuk's arguments to improve the above result with the inequality $f(n) \leqslant \mathfrak{G}(n)$ for all but finitely many n. In the above examples, the exact growth function was unknown. However, Laurent Bartholdi and Anna Erschler [**BE12**] constructed examples of groups of intermediate growth, where they actually compute $\mathfrak{G}(n)$, up to an appropriate equivalence relation. Note, however, that the Milnor–Wolf Problem is still open for finitely presented groups.

On the other hand, Mikhael Gromov proved an even more striking result:

THEOREM (Gromov's Polynomial Growth Theorem, [**Gro81a**]). *Every finitely generated group of polynomial growth is virtually nilpotent.*

This is a typical example of an algebraic property that may be recognized *via* a, seemingly, weak geometric information. A corollary of Gromov's theorem is *quasiisometric rigidity* for virtually nilpotent groups:

COROLLARY. *Suppose that G is a group quasiisometric to a nilpotent group. Then G itself is* virtually nilpotent*, i.e. it contains a nilpotent subgroup of finite index.*

Gromov's theorem and its corollary will be proven in Chapter 14. Since the first version of these notes was written, Bruce Kleiner [**Kle10**] and, later, Narutaka Ozawa [**Oza15**] gave completely different (and much shorter) proofs of Gromov's polynomial growth theorem, using harmonic functions on graphs (Kleiner) and functional-analytic tools (Ozawa). Both proofs still require the Tits' Alternative. Kleiner's techniques provided the starting point for Y. Shalom and T. Tao, who proved the following effective version of Gromov's Theorem [**ST10**]:

THEOREM (Shalom–Tao Effective Polynomial Growth Theorem). *There exists a constant C such that for any finitely generated group G and $d > 0$, if for some $R \geqslant \exp\left(\exp\left(Cd^C\right)\right)$, the ball of radius R in G has at most R^d elements, then G has a finite index nilpotent subgroup of class less than C^d.*

We decided to retain, however, Gromov's original proof since it contains a wealth of ideas that generated in their turn new areas of research. Remarkably, the same piece of logic (a weak version of the axiom of choice) that makes the Banach-Tarski paradox possible also allows to construct *ultralimits*, a powerful tool in the proof of Gromov's theorem and that of many rigidity theorems (e.g, quasiisometric rigidity theorems of Kapovich, Kleiner and Leeb) as well as in the investigation of fixed point properties.

Regarding Questions (A) and (B), the best one can hope for is that the geometry of a group (up to quasiisometric equivalence) allows to recover, not just some of its algebraic features, but the group itself, *up to virtual isomorphism.* Two groups G_1 and G_2 are said to be *virtually isomorphic* if there exist subgroups

$$F_i \triangleleft H_i \leqslant G_i, i = 1, 2,$$

so that H_i has finite index in G_i, F_i is a finite normal subgroup in H_i, $i = 1, 2$, and H_1/F_1 is isomorphic to H_2/F_2. Virtual isomorphism implies quasiisometry but, in general, the converse is false, see Example 5.46. In the situation when the converse implication also holds, one says that the group G_1 is *quasiisometrically rigid.*

An example of quasiisometric rigidity is given by the following theorem proven by Richard Schwartz [**Sch96b**]:

THEOREM (Schwartz QI rigidity theorem). *Suppose that Γ is a nonuniform lattice of isometries of the hyperbolic space $\mathbb{H}^n, n \geqslant 3$. Then each group quasiisometric to Γ must be virtually isomorphic to Γ.*

We will present a proof of this theorem in Chapter 22. In the same chapter we use similar "zooming" arguments to prove the special case of *Mostow' Rigidity Theorem*:

THEOREM (The Mostow Rigidity Theorem). *Let Γ_1 and Γ_2 be lattices of isometries of $\mathbb{H}^n, n \geqslant 3$, and let $\varphi : \Gamma_1 \to \Gamma_2$ be a group isomorphism. Then φ is given by conjugation* via *an isometry of $\mathbb{H}^n$.*

Note that the proof of Schwartz' theorem fails for $n = 2$, where non-uniform lattices are virtually free. (Here and in what follows when we say that a group has a certain property *virtually* we mean that it has a finite index subgroup with that property.) However, in this case, quasiisometric rigidity still holds as a corollary of the Stallings Theorem on ends of groups:

THEOREM. *Let Γ be a group quasiisometric to a free group of finite rank. Then Γ is itself virtually free.*

This theorem will be proven in Chapter 18. We also prove:

THEOREM (Stallings "Ends of groups" theorem). *If G is a finitely generated group with infinitely many ends, then G splits as a graph of groups with finite edge–groups.*

In this book we give two proofs of the above theorem, which, while quite different, are both inspired by the original argument of Stallings. In Chapter 18 we prove Stallings' theorem for *almost finitely presented* groups. This proof follows the ideas of Dunwoody, Jaco and Rubinstein: We will be using *minimal Dunwoody tracks*, where minimality is defined with respect to a certain hyperbolic metric on the presentation complex (unlike combinatorial minimality used by Dunwoody). In Chapter 19, we will give another proof, which works for all finitely generated groups and follows a proof sketched by Gromov in **[Gro87]**, using least energy harmonic functions. We decided to present both proofs, since they use different machinery (the first is more geometric and the second more analytical) and different (although related) geometric ideas.

In Chapter 18 we also prove:

THEOREM (Dunwoody's Accessibility Theorem). *Let G be an almost finitely presented group. Then G is* accessible*, i.e. the decomposition process of G as a graph of groups with finite edge groups eventually terminates.*

In Chapter 21 we prove Tukia's theorem, which establishes quasiisometric rigidity of the class of fundamental groups of compact hyperbolic n-manifolds, and, thus, complements Schwartz' Theorem above:

THEOREM (Tukia's QI Rigidity Theorem). *If a group Γ is quasiisometric to the hyperbolic n-space, then Γ is virtually isomorphic to the fundamental group of a compact hyperbolic n-manifold.*

Note that the proofs of the theorems of Mostow, Schwartz and Tukia all rely upon the same analytical tool: Quasiconformal mappings of Euclidean spaces. In contrast, the analytical proofs of Stallings' theorem presented in the book are mostly motivated by another branch of geometric analysis, namely, the theory of minimal submanifolds and harmonic functions.

In regard to Question (B), we investigate two closely related classes of groups: Hyperbolic and relatively hyperbolic groups. These classes generalize fundamental groups of compact negatively curved Riemannian manifolds and, respectively, complete Riemannian manifolds of finite volume. To this end, in Chapters 8, 9 we cover basics of hyperbolic geometry and theory of hyperbolic and relatively hyperbolic groups.

Other sources. Our choice of topics in geometric group theory is far from exhaustive. We refer the reader to [**Aea91**],[**Bal95**], [**Bow91a**], [**VSCC92**], [**Bow06a**], [**BH99**], [**CDP90**], [**Dav08**], [**Geo08**], [**GdlH90**], [**dlH00**], [**NY11**], [**Pap03**], [**Roe03**], [**Sap14**], [**Väi05**], for the discussion of other parts of the theory.

Work on this book started in 2002 and the material which we cover mostly concerns developments in the Geometric Group Theory from 1960s through 1990s. In the meantime, while we were working on the book, some major exciting developments in the field have occured which we did not have a chance to discuss in the book. To name a few, these developments are subgroup separability and connections with the 3-dimensional topology [**Ago13, KM12, HW12, Bes14**], applications of the geometric group theory to the higher dimensional topology [**Yu00, MY02, BLW10, BL12**], theory of Kleinian groups [**Min10, BCM12, Mj14b, Mj14a**], quasiconformal analysis on boundaries of hyperbolic groups and Cannon Conjecture [**BK02a, BK05, Bon11, BK13b, Mar13, Haï15**], the theory of approximate groups [**BG08a, Tao08, BGT12, Hru12**], the first-order logic of free groups [**Sel01, Sel03, Sel05a, Sel04, Sel05b, Sel06a, Sel06b, Sel09, Sel13**], the theory of systolic groups [**JŚ03, JŚ06, HŚ08, Osa13**], probabilistic aspects of the Geometric Group Theory [**Gro03, Ghy04, Oll04a, Oll05, KSS06, Oll07, KS08, OW11, AŁŚ15**].

Requirements. The book is intended as a reference for graduate students and more experienced researchers, it can be used as a basis for a graduate course and as a first reading for a researcher wishing to learn more about geometric group theory. This book is partly based on lectures which we were teaching at Oxford University (C.D.) and University of Utah and University of California, Davis (M.K.). We expect the reader to be familiar with basics of group theory, algebraic topology (fundamental groups, covering spaces, (co)homology, Poincaré duality) and elements of differential topology and Riemannian geometry. Some of the background material is covered in Chapters 1, 2 and 3. We tried to make the book as self-contained as possible, but some theorems are stated without a proof, they are marked as **Theorem**.

Acknowledgments. The work of the first author was supported in part by the ANR projects "Groupe de recherche de Géométrie et Probabilités dans les Groupes", by the EPSRC grant "Geometric and analytic aspects of infinite groups" and by the ANR-10-BLAN 0116, acronym GGAA. The second author was supported by the NSF grants DMS-02-03045, DMS-04-05180, DMS-05-54349, DMS-09-05802 and DMS-12-05312. During the work on this book the authors visited the Max Planck Institute for Mathematics (Bonn), and wish to thank this institution for its hospitality. We are grateful to Thomas Delzant, Bruce Kleiner and Mohan Ramachandran for supplying some proofs, to Ilya Kapovich, Bernhard Leeb, Andrew Sale, Mark Sapir, Alessandro Sisto and the anonymous referees of this book for the numerous corrections and suggestions. We also thank Yves de Cornulier, Rostislav Grigorchuk, Pierre Pansu and Romain Tessera for useful references.

Addresses:

Cornelia Druţu: Mathematical Institute,
Andrew Wiles Building,
University of Oxford,
Woodstock Road,
Oxford OX2 6GG, United Kingdom
email: Cornelia.Drutu@maths.ox.ac.uk

Michael Kapovich: Department of Mathematics,
University of California,
Davis, CA 95616, USA
email: kapovich@math.ucdavis.edu

Contents

CHAPTER 1

Geometry and Topology

Treating groups as *geometric objects* is the major theme and defining feature of the geometric group theory. In this chapter we discuss basics of metric (and topological) spaces, while in Chapter 2 we will have a brief review of Riemannian geometry. We refer the reader interested in in-depth discussion of metric geometry to [**BBI01**]. We expect the reader to be familiar with basics of algebraic topology which can be found, for instance in [**Hat02**] or [**Mas91**].

1.1. Sets-theoretic preliminaries

1.1.1. General notation. Given a set X we denote by $\mathcal{P}(X) = 2^X$ the *power set* of X, i.e., the set of all subsets of X. If two subsets A, B in X have the property that $A \cap B = \emptyset$ then we denote their union by $A \sqcup B$, and we call it the *disjoint union.* For a subset E of a set X we let E^c denote the complement $X \setminus E$ of E in X. A *pointed set* is a pair (X, x), where x is an element of X. The composition of two maps $f : X \to Y$ and $g : Y \to Z$ is denoted either by $g \circ f$ or by gf . We will use the notation Id_X or simply Id (when X is clear) to denote the identity map $X \to X$. For a map $f : X \to Y$ and a subset $A \subset X$, we let $f\big|_A$ denote the restriction of f to A. We will use the notation $|E|$ or $\mathrm{card}\,(E)$ to denote cardinality of a set E. (Sometimes, however, $|E|$ will denote the Lebesgue measure of a subset of the Euclidean space.) We will use the notation $\mathbb{D}^n$ for the closed unit ball centered at the origin in the n-dimensional Euclidean space and $\mathbb{S}^{n-1}$ for the corresponding unit sphere. In contrast, we will use the notation $B(x, r)$ for the open metric ball (in a general metric space) centered at x, of radius r. Accordingly, $\mathbb{B}^n$ will denote the open unit ball in $\mathbb{R}^n$.

The Axiom of Choice (AC) plays a prominent part in many of the arguments in this book. We discuss AC in more detail in section 7.1, where we also list equivalent and weaker forms of AC. Throughout the book we make the following convention:

Convention 1.1. We always assume ZFC: The Zermelo–Fraenkel axioms of set theory and the Axiom of Choice.

Given a non-empty set X, we denote by $\mathrm{Bij}(X)$ the group of bijections $X \to X$, with composition as the binary operation.

Convention 1.2. Throughout the paper we denote by $\mathbf{1}_A$ the characteristic function of a subset A in a set X, i.e., the function $\mathbf{1}_A : X \to \{0, 1\}$, $\mathbf{1}_A(x) = 1$ if and only if $x \in A$.

By the *codimension of a subspace X in a space Y* we mean the difference between the dimension of Y and the dimension of X, whatever the notion of dimension that we use.

We will use the notation $\cong$ to denote an isometry of metric spaces and $\simeq$ to denote an isomorphism of groups.

Throughout the book, $\mathbb{N}$ will denote the set of natural numbers and $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$.

1.1.2. Growth rates of functions. We will be using in this book two different *asymptotic* inequalities and equivalences for functions: One is used to compare Dehn functions of groups and the other to compare growth rates of groups.

DEFINITION 1.3. Let X be a subset of $\mathbb{R}$. Given two functions $f, g : X \to \mathbb{R}$, we say that *the order of the function f is at most the order of the function g* and we write $f \precsim g$, if there exist real numbers $a, b, c, d, e > 0$ and x_0 such that for all $x \in X, x \geq x_0$, we have: $bx + c \in X$ and

$$f(x) \leqslant ag(bx + c) + dx + e.$$

If $f \precsim g$ and $g \precsim f$ then we write $f \approx g$ and we say that f and g are *approximately equivalent.*

This definition will be typically used with $X = \mathbb{R}_+$ or $X = \mathbb{N}$, in which case a, b, c, d, e will be natural numbers.

The equivalence class of a function with respect to equivalence relation $\approx$ is called *the order of the function*. If a function f has (at most) the same order as the function x, x^2, x^3, x^d or $\exp(x)$ it is said that *the order of the function f is (at most) linear, quadratic, cubic, polynomial, or exponential,* respectively. A function f is said to have *subexponential* order if it has order at most $\exp(x)$ and is not approximately equivalent to $\exp(x)$. A function f is said to have *intermediate order* if it has subexponential order and $x^n \precsim f(x)$ for every n.

DEFINITION 1.4. We introduce the following *asymptotic inequality* between functions $f, g : X \to \mathbb{R}$ with $X \subset \mathbb{R}$: We write $f \preceq g$ if there exist $a, b > 0$ and $x_0 \in \mathbb{R}$ such that for all $x \in X$, $x \geq x_0$, we have: $bx \in X$ and

$$f(x) \leq ag(bx).$$

If $f \preceq g$ and $g \preceq f$ then we write $f \asymp g$ and we say that f and g are *asymptotically equal.*

Note that this definition is more refined than the *order notion* $\approx$. For instance, $x \approx 0$ while these functions are not asymptotically equal. This situation arises, for instance, in the case of free groups (which are given free presentation): The Dehn function is zero, while the area filling function of the Cayley graph is $A(\ell) \asymp \ell$. The equivalence relation $\approx$ is more appropriate for Dehn functions than the relation $\asymp$, because in the case of a free group one may consider either a presentation with no relations, in which case the Dehn function is zero, or another presentation that yields a linear Dehn function.

EXERCISE 1.5. 1. Show that $\approx$ and $\asymp$ are equivalence relations.

2. Suppose that $x \preceq f$, $x \preceq g$. Then $f \approx g$ if and only if $f \asymp g$.

1.1.3. Jensen's inequality. Suppose that (X, μ) is a space equipped with the probability measure μ, $f : X \to \mathbb{R}$ is a measurable function and φ is a convex function on the range of f. Then Jensen's inequality [**Rud87**, Theorem 3.3] reads:

$$\varphi\left(\int_X f\mathrm{d}\mu\right) \leqslant \int_X \varphi \circ f\mathrm{d}\mu.$$

We will be using this inequality in when the function f is strictly positive and $\varphi(t) = t^{-1}$: The function $\varphi(t)$ is convex for $t > 0$. It will be convenient to eliminate the probability measure assumption. We will be working with spaces (X, μ) of finite (but nonzero) measure. Instead of normalizing the measure μ to be a probability measure, we can as well replace integrals $\int_X h\mathrm{d}\mu$ with *averages*

$$⨍_X h\mathrm{d}\mu = \frac{1}{M}\int_X h\mathrm{d}\mu,$$

where $M = \int_X \mathrm{d}\mu$. With this in mind, Jensen's inequality becomes

$$\left(⨍_X f\mathrm{d}\mu\right)^{-1} \leqslant ⨍_X \frac{1}{f}\mathrm{d}\mu.$$

Replacing f with $\frac{1}{f}$ we also obtain:

$$\left(⨍_X \frac{1}{f}\mathrm{d}\mu\right)^{-1} \leqslant ⨍_X f\mathrm{d}\mu. \tag{1.1}$$

1.2. Measure and integral

DEFINITION 1.6. An *algebra of subsets* of a set X is a non-empty collection $\mathcal{A}$ of subsets of X such that:

(1) $X \in \mathcal{A}$;
(2) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}\,,\ A \cap B \in \mathcal{A}$;
(3) $A \in \mathcal{A} \Rightarrow A^c = X \setminus A \in \mathcal{A}$.

More generally, a σ-*algebra* of subsets of X is an algebra of subsets closed under countable intersections and countable unions.

Given a topological space X, the smallest σ-algebra of subsets of X containing all open subsets of X is called the *Borel σ-algebra* of X. Elements of this σ-algebra are called *Borel subsets* of X.

DEFINITION 1.7. A *finitely additive (f. a.) measure* μ on an algebra $\mathcal{A}$ of subsets of X is a function $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(A \sqcup B) = \mu(A) + \mu(B)$ for all $A, B \in \mathcal{A}$.

An immediate consequence of the f. a. property is that for any two sets $A, B \in \mathcal{A}$,

$$\mu(A\cup B) = \mu((A\setminus B)\sqcup(A\cap B)\sqcup(B\setminus A)) = \mu(A\setminus B)+\mu(A\cap B)+\mu(B\setminus A) \leqslant \mu(A)+\mu(B).$$

In some texts the f. a. measures are called simply 'measures'. We prefer the terminology above, since in other texts a 'measure' is meant to be countably additive as defined below.

Definition 1.8. A *countably additive* (c. a.) measure on a σ-algebra $\mathcal{A}$ of subsets of X is a function $\mu : \mathcal{A} \to [0, \infty]$ such that

$$\mu\left(\bigsqcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$$

for any countable subset $\{A_i, i \in I\} \subset \mathcal{A}$, whose elements are pairwise disjoint.

In order to simplify the terminology, we will suppress dependence of f.a. (resp. c.a) measure μ on the algebra (resp. σ-algebra) of subsets of X and will refer to such μ simply as f.a. (resp. c.a.) measure on X.

Definition 1.9. If μ is finitely (resp. countably) additive measure X, such that $\mu(X) = 1$, then μ is called an f.a. (resp. c.a.) *probability measure* on X, which is abbreviated as f.a.p. measure (resp. c.a.p. measure).

Suppose that G be a group acting on X preserving an algebra (resp. σ-algebra) $\mathcal{A}$. If μ is an f.a. (resp. c.a.) measure on $\mathcal{A}$, such that $\mu(\gamma A) = \mu(A)$ for all $\gamma \in G$ and $A \in \mathcal{A}$, then μ is called *G–invariant*.

We let $B(X)$ denote the vector space of real-valued bounded functions on a set X. In addition to measures we will need the notion of a *finitely additive integral* integral. We discuss integrals of functions $f \in B(X)$ and only in the simpler case of finitely additive probability measures μ, and the algebra $\mathcal{A}$ consisting of all subsets of X (the setting where we will use finitely additive integrals in Chapter 16). We refer the reader to **[DS88]** for development of finitely additive integrals in greater generality.

For $A \in \mathcal{A}$ we let $\mathbf{1}_A$ denote the characteristic (or indicator) function of A,

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

A *finitely additive integral* on $(X, \mathcal{A}, \mu, B(X))$ is a linear functional

$$\int_X : f \mapsto \int_X f \, d\mu, \quad f \in B(X), \quad \int_X f \, d\mu \in \mathbb{R},$$

satisfying the following properties:

- If $f(x) \geqslant 0$ for all $x \in X$, then $\int_X f d\mu \geqslant 0$.
- $\int_X \mathbf{1}_A \, d\mu = \mu(A)$ for all $A \in \mathcal{A}$.

For a subgroup $G \leqslant Bij(X)$, the integral $\int_X$ is said to be *G-invariant* if

$$\int_X f \circ \gamma \, d\mu = \int_X f \, d\mu$$

for every $\gamma \in G$ and every $f \in B(X)$.

Theorem 1.10. *If μ is a G-invariant f.a.p. measure on the algebra $\mathcal{A} = 2^X$ of all subsets of X, then there exists a G-invariant integral $\int_X$ on $(X, \mathcal{A}, \mu, B(X))$ such that*

$$\int_X \mathbf{1}_A \, d\mu = \mu(A)$$

for every $A \in \mathcal{A}$.

PROOF. We let $B_+(X)$ denote the subset of $B(X)$ consisting of all nonnegative functions $f \in B(X)$. Observe that the linear span of $B_+(X)$ is the entire $B(X)$. First of all, for each $A \in \mathcal{A}$ we have the integral

$$\int_X \mathbf{1}_A \, d\mu := \mu(A).$$

We next extend the integral from the set of characteristic functions $\mathbf{1}_A$, $A \in \mathcal{A}$ to the linear subspace $S(X) \subset B(X)$ of *simple functions*, i.e., the linear span of the set of characteristic functions. We also define $S_+(X)$ as $B_+(X) \cap S(X)$. In order to construct an extension of $\int_X$ to $S(X)$, we observe that each $f \in S(X)$ can be written in the form

$$f = \sum_{i=1}^{n} s_i \mathbf{1}_{A_i} \tag{1.2}$$

where the subsets A_i are pairwise disjoint. Moreover, we can choose A_i's such that for each A_i either $f\big|_{A_i} \geqslant 0$ or $f\big|_{A_i} < 0$. (Here we are helped by the fact that $\mathcal{A} = 2^X$.) Next, for $s_i \in \mathbb{R}, A_i \in \mathcal{A}, i = 1, \ldots, n$ and $t_j \in \mathbb{R}, B_j \in \mathcal{A}, j = 1, \ldots, m$, finite additivity of μ implies that if

$$\sum_{i=1}^{n} s_i \mathbf{1}_{A_i} = \sum_{j=1}^{m} t_j \mathbf{1}_{B_j}$$

then

$$\sum_{i=1}^{n} s_i \mu(A_i) = \sum_{j=1}^{m} t_j \mu(B_j).$$

Therefore, we can extend $\int$ to a linear functional on $S(X)$ by linearity:

$$\int_X \left(\sum_{i=1}^{n} s_i \mathbf{1}_{A_i} \right) d\mu = \sum_{i=1}^{n} s_i \mu(A_i).$$

Since for every $f \in S_+(X)$ we can assume that in (1.2), each $s_i \geqslant 0$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, it follows that

$$\int_X f \, d\mu \geqslant 0.$$

Next, given a function $f \in B_+(X)$ we set

$$\int_X f \, d\mu := \sup\{\int_X g \, d\mu : g \in S_+(X), g \leqslant f\}.$$

It is clear with this definition that, since μ is G-invariant, so is the map

$$\int_X : B_+(X) \to \mathbb{R}.$$

Furthermore, it is clear that

$$\int_X af \, d\mu = a \int_X f \, d\mu$$

for all $a \geqslant 0$ and $f \in B_+(X)$. However, additivity of $\int_X$ thus defined is not obvious. We leave it to the reader to verify the simpler fact that for all functions $f, g \in B_+(X)$ we have

$$\int_X f + g \, d\mu \geqslant \int_X f \, d\mu + \int_X g \, d\mu.$$

We will prove the reverse inequality. For each subset $A \subset X$ and $f \in B_+(X)$ define the integral

$$\int_A f\,d\mu := \int_X f\mathbf{1}_A\,d\mu.$$

Given a simple function h, $0 \leqslant h \leqslant f+g$, we need to show that

$$\int_X h\,d\mu \leqslant \int_X f\,d\mu + \int_X g\,d\mu.$$

The function h can be written as

$$h = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

with pairwise disjoint $A_i \in \mathcal{A}$ and $a_i > 0$. Therefore, in view of linearity of $\int_X$ on $S(X)$, it suffices to prove that

$$a_i\mu(A_i) \leqslant \int_{A_i} f\,d\mu + \int_{A_i} g\,d\mu$$

for each i. Thus, the problem reduces to the case $n = 1$, $A_1 = A$ and (by dividing by a_1) to proving the inequality

$$\mu(A) \leqslant \int_A f\,d\mu + \int_A g\,d\mu,$$

for functions $f, g \in B_+(X)$ satisfying

$$\mathbf{1}_A \leqslant f+g. \tag{1.3}$$

Let c be an integer upper bound for f and g. For each $N \in \mathbb{N}$, consider the following simple functions f_N, g_N:

$$f - \frac{c}{N}\mathbf{1}_A \leqslant f_N := \sum_{j=0}^{N} \mathbf{1}_{f^{-1}((cj/N, c(j+1)/N])} \frac{cj}{N} \leqslant f,$$

$$g - \frac{c}{N}\mathbf{1}_A \leqslant g_N := \sum_{j=0}^{N} \mathbf{1}_{g^{-1}((cj/N, c(j+1)/N])} \frac{cj}{N} \leqslant g.$$

In view of the inequality (1.3) we have

$$\left(1 - \frac{2c}{N}\right)\mathbf{1}_A \leqslant f_N + g_N.$$

The latter implies (by the definition of $\int_X$) that

$$\left(1 - \frac{2c}{N}\right)\mu(A) \leqslant \int_A f_N\,d\mu + \int_A g_N\,d\mu \leqslant \int_A f\,d\mu + \int_A g\,d\mu.$$

Since this inequality holds for all $N \in \mathbb{N}$, we conclude that

$$\mu(A) \leqslant \int_A f\,d\mu + \int_A g\,d\mu$$

as required. Thus, $\int_X$ is an additive functional on $B_+(X)$. Since $B_+(X)$ spans $B(X)$, $\int_X$ extends uniquely (by linearity) to a linear functional on $B(X)$. Clearly, the result is a G-invariant integral on $B(X)$. $\square$

1.3. Topological spaces. Lebesgue covering dimension

In this section we review some topological notions that shall be used in the book.

Notation and terminology. A *neighborhood* of a point in a topological space will always mean an *open neighborhood*. We will use the notation $\overline{A}$ and $\mathrm{cl}(A)$ for the closure of a subset A in a topological space X. We will denote by $\mathrm{int}\, A$ the interior of A in X. A subset of a topological space X is called *clopen* if it is both closed and open. We will use the notation $\mathcal{C}_X$ and $\mathcal{K}_X$ for the sets of all closed and of compact subsets in X respectively.

A topological space X is said to be *locally compact* if there is a basis of topology of X consisting of relatively compact subsets of X, i.e., subsets of X with compact closures. A space X is called *σ-compact* if there exists a sequence of compact subsets $(K_n)_{n\in\mathbb{N}}$ in X such that $X = \bigcup_{n\in\mathbb{N}} K_n$. A *second countable topological space* is a topological space which admits a countable base of topology (this is sometimes called *the second axiom of countability*). A second countable space is *separable* (i.e., contains a countable dense subset) and *Lindelöf* (i.e., every open cover has a countable sub-cover). A locally compact second countable space is σ-compact.

The *wedge of* a family of pointed topological spaces $(X_i, x_i), i \in I$, denoted by $\vee_{i\in I} X_i$, is the quotient of the disjoint union $\sqcup_{i\in I} X_i$, where we identify all the points x_i. The wedge of two pointed topological spaces is denoted $X_1 \wedge X_2$.

If $f : X \to \mathbb{R}$ is a function on a topological space X, then we will denote by $\mathrm{Supp}(f)$ the *support* of f, i.e., the set

$$cl\left(\{x \in X : f(x) \neq 0\}\right).$$

Given two topological spaces X, Y, we let $C(X;Y)$ denote the space of all continuous maps $X \to Y$; we also set $C(X) := C(X;\mathbb{R})$. For a function $f \in C(X)$ we define its norm

$$\|f\| = \sup_{x\in X} |f(x)|.$$

We always endow the space $C(X;Y)$ with the *compact-open topology.* A sub-basis of this topology consists of the subsets

$$U_{K,V} = \{f : X \to Y : f(K) \subset V\} \subset C(X;Y),$$

where $K \subset X$ is compact and $V \subset Y$ is open.

If Y is a metric space then the compact-open topology is equivalent to the topology of *uniform convergence on compacts*: A sequence of functions $f_i : X \to Y$ converges to a function $f : X \to Y$ if and only if for every compact subset $K \subset X$ the sequence of restrictions $f_i\big|_K : K \to Y$ converges to $f\big|_K$ uniformly.

A *homotopy* between the maps

$$f_0, f_1 : X \to Y$$

of two topological spaces, is a continuous map $F : X \times [0,1] \to Y$ such that $F(x,0) = f_0(x), F(x,1) = f_1(x), x \in X$. *Tracks* of this homotopy are paths $F(x,t), t \in [0,1]$ in Y, for various (fixed) points $x \in X$.

A continuous map $f : X \to Y$ of topological spaces is called *proper* if preimages of compact sets under f are again compact. In line with this, one defines a *proper homotopy* between two maps

$$f_0, f_1 : X \to Y,$$

by requiring the homotopy F between these maps to be a proper map $F : X \times [0,1] \to Y$.

A topological space is called *perfect* if it is nonempty and contains no isolated points, i.e., points $x \in X$ such that the singleton $\{x\}$ is open in X. A *neighborhood* of a subset $A \subset X$ is an open subset $U \subset X$ containing A.

DEFINITION 1.11. A topological space X is *regular* if every closed subset $A \subset X$ and a singleton $\{x\} \subset X \setminus A$, have disjoint neighborhoods. A topological space X is called *normal* if every pair of disjoint closed subsets $A, B \subset X$ have disjoint open neighborhoods, i.e., there exist disjoint open subsets $U, V \subset X$ such that $A \subset U, B \subset V$.

EXERCISE 1.12. 1. Every normal Hausdorff space is regular.
2. Every compact Hausdorff space is normal.

We will also need a minor variation on the notion of normality:

DEFINITION 1.13. Two subsets A, B of a topological space X are said to be *separated by a function* if there exists a continuous function $\rho = \rho_{A,B} : X \to [0,1]$ so that
1. $\rho\big|_A \equiv 0$
2. $\rho\big|_B \equiv 1$.
A topological space X is called *perfectly normal* if every two disjoint closed subsets of X can be separated by a function.

We will see below (Lemma 1.49) that every metric space is perfectly normal. A much harder result is

THEOREM 1.14 (Tietze–Urysohn extension theorem). *Every normal topological space X is perfectly normal. Moreover, for every closed subset $C \subset X$ and a continuous function $f : C \to \mathbb{R}$, the function f admits a continuous extension to X.*

A proof of this extension theorem can be found in [**Eng95**]. In view of this theorem, every normal topological space is perfectly normal, since one can take $C = A \cup B$ and let ρ be a continuous extension of the function

$$f : C \to \mathbb{R}, \quad f\big|_A \equiv 0, f\big|_B \equiv 1.$$

COROLLARY 1.15. *In the definition of a normal topological space, one can take U and V to have disjoint closures.*

PROOF. Let $f : C = A \cup B \to \{0,1\}$ be the function as above. define $\rho : X \to \mathbb{R}$ to be a continuous extension of f. Then take

$$U := \rho^{-1}((-\infty, 1/3)), \quad V := \rho^{-1}((2/3, \infty)). \quad \square$$

LEMMA 1.16. *[Extension lemma] Suppose that X, Y topological spaces, where Y is regular and X contains a dense subset A.*

1. If $f : X \to Y$ is a mapping satisfying the property that for each $x \in X$ the restriction of f to $A \cup \{x\}$ is continuous, then f is continuous.

2. Assume now that A is open and set $X \setminus A = Z$. Suppose that $f : X \to Y$ is such that the restriction $f\big|_{A \cup \{z\}}$ is continuous at z for every point $z \in Z \subset X$. Then $f : X \to Y$ is continuous at each point $z \in Z$.

PROOF. 1. We will verify continuity of f at each $x \in X$. Let $y = f(x)$ and let V be an (open) neighborhood of y in Y; the complement $C = Y \setminus V$ is closed. Since Y is regular, there exist disjoint open neighborhoods $V_1 \subset V$ of y and V_2 of C. Therefore, the closure W of V_1 is contained in V. By continuity of the map

$$f|_{A\cup\{x\}},$$

there exists an (open) neighborhood U of x in X, such that

$$f(U \cap (A \cup \{x\})) \subset V_1 \subset W.$$

Let us verify that $f(U) \subset W \subset V$. Take $z \in U$. By continuity of

$$f|_{A\cup\{z\}},$$

the preimage

$$D = f^{-1}(W) \cap (A \cup \{z\})$$

is closed in $A \cup \{z\}$. This preimage contains $U \cap A$; the latter is dense in U, since A is dense in X and $U \subset X$ is open. Therefore, D contains the closure of $U \cap A$ in $A \cup \{z\}$. This closure contains the point z since $z \in U$ and $U \cap A$ is dense in U. It follows that z is in D and, hence, $f(z) \in W$. We conclude that $f(U) \subset V$ and, therefore, f is continuous at x.

2. We change the topology $\mathcal{T}_X$ on X to a new topology $\mathcal{T}'_X$ whose basis is the union of $\mathcal{T}_X$ and the power set 2^A. Since $A \in \mathcal{T}_X$, it follows that the map $f : X \to Y$ is continuous at each point $a \in A$ with respect to the new topology. Part 1 now implies continuity of the map $f : (X, \mathcal{T}'_X) \to (Y, \mathcal{T}_Y)$. It follows that $f : A \cup \{z\} \to Y$ is continuous at each $z \in Z$ with respect to the original topology. □

A topological space X is said to be *locally path-connected* if for each $x \in X$ and each neighborhood U of x, there exists a neighborhood $V \subset U$ of x, such that every point $y \in V$ can be connected to x by a path contained in U. In other words, the inclusion $V \hookrightarrow U$ induces the map

$$\pi_0(V) \to \pi_0(U)$$

whose image is a singleton.

An open covering $\mathcal{U} = \{U_i : i \in I\}$ of a topological space X is called *locally finite* if every subset $J \subset I$ such that

$$\bigcap_{i\in J} U_i \neq \emptyset$$

is finite. Equivalently, every point $x \in X$ has a neighborhood which intersects only finitely many U_i's.

The *multiplicity* of an open cover $\mathcal{U} = \{U_i : i \in I\}$ of a space X is the supremum of cardinalities of subsets $J \subset I$ so that

$$\bigcap_{i\in J} U_i \neq \emptyset.$$

A cover $\mathcal{V}$ is called a *refinement* of a cover $\mathcal{U}$ if every $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$.

DEFINITION 1.17. The *(Lebesgue) covering dimension* of a topological space Y is the least number n such that the following holds: Every open cover $\mathcal{U}$ of Y admits a refinement $\mathcal{V}$ which has multiplicity at most $n + 1$.

The following example shows that covering dimension is consistent with our "intuitive" notion of dimension:

EXAMPLE 1.18. If M is a n-dimensional topological manifold, then n equals the covering dimension of M. See e.g. [**Nag83**].

1.4. Exhaustions of locally compact spaces

DEFINITION 1.19. A family of compact subsets $\{K_i : i \in I\}$ of a topological space X is said to be an *exhaustion* of X if:

1. $\bigcup_{i\in I} K_i = X$.
2. For each $i \in I$ there exists $j \in I$ such that

$$K_i \subset \text{int}(K_j).$$

PROPOSITION 1.20. *If X is locally compact, Hausdorff and 2nd countable, it admits an exhaustion by a countable collection of compact subsets. Moreover, there exists a countable exhaustion $\{K_n : n \in \mathbb{N}\}$ of X such that $K_n \subset \text{int}\, K_{n+1}$ for each n.*

PROOF. If X is empty, there is nothing to prove, therefore, we will assume that $X \neq \emptyset$. Let $\mathcal{B}$ be a countable basis of X. Define $\mathcal{U} \subset \mathcal{B}$ to be a subset of $\mathcal{B}$ consisting of relatively open sets.

LEMMA 1.21. *$\mathcal{U}$ is a basis of X.*

PROOF. Let $x \in X$ be a point and V a neighborhood of x. Since X is locally compact, there exists a compact subset $K \subset X$ with

$$x \in \text{int}(K) \subset V.$$

Then the boundary ∂K of K in X is disjoint from $\{x\}$. Since K is regular and $\mathcal{B}$ is a basis, there exists a neighborhood W of ∂K in K and $B \in \mathcal{B}$, a neighborhood of x in X, such that $B \subset \text{int}(K)$ and $B \cap W = \emptyset$. Then the closure $\bar{B}$ of B is compact (and, thus, $B \in \mathcal{U}$) and contained in $K \setminus W \subset \text{int}(K)$. □

We define an exhaustion of X inductively. Set K_1 be the closure of any $U_1 \in \mathcal{U}$. Given a compact subset K_n, we consider its cover by elements of $\mathcal{U}$. By compactness of K_n, there exists a finite subcollection $U_1, \ldots U_k \in \mathcal{U}$ covering K_n. Set

$$K_{n+1} := \bigcup_{i=1}^{k} \overline{U_i}.$$

This is the required exhaustion. □

1.5. Direct and inverse limits

Let I be a *directed set*, i.e., a partially ordered set, where every two elements i, j have an *upper bound*, which is some $k \in I$ such that $i \leqslant k, j \leqslant k$. The reader should think of the set of real numbers, or positive real numbers, or natural numbers, as the main examples of directed sets. A *directed system* of sets (or topological spaces, or groups) indexed by I is a collection of sets (or topological spaces, or groups) $A_i, i \in I$, and maps (or continuous maps, or homomorphisms) $f_{ij} : A_i \to A_j, i \leqslant j$, satisfying the following compatibility conditions:

(1) $f_{ik} = f_{jk} \circ f_{ij}, \forall i \leqslant j \leqslant k,$

(2) $f_{ii} = Id$.

An *inverse system* is defined similarly, except $f_{ij} : A_j \to A_i, i \leqslant j$, and, accordingly, in the first condition we use $f_{ij} \circ f_{jk} = f_{ik}$.

We will use the notation $(A_i, f_{ij}, i, j \in I)$ for direct and inverse systems of sets, spaces and groups.

The *direct limit* of the direct system of sets is the set

$$A = \varinjlim A_i = \left(\coprod_{i \in I} A_i \right) / \sim,$$

where $a_i \sim a_j$ whenever $f_{ik}(a_i) = f_{jk}(a_j)$ for some $k \in I$. In particular, we have maps $f_m : A_m \to A$ given by $f_m(a_m) = [a_m]$, where $[a_m]$ is the equivalence class in A represented by $a_m \in A_m$. Note that

$$A = \bigcup_{i \in I} f_m(A_m).$$

If A_i's are groups, then we equip the direct limit with the group operation:

$$[a_i] \cdot [a_j] = [f_{ik}(a_i)] \cdot [f_{jk}(a_j)],$$

where $k \in I$ is an upper bound for i and j.

If A_i's are topological spaces, we equip the direct limit with the *final topology*, i.e., the topology where $U \subset \varinjlim A_i$ is open if and only if $f_i^{-1}(U)$ is open for every i. In other words, this is the quotient topology descending from the disjoint union of A_i's.

Similarly, the *inverse limit* of an inverse system is

$$\varprojlim A_i = \left\{ (a_i) \in \prod_{i \in I} A_i : a_i = f_{ij}(a_j), \forall i \leqslant j \right\}.$$

If A_i's are groups, we equip the inverse limit with the group operation induced from the direct product of the groups A_i. If A_i's are topological spaces, we equip the inverse limit the *initial topology*, i.e., the subset topology of the Tychonoff topology on the direct product. Explicitly, this is the topology generated by the open sets of the form $f_m^{-1}(U_m)$, $U_m \subset X_m$ are open subsets and $f_m : \varprojlim A_i \to A_m$ is the restriction of the coordinate projection.

EXERCISE 1.22. 1. Show that $\varprojlim A_i$ is closed in $\prod_{i \in I} A_i$. 2. Conclude that if each A_i is compact, so is $\varprojlim A_i$.

Given a subset $J \subset I$, we have the *restriction map*

$$\rho : \prod_{i \in I} A_i \to \prod_{j \in J} A_j, \quad \lambda \mapsto \lambda\big|_J$$

where we treat elements of the product spaces as functions $I \to \bigcup_{i \in I} A_i$ and $J \to \bigcup_{j \in J} A_j$ respectively. A subposet $J \subset I$ is called *cofinal* if for each $i \in I$ there exists $j \in J$ such that $i \leq j$.

EXERCISE 1.23. Show that if $J \subset I$ is cofinal then the restriction map ρ is a bijection.

Exercise 1.24. Assuming that each A_i is a Hausdorff topological space, show that $\varprojlim A_i$ is a closed subset of the product space $\prod_{i\in I} A_i$. Conclude that the inverse limit of a directed system of compact Hausdorff topological spaces is again compact and Hausdorff. Conclude, furthermore, that if each A_i is totally disconnected, so is the inverse limit.

Suppose that each A_i is a topological space and we are given a subset $A_i' \subset A_i$ with the subspace topology. Then we have a natural continuous embedding

$$\iota : \prod_{i\in I} A_i' \to \prod_{i\in I} A_i$$

Exercise 1.25. Suppose that for each $i \in I$ there exists $j \in I$ such that $f_{ij}(A_j) \subset A_i'$. Verify that the map ι is a bijection.

We now turn from topological spaces to groups.

Exercise 1.26. Every group G is the direct limit of a direct system $G_i, i \in I$, consisting of all finitely generated subgroups of G. Here the partial order on I is given by inclusion and homomorphisms $f_{ij} : G_i \to G_j$ are tautological embeddings.

Exercise 1.27. Suppose that G is the direct limit of a direct system of groups $\{G_i, f_{ij} : i, j \in I\}$. Assume also that for every i we are given a subgroup $H_i \leqslant G_i$ satisfying

$$f_{ij}(H_i) \leqslant H_j, \quad \forall i \leqslant j.$$

Then the family of groups and homomorphisms

$$\mathcal{H} = \{H_i, f_{ij}|_{H_i} : i, j \in I\}$$

is again a direct system; let H denote the direct limit of this system. Show that there exists a monomorphism $\phi : H \to G$, so that for every $i \in I$,

$$f_i|_{H_i} = \phi \circ h_i : H_i \to G,$$

where $h_i : H_i \to H$ are the homomorphisms associated with the direct limit of the system $\mathcal{H}$.

Exercise 1.28. 1. Let $H \leqslant G$ be a subgroup. Then $|G : H| \leqslant n$ if and only if the following holds: For every subset $\{g_0, \ldots, g_n\} \subset G$, there exist $i \neq j$ so that $g_i g_j^{-1} \in H$.

2. Suppose that G is the direct limit of a system of groups $\{G_i, f_{ij}, i, j \in I\}$. Assume also that there exist $n \in \mathbb{N}$ so that for every $i \in I$, the group G_i contains a subgroup H_i of index $\leqslant n$ and the assumptions of Exercise 1.27 are satisfied. Let the group H be the direct limit of the system

$$\{H_i, f_{ij}|_{H_i} : i, j \in I\}$$

and $\phi : H \to G$ be the monomorphism as in Exercise 1.27. Show that

$$|G : \phi(H)| \leqslant n.$$

1.6. Graphs

An *unoriented graph* Γ consists of the following data:

- a set V called the *set of vertices* of the graph;
- a set E called the *set of edges* of the graph;
- a map ι called *incidence map* defined on E and taking values in the set of subsets of V of cardinality one or two.

We will use the notation $V = V(\Gamma)$ and $E = E(\Gamma)$ for the vertex and edge sets of the graph Γ. Two vertices u, v such that $\{u, v\} = \iota(e)$ for some edge e, are called *adjacent*. In this case, u and v are called the *endpoints* of the edge e.

An unoriented graph can also be seen as a 1-dimensional cell complex (see section 1.7), with 0-skeleton V and with 1-dimensional cells/edges labeled by elements of E, such that the boundary of each 1-cell $e \in E$ is the set $\iota(e)$.

Note that in the definition of a graph we allow for *monogons*[1] (i.e., edges connecting a vertex to itself) and *bigons*[2] (pairs of distinct edges connecting the same pair of vertices). A graph is *simplicial* if the corresponding cell complex is a simplicial complex. In other words, a graph is simplicial if and only if it contains no monogons and bigons.

The incidence map ι defining a graph Γ is set-valued; converting ι to a pair of ordinary maps $E \to V$ is the choice of an *orientation* of Γ: An orientation of Γ is a choice of two maps

$$o : E \to V, \quad t : E \to V$$

such that $\iota(e) = \{o(e), t(e)\}$ for every $e \in E$. In view of the Axiom of Choice, every graph can be oriented.

Definition 1.29. An *oriented* or *directed* graph is a graph Γ equipped with an orientation. The maps o and t are called the *head (or origin) map* and the *tail map* respectively.

Will denote an oriented graphs $\overline{\Gamma}$, their edge-sets $\overline{E}$ and oriented edges $\bar{e}$.

Convention 1.30. In this book, unless we state otherwise, all graphs are assumed to be unoriented.

Examples of graphs. Below we describe several examples of graphs which will appear in this book.

Example 1.31 (n-rose). This graph, denoted R_n, has one vertex and n edges connecting this vertex to itself.

Example 1.32. [i-star or i-pod] This graph, denoted $\pm_i$, has $i + 1$ vertices, $v_0, v_1, \ldots, v_i$. Two vertices are connected by a unique edge if and only if one of these vertices is v_0. The vertex v_0 is the *center* of the star and the edges are called its *legs*.

Example 1.33 (n-circle). This graph, denoted C_n, has n vertices which are identified with the nth roots of unity:

$$v_k = e^{2\pi i k/n}.$$

[1]Not to be confused with *unigons*, which are hybrids of unicorns and dragons.

[2]Also known as *digons*.

Two vertices u, v are connected by a unique edge if and only if they are adjacent to each other on the unit circle:

$$uv^{-1} = e^{\pm 2\pi i/n}.$$

EXAMPLE 1.34 (n-interval). This graph, denoted I_n, has the vertex set equal to $[1, n+1] \cap \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. Two vertices n, m of this graph are connected by a unique edge if and only if

$$|n - m| = 1.$$

Thus, I_n has n edges.

EXAMPLE 1.35 (Half-line). This graph, denoted H, has the vertex set equal to $\mathbb{N}$ (the set of natural numbers). Two vertices n, m are connected by a unique edge if and only if

$$|n - m| = 1.$$

The subset $[n, \infty) \cap \mathbb{N} \subset V(H)$ is the vertex set of a subgraph of H also isomorphic to the half-line H. We will use the notation $[n, \infty)$ for this subgraph.

EXAMPLE 1.36 (Line). This graph, denoted I_n, has the vertex set equal to $\mathbb{Z}$, the set of integerss. Two vertices n, m of this graph are connected by a unique edge if and only if

$$|n - m| = 1.$$

As with general cell complexes and simplicial complexes, we will frequently conflate a graph with its *geometric realization*:

DEFINITION 1.37. The *geometric realization* or *underlying topological space* of an oriented graph $\overline{\Gamma}$ is the quotient space of the topological space

$$U = \bigsqcup_{v \in V} \{v\} \sqcup \bigsqcup_{\bar{e} \in \bar{E}} \{\bar{e}\} \times [0, 1]$$

by the equivalence relation

$$\bar{e} \times \{0\} \sim o(\bar{e}), \quad \bar{e} \times \{1\} \sim t(\bar{e}).$$

One defines the geometric realization of an undirected graph Γ by converting Γ to an oriented graph $\overline{\Gamma}$; the topology of the resulting space is independent of the orientation.

An edge connecting vertices u, v of Γ is denoted $[u, v]$: This is unambiguous if Γ is simplicial. A finite ordered set $[v_1, v_2], [v_2, v_3], \ldots, [v_n, v_{n+1}]$ is called an *edge-path* in Γ. The number n is called the *combinatorial length* of the edge-path. An edge-path in Γ is a *cycle* if $v_{n+1} = v_1$. A *simple* cycle (or a *circuit*), is a cycle where all vertices $v_i, i = 1, \ldots, n$, are distinct. In other words, a simple cycle is a subgraph isomorphic to the n-circle for some n. A graph Γ is *connected* if any two vertices of Γ are connected by an edge-path. Equivalently, the topological space underlying Γ is path-connected.

A subgraph $\Gamma' \subset \Gamma$ is called a *connected component* of Γ if Γ' is a maximal (with respect to the inclusion) connected subgraph of Γ.

A *simplicial tree* is a connected graph without circuits.

A *morphism* of graphs $f : \Gamma \to \Gamma'$ is a pair of maps $f_V : V(\Gamma) \to V(\Gamma')$, $f_E : E(\Gamma) \to E(\Gamma')$ such that

$$\iota' \circ f_E = f_V \circ \iota$$

where ι and ι' are incidence maps of the graphs Γ, Γ' respectively. Thus, every morphism of graphs induces a (nonunique) continuous map $f : \Gamma \to \Gamma'$ of geometric realizations sending vertices of vertices and edges to edges. A *monomorphism* of graphs is a morphism such that the corresponding maps f_V, f_E are injective. The image of a monomorphism $\Gamma \to \Gamma'$ is a *subgraph* of Γ'. In other words, a subgraph in a graph Γ' is defined by subsets $V \subset V(\Gamma'), E \subset E(\Gamma')$ such that

$$\iota'(e) \subset V$$

for every $e \in E$. A subgraph Γ' of Γ is called *full* if every $e = [v, w] \in E(\Gamma)$ connecting vertices of Γ', is an edge of Γ'.

EXERCISE 1.38. Simple circuits in a graph Γ' are precisely subgraphs whose underlying spaces are homeomorphic to the circle.

A morphism $f : \Gamma \to \Gamma'$ of graphs which is invertible (as a morphism) is called an *isomorphism* of graphs: More precisely, we require that the maps f_V, f_E are invertible and the inverse maps define a morphism $\Gamma' \to \Gamma$. In other words, an isomorphism of graphs is an isomorphism of the corresponding cell complexes.

EXERCISE 1.39. 1. For every isomorphism of graphs there exists a (nonunique) homeomorphism $f : \Gamma \to \Gamma'$ of geometric realizations, such that the images of the edges of Γ are edges of Γ' and images of vertices are vertices.

2. Isomorphisms of graphs are morphisms such that the corresponding vertex and edge maps are bijective.

We use the notation $\mathrm{Aut}(\Gamma)$ for the group of automorphisms of a graph Γ.

Maps of graphs. Sometimes, it is convenient to consider maps of graphs which are not morphisms. A *map of graphs* $f : \Gamma \to \Gamma'$ consists of a pair of maps (g, h):

1. $g : V(\Gamma) \to V(\Gamma')$ is a map which sends adjacent vertices map to adjacent or equal vertices.

2. A *partially defined* map of the edge-sets:

$$h : E_o \to E(\Gamma'),$$

where E_o consists of only of edges e of Γ, which connect vertices $v, w \in V(\Gamma)$ with distinct images:

$$g(v) \neq g(w).$$

For each $e \in E_o$, we require the edge $e' = h(e)$ to connect the vertices $g(o(e)), g(t(e))$. In other words, f amounts to a morphism of graph $\Gamma_o \to \Gamma'$, where the vertex set of Γ_o is $V(\Gamma)$ and the edge-set of Γ_p is E_o.

Collapsing a subgraph. Given a graph Γ and its (nonempty) subgraph Λ, we define a new graph, $\Gamma' = \Gamma/\Lambda$ by "collapsing" the subgraph Λ to a vertex. Here is the precise definition. Define the partition $V(\Gamma) = W \sqcup W^c$,

$$W = V(\Lambda), \quad W^c = V(\Gamma) \setminus V(\Lambda).$$

The vertex set of Γ' equals

$$W^c \sqcup \{v_o\}.$$

Thus, we have a natural surjective map $V(\Gamma) \to V(\Gamma')$ sending each $v \in W^c$ to itself and each $v \in W$ to the vertex v_o. The edge-set of Γ' is the set of edges in Γ which *do not* connect vertices of Λ' to each other. Each edge $e \in E(\Gamma)$ connecting $v \in W^c$ to $w \in W$ projects to an edge, also called e, connecting v to v_0. If an edge e connects two vertices in W^c, it is also retained and connects the same vertices in Γ'.

The map $V(\Gamma) \to V(\Gamma')$ extends to a *collapsing* map of graphs $\kappa : \Gamma \to \Gamma'$.

EXERCISE 1.40. If Γ is a tree and Λ is a subtree, then Γ' is again a tree.

The *valency (or valence, or degree) of a vertex* v of a graph Γ is the number of edges having v as one of its endpoints, where every monogon with both vertices equal to v is counted twice. The *valence* of Γ is the supremum of valences of its vertices.

DEFINITION 1.41. Let $F \subset V = V(\Gamma)$ be a set of vertices in a graph Γ. The *vertex-boundary* of F, denoted by $\partial_V F$, is the set of vertices in F each of which is adjacent to a vertex in $V \setminus F$. Similarly, the *exterior vertex-boundary* of F is

$$\partial^V F = \partial_V F^c,$$

where $F^c = V \setminus F$. The *edge-boundary* of F, denoted by $E(F, F^c)$, is the set of edges e such that the set of endpoints $\iota(e)$ intersects both F and its complement $F^c = V \setminus F$ in exactly one element.

Unlike the vertex-boundary, the edge boundary is the same for F, as for its complement F^c. There is a natural surjective map from the edge-boundary $E(F, F^c)$ to the vertex-boundary $\partial_V F$, sending each edge $e \in E(F, F^c)$ to the vertex v, $\{v\} = \iota(e) \cap F^c$. This map is at most C-to-1, where C is valence Γ. Hence, assuming that C is finite, cardinalities of the two types of boundaries are *comparable*:

$$|\partial_V F| \leqslant |E(F, F^c)| \leqslant C |\partial_V F| \,, \tag{1.4}$$

$$|\partial^V F| \leqslant |E(F, F^c)| \leqslant C |\partial^V F| \,, \tag{1.5}$$

DEFINITION 1.42. A simplicial graph Γ is *bipartite* if the vertex set V splits as $V = Y \sqcup Z$, so that each edge $e \in E$ has one endpoint in Y and one endpoint in Z. In this case, we write $\Gamma = Bip(Y, Z; E)$.

EXERCISE 1.43. Let W be an n-dimensional vector space over a field K ($n \geqslant 3$). Let Y be the set of 1-dimensional subspaces of W and let Z be the set of 2-dimensional subspaces of W. Define the bipartite graph $\Gamma = Bip(Y, Z, E)$, where $y \in Y$ is adjacent to $z \in Z$ if, as subspaces in W, $y \subset z$.

1. Compute (in terms of K and n) the valence of Γ, the (combinatorial) length of the shortest circuit in Γ, and show that Γ is connected. 2. Estimate from above the length of the shortest path between any pair of vertices of Γ. Can you get a bound independent of K and n?

1.7. Complexes and homology

Complexes are higher-dimensional generalizations of graphs. In this book, we will primarily use two notions of complexes:

- Simplicial complexes.
- Cell complexes.

As we expect the reader to be familiar with basics of algebraic topology, we will discuss simplicial and cell complexes only briefly.

1.7.1. Simplicial complexes.

DEFINITION 1.44. A simplicial complex X consists of a set $V = V(X)$, called the *vertex set* of X, and a collection $S(X)$ of finite nonempty subsets of V; members of $S(X)$ of cardinality $n+1$ are called n-dimensional simplices. A simplicial complex is required to satisfy the following axioms:

(1) For every simplex $\sigma \in S(X)$, every nonempty subset $\tau \subset \sigma$ is also a simplex. The subset τ is called a *face* of σ. *Vertices* of σ are the 0-faces of σ.
(2) Every singleton $\{v\} \subset V$ is an element of $S(X)$.

A *simplicial map* or *morphism* of two simplicial complexes $f : X \to Y$ is a map $f : V(X) \to V(Y)$ which sends simplices to simplices (dimension of a simplex might decrease under f).

Products of simplicial complexes. Let X, Y be simplicial complexes. We order all the vertices of X and Y. The *product* $Z = X \times Y$, is defined as a simplicial complex whose vertex set is $V(X) \times V(Y)$. A simplex in Z is a tuple

$$(x_0, y_0), \ldots, (x_n, y_n),$$

where $x_0 \leqslant x_1 \leqslant \ldots \leqslant x_n$, $y_0 \leqslant y_1 \leqslant \ldots \leqslant y_n$ and

$$\{x_0, \ldots, x_n\}, \{y_0, \ldots, y_n\}$$

are simplices in X, Y respectively. The product complex Z depends on the orderings of X and Y; however, which orderings to choose will be irrelevant for our purposes.

Simplicial volume. Suppose that X, Y are simplicial complexes, X is finite and $f : X \to Y$ is a simplicial/cellular map. Then the *n-dimensional simplicial/cellular volume* $Vol_n(f)$ of f is just the number of n-dimensional simplices in its domain X. Note that this, somewhat strange, concept, is independent of the map f but is, nevertheless, quite useful, see §5.10.

The more intuitive concept is the one of the *combinatorial volume* of the map f. Define the *combinatorial volume* $cVol_n(f)$ of f as the number of n-dimensional simplices in X which are mapped by f to n-dimensional simplices.

We will use the notation $X^{(i)}$ to denote the i-th skeleton of the simplicial complex X, which is the subset of simplices of dimension $\leq i$ in X. The *geometric realization* of an n-simplex is the *standard simplex*

$$\Delta^n = \{x \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} x_i = 0, x_i \geq 0, i = 1, \ldots, n+1\}.$$

Faces of Δ^n are intersections of the standard simplex with coordinate subspaces of $\mathbb{R}^{n+1}$. Given a simplicial complex X, by gluing copies of standard simplices, one obtains a topological space, which is a *geometric realization* of X.

A *gallery* in an n-dimensional simplicial complex X is a chain of n-simplices $\sigma_1, \ldots, \sigma_k$ so that $\sigma_i \cap \sigma_{i+1}$ is an $n-1$-simplex for every $i = 1, \ldots, k-1$.

Let σ, τ be simplices of dimensions m and n respectively with the vertex sets

$$\sigma^{(0)} = \{v_0, \ldots, v_m\}, \quad \tau^{(0)} = \{w_0, \ldots, w_n\}$$

The product $\sigma \times \tau$, of course, is not a simplex (unless $nm = 0$), but it admits a standard triangulation, whose vertex set is

$$\sigma^{(0)} \times \tau^{(0)}.$$

This triangulation is defined as follows. Pairs $u_{ij} = (v_i, w_j)$ are the vertices of $\sigma \times \tau$. Distinct vertices

$$(u_{i_0,j_0}, \ldots, u_{i_k,j_k})$$

span a k-simplex in $\sigma \times \tau$ if and only if $j_0 \leqslant \ldots \leqslant j_k$.

A homotopy between simplicial maps $f_0, f_1 : X \to Y$ is a simplicial map $F : X \times I \to Y$ which restricts to f_i on $X \times \{i\}, i = 0, 1$. The *tracks* of the homotopy F are the paths $\mathfrak{p}(t) = F(x, t), x \in X$.

Cohomology with compact support. Let X be a simplicial complex. Recall that besides the usual cohomology groups $H^*(X; A)$ (with coefficients in a ring A that the reader can assume to be $\mathbb{Z}$ or $\mathbb{Z}_2$), we also have *cohomology with compact support* $H_c^*(X, A)$ which are defined as follows. Consider the usual cochain complex $C^*(X; A)$. We say that a cochain $\sigma \in C^*(X; A)$ has *compact support* if it vanishes outside of a finite subcomplex in X. Thus, in each chain group $C^k(X; A)$ we have the subgroup $C_c^k(X; A)$ consisting of compactly supported cochains. Then the usual coboundary operator δ satisfies

$$\delta : C_c^k(X; A) \to C_c^{k+1}(X; A).$$

The cohomology of the new cochain complex $(C_c^*(X; A), \delta)$ is denoted $H_c^*(X; A)$ and is called *cohomology of X with compact support*. Maps of simplicial complexes no longer induce homomorphisms of $H_c^*(X; A)$ since they do not preserve the compact support property of cochains; however, *proper* maps of simplicial complexes do induce natural maps on H_c^*. Similarly, maps which are *properly* homotopic induce equal homomorphisms of H_c^* and *proper homotopy equivalences* induce isomorphisms of H_c^*. In other words, H_c^* satisfies the functoriality property of the usual cohomology groups as long as we restrict to the category of proper maps.

Bounded cohomology. Another variation on this construction, which has many applications in geometric group theory, is the concept of *bounded cohomology*.

Let A be a subgroup in $\mathbb{R}$ (the groups $A = \mathbb{Z}$ and $A = \mathbb{R}$ will be the main examples here). One defines the group of *bounded cochains* $C_b^k(X; A) \subset C^k(X; A)$ as the group consisting of cochains which are bounded as functions on $C_k(X)$. It is immediate that the usual coboundary operator satisfies

$$\delta_k : C_b^k(X; A) \to C_b^{k+1}(X; A)$$

for every k. This allows one to define groups of bounded cocycles $Z_b^k(X; A)$ and coboundaries $B_b^{k+1}(X; A)$ as the kernel and image of the coboundary operator restricted to $C_b^k(X; A)$. Hence, one defines the *bounded cohomology groups*

$$H_b^k(X; A) = Z_b^k(X; A)/B_b^k(X; A).$$

The inclusion $C_b^*(X) \to C^*(X)$ induces the group homomorphism

$$H_b^k(X; A) \to H^k(X; A).$$

1.7.2. Cell complexes. A CW complex X is defined as the increasing union of subspaces denoted $X^{(n)}$ (or X^n), called *n-skeleta* of X. The 0-skeleton $X^{(0)}$ of X is a set with discrete topology. Assume that $X^{(n-1)}$ is defined. Let

$$U_n := \coprod_{\alpha \in J} \mathbb{D}^n_\alpha,$$

a (possibly empty) disjoint union of closed n-balls $\mathbb{D}^n_\alpha$. Suppose that for each $\mathbb{D}^n_\alpha$ we have a continuous *attaching map* $e_\alpha : \partial \mathbb{D}^n_\alpha \to X^{(n-1)}$. This defines a map

$$e^n : \partial U_n \to X^{(n-1)}$$

and an equivalence relation $x \sim y = e^n(x)$, $x \in U, y \in X^{(n-1)}$. The space $X^{(n)}$ is the quotient space of $X^{(n-1)} \sqcup U_n$ with the quotient topology with respect to this equivalence relation. Each attaching map e_α extends to the map $\hat{e}_\alpha : \mathbb{D}^n_\alpha \to X^{(n)}$. We will use the notation $\sigma = \mathbb{D}^n_\alpha / e_\alpha$ for the image of $\mathbb{D}^n$ in X^n, it is homeomorphic to the quotient $\mathbb{D}^n_\alpha / \sim$. We will also use the notation $e_\alpha = \partial \hat{e}_\alpha$ and refer to the image of e_α as the *boundary* of the cell $\sigma = \hat{e}_\alpha(\mathbb{D}^n)$, $\partial\sigma = e_\alpha(\partial \mathbb{D}^n)$. The set

$$X := \coprod_{n \in \mathbb{N}} X_n$$

is equipped with the *weak topology*, where a subset $C \subset X$ is closed if and only if the intersection of C with each skeleton is closed (equivalently, intersection of C with each $\hat{e}_\alpha(\mathbb{D}^n_\alpha)$ in X is closed). The space X, together with the collection of maps e_α, is called a *CW complex*. By abuse of terminology, the maps $\hat{e}_\alpha$, the balls $\mathbb{D}^n_\alpha$, and their projections to X are called *n-cells* in X. Similarly, we will conflate the cell complex X and its underlying topological space $|X|$. We will also refer to CW complexes simply as *cell complexes*, even though the usual notion of a cell complex is less restrictive than the one of a CW complex.

We use the terminology an *open* n-cell in X for the open ball $\mathbb{D}^n_\alpha \setminus \partial \mathbb{D}^n_\alpha$, and well as for the restriction of the map $\hat{e}_\alpha$ to the open ball $\mathbb{D}^n_\alpha \setminus \partial \mathbb{D}^n_\alpha$ and for the image of this restriction. We will refer to this open n-cell as the *interior* of the corresponding n-cell.

We will use the notation σ for cells and σ° for their interiors.

The *dimension* of a cell complex X is the supremum of n's such that X has an n-cell. Equivalently, dimension of X is its topological dimension.

EXERCISE 1.45. A subset $K \subset X$ is compact if and only if it is closed and contained in a finite union of cells.

A map $X \to Y$ of cell complexes is called *cellular* if $f(X^{(i)})$ is contained in $Y^{(i)}$ for every i. Similarly to the definition of the n-dimensional simplicial volume of simplicial maps, one defines $Vol_n(f)$, the *cellular volume* of a cellular map $f : X \to Y$, as the number of n-cells in X, provided that X has only finitely many n-cells.

Regular and almost regular cell complexes. A cell complex X is said to be *regular* if every attaching map e_α is one-to-one. For instance, every simplicial complex is a regular cell complex. If $D \subset \mathbb{R}^n$ is a bounded convex polyhedron, then D has natural structure of a regular cell complex X, faces of D are the cells of X.

A regular cell complex is called *triangular* if every cell is naturally isomorphic to a simplex. (Note that a triangular cell complex need not be simplicial since a, nonempty, intersections of two cells need not be a single cell.)

A slightly more general notion is the one of an *almost regular* cell complex. (We could not find this notion in the literature and the terminology is ours.) A cell complex X is *almost regular* if the boundary $\mathbb{S}^{n-1}$ of every n-cell D^n_α is given structure of a regular cell complex K_α, so that the attaching map e_α is one-to-one on every open cell in $\mathbb{S}^{n-1}$.

EXAMPLE 1.46. 1. Consider the 2-dimensional complex X, which has single vertex, single 1-cell (thus, $X^{(1)}$ is homeomorphic to $\mathbb{S}^1$). Let $e : \mathbb{S}^1 \to \mathbb{S}^1$ be a k-fold covering. Attach the 2-cell D^2 to $X^{(1)}$ via the map e. The result is an almost regular (but not regular) cell complex.

2. Let X be the 2-dimensional cell complex obtained by attaching the single 2-cell to the single vertex by the constant map. Then X is not an almost regular cell complex.

Almost regular 2-dimensional cell complexes (with a single vertex) appear naturally in the context of group presentations, see Definition 4.90. For instance, suppose that X is a simplicial complex and $Y \subset X^{(1)}$ is a *forest*, i.e., a subcomplex isomorphic to a disjoint union of simplicial trees. Then the quotient X/Y is an almost regular cell complex.

Barycentric subdivision of an almost regular cell complex. Our goal is to (canonically) subdivide an almost regular cell complex X so that the result is a triangular regular cell complex $X' = Y$. We define Y as an increasing union of regular subcomplexes Y_n (where $Y_n \subset Y^{(n)}$ but, in general, is smaller).

First, set $Y_0 := X^{(0)}$. Suppose that $Y_{n-1} \subset Y^{(n-1)}$ is defined, so that $|Y_{n-1}| = X^{(n-1)}$. Consider attaching maps $e_\alpha : \partial\mathbb{D}^n_\alpha \to X^{(n-1)}$. We take the preimage of the regular cell complex structure of Y_{n-1} under e_α to be a *refinement* L_α of the regular cell complex structure K_α on $\mathbb{S}^{n-1}$. We then define a regular cell complex M_α on D^n_α by conning off every cell in L_α from the origin $o \in \mathbb{D}^n_\alpha$. Then cells in M_α are the cones $Cone_{o_\alpha}(s)$, where s's are the cells in L_α.

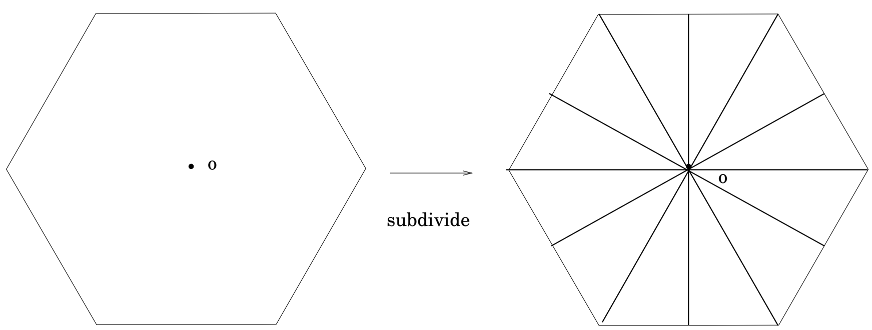


FIGURE 1.1. *B*arycentric subdivision of a 2-cell.

Since, by the induction assumption, every cell in Y_{n-1} is a simplex, its preimage s in $\mathbb{S}^{n-1}$ is also a simplex, this $Cone_o(s)$ is a simplex as well. We then attach each cell $\mathbb{D}^n_\alpha$ to Y_n by the original attaching map e_α. It is clear that the new cells $Cone_{o_j}(s)$ are embedded in Y_n and each is naturally isomorphic to a simplex. Lastly, we set

$$Y := \bigcup_n Y_n.$$

Second barycentric subdivision. Note that the complex X' constructed above may not be a simplicial complex. The problem is that if x, y are distinct vertices of L_j, their images under the attaching map e_α could be the same (a point z). Thus the edges $[o_j, x], [o_j, y]$ in Y_{n+1} will intersect in the set $\{o_j, z\}$. However, if the complex X was regular, this problem does not arise and X' is a simplicial complex. Thus in order to promote X to a simplicial complex (whose geometric realization is homeomorphic to $|X|$), we take the *second barycentric subdivision* X'' of X: Since X' is a regular cell complex, the complex X'' is naturally isomorphic to a simplicial complex.

Mapping cylinders. Let $f : X \to Y$ be a map of topological spaces. The *mapping cylinder* M_f of f is the quotient space of

$$X \times [0, 1] \sqcup Y$$

by the equivalence relation:

$$(x, 1) \sim f(x).$$

Similarly, given two maps $f_i : X \to Y_i, i = 0, 1$, we form the *double mapping cylinder* M_{f_1, f_2}, which is the quotient space of

$$X \times [0, 1] \sqcup Y_0 \sqcup Y_1$$

by the equivalence relation:

$$(x, i) \sim f_i(x), i = 0, 1.$$

If $f : X \to Y, f_i : X \to Y_i, i = 0, 1$, are cellular map of cell complexes, then the corresponding mapping cylinders and double mapping cylinders also have natural structure of cell complexes.

Morphisms of almost regular complexes.

Definition 1.47. Let X and Y almost regular cell complexes. A cellular map $f : X \to Y$ is said to be *almost regular* if for every n-cell σ in X either:

(a) f *collapses* σ, i.e., $f(\sigma) \subset Y^{(n-1)}$, or

(b) f maps the interior of σ homeomorphically onto the interior of an n-cell in Y.

An almost regular map is *regular* or *noncollpasing* if only (b) occurs.

For instance, a simplicial map of simplicial complexes is always almost regular, while a simplicial topological embedding of simplicial complexes is noncollapsing.

1.8. Metric spaces

1.8.1. General metric spaces. A *metric space* is a set X endowed with a function $\text{dist} : X \times X \to \mathbb{R}$ satisfying the following properties:

(M1) $\text{dist}(x, y) \geqslant 0$ for all $x, y \in X$; $\text{dist}(x, y) = 0$ if and only if $x = y$;

(M2) (Symmetry) for all $x, y \in X$, $\text{dist}(y, x) = \text{dist}(x, y)$;

(M3) (Triangle inequality) for all $x, y, z \in X$, $\text{dist}(x, z) \leqslant \text{dist}(x, y) + \text{dist}(y, z)$.

The function dist is called *metric* or *distance function.* Occasionally, we will relax the axiom (M1) and allow $\text{dist}(x, y) = 0$ even for distinct points $x, y \in X$; we will also allow dist to take infinite values, in this case, we interpret triangle inequalities following the usual calculus conventions ($a + \infty = \infty$ for every $a \in \mathbb{R} \cup \infty$, etc.). In this case, we will refer to dist as a *pseudo-distance* or *pseudo-metric.*

Notation. We will use the notation d or dist to denote the metric on a metric space X. For $x \in X$ and $A \subset X$ we will use the notation $\operatorname{dist}(x, A)$ for the *minimal distance* from x to A, i.e.,

$$\operatorname{dist}(x, A) = \inf\{d(x, a) : a \in A\}.$$

Similarly, given two subsets $A, B \subset X$, we define their *minimal distance*

$$\operatorname{dist}(A, B) = \inf\{d(a, b) : a \in A, b \in B\}.$$

For subsets $A, B \subset X$ we let

$$\operatorname{dist}_{Haus}(A, B) = \max\left(\sup_{a \in A} \operatorname{dist}(a, B), \sup_{b \in B} \operatorname{dist}(b, A)\right)$$

denote the *Hausdorff (pseudo) distance* between A and B in X. Two subsets of X are called *Hausdorff-close* if they are within finite Hausdorff distance from each other. See §1.9 for further details on this distance and its generalizations.

Given two maps $f_i : (X, \operatorname{dist}_X) \to (Y, \operatorname{dist}_Y), i = 1, 2$, we define the *distance between these maps*

$$\operatorname{dist}(f_1, f_2) := \sup_{x \in X} \operatorname{dist}(f_1(x), f_2(x)) \in [0, \infty].$$

Let (X, dist) be a metric space. We will use the notation $\mathcal{N}_R(A)$ to denote the *open R-neighborhood* of a subset $A \subset X$, i.e., $\mathcal{N}_R(A) = \{x \in X : \operatorname{dist}(x, A) < R\}$. In particular, if $A = \{a\}$ then $\mathcal{N}_R(A) = B(a, R)$ is the open R-ball centered at a.

We will use the notation $\overline{\mathcal{N}}_R(A)$, $\overline{B}(a, R)$ to denote the corresponding *closed neighborhoods* and *closed balls* defined by non-strict inequalities.

We denote by $S(x, r)$ the *sphere with center x and radius r*, i.e., the set

$$\{y \in X \ : \ \operatorname{dist}(y, x) = r\}.$$

We will use the notation AB to denote a geodesic segment connecting the point A to the point B in X: Note that such segment may be non-unique, so our notation is slightly ambiguous. Similarly, we will use the notation $\triangle(A, B, C)$ or $T(A, B, C)$ for a geodesic triangle in X with the vertices A, B, C. The *perimeter* of a triangle is the sum of its side-lengths (lengths of its edges). Lastly, we will use the notation $\blacktriangle(A, B, C)$ for a solid triangle in a surface with the given vertices A, B and C. Precise definitions of geodesic segments and triangles will be given in section 1.8.2.

A metric space is said to satisfy the *ultrametric inequality* if

$$\operatorname{dist}(x, z) \leqslant \max(\operatorname{dist}(x, y), \operatorname{dist}(y, z)), \forall x, y, z \in X.$$

We will see some examples of ultrametric spaces in section 1.14.

Every norm $|\cdot|$ on a vector space V defines a metric on V:

$$\operatorname{dist}(u, v) = |u - v|.$$

The standard examples of norms on the n-dimensional real vector space V are:

$$|v|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, 1 \leqslant p < \infty,$$

and

$$|v|_{max} = |v|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

In what follows, our default assumption, unless stated otherwise, is that $\mathbb{R}^n$ is equipped with the Euclidean metric, defined by the ℓ_2-norm $|v|_2$; we will also use the notation $\mathbb{E}^n$ for the Euclidean n-space.

EXERCISE 1.48. Show that the Euclidean plane $\mathbb{E}^2$ satisfies the *parallelogram identity*: If A, B, C, D are vertices of a parallelogram P in $\mathbb{E}^2$ with the diagonals AC and BD, then

$$d^2(A,B) + d^2(B,C) + d^2(C,D) + d^2(D,A) = d^2(A,C) + d^2(B,D), \tag{1.6}$$

i.e., sum of squares of the lengths of the sides of P equals the sum of squares of the length of the diagonals of P.

If X, Y are metric spaces, the *product metric* on the direct product $X \times Y$ is defined by the formula

$$d^2((x_1, y_1), (x_2, y_2)) = d^2(x_1, x_2) + d^2(y_1, y_2). \tag{1.7}$$

We will need a *separation* lemma, which is standard (see for instance [**Mun75**, §32]), but we include a proof for the convenience of the reader.

LEMMA 1.49. *Every metric space X is perfectly normal.*

PROOF. Let $A, V \subset X$ be disjoint closed subsets. Both functions dist_A, dist_V, which assign to $x \in X$ its minimal distance to A and to V respectively, are clearly continuous. Therefore, the ratio

$$\sigma(x) := \frac{\text{dist}_A(x)}{\text{dist}_V(x)}, \quad \sigma : X \to [0, \infty],$$

is continuous as well. Let $\tau : [0, \infty] \to [0, 1]$ be a continuous monotone function such that $\tau(0) = 0, \tau(\infty) = 1$, e.g.

$$\tau(y) = \frac{2}{\pi} \arctan(y), \quad y \neq \infty, \quad \tau(\infty) := 1.$$

Then the composition $\rho := \tau \circ \sigma$ satisfies the required properties. □

A metric space (X, dist) is called *proper* if for every $p \in X$ and $R > 0$ the closed ball $\overline{B}(p, R)$ is compact. In other words, the distance function $d_p(x) = d(p, x)$ is proper.

DEFINITION 1.50. Given a function $\phi : \mathbb{R}_+ \to \mathbb{N}$, a metric space X is called ϕ–*uniformly discrete* if each ball $\overline{B}(x, r) \subset X$ contains at most $\phi(r)$ points. A metric space is called *uniformly discrete* if it is ϕ–uniformly discrete for some function ϕ.

Note that every uniformly discrete metric space necessarily has discrete topology.

Given two metric spaces (X, dist_X), (Y, dist_Y), a map $f : X \to Y$ is an *isometric embedding* if for every $x, x' \in X$

$$\text{dist}_Y(f(x), f(x')) = \text{dist}_X(x, x').$$

The image $f(X)$ of an isometric embedding is called an *isometric copy of X in Y*.

A surjective isometric embedding is called an *isometry*, and the metric spaces X and Y are called *isometric*. A surjective map $f : X \to Y$ is called *a similarity with the factor λ* if for all $x, x' \in X$,

$$\text{dist}_Y(f(x), f(x')) = \lambda \text{dist}_X(x, x').$$

The group of isometries of a metric space X is denoted $\mathrm{Isom}(X)$. A metric space is called *homogeneous* if the group $\mathrm{Isom}(X)$ acts transitively on X, i.e., for every $x, y \in X$ there exists an isometry $f : X \to X$ such that $f(x) = y$.

1.8.2. Length metric spaces. Throughout this book, by a *path* in a topological space X we mean a continuous map $\mathfrak{p} : [a, b] \to X$. A path is said to *join* (or *connect*) two points x, y if $\mathfrak{p}(a) = x$, $\mathfrak{p}(b) = y$. We will frequently conflate a path and its image.

Given a path $\mathfrak{p}$ in a metric space X, one defines the *length* of $\mathfrak{p}$ as follows. A partition

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$$

of the interval $[a, b]$ defines a finite collection of points $\mathfrak{p}(t_0), \mathfrak{p}(t_1), \ldots, \mathfrak{p}(t_{n-1}), \mathfrak{p}(t_n)$ in the space X. The *length of* $\mathfrak{p}$ is then defined to be

$$\text{(1.8)} \qquad \mathrm{length}(\mathfrak{p}) = \sup_{a=t_0<t_1<\cdots<t_n=b} \sum_{i=0}^{n-1} \mathrm{dist}(\mathfrak{p}(t_i), \mathfrak{p}(t_{i+1}))$$

where the supremum is taken over all possible partitions of $[a, b]$ and all integers n.

If the length of $\mathfrak{p}$ is finite then $\mathfrak{p}$ is called *rectifiable*, and the path $\mathfrak{p}$ is called *non-rectifiable* otherwise.

Exercise 1.51. Consider a C^1-smooth path in the Euclidean space $\mathfrak{p} : [a, b] \to \mathbb{R}^n$, $\mathfrak{p}(t) = (x_1(t), \ldots, x_n(t))$. Prove that its length (defined above) is given by the familiar formula

$$\mathrm{length}(\mathfrak{p}) = \int_a^b \sqrt{[x_1'(t)]^2 + \ldots + [x_n'(t)]^2}\, dt.$$

Similarly, if (M, g) is a connected Riemannian manifold and dist is the Riemannian distance function (see section 2.3), then the two notions of length, given by equations (2.1) and by (1.8), coincide for smooth paths.

Exercise 1.52. Prove that the graph of the function $f : [0, 1] \to \mathbb{R}$,

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if} \quad 0 < x \leqslant 1\,, \\ 0 & \text{if} \quad x = 0\,, \end{cases}$$

is a non-rectifiable path joining $(0, 0)$ and $(1, \sin(1))$.

Let (X, dist) be a metric space. We define a new metric dist_ℓ on X, known as the *induced intrinsic metric:* $\mathrm{dist}_\ell(x, y)$ is the infimum of the lengths of all rectifiable paths joining x to y.

Exercise 1.53. Show that:
1. dist_ℓ is a metric on X with values in $[0, \infty]$.
2. $\mathrm{dist} \leqslant \mathrm{dist}_\ell$.

Suppose that $\mathfrak{p} : [0, b] \to X$ is a path joining x to y and realizing finite infimum in the definition of distance $D = \mathrm{dist}_\ell(x, y)$. We will (re)parameterize such $\mathfrak{p}$ by its arc-length:

$$\mathfrak{q}(s) = \mathfrak{p}(t),$$

where

$$s = \mathrm{length}(\mathfrak{p}|_{[0,t]})).$$

The resulting path $\mathfrak{q} : [0, D] \to (X, \mathrm{dist}_\ell)$ is called a *geodesic segment* in (X, dist_ℓ).

Note that in a path metric space, *a priori*, not every two points are connected by a geodesic. We extend the notion of geodesic to general metric spaces: A *geodesic* in a metric space (X, dist) is an isometric embedding $\mathfrak{g}$ of an interval in $\mathbb{R}$ into X. Note that this notion is different from the one in Riemannian geometry, where geodesics are isometric embeddings only *locally*, and need not be arc-length parameterized. A geodesic is called a *geodesic ray* if it is defined on an interval $(-\infty, a]$ or $[a, +\infty)$, and it is called *bi-infinite* or *complete* if it is defined on $\mathbb{R}$. As with paths, we will frequently conflate geodesics and their images.

EXERCISE 1.54. Prove that for (X, dist_ℓ) the two notions of geodesics (for maps of finite intervals) agree.

DEFINITION 1.55. A metric space (X, dist) such that $\mathrm{dist} = \mathrm{dist}_\ell$ is called a *length (or path) metric space.*

DEFINITION 1.56. A metric space X is called *geodesic* if every two points in X are connected by a geodesic path. A subset A in a metric space X is called *convex* if for every two points $x, y \in A$ there exists a geodesic $\gamma \subset X$ connecting x and y.

EXERCISE 1.57. Each geodesic metric space is locally path-connected.

A *geodesic triangle* $T = T(A, B, C)$ or $\Delta(A, B, C)$ with vertices A, B, C in a metric space X is a collection of geodesic segments AB, BC, CA in X. These segments are called *edges* of T. We would like to empathize that *triangles* in this book are 1-dimensional objects; we will use the terminology *a solid triangle*, to denote the corresponding 2-dimensional object.

Later on, in Chapters 8 and 9 we will use *generalized* triangles, where some edges are geodesic rays or, even, complete geodesics. The corresponding vertices of the generalized triangles will be *points of the ideal boundary* of X.

EXAMPLES 1.58.
(1) $\mathbb{R}^n$ with the Euclidean metric is a geodesic metric space.
(2) $\mathbb{R}^n \setminus \{0\}$ with the Euclidean metric is a length metric space, but not a geodesic metric space.
(3) The unit circle $\mathbb{S}^1$ with the metric inherited from the Euclidean metric of $\mathbb{R}^2$ (the chordal metric) is not a length metric space. The induced intrinsic metric on $\mathbb{S}^1$ is the one that measures distances as angles in radians, it is the distance function of the Riemannian metric induced by the embedding $\mathbb{S}^1 \to \mathbb{R}^2$.
(4) The Riemannian distance function dist defined for a connected Riemannian manifold (M, g) (see section 2.3) is a path-metric. If this metric is complete, then the path-metric is geodesic.
(5) Every connected graph equipped with the standard distance function (see section 1.8.3) is a geodesic metric space.

EXERCISE 1.59. If X, Y are geodesic metric spaces, so is $X \times Y$. If X, Y are path-metric spaces, so is $X \times Y$. Here $X \times Y$ is equipped with the product metric defined by the formula (1.7).

THEOREM 1.60 (Hopf–Rinow Theorem [**Gro07**]). *If a length metric space is complete and locally compact, then it is geodesic and proper.*

EXERCISE 1.61. Construct an example of a metric space X which is not a length metric space, so that X is complete, locally compact, but is not proper.

1.8.3. Graphs as length spaces. Let Γ be a connected graph. Recall that we are conflating Γ and its geometric realization; the notation $x \in \Gamma$ below will simply mean that x is a point of the geometric realization.

We introduce a path-metric dist on the geometric realization of Γ as follows. We declare every edge of Γ to be isometric to the unit interval in $\mathbb{R}$. Then, the distance between any vertices of Γ is the combinatorial length of the shortest edge-path connecting these vertices. Of course, points of the interiors of edges of Γ are not connected by any edge-paths. Thus, we consider *fractional* edge-paths, where in addition to the edges of Γ we allow intervals contained in the edges. The length of such a fractional path is the sum of lengths of the intervals in the path. Then, for $x, y \in \Gamma$,

$$\mathrm{dist}(x, y) = \inf_{\mathfrak{p}} (\mathrm{length}(\mathfrak{p})),$$

where the infimum is taken over all fractional edge-paths $\mathfrak{p}$ in Γ connecting x to y. The metric dist is called the *standard* metric on Γ.

EXERCISE 1.62. a. Show that infimum is the same as minimum in this definition.

b. Show that every edge of Γ (treated as a unit interval) is isometrically embedded in (Γ, dist).

c. Show that dist is a path-metric.

d. Show that dist is a complete metric.

The notion of a standard metric on a graph generalizes to the concept of a *metric graph*, which is a connected graph Γ equipped with a path-metric $dist_\ell$. Such path-metric is, of course, uniquely determined by the lengths of edges of Γ with respect to the metric d.

EXAMPLE 1.63. Consider Γ which is the complete graph on 3 vertices (a triangle) and declare that two edges e_1, e_2 of Γ are unit intervals and the remaining edge e_3 of Γ has length 3. Let dist_ℓ be the corresponding path-metric on Γ. Then e_3 is not isometrically embedded in $(\Gamma, \mathrm{dist}_\ell)$.

Diameters in graphs. Recall that the diameter of a metric space is the supremum of distances between its points. Suppose that Γ is a connected graph equipped with the standard metric. A subgraph Γ' of Γ is called *a diameter* of Γ, if Γ' is isomorphic to the n-interval I_n, where $n = \mathrm{diam}(\Gamma)$ is the diameter of Γ. This should not cause a confusion since one *diameter* is a number, while the other *diameter* is a subgraph.

EXERCISE 1.64. Suppose that T is a tree of finite diameter. Then:

1. Any two diameters of T have nonempty intersection.

2. The intersection C of all diameters of T is nonempty. The subtree C is the *core* of T.

3. Each connected component of $T \setminus C$ has diameter strictly less than n.

EXERCISE 1.65. Show that each connected graph of finite valence and infinite diameter contains an isometrically embedded copy of $\mathbb{R}_+$.

LEMMA 1.66. *Suppose that $f : H \to T$ is a map of graphs, where H is the half-line and T is a tree, such that* $\mathrm{diam}(f(H))$ *is finite. Then there exists a vertex $v \in T$ such that $f^{-1}(v)$ is unbounded.*

PROOF. The proof is by induction on $D = \text{diam}(f(H))$. If $D = 1$, there is nothing to prove. Suppose that D is at least 2. The image subgraph $f(H)$ is connected and, hence, is a subtree $A \subset T$. Let $C \subset A$ be the core of A as in Exercise 1.64. If there exists a vertex $a \in V(C)$ with infinite preimage $f^{-1}(a)$, we are done. Otherwise, there exists n such that

$$f([n,\infty)),$$

is disjoint from C. Since the subgraph $H' = [n,\infty)$ is isomorphic to the half-line H, we obtain a new map of graph $f\big|_{H'} : H' \to T$. The diameter of the image of this map is strictly less than D. Lemma follows from the induction hypothesis. □

1.9. Hausdorff and Gromov–Hausdorff distances. Nets

The Hausdorff distance between two distinct spaces (for instance, between a space and a dense subspace in it) can be zero. The Hausdorff distance becomes a genuine distance only when restricted to certain classes of subsets, for instance, to the class of compact subsets of a metric space. Still, for simplicity, we call it a *distance* or a *metric* in all cases.

Hausdorff distance defines the topology of *Hausdorff–convergence* on the set $\mathcal{K}_X$ of compact subsets of a metric space X. This topology extends to the set $\mathcal{C}_X$ of closed subsets of X as follows. Given $\epsilon > 0$ and a compact $K \subset X$ we define the neighborhood $U_{\epsilon,K}$ of a closed subset $C \in \mathcal{C}_X$ to be

$$\{Z \in \mathcal{C}_X : \text{dist}_{Haus}(Z \cap K, C \cap K) < \epsilon\}.$$

This system of neighborhoods generates a topology on $\mathcal{C}_X$, called *Chabauty topology*. Thus, a sequence $C_i \in \mathcal{C}_X$ converges to a closed subset $C \in \mathcal{C}_X$ if and only if for every compact subset $K \subset X$,

$$\lim_{i\to\infty} (C_i \cap K) = C \cap K,$$

where the limit is in the topology of Hausdorff–convergence.

M. Gromov defined in [**Gro81a**, section 6] the *modified Hausdorff pseudo-distance* (also called the *Gromov–Hausdorff pseudo-distance*) on the class of proper metric spaces:

$$\text{(1.9)} \qquad \text{dist}_{GH}((X, d_X), (Y, d_Y)) = \inf_{(x,y)\in X\times Y} \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric}$$

$$\text{dist on } M = X \sqcup Y, \text{ such that } \text{dist}(x, y) < \varepsilon, \text{dist}\big|_X = d_X, \text{dist}\big|_Y = d_Y \text{ and}$$

$$B(x, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}.$$

For homogeneous metric spaces the modified Hausdorff pseudo-distance coincides with the pseudo-distance for the pointed metric spaces:

$$\text{(1.10)} \qquad \text{dist}_H((X, d_X, x_0), (Y, d_Y, y_0)) = \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric}$$

$$\text{dist on } M = X \sqcup Y \text{ such that } \text{dist}(x_0, y_0) < \varepsilon,\ \text{dist}|_X = d_X, \text{dist}|_Y = d_Y,$$

$$B(x_0, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y_0, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}.$$

This pseudo-distance becomes a metric when restricted to the class of proper pointed metric spaces. Note that since we use pseudo-metrics in order to define

d_{GH} and d_H, instead of considering pseudo-metrics on the disjoint union $X \sqcup Y$, we can as well consider pseudo-metrics on spaces Z such that X, Y embed isometrically in Z.

In order to simplify the terminology we shall refer to all three pseudo-distances as 'distances' or 'metrics.'

One can associate to every metric space (X, dist) a discrete metric space that is at finite Hausdorff distance from X, as follows.

DEFINITION 1.67. An *ε–separated subset* A in X is a subset such that

$$\mathrm{dist}(a_1, a_2) \geqslant \varepsilon\,,\ \forall a_1, a_2 \in A,\ a_1 \neq a_2\,.$$

A subset S of a metric space X is said to be *r-dense* in X if the Hausdorff distance between S and X is at most r. In other words, for every $x \in X$, we have the inequality $\mathrm{dist}(x, S) \leq r$.

DEFINITION 1.68. An *ε-separated δ–net* in a metric space X is a subset of X that is ε–separated and δ–dense.

An *ε-separated net* in X is a subset that is ε–separated and 2ε–dense.

When the constants ε and δ are not relevant we shall not mention them and simply speak of *separated nets*.

LEMMA 1.69. *A maximal (with respect to inclusion) δ–separated set in X is a δ–separated net in X.*

PROOF. Let N be a maximal δ–separated subset in X. For every $x \in X \setminus N$, the set $N \cup \{x\}$ is no longer δ–separated, by maximality of N. Hence there exists $y \in N$ such that $\mathrm{dist}(x, y) < \delta$. □

By Zorn's lemma a maximal δ–separated subset always exists. Thus, every metric space contains a δ–separated net, for any $\delta > 0$.

EXERCISE 1.70. Prove that if (X, dist) is compact then every separated net in X is finite; hence, every separated subset in X is finite.

DEFINITION 1.71 (Rips complex). Let (X, dist) be a metric space. For $R \geqslant 0$ we define a simplicial complex $\mathrm{Rips}_R(X)$: Its vertices are points of X; vertices $x_0, x_1, ..., x_n$ span a simplex if and only if for all i, j,

$$\mathrm{dist}(x_i, x_j) \leqslant R.$$

The simplicial complex $\mathrm{Rips}_R(X)$ is called the *R-Rips complex of X*.

We will discuss Rips complexes in more detail in §6.2.1.

REMARK 1.72. The complex $\mathrm{Rips}_r(X)$ was first introduced by Leopold Vietors in **[Vie27]**, who was primarily interested in the case of compact metric spaces X and small values of r. This complex was reinvented by Eliyahu Rips in 1980s with the primary goal of studying hyperbolic groups, where X is a hyperbolic group equipped with the word metric and r was large. Accordingly, the complex $\mathrm{Rips}_r(X)$ is also known as the *Vietoris complex* and the *Vietoris-Rips complex.*

1.10. Lipschitz maps and Banach-Mazur distance

If one attempts to think of metrics spaces *categorically*, one wonders what is the right notion of *morphism* in metric geometry. It turns out that depending on the situation, one has to use different notion of morphisms. Lipschitz (especially 1-Lipschitz) and locally Lipschitz maps appear to be the most useful. However, as we will see throughout the book, other classes of maps are also important, especially quasiisometric, quasisymmetric and uniformly proper maps.

1.10.1. Lipschitz and locally Lipschitz maps. A map $f : X \to Y$ between two metric spaces (X, dist_X), (Y, dist_Y) is L-*Lipschitz* if for all $x, x' \in X$

$$\mathrm{dist}_Y(f(x), f(x')) \leqslant L\mathrm{dist}_X(x, x')\,.$$

A map which is L-Lipschitz for some L is called simply *Lipschitz*.

EXERCISE 1.73. 1. Show that every L-Lipschitz path $\mathfrak{p} : [0, 1] \to X$ is rectifiable and $\mathrm{length}(\mathfrak{p}) \leqslant L$.

2. Show that a map $f : X \to Y$ is an isometry if and only if f is 1-Lipschitz and admits a 1-Lipschitz inverse.

For a Lipschitz function $f : X \to \mathbb{R}$ let $\mathrm{Lip}(f)$ denote

$$\mathrm{Lip}(f) := \inf\{L : f \text{ is } L\text{–Lipschitz}\} \tag{1.11}$$

EXERCISE 1.74. Suppose that f, g are Lipschitz functions on X. Let $\|f\|, \|g\|$ denote the sup-norms of f and g on X.Show that

1. "Sum rule": $\mathrm{Lip}(f + g) \leqslant \mathrm{Lip}(f) + \mathrm{Lip}(g)$.
2. "Product rule": $\mathrm{Lip}(fg) \leqslant \mathrm{Lip}(f)\|g\| + \mathrm{Lip}(g)\|f\|$.
3. "Ratio rule":

$$\mathrm{Lip}\left(\frac{f}{g}\right) \leqslant \frac{\mathrm{Lip}(f)\|g\| + \mathrm{Lip}(g)\|f\|}{\inf_{x\in X} g^2(x)}.$$

Note that in case when f is a smooth function on a Riemannian manifold (e.g., on $\mathbb{R}^n$), these formulae follow from the formulae for the derivatives of the sum, product and ratio of two functions.

The following is a fundamental theorem about Lipschitz maps between Euclidean spaces:

THEOREM 1.75 (Rademacher Theorem, see Theorem 3.1 in [**Hei01**]). *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be Lipschitz. Then f is differentiable at almost every point in U.*

A map $f : X \to Y$ is called *locally Lipschitz* if for every $x \in X$ there exists $\epsilon > 0$ so that the restriction $f\big|_{B(x,\epsilon)}$ is Lipschitz. We let $\mathrm{Lip}_{\mathrm{loc}}(X; Y)$ denote the space of locally Lipschitz maps $X \to Y$. We set $\mathrm{Lip}_{\mathrm{loc}}(X) := \mathrm{Lip}_{\mathrm{loc}}(X; \mathbb{R})$.

EXERCISE 1.76. Fix a point p in a metric space (X, dist) and define the function dist_p by $\mathrm{dist}_p(x) := \mathrm{dist}(x, p)$. Show that this function is 1-Lipschitz. Prove the same for the function $\mathrm{dist}_A(x) = \mathrm{dist}(x, A)$, where $A \subset X$ is a nonempty subset.

LEMMA 1.77 (Lipschitz bump-function). *Let $0 < R < \infty$. Then there exists a $\frac{1}{R}$–Lipschitz function $\varphi = \varphi_{p,R}$ on X such that*

1. $\mathrm{Supp}(\varphi) = \overline{B}(p, R)$.

2. $\varphi(p) = 1$.
3. $0 \leqslant \varphi \leqslant 1$ *on* X.

PROOF. We first define the function $\zeta : \mathbb{R}_+ \to [0, 1]$ which vanishes on the interval $[R, \infty)$, is linear on $[0, R]$ and equals 1 at 0. Then ζ is $\frac{1}{R}$–Lipschitz. Now take $\varphi := \zeta \circ \operatorname{dist}_p$. □

LEMMA 1.78 (Lipschitz partition of unity). *Suppose that we are given a locally finite covering of a metric space* X *by a countable set of open* R_i*-balls* $B_i := B(x_i, R_i)$, $i \in I \subset \mathbb{N}$. *Then there exists a collection of Lipschitz functions* η_i, $i \in I$, *so that:*
1. $\sum_i \eta_i \equiv 1$.
2. $0 \leqslant \eta_i \leqslant 1, \quad \forall i \in I$.
3. $\operatorname{Supp}(\eta_i) \subset \overline{B}(x_i, R_i), \quad \forall i \in I$.

PROOF. For each i define the bump-function using Lemma 1.77:

$$\varphi_i := \varphi_{x_i, R_i}.$$

Then the function

$$\varphi := \sum_{i \in I} \varphi_i$$

is positive on X. Finally, define

$$\eta_i := \frac{\varphi_i}{\varphi}.$$

It is clear that the functions η_i satisfy all the required properties. □

REMARK 1.79. Since the collection of balls $\{B_i\}$ is locally finite, it is clear that the function

$$L(x) := \sup_{i \in I, \eta_i(x) \neq 0} \operatorname{Lip}(\eta_i)$$

is bounded on compact sets in X, however, in general, it is unbounded on X. We refer the reader to the equation (1.11) for the definition of $\operatorname{Lip}(\eta_i)$.

From now on, we assume that X is a proper metric space.

PROPOSITION 1.80. $\operatorname{Lip}_{\mathrm{loc}}(X)$ *is a dense subset in* $C(X)$, *the space of continuous functions* $X \to \mathbb{R}$, *equipped with the compact-open topology.*

PROOF. Fix a base-point $o \in X$ and let $A_n, n \in \mathbb{N}$, denote the annulus

$$\{x \in X : n - 1 \leqslant \operatorname{dist}(x, o) \leqslant n\}.$$

Let f be a continuous function on X. Pick $\epsilon > 0$. Our goal is to find a locally Lipschitz function g on X so that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. Since f is uniformly continuous on compact sets, for each $n \in \mathbb{N}$ there exists $\delta = \delta(n, \epsilon)$ such that

$$\forall x, x' \in A_n, \quad \operatorname{dist}(x, x') < \delta \Rightarrow |f(x) - f(x')| < \epsilon.$$

Therefore for each n we find a finite subset

$$X_n := \{x_{n,1}, \ldots, x_{n,m_n}\} \subset A_n$$

so that for $r := \delta(n, \epsilon)/4, R := 2r$, the open balls $B_{n,j} := B(x_{n,j}, r)$ cover A_n. We reindex the set of points $\{x_{n,j}\}$ and the balls $B_{n,j}$ with a countable set I. Thus, we obtain an open locally finite covering of X by the balls $B_j, j \in I$. Let $\{\eta_j, j \in I\}$ denote the corresponding Lipschitz partition of unity. It is then clear that

$$g(x) := \sum_{i \in I} \eta_i(x) f(x_i)$$

is a locally Lipschitz function. For $x \in B_i$ let $J \subset I$ be such that

$$x \notin B(x_j, R_j), \quad \forall j \notin J.$$

Then $|f(x) - f(x_j)| < \epsilon$ for all $j \in J$. Therefore

$$|g(x) - f(x)| \leqslant \sum_{j \in J} \eta_j(x) |f(x_j) - f(x)| < \epsilon \sum_{j \in J} \eta_j(x) = \epsilon \sum_{i \in I} \eta_j(x) = \epsilon.$$

It follows that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. □

A relative version of Proposition 1.80 also holds:

PROPOSITION 1.81. *Let $A \subset X$ be a closed subset contained in a subset U which is open in X. Then, for every $\epsilon > 0$ and every continuous function $f \in C(X)$ there exists a function $g \in C(X)$ so that:*
1. g is locally Lipschitz on $X \setminus U$.
2. $\|f - g\| < \epsilon$.
3. $g\big|_A = f\big|_A$.

PROOF. For the closed set $V := X \setminus U$ pick a continuous function $\rho = \rho_{A,V}$ separating the sets A and V. Such a function exists, by Lemma 1.49. According to Proposition 1.80, there exists $h \in \mathrm{Lip}_{\mathrm{loc}}(X)$ such that $\|f - h\| < \epsilon$. Then take

$$g(x) := \rho(x) h(x) + (1 - \rho(x)) f(x).$$

We leave it to the reader to verify that g satisfies all the requirements of the proposition. □

1.10.2. Bi–Lipschitz maps. The Banach-Mazur distance. A map of metric spaces $f : X \to Y$ is $L-$*bi-Lipschitz* if it is a bijection and both f and f^{-1} are L-Lipschitz for some L; equivalently, f is surjective and there exists a constant $L \geqslant 1$ such that for every $x, x' \in X$

$$\frac{1}{L} \mathrm{dist}_X(x, x') \leqslant \mathrm{dist}_Y(f(x), f(x')) \leqslant L \mathrm{dist}_X(x, x')\,.$$

A *bi-Lipschitz embedding* is defined by dropping surjectivity assumption.

EXAMPLE 1.82. Suppose that X, Y are connected Riemannian manifolds (M, g), (N, h) (see section 2.3). Then a diffeomorphism $f : M \to N$ is L-bi-Lipschitz if and only if

$$L^{-1} \leqslant \sqrt{\frac{f^* h}{g}} \leqslant L.$$

In other words, for every tangent vector $v \in TM$,

$$L^{-1} \leqslant \frac{|df(v)|}{|v|} \leqslant L.$$

If there exists a bi-Lipschitz map $f : X \to Y$, the metric spaces (X, dist_X) and (Y, dist_Y) are called *bi-Lipschitz equivalent* or *bi-Lipschitz homeomorphic*. If dist_1 and dist_2 are two distances on the same metric space X such that the identity map $\mathrm{id} : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$ is bi-Lipschitz, then we say that dist_1 and dist_2 are *bi-Lipschitz equivalent*.

EXAMPLES 1.83.

(1) Any two metrics d_1, d_2 on $\mathbb{R}^n$ defined by two norms on $\mathbb{R}^n$, are bi-Lipschitz equivalent.
(2) Any two left-invariant Riemannian metrics on a connected real Lie group define bi-Lipschitz equivalent distance functions.

EXAMPLE 1.84. If $T : V \to W$ is a continuous linear map between Banach spaces, then

$$Lip(T) = \|T\|,$$

the operator norm of T.

The *Banach-Mazur distance* $\mathrm{dist}_{BM}(V, W)$ between two Banach spaces V and W is

$$\log\left(\inf_{T:V\to W}\left(\|T\| \cdot \|T^{-1}\|\right)\right),$$

where the infimum is taken over all bounded invertible linear maps $T : V \to W$ with bounded inverse. The reader can think of dist_{BM} as a Banach-space analogue (and precursor) of Gromov–Hausdorff distance between metric spaces.

EXERCISE 1.85. Show that dist_{BM} is a metric on the set of n-dimensional Banach spaces.

<u>THEOREM</u> 1.86 (John's Theorem, see e.g. [**MS86**], Theorem 3.3). *For every pair of n-dimensional normed vector spaces V, W,* $\mathrm{dist}_{BM}(V, W) \leqslant \log(n)$.

1.11. Hausdorff dimension

In this section we review the concept of *Hausdorff dimension* for metric spaces.

We let ω_n denote the volume of the unit Euclidean n-ball. The function ω_n extends to $\mathbb{R}_+$ by the formula

$$\omega_\alpha = \frac{\pi^{\alpha/2}}{\Gamma(1 + \alpha/2)},$$

where Γ is the Gamma-function.

Let K be a metric space and $\alpha > 0$. The *α–Hausdorff measure* $\mu_\alpha(K)$ is defined as

$$\omega_\alpha \lim_{r\to 0} \inf \sum_i r_i^\alpha, \tag{1.12}$$

where the infimum is taken over all countable coverings of K by balls $B(x_i, r_i)$, $r_i \leqslant r$. The motivation for this definition is that the volume of the Euclidean r-ball of dimension $n \in \mathbb{N}$ is $\omega_n r^n$; hence, Lebesgue measure of a (measurable) subset of $\mathbb{R}^n$ equals its n-Hausdorff measure. Euclidean spaces, of course, have integer dimension, the point of Hausdorff measure and dimension is to extend the definition to the non-integer case.

EXERCISE 1.87. Suppose that $f : X \to Y$ is an L-Lipschitz map between metric spaces. Show that

$$\mu_\alpha(f(X)) \leqslant L^\alpha \mu_\alpha(X).$$

The *Hausdorff dimension* of the metric space K is defined as:

$$\dim_H(K) := \inf\{\alpha : \mu_\alpha(K) = 0\}.$$

EXERCISE 1.88. Verify that the Euclidean space $\mathbb{R}^n$ has Hausdorff dimension n.

We will need the following theorem:

THEOREM 1.89 (L. Sznirelman; see [**HW41**]). *The covering dimension* $\dim(X)$ *of a proper metric space* X *is at most the Hausdorff dimension* $\dim_H(X)$.

Let $A \subset X$ be a closed subset. Recall that $\mathbb{D}^n := \overline{B}(0,1) \subset \mathbb{R}^n$ denotes the closed unit ball in $\mathbb{R}^n$. Define

$$C(X, A; B^n) := \{f \in C(X, \mathbb{D}^n) \ : \ f(A) \subset \mathbb{S}^{n-1} = \partial \mathbb{D}^n\}.$$

An immediate consequence of Proposition 1.81 is the following.

COROLLARY 1.90. *For every function* $f \in C(X, A; \mathbb{D}^n)$ *and an open set* $U \subset X$ *containing* A*, there exists a sequence of functions* $g_i \in C(X, A; \mathbb{D}^n)$ *so that for all* $i \in \mathbb{N}$*:*

1. $g_i\big|_A = f\big|_A$.
2. $g_i \in \mathrm{Lip}(X \setminus U; \mathbb{R}^n)$.

For a continuous map $f : X \to \mathbb{D}^n$ define $A = A_f$ as

$$A := f^{-1}(\mathbb{S}^{n-1}).$$

DEFINITION 1.91. The map f is *inessential* if it is homotopic rel. A to a map $f' : X \to \mathbb{S}^{n-1}$. An *essential map* is the one which is not inessential.

We will be using the following characterization of the covering dimension due to Alexandrov:

THEOREM 1.92 (P. S. Alexandrov, see Theorem III.5 in [**Nag83**]). $\dim(X) < n$ *if and only if every continuous map* $f : X \to \mathbb{D}^n$ *is inessential.*

We are now ready to prove Theorem 1.89. Suppose that $\dim_H(X) < n$. We will prove that $\dim(X) < n$ as well. We need to show that every continuous map $f : X \to \mathbb{D}^n$ is inessential. Let D denote the annulus $\{x \in \mathbb{R}^n : 1/2 \leqslant |x| < 1\}$. Set $A := f^{-1}(\mathbb{S}^{n-1})$ and $U := f^{-1}(D)$.

Take the sequence g_i given by Corollary 1.90. Since each g_i is homotopic to f rel. A, it suffices to show that some g_i is inessential. Since $f = \lim_i g_i$, it follows that for all sufficiently large i,

$$g_i(U) \cap B\left(0, \frac{1}{3}\right) = \emptyset.$$

We claim that the image of every such g_i misses a point in $B\left(0, \frac{1}{3}\right)$. Indeed, since $\dim_H(X) < n$, the n-dimensional Hausdorff measure of X is zero. However, $g_i|X \setminus U$ is locally Lipschitz. Therefore $g_i(X \setminus U)$ has zero n-dimensional Hausdorff (and hence Lebesgue) measure, see Exercise 1.87. It follows that $g_i(X)$ misses a point y in $B\left(0, \frac{1}{3}\right)$. Composing g_i with the retraction $\mathbb{D}^n \setminus \{y\} \to \mathbb{S}^{n-1}$ we get a map $f' : X \to \mathbb{S}^{n-1}$ which is homotopic to f rel. A. Thus f is inessential and, therefore, $\dim(X) < n$. □

1.12. Norms and valuations

In this and the following section we describe certain metric spaces of algebraic origin that will be used in the proof of the Tits alternative. We refer the reader to [**Lan02**, Chapter XII] for more detail.

A *norm* or an *absolute value* on a ring R is a function $|\cdot|$ from R to $\mathbb{R}_+$, which satisfies the following axioms:

1. $|x| = 0 \iff x = 0$.
2. $|xy| = |x| \cdot |y|$.
3. $|x + y| \leqslant |x| + |y|$.

An element $x \in R$ such that $|x| = 1$ is called a *unit*. A norm $|\cdot|$ is called *nonarchimedean* if it satisfies the *ultrametric* inequality

$$|x + y| \leqslant \max(|x|, |y|).$$

According to Ostrowski's theorem, if $(F, ||\cdot||)$ is a normed field which is *not nonarchimedean*, then there exists a bi-Lipschitz monomorphism $f : R \hookrightarrow \mathbb{C}$, where $\mathbb{C}$ is equipped with the Euclidean norm $|\cdot|$ given by the absolute value of complex numbers. More precisely, there exists $\alpha \in \mathbb{R}_+$ such that

$$||x|| = |f(x)|^\alpha,$$

for all $x \in F$. Such norms $||\cdot||$ are called *archimedean*. We will be primarily interested in normed archimedean fields which are $\mathbb{R}$ and $\mathbb{C}$ with the usual norms given by the absolute value. In the case $F = \mathbb{Q}$, Ostrowski's theorem can be made even more precise: Every norm $||\cdot||$ on $\mathbb{Q}$ arises as a power of the Euclidean norm or of a p-adic norm.

Below is an alternative approach to nonarchimedean normed rings R. A function $\nu : R \to \mathbb{R} \cup \{\infty\}$ is called a *valuation* if it satisfies the following axioms:

1. $v(x) = \infty \iff x = 0$.
2. $v(xy) = v(x) + v(y)$.
3. $v(x + y) \geqslant \min(v(x), v(y))$.

Therefore, one converts a valuation to a nonarchimedean norm by setting

$$|x| = c^{-v(x)}, x \neq 0, \quad |0| = 0,$$

where $c > 0$ is a fixed real number.

REMARK 1.93. More generally, one also considers valuations with values in arbitrary ordered abelian groups, but we will not need this.

A normed ring R is said to be *local* if it is locally compact as a metric space; a normed ring R is said to be *complete* if it is complete as a metric space. A norm on a field F is said to be *discrete* if the image Γ of $|\cdot| : F^\times = F \setminus \{0\} \to \mathbb{R}^\times$ is an infinite cyclic group. If the norm is discrete, then an element $\pi \in F$ such that $|\pi|$ is a generator of Γ satisfying $|\pi| < 1$, is called a *uniformizer* of F. If F is a field with valuation v, then the subset

$$O_v = \{x \in F : v(x) \geqslant 0\}$$

is a subring in F, the *valuation ring* or the *ring of integers* in F.

EXERCISE 1.94. 1. Verify that every nonzero element of a field F with discrete norm has the form $\pi^k u$, where u is a unit.

2. Verify that every discrete norm is nonarchimedean.

Below are the two main examples of fields with discrete norms:

1. Field $\mathbb{Q}_p$ of p-adic numbers. Fix a prime number p. For each number $x = q/p^n \in \mathbb{Q}$ (where both numerator and denominator of q are not divisible by p) set $|x|_p := p^n$. Then $|\cdot|_p$ is a nonarchimedean norm on $\mathbb{Q}$, called the *p-adic norm*. The completion of $\mathbb{Q}$ with respect to the p-adic norm is the field of p-adic numbers $\mathbb{Q}_p$. The ring of p-adic integers O_p intersects $\mathbb{Q}$ along the subset consisting of (reduced) fractions $\frac{n}{m}$ where $m, n \in \mathbb{Z}$ and m is not divisible by p. Note that p is a uniformizer of $\mathbb{Q}_p$.

REMARK 1.95. We will not use the common notation $\mathbb{Z}_p$ for O_p, in order to avoid the confusion with finite cyclic groups.

EXERCISE 1.96. Verify that O_p is open in $\mathbb{Q}_p$. Hint: Use the fact that $|x+y|_p \leqslant 1$ provided that $|x|_p \leqslant 1, |y_p| \leqslant 1$.

Recall that one can describe real numbers using infinite decimal sequences. There is a similar description of p-adic numbers using "base p arithmetic." Namely, we can identify p-adic numbers with semi-infinite Laurent series

$$\sum_{k=-n}^{\infty} a_k p^k,$$

where $n \in \mathbb{Z}$ and $a_k \in \{0, \ldots, p-1\}$. Operations of addition and multiplication here are the usual operations with power series where we treat p as a formal variable, the only difference is that we still have to "carry to the right" as in the usual decimal arithmetic.

With this identification, $|x|_p = p^n$, where a_{-n} is the first nonzero coefficient in the power series. The corresponding valuation is $v(x) = -n$, $c = p$. In particular, the ring O_p is identified with the set of series

$$\sum_{k=0}^{\infty} a_k p^k.$$

REMARK 1.97. In other words, one can describe p-adic numbers as left-infinite sequences of (base p) digits

$$\cdots a_m a_{m-1} \ldots a_0 . a_{-1} \cdots a_{-n}$$

where $\forall i, a_i \in \{0, \ldots, p-1\}$, and the algebraic operations require "carrying to the left" instead of carrying to the right.

EXERCISE 1.98. Show that in $\mathbb{Q}_p$,

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}.$$

2. Let A be a field. Consider the ring $R = A[t, t^{-1}]$ of Laurent polynomials

$$f(t) = \sum_{k=n}^{m} a_k t^k.$$

Set $v(0) = \infty$ and for nonzero f let $v(f)$ be the least n so that $a_n \neq 0$. In other words, $v(f)$ is the order of vanishing of f at $0 \in R$.

EXERCISE 1.99. 1. Verify that v is a valuation on R. Define $|f| := e^{-v(f)}$.

2. Verify that the completion $\widehat{R}$ of R with respect to the above norm is naturally isomorphic to the ring of semi-infinite formal Laurent series

$$f = \sum_{k=n}^{\infty} a_k t^k,$$

where $v(f)$ is the minimal n such that $a_n \neq 0$.

Let $A(t)$ be the field of rational functions in the variable t. We embed A in $\widehat{R}$ by the rule

$$\frac{1}{1-at} = 1 + \sum_{n=1}^{\infty} a^n t^n.$$

If A is algebraically closed, every rational function is a product of a polynomial function and several functions of the form

$$\frac{1}{a_i - t},$$

so we obtain an embedding $A(t) \hookrightarrow \widehat{R}$ in this case. If A is not algebraically closed, proceed as follows. First, construct, as above, an embedding ι of $\bar{A}(t)$ to the completion of $\bar{A}[t, t^{-1}]$, where $\bar{A}$ is the algebraic closure of A. Next, observe that this embedding is equivariant with respect to the Galois group $Gal(\bar{A}/A)$, where $\sigma \in Gal(\bar{A}/A)$ acts on Laurent series

$$f = \sum_{k=n}^{\infty} a_k t^k, a \in \bar{A},$$

by

$$f^{\sigma} = \sum_{k=n}^{\infty} a_k^{\sigma} t^k.$$

Therefore, $\iota(A(t)) \subset \widehat{R}, R = A[t, t^{-1}]$.

In any case, we obtain a norm on $A(t)$ by restricting the norm in $\widehat{R}$. Since $R \subset \iota A(t)$, it follows that $\widehat{R}$ is the completion of $\iota A(t)$. In particular, $\widehat{R}$ is a complete normed field.

EXERCISE 1.100. 1. Verify that $\widehat{R}$ is local if and only if A is finite.

2. Show that t is a uniformizer of $\widehat{R}$.

3. At the first glance, it looks like $\mathbb{Q}_p$ is the same as $\widehat{R}$ for $A = \mathbb{Z}_p$, since elements of both are described using formal power series with coefficients in $\{0, \ldots, p-1\}$. What is the difference between these fields?

The same constructrion works with Laurent polynomials of several variables. We let A be a field, $T = \{t_1, \ldots, t_l\}$ is a finite set of variables and consider first the ring of Laurent polynomials:

$$A[t_1^{\pm}1, \ldots, t_n^{\pm 1}]$$

in these variables. The degree of a monomial

$$at_1^{k_1} \ldots t_n^{k_n}$$

with $a \neq 0$ is defined as the sum

$$k_1 + \ldots + k_n.$$

For a general Laurent polynomial p in the variables T, set $v(p) = d$ iff d is the lowest degree of all nonzero monomials in p. With this definition, one again gets a complete norm $|\cdot|$ on the field $A(t_1, ..., t_n)$ of rational functions in the variables t_i, where

$$|p| = e^{-v(p)}.$$

EXERCISE 1.101. If A is finite, then the normed field $(A(t_1, ..., t_n), |\cdot|)$ is local.

Similarly, we have the following lemma:

LEMMA 1.102. *$\mathbb{Q}_p$ is a local field.*

PROOF. It suffices to show that the ring O_p of p-adic integers is compact. Since $\mathbb{Q}_p$ is complete, we only need to show that O_p is closed and totally bounded, i.e., for every $\epsilon > 0$, O_p has a finite cover by closed ϵ-balls. The fact that O_p is closed follows from the fact that $|\cdot|_p : \mathbb{Q}_p \to \mathbb{R}$ is continuous and O_p is given by the inequality $O_p = \{x : |x|_p \leqslant 1\}$.

Let us check that O_p is totally bounded. For $\epsilon > 0$ pick $k \in \mathbb{N}$ such that $p^{-k} < \epsilon$. The ring $\mathbb{Z}/p^k\mathbb{Z}$ is finite, let $z_1, \ldots, z_N \in \mathbb{Z} \setminus \{0\}$ (where $N = p^k$) denote representatives of the cosets in $\mathbb{Z}/p^k\mathbb{Z}$. We claim that the set of fractions

$$w_{ij} = \frac{z_i}{z_j}, \quad 1 \leqslant i, j \leqslant N,$$

forms a p^{-k}-net in $O_p \cap \mathbb{Q}$. Indeed, for a rational number $\frac{m}{n} \in O_p \cap \mathbb{Q}$, find $s, t \in \{z_1, \ldots, z_N\}$ such that

$$s \equiv m, t \equiv n, \mod p^k.$$

Then

$$\frac{m}{n} - \frac{s}{t} \in p^k O_p$$

and, hence,

$$\left|\frac{m}{n} - \frac{s}{t}\right|_p \leqslant p^{-k}.$$

Since $O_p \cap \mathbb{Q}$ is dense in O_p, it follows that

$$O_p \subset \bigcup_{i,j=1}^{N} \overline{B}\left(w_{ij}, \epsilon\right). \quad \square$$

EXERCISE 1.103. Show that O_p is homeomorphic to the Cantor set. Hint: Verify that O_p is totally disconnected and perfect.

1.13. Norms on field extensions. Adeles

A proof of the following theorem can be found e.g., in [**Lan02**, Chapter XII.2, Proposition 2.5].

THEOREM 1.104. *Suppose that $(\mathbb{E}, |\cdot|)$ is a normed field and $\mathbb{E} \subset \mathbb{F}$ is a finite extension. Then the norm $|\cdot|$ extends to a norm $|\cdot|$ on $\mathbb{F}$ and this extension is unique. If $|\cdot|$ and if $(\mathbb{E}, |\cdot|)$ is a local field, then so is $(\mathbb{F}, |\cdot|)$.*

We note that the statement about local fields follows from the fact that if V is a finite-dimensional normed vector space over a local field, then V is locally compact.

Norms on number fields are used to define *rings of adeles* of these fields. We refer the reader to [**Lan64**, Chapter 6] for the detailed treatment of adeles.

We let $\mathrm{Nor}(\mathbb{Q})$ denote the set of *norms* on $\mathbb{Q}$, $|\cdot| : \mathbb{Q} \to \mathbb{R}_+$, see §1.12. If $\mathbb{F}$ is an algebraic number field (a finite algebraic extension of $\mathbb{Q}$), then we let $\mathrm{Nor}(\mathbb{F})$ denote the set of norms on $\mathbb{F}$ extending the ones on $\mathbb{Q}$. We will use the notation ν and $|\cdot|_\nu$ for the elements of $\mathrm{Nor}(\mathbb{F})$ and O_ν for the corresponding rings of integers; we let ν_p denote the p-adic norm and its unique extension to $\mathbb{F}$. Note that for each $x \in \mathbb{Q}$, $x \in O_p$ (the ring of p-adic integers) for all but finitely many p's, since x has only finitely many primes in its denominator.

For each ν we let $\mathbb{F}_\nu$ denote the completion of $\mathbb{F}$ with respect to ν and set $N_\nu = [\mathbb{F}_\nu : \mathbb{Q}_\nu]$.

LEMMA 1.105 (Product formula). *For each $x \in \mathbb{F} \setminus \{0\}$ we have*

$$\prod_{\nu \in \mathrm{Nor}(\mathbb{F})} (\nu(x))^{N_\nu} = 1.$$

PROOF. We will prove this in the case $\mathbb{F} = \mathbb{Q}$; the reader can find the proof for general number fields in [**Lan64**, Chapter 6]. If $x = p$ is prime, then $|p| = p$ for the archimedean norm, $\nu(p) = 1$ if $\nu \neq \nu_p$ is a nonarchimedean norm and $\nu_p(p) = 1/p$. Thus, the product formula holds for prime numbers x. Since norms are multiplicative functions from $\mathbb{Q}^\times$ to $\mathbb{R}_+$, the product formula holds for arbitrary $x \neq 0$. □

For a nonarchimedean norm ν we let

$$O_\nu = \{x \in \mathbb{F}_\nu : |x|_\nu \leq 1\}$$

denote the ring of integers in $\mathbb{F}_\nu$. Since ν is nonarchimedean, O_ν is both closed and open in $\mathbb{F}_\nu$.

DEFINITION 1.106. For a finitely generated algebraic number field $\mathbb{F}$, the ring of *adeles* is the *restricted product*

$$\mathbb{A}(\mathbb{F}) := \prod_{\nu \in \mathrm{Nor}(\mathbb{F})}{}' \mathbb{F}_\nu,$$

i.e., the subset of the direct product

$$\prod_{\nu \in \mathrm{Nor}(\mathbb{F})} \mathbb{F}_\nu \tag{1.13}$$

which consists of sequences whose projection to $\mathbb{F}_\nu$ belongs to O_ν for all but finitely many ν's. The ring operations on $\mathbb{A}(\mathbb{F})$ are defined first on sequences in the infinite product which have only finitely many nonzero terms and then extends to the rest of $\mathbb{A}(\mathbb{F})$ by taking suitable limits.

Note that in the case $F = \mathbb{Q}$, the $\mathbb{A}(\mathbb{Q})$ is the restricted product

$$\mathbb{R} \times \prod_{p \text{ is prime}}{}' \mathbb{Q}_p.$$

Adelic topology. Open subsets in the adelic topology on $\mathbb{A}(F)$ are products of open sets of F_ν for finitely many ν's (including all archimedean ones) and of O_ν's for the rest of ν's. Then the ring operations are continuous with respect to this topology. Accordingly, we topologize the group $GL(n, \mathbb{A}(\mathbb{F}))$ using the product topology on $\mathbb{A}(F)^{n^2}$. With this topology, $GL(n, \mathbb{A}(\mathbb{F}))$ becomes a topological group. Tychonoff's theorem implies compactness of product sets of the form

$$\prod_{\nu \in \mathrm{Nor}(\mathbb{F})} C_\nu,$$

where $C_\nu \subset \mathbb{F}_\nu$ is compact for each ν, which equals to O_ν for all but finitely many ν's,

Theorem 1.107 (See e.g. Chapter 6, Theorem 1 in [**Lan64**]). *The image $\iota(\mathbb{F})$ of the diagonal embedding $\mathbb{F} \hookrightarrow \mathbb{A}(\mathbb{F})$ is a discrete subset in $\mathbb{A}(\mathbb{F})$.*

Proof. Since $\iota(F)$ is an additive subgroup of the topological group $\mathbb{A}(F)$, it suffices to verify that 0 is an isolated point of $\iota(\mathbb{F})$. Take the archimedean norms $\nu_1, \ldots, \nu_m \in \mathrm{Nor}(\mathbb{F})$ (there are only finitely many of them since the Galois group $Gal(\mathbb{F}/\mathbb{Q})$ is finite) and consider the open subset

$$U = \prod_{i=1}^{m} \{x \in \mathbb{F}_{\nu_i} : \nu_i(x) < 1/2\} \times \prod_{\mu \in \mathrm{Nor}(\mathbb{F}) \setminus \{\nu_1, \ldots, \nu_m\}} O_\mu$$

of $\mathbb{A}(\mathbb{F})$. Then for each $(x_\nu) \in U$,

$$\prod_{\nu \in \mathrm{Nor}(\mathbb{F})} \nu(x_\nu) < 1/2 < 1.$$

Hence, by the product formula, the intersection of U with the image of $\mathbb{F}$ in $\mathbb{A}(\mathbb{F})$ consists only of $\{0\}$. ☐

In order to appreciate this theorem, note that $\mathbb{F} = \mathbb{Q}$ is dense in the completion of $\mathbb{Q}$ with respect to every norm. We also observe that this theorem fails if we equip $\mathbb{A}(\mathbb{F})$ with the topology induced from the product topology on the product of all $\mathbb{F}_\nu$'s.

Corollary 1.108. *The image of $GL(n, \mathbb{F})$ under the embedding $\iota : GL(n, \mathbb{F}) \to GL(n, \mathbb{A}(\mathbb{F}))$ (induced by the diagonal embedding $\mathbb{F} \to \mathbb{A}(\mathbb{F})$) is a discrete subgroup.*

Even though, the adelic topology on $\mathbb{A}(F)$ is strictly stronger than the product topology, we note, nevertheless, that for a finitely generated subgroup $L \leqslant \mathbb{F}$, the image of L under the diagonal embedding $\iota : \mathbb{F} \to \mathbb{A}(\mathbb{F})$ projects to O_ν for all but finitely many ν's (since the generators of L have only finitely many denominators). Thus, the restriction of the adelic topology to $\iota(L)$ coincides with the restriction of the product topology. The same applies for finitely generated subgroups $\Gamma \leqslant GL(n, \mathbb{F})$, since such Γ is contained in $GL(n, \mathbb{F}')$ for a finitely generated subfield $\mathbb{F}'$ of $\mathbb{F}$. We, thus, obtain:

Corollary 1.109. *Suppose that $\Gamma \leqslant GL(n, \mathbb{F})$ is a finitely generated subgroup which project to a relatively compact subgroup of $GL(n, \mathbb{F}_\nu)$ for every norm ν. Then Γ is finite.*

PROOF. Since Γ is finitely generated, the restriction of the product topology on

$$\prod_{\nu \in \mathrm{Nor}(F)} GL(n, \mathbb{F}_\nu)$$

to $\iota(\Gamma)$ coincides with the adelic topology, since Γ projects to $GL(n, O_\nu)$ for all but finitely many ν's. In the adelic topology, $\iota(\Gamma)$ is discrete, while, in the product topology, it is a closed subset of a set C, which is the product of compact subsets of the groups $GL(n, \mathbb{F}_\nu)$. Hence, by Tychonoff's Theorem, C is compact. Thus, $\iota(\Gamma)$ is a discrete compact topological space, which implies that $\iota(\Gamma)$ is finite. Since ι is injective, it follows that Γ is finite as well. $\square$

COROLLARY 1.110. *Suppose that $\alpha \in \bar{\mathbb{Q}}$ is an algebraic integer, i.e., a root of a monic polynomial $p(x)$ with integer coefficients. Then either α is a root of unity or $p(x)$ has a root β such that $|\beta| > 1$. In other words, there exists an element $\sigma \in Gal(\bar{\mathbb{Q}}/\mathbb{Q})$ which sends α to an algebraic number β with non-unit absolute value.*

PROOF. Let $\mathbb{F} = \mathbb{Q}(\alpha)$ and consider the cyclic subgroup $\Gamma < \mathbb{Q}^\times$ generated by α. Since $\alpha \in O_{\mathbb{F}}$, we conclude that α belongs to the ring of integers O_α for each nonarchimedean norm ν of $\mathbb{F}$. Thus, Γ projects to the compact subgroup O_ν of $\mathbb{A}(\mathbb{F})$. For each archimedean norm ν, we have $|\alpha|_\nu = |\sigma(\alpha)|$, where $\sigma \in Gal(\mathbb{F}/\mathbb{Q})$. However, $\sigma(\alpha)$ is another root of $p(x)$. Therefore, either there exists a root β of $p(x)$ such that $|\beta| \neq 1$, or Γ projects to a compact subgroup of F_ν for each ν, both archimedean and nonarchimedean. In the latter case, by Corollary 1.109, the group Γ is finite. Hence, α is a root of unity. $\square$

Our next goal is to extend this corollary to the case of general finitely generated fields, including transcendental extensions of $\mathbb{Q}$, as well as fields of positive characteristic. The following theorem is Lemma 4.1 in [**Tit72**]:

THEOREM 1.111. *Let $\mathbb{E}$ be a finitely generated field and suppose that $\alpha \in \mathbb{E}$ is not a root of unity. Then there exists an extension $(\mathbb{F}, |\cdot|)$ of $\mathbb{E}$, which is a local field with the norm $|\cdot|$, such that $|\alpha| \neq 1$.*

PROOF. Let $\mathbb{P} \subset \mathbb{E}$ denote the prime subfield of $\mathbb{E}$. Since $\mathbb{E}$ is finitely generated over $\mathbb{P}$, there is a finite transcendence basis $T = \{t_1, \ldots, t_n\}$ for $\mathbb{E}$ over $\mathbb{E}$ and $\mathbb{E}$ is a finite extension of $\mathbb{P}(T) = \mathbb{P}(t_1, \ldots, t_n)$ (cf. [Chapter VI.1][**Hun80**]). Here $\mathbb{P}(t_1, \ldots, t_n)$ is isomorphic to the field of rational functions with coefficients in $\mathbb{P}$ and variables in T. We also let T' denote a (finite) transcendence basis of $\mathbb{E}$ over $\mathbb{P}(\alpha)$.

There are two main cases to consider.

Case 1: $\mathbb{E}$ has characteristic $p > 0$, equivalently, $\mathbb{P} \cong \mathbb{Z}_p$ for some p. If α were to be algebraic over $\mathbb{P}$, then $\mathbb{P}(\alpha)$ would be finite and, hence, $\alpha^i = \alpha^k$ for some $i \neq k$, implying that α is a root of unity. This is a contradiction. Therefore, α is transcendental over $\mathbb{P}$ and, hence, we can assume that α is an element of T. Define the ring $A = \mathbb{P}[T]$ and let $I \subset A$ denote the ideal generated by T. As we explained in the previous section, there is a (unique) valuation v on $\mathbb{P}(T)$ (with the norm $|\cdot|$) such that

$$v(a) = k \iff a \in I^k \setminus I^{k+1}.$$

By the construction, $v(\alpha) = 1$ and, hence, $|\alpha| \neq 1$. The completion of $\mathbb{P}(T)$ with respect to the norm $|\cdot|$ is a local field, since $\mathbb{P}$ is finite. We then extend the norm to a norm on $\mathbb{E}$; the completion with respect to this norm is again a local field.

Case 2: $\mathbb{E}$ has zero characteristic, equivalently, $\mathbb{P} \cong \mathbb{Q}$. Suppose, first, that α is a transcendental number. Then there exists an embedding

$$\mathbb{Q}(\alpha) \to \mathbb{C}$$

which sends α to a transcendental number whose absolute value is > 1, e.g., $\alpha \mapsto \pi$. This embedding extends to an embedding $\mathbb{E} \to \mathbb{C}$, thereby finishing the proof.

Suppose, therefore, that α is an algebraic number, $\alpha \in \mathbb{Q}$. Let $p(x)$ be the minimal monic polynomial of α.

Subcase 2a. Assume first that p has integer coefficients. Then, since α is not a root of unity, by Corollary 1.110, one of the roots β of p has absolute value > 1. Consider the Galois automorphism $\phi : \mathbb{Q}(\alpha) \to \mathbb{Q}(\alpha)$ sending α to β. We then extend ϕ to an embedding

$$\psi : \mathbb{E}(T' \cup \{\alpha\}) \to \mathbb{C},$$

by sending the elements of T to complex numbers which are algebraically independent over $\mathbb{Q}(\alpha)$. Lastly, since $\mathbb{E}(T' \cup \{\alpha\}) \subset \mathbb{E}$ is an algebraic extension and $\mathbb{C}$ is algebraically closed, the embedding ψ extends to the required embedding $\mathbb{E} \to \mathbb{C}$.

Subcase 2b. Lastly, we consider the case when $p \in \mathbb{Q}[x]$ has a non-integer coefficient. We consider the infinite cyclic subgroup generated by α in $\mathbb{Q}(\alpha)^\times$ and the embedding

$$\langle \alpha \rangle \to \mathbb{Q}(\alpha) \to \mathbb{A}_\alpha,$$

where $\mathbb{A}_\alpha$ in the ring of adeles $\mathbb{A}_\alpha$ of the field $\mathbb{Q}(\alpha)$; here $\mathbb{Q}(\alpha) \to \mathbb{A}_\alpha$ is the diagonal embedding. Since the subgroup $\mathbb{Q}(\alpha) < \mathbb{A}_\alpha$ is discrete and $\langle \alpha \rangle$ is an infinite subgroup, Tychonoff compactness theorem implies that the projection of $\langle \alpha \rangle$ to at least one of the factors of $\mathbb{A}_\alpha$ is unbounded. If this factor were archimedean, we would obtain a Galois embedding $\mathbb{Q}(\alpha) \to \mathbb{C}$ sending α to $\beta \in \mathbb{C}$ whose absolute value is different from 1. This situation is already handled in the Subcase 2a. Suppose, therefore, that there is a prime number p such that $\langle \alpha \rangle$ is an unbounded subgroup of the p-adic completion of $\mathbb{Q}(\alpha)$, which means that $|\alpha|_p \neq 1$, where $|\cdot|_p$ is the extension of the p-adic norm to $\mathbb{Q}(\alpha)$. Next, extending the norm $|\cdot|_p$ from $\mathbb{Q}(\alpha)$ to $\mathbb{E}$ and then taking the completion, we obtain an embedding to $\mathbb{E}$ to a local field, α has non-unit norm. □

1.14. Metrics on affine and projective spaces

In this section we will use normed fields to define metrics on affine and projective spaces. Consider the vector space $V = \mathbb{F}^n$ over a normed field $\mathbb{F}$, with the standard basis $e_1, \ldots, e_n$. We equip V with the usual Euclidean/hermitian norm in the case $\mathbb{F}$ is archimedean and with the max-norm

$$|(x_1, \ldots, x_n)| = \max_i |x_i|$$

if $\mathbb{F}$ is nonarchimedean. We let $\langle \cdot, \cdot \rangle$ denote the standard inner/hermitian product on V in the archimedean case.

Exercise 1.112. Suppose that $\mathbb{F}$ is nonarchimedean. Show that the metric $|v - w|$ on V satisfies the ultrametric triangle inequality.

If $\mathbb{F}$ is nonarchimedean, define the group $K = GL(n, O)$, consisting of matrices A such that $A, A^{-1} \in Mat_n(O)$.

EXERCISE 1.113. If $\mathbb{F}$ is a nonarchimedean local field, show that the group K is compact with respect to the subset topology induced from $Mat_n(\mathbb{F}) = \mathbb{F}^{n^2}$.

LEMMA 1.114. *The group K acts isometrically on V.*

PROOF. It suffices to show that elements $g \in K$ do not increase the norm on V. Let a_{ij} denote the matrix coefficients of g. Then, for a vector $v = \sum_i v_i e_i \in V$, the vector $w = g(v)$ has coordinates

$$w_j = \sum_i a_{ji} v_i.$$

Since $|a_{ij}| \leqslant 1$, the ultrametric inequality implies

$$|w| = \max_j |w_j|, \quad |w_j| \leqslant \max_i |a_{ji} v_i| \leqslant |v|.$$

Thus, $|g(v)| \leqslant |v|$. □

If $\mathbb{F}$ is archimedean, we let $K < GL(V)$ denote the orthogonal/hermitian subgroup preserving the inner/hermitian product on V. The following is a standard fact from the elementary linear algebra, which follows from the spectral theorem, see e.g. [**Str06**, §6.3]:

THEOREM 1.115 (Singular Value Decomposition Theorem). *If $\mathbb{F}$ is archimedean, then every matrix $M \in End(V)$ admits a* singular value decomposition

$$M = UDV,$$

where $U, V \in K$ and D is a diagonal matrix with nonnegative entries arranged in the descending order. The diagonal entries of D are called the singular values *of M.*

We will also need a (slightly less well-known) analogue of the singular value decomposition in the case of nonarchimedean normed fields, see e.g. [**DF66**, §12.2, Theorem 21]:

THEOREM 1.116 (Smith Normal Form Theorem). *Let $\mathbb{F}$ be a field with discrete norm and uniformizer π and ring of integers O. Then every matrix $M \in Mat_n(\mathbb{F})$ admits a* Smith Normal Form *decomposition*

$$M = LDU,$$

where D is diagonal with diagonal entries $(d_1, \ldots, d_n)$, $d_i = \pi^{k_i}$, $i = 1, \ldots, n$,

$$k_1 \geqslant k_2 \geqslant \ldots \geqslant k_n,$$

and $L, U \in K = GL(n, O)$. The diagonal entries $d_i \in \mathbb{F}$ are called the invariant factors *of M.*

PROOF. First, note that permutation matrices belong to K; the group K also contains upper and lower triangular matrices with coefficients in O, whose diagonal entries are units in $\mathbb{F}$. We then apply Gauss Elimination Algorithm to the matrix M. Note that the row operation of adding the z-multiple of the i-th row to the j-th row amounts to multiplication on the left with the lower-triangular elementary matrix $E_{ij}(z)$ with the ij-entry equal z. If $z \in O$, then $E_{ij} \in K$. Similarly, column operations amount to multiplication on the right by an upper-triangular elementary matrix. Observe also that dividing a row (column) by a unit in $\mathbb{F}$ amounts to multiplying a matrix on left (right) by an appropriate diagonal matrix with unit entries on the diagonal.

We now describe row operations for the Gauss Elimination in detail (column operations will be similar). Consider (nonzero) i-th column of a matrix $A \in End(\mathbb{F}^n)$. We first multiply M on left and right by permutation matrices so that a_{ii} has the largest norm in the i-th column. By dividing rows on A by units in $\mathbb{F}$, we achieve that every entry in the i-th column is a power of π. Now, eliminating nonzero entries in the i-th column will require only row operations involving $\pi^{s_{ij}}$-multiples of the i-th row, where $s_{ij} \geqslant 0$, i.e., $\pi^{s_{ij}} \in O$. Applying this form of Gauss Algorithm to M, we convert M to a diagonal matrix A, whose diagonal entries are powers of π and

$$A = L'MU', \quad L', M' \in GL(n, O).$$

Multiplying A on left and right by permutation matrices, we rearrange the diagonal entries to have weakly decreasing exponents. □

Note that both singular value decomposition and Smith normal form decomposition both have the form:

$$M = UDV, \quad U, V \in K,$$

and D is diagonal. Such decomposition of the $Mat_n(\mathbb{F})$ is called the *Cartan decomposition.* To simplify the terminology, we will refer to the diagonal entries of D as *singular values* of M in both archimedean and nonarchimedean cases.

EXERCISE 1.117. Deduce the Cartan decomposition in $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, from the statement that given any Euclidean/hermitian bilinear form q on $V = \mathbb{F}^n$, there exists a basis orthogonal with respect to q and orthonormal with respect to the standard inner product

$$x_1\overline{y}_1 + \ldots + x_n\overline{y}_n.$$

We now turn our discussion to projective spaces. The $\mathbb{F}$-projective space $P = \mathbb{F}P^{n-1}$ is the quotient of $\mathbb{F}^n \setminus \{0\}$ by the action of $\mathbb{F}^\times$ *via* scalar multiplication.

NOTATION 1.118. Given a nonzero vector $v \in V$ let $[v]$ denote the projection of v to the projective space $P(V)$; similarly, for a subset $W \subset V$ we let $[W]$ denote the image of $W \setminus \{0\}$ under the canonical projection $V \to P(V)$. Given an invertible linear map $g : V \to V$, we will retain the notation g for the induced projective map $P(V) \to P(V)$.

Suppose now that $\mathbb{F}$ is a normed field. Our next goal is to define the *chordal metric* on $P(V) = \mathbb{F}P^{n-1}$. In the case of an archimedean field $\mathbb{F}$, we define the Euclidean or hermitian norm on $V \wedge V$ by declaring basis vectors

$$e_i \wedge e_j, 1 \leqslant i < j \leqslant n$$

to be orthonormal. Then

$$|v \wedge w|^2 = |v|^2|w|^2 - \langle v, w\rangle \langle w, v\rangle = |u| \cdot |v| \cdot |\sin(\varphi)|$$

where $\varphi = \angle(v, w)$. In other words, this is the area of the parallelogram spanned by the vectors u and v.

In the case when $\mathbb{F}$ is nonarchimedean, we equip $V \wedge V$ with the max-norm so that

$$|v \wedge w| = \max_{i,j} |x_i y_j - x_j y_i|$$

where $v = (x_1, \ldots, x_n)$, $w = (y_1, \ldots, y_n)$.

DEFINITION 1.119. The *chordal metric* on $P(V)$ is defined by

$$d([v],[w]) = \frac{|v \wedge w|}{|v| \cdot |w|}.$$

In the nonarchimedean case this definition is due to A. Néron **[Nér64]**.

EXERCISE 1.120. 1. If $\mathbb{F}$ is nonarchimedean, show that the group $GL(n, O)$ preserves the chordal metric.

2. If $\mathbb{F} = \mathbb{R}$, show that the orthogonal group preserves the chordal metric.

3. If $\mathbb{F} = \mathbb{C}$, show that the unitary group preserves the chordal metric.

4. Show that each $g \in GL(n, \mathbb{K})$ is a Lispchitz homeomorphism with respect to the chordal metric.

It is clear that $d(\lambda v, \mu w) = d(v, w)$ for all nonzero scalars λ, μ and nonzero vectors v, w. It is also clear that $d(v, w) = d(w, v)$ and $d(v, w) = 0$ if and only if $[v] = [w]$. What is not so obvious is why d satisfies the triangle inequality. Note, however, that in the case of a nonarchimedean field $\mathbb{F}$,

$$d([v],[w]) \leqslant 1$$

for all $[v], [w] \in P$. Indeed, pick unit vectors v, w representing $[v], [w]$; in particular, v_i, w_j belong to O for all i, j. Then, the denominator in the definition of $d([v],[w])$ equals 1, while the numerator is $\leqslant 1$, since O is a ring.

PROPOSITION 1.121. *If $\mathbb{F}$ is nonarchimedean, then d satisfies the triangle inequality.*

PROOF. We will verify the triangle inequality by giving an alternative description of the function d. We define *affine patches* on P to be the affine hyperplanes

$$A_j = \{x \in V : x_j = 1\} \subset V$$

together with the (injective) projections $A_j \to P$. Every affine patch is, of course, just a translate of $\mathbb{F}^{n-1}$, so that e_j is the translate of the origin. We, then, equip A_j with the restriction of the metric $|v - w|$ from V. Let $B_j \subset A_j$ denote the closed unit ball centered at e_j. In other words,

$$B_j = A_j \cap O^{n+1}.$$

We now set $d_j(x, y) = |x - y|$ if $x, y \in B_j$ and $d_j(x, y) = 1$ otherwise. It follows immediately from the ultrametric triangle inequality that d_j is a metric. Define for $[x], [y] \in P$ the function $\mathrm{dist}([x],[y])$ by:

1. If there exists j so that $x, y \in B_j$ project to $[x], [y]$, then $\mathrm{dist}([x],[y]) := d_j(x, y)$.

2. Otherwise, set $\mathrm{dist}([x],[y]) = 1$.

If we knew that dist is well-defined (*a priori*, different indices j give different values of dist), it would be clear that dist satisfies the ultrametric triangle inequality. Proposition will, now, follow from:

LEMMA 1.122. *$d([x],[y]) = \mathrm{dist}([x],[y])$ for all points in P.*

Proof. The proof will break in two cases:

1. There exists k such that $[x],[y]$ lift to $x,y\in B_k$. To simplify the notation, we will assume that $k=n+1$. Since $x,y\in B_{n+1}$, $|x_i|\leqslant 1, |y_i|\leqslant 1$ for all i, and $x_{n+1}=y_{n+1}=1$. In particular, $|x|=|y|=1$. Hence, for every i,

$$|x_i-y_i|=|x_iy_{n+1}-x_jy_{n+1}|\leqslant \max_j |x_iy_j-x_jy_i|\leqslant d([x],[y]),$$

which implies that

$$\operatorname{dist}([x],[y])\leqslant d([x],[y]).$$

We will now prove the opposite inequality:

$$\forall i,j \quad |x_iy_j-x_jy_i|\leqslant a:=|x-y|.$$

There exist $z_i,z_j\in\mathbb{F}$ so that

$$y_i=x_i(1+z_i),\quad y_j=x_j(1+z_j),$$

where, if $x_i\neq 0, x_j\neq 0$,

$$z_i=\frac{y_i-x_i}{x_i},\quad z_j=\frac{y_j-x_j}{x_j}.$$

We will consider the case $x_ix_j\neq 0$, leaving the exceptional cases to the reader. Then,

$$|z_i|\leqslant\frac{a}{|x_i|},\quad |z_i|\leqslant\frac{a}{|x_j|}.$$

Computing $x_iy_j-x_jy_i$ using the new variables z_i,z_j, we obtain:

$$|x_iy_j-x_jy_i|=|x_ix_j(1+z_j)-x_ix_j(1+z_i)|=|x_jx_j(z_j-z_i)|\leqslant$$

$$|x_ix_j|\max\left(|z_i|,|z_i|\right)\leqslant |x_ix_j|\max\left(\frac{a}{|x_i|},\frac{a}{|x_j|}\right)\leqslant a\max\left(|x_i|,|x_j|\right)\leqslant a,$$

since $x_i,x_j\in O$.

2. Suppose that (1) does not happen. Since $d([x],[y])\leqslant 1$ and $\operatorname{dist}([x],[y])=1$ (in the second case), we just have to prove that

$$d([x],[y])\geqslant 1.$$

Consider representatives x,y of points $[x],[y]$ and let i,j be the indices such that

$$|x_i|=|x|,\quad |y_j|=|y|.$$

Clearly, i,j are independent of the choices of the vectors x,y representing $[x],[y]$. Therefore, we choose x so that $x_i=1$, which implies that $x_k\in O$ for all k. If $y_i=0$ then

$$|x_iy_j-x_jy_i|=|y_j|$$

and

$$d([x],[y])\geqslant\frac{\max_j|1\cdot y_j|}{|y_j|}=1.$$

Thus, we assume that $y_i\neq 0$. This allows us to choose $y\in A_i$ as well. Since (1) does not occur, $y\notin O^{n+1}$, which implies that $|y_j|>1$. Now,

$$d([x],[y])\geqslant\frac{|x_iy_j-x_jy_i|}{|x_i|\cdot|y_j|}=\frac{|y_j-x_j|}{|y_j|}.$$

Since $x_j \in O$ and $y_j \notin O$, the ultrametric inequality implies that $|y_j - x_j| = |y_j|$. Therefore,

$$\frac{|y_j - x_j|}{|y_j|} = \frac{|y_j|}{|y_j|} = 1$$

and $d([x],[y]) \geqslant 1$. This concludes the proof of lemma and proposition. $\square$

COROLLARY 1.123. *If $\mathbb{K}$ is nonarchimedean, then the metric d on P is locally isometric to the metric $|x-y|$ on the affine space $\mathbb{F}^{n-1}$.*

We now consider real and complex projective spaces. Choosing unit vectors u, v as representatives of points $[u],[v] \in P$, we get:

$$d([u],[v]) = \sin(\angle(u,v)),$$

where we normalize the angle to be in the interval $[0,\pi]$. Consider now three points $[u],[v],[w] \in P$; our goal is to verify the triangle inequality

$$d([u],[w]) \leqslant d([u],[v]) + d([v],[w]).$$

We choose unit vectors u, v, w representing these points so that

$$0 \leqslant \alpha = \angle(u,v) \leqslant \frac{\pi}{2}, \quad 0 \leqslant \beta = \angle(v,w) \leqslant \frac{\pi}{2}.$$

Then,

$$\gamma = \angle(u,w) \leqslant \alpha + \beta$$

and the triangle inequality for the metric d is equivalent to the inequality

$$\sin(\gamma) \leqslant \sin(\alpha) + \sin(\beta).$$

We leave verification of the last inequality as an exercise to the reader. Thus, we obtain

THEOREM 1.124. *The chordal metric is a metric on P in both archimedean and nonarchimedean cases.*

EXERCISE 1.125. Suppose that $\mathbb{F}$ is a normed field (either nonarchimedean or archimedean).

1. Verify that metric d determines the topology on P which is the quotient topology induced from $V \setminus \{0\}$.
2. Assuming that $\mathbb{F}$ is local, verify that P is compact.
3. If the norm on $\mathbb{F}$ is complete, show that the metric space (P,d) is complete.
4. If H is a hyperplane in $V = \mathbb{F}^n$, given as $\operatorname{Ker} f$, where $f : V \to \mathbb{F}$ is a linear function, show that

$$\operatorname{dist}([v],[H]) = \frac{|f(v)|}{\|v\| \, \|f\|}.$$

1.15. Quasiprojective transformations. Proximal transformations

In what follows, V is a finite-dimensional vector space of dimension n, over a local field $\mathbb{F}$. W Each automorphism $g \in GL(V)$ of the vector space V projects to a projective transformation $g \in PGL(V)$, $g : P(V) \to P(V)$. Given g, we will always extend the norm from $\mathbb{F}$ to the splitting field $\mathbb{E}$ of the characteristic polynomial of g, in order to define norms of eigenvalues of g.

On the other hand, endomorphisms of V (i.e., linear maps $V \to V$) do not project, in general, to self-maps $P(V) \to P(V)$. Nevertheless, if $g \in End(V)$ is a linear transformation of rank $r > 0$ with kernel $\operatorname{Ker}(g)$ and image $\operatorname{Im}(g)$, then

g determines a *quasiprojective transformation* g of $P(V)$, whose domain dom_g is the complement of $P(\mathrm{Ker}(g))$ and whose image is $\mathrm{Im}_g := P(\mathrm{Im}(g))$. The number $r = \mathrm{rank}\,(g)$ is called the rank of this quasiprojective transformation. The subspace $\mathrm{Ker}_g := P(\mathrm{Ker}(g))$ is the *kernel*, or the *indeterminacy set* of g. We let $End(P(V))$ denote the semigroup of quasiprojective transformations of $P(V)$. Rank 1 quasiprojective transformations are *quasiconstant maps*: Each quasiconstant map is undefined on a hyperplane in $P(V)$ and its image is a single point.

EXERCISE 1.126. For $h \in GL(V)$ and $g \in End(V)$ we have:

$$\mathrm{Im}_{hg} = h(\mathrm{Im}_g), \quad \mathrm{Ker}_{hg} = \mathrm{Ker}_g,$$

$$\mathrm{Im}_{gh} = \mathrm{Im}_g, \quad \mathrm{Ker}_{gh} = h^{-1}(\mathrm{Ker}_g).$$

Rank of a quasiprojective transformation can be detected locally:

EXERCISE 1.127. Suppose that $g \in End(P(V))$ and $U \subset \mathrm{dom}_g \subset P(V)$ is a nonempty open subset. Then

$$\mathrm{rank}\,(g) = \dim(g(U)) + 1.$$

We will topologize $End(P(V))$ using the operator norm topology on $End(V)$.

EXERCISE 1.128. 1. rank is a lower semicontinuous function on $End(P(V))$.

2. A sequence g_i converges to g in $End(P(V))$ if and only if it converges to g uniformly on compacts in dom_g. (In particular, each compact $C \subset \mathrm{dom}_g$ is contained in dom_{g_i} for all but finitely many i's.)

3. Suppose that $g \in End(V)$ is such that the dominant eigenvalue λ_1 of g satisfies $|\lambda_1| < 1$. Show that the sequence $g^k \in End(V)$ converges to $0 \in End(V)$. Hint: Show that

$$||g^k|| \leqslant |\lambda_1|^k p(k),$$

where $p(k)$ is a polynomial in k of degree $\leqslant n-1$.

THEOREM 1.129 (A convergence property). *The semigroup $End(P(V))$ is compact: Each sequence $g_i \in End(P(V))$ subconverges to a quasiprojective transformation.*

PROOF. We will use the Cartan decomposition $End(V) = K \cdot Diag(V) \cdot K$: Each $g \in End(V)$ has the form $g = k_g a_g, k'_g$, where k_g, k'_g belong to the compact subgroup $K < GL(V)$ and a_g is a diagonal transformation whose diagonal entries are the singular values of g. Assuming $g \neq 0$, $a_g \neq 0$ as well and, by replacing g with its scalar multiple (which does not affect the corresponding quasiprojective endomorphism), we can assume that the dominant eigenvalue of a_g equals 1, i.e. $||a_g|| = 1$. We apply this to elements of a sequence $g_i \in End(P(V))$ and obtain:

$$g_i = k_{g_i} a_{g_i} k'_{g_i}, \quad ||a_{g_i}|| = 1.$$

Since $\mathbb{F}$ is a locally compact field, the sequences $k_{g_i}, a_{g_i}, k'_{g_i}$ subconverge in $End(V)$: The limits k, k' of convergent subsequences of k_{g_i}, k'_{g_i} belong to the group K, while a_i subconverges to an endomorphism a of the unit norm, in particular, this limit is different from 0. Thus, the sequence (g_i) subconverges to the nonzero endomorphism kak'. $\square$

LEMMA 1.130. *Suppose that* (g_i) *is a sequence in* $GL(V)$ *converging to a quasiconstant map* $\hat{g}$. *Then*

$$\lim_{i\to\infty} Lip(g_i) = 0$$

uniformly on compacts in $\dim(\hat{g})$. *In other words, for every compact*

$$C \subset \mathrm{dom}_{\hat{g}}$$

we have

$$\sup_{x,y\in C, x\neq y} \frac{d(g_i(x), g_i(y))}{d(x,y)} = 0$$

PROOF. In view of the Cartan decomposition, it suffices to analyze the case when each g_i is a diagonal matrix with the diagonal entries $\lambda_1 = 1, \lambda_2, \ldots, \lambda_n$, satisfying

$$1 \geqslant |\lambda_{2,i}| \geqslant \ldots \geqslant |\lambda_{n,i}|$$

such that

$$\lim_{i\to\infty} \lambda_{2,i} = 0.$$

In particular, the maps g_i preserve the affine hyperplane

$$A_1 = \{(1, x_2, \ldots, x_n) : x_2, \ldots, x_n \in \mathbb{F}\}$$

in V. We will identify A_1 with

$$\mathrm{dom}_{\hat{g}} \subset P(V)$$

and, accordingly, lift C to a compact subset (again denoted by C) in A_1.

We first consider the case of a nonarchimedean field $\mathbb{F}$. We will use the action of g_i on A_1 in order to analyze the contraction properties of g_i. Since the sequence g_i restricted to C converges uniformly to e_1, for all sufficiently large i, $g_i(C)$ is contained in the unit ball centered at e_1. In view of Lemma 1.122, it is clear that the maps g_i do not increase distances between points in A_1 (measured in the metric d_1 on A_1). Furthermore, for all $x, y \in B(e_1, 1) \subset A_1$, we have

$$d_1(g_i(x), g_i(y)) \leqslant \lambda_{2,i} d_1(x,y) \tag{1.14}$$

and, hence, the Lipschitz constant of g_i converges to zero.

Consider now the archimedean case. As in the nonarchimedean case, we let d_1 denote the restriction to A_1 of the metric defined via the maximum-norm on V. We leave it to the reader to check the inequalities:

$$D^2 d_1(x,y)\frac{|x-y|}{|x|\ |y|} \leqslant d([x],[y]) \leqslant n\|x-y\|_{max} = n d_1(x,y)$$

for all points $x, y \in A_1$ satisfying $\max(|x|, |y|) \leqslant D$. This shows that the map $(A_1, d_1) \to (P(V), d)$ is uniformly bilipschitz on each compact in A_1. On the other hand, the map

$$g_i : (A_1, d_1) \to (A_1, d_1)$$

satisfies the inequality (1.14) for all $x, y \in A_1$. Lemma follows. □

REMARK 1.131. It is useful to note here that while singular values depend on the choice of a basis in V, the limit quasiprojective transformation of the sequence (g_i) is, of course, independent of the basis. The same applies to the notion of proximality below.

The most important, for us, example of convergence to a quasiprojective transformation comes from iterations of a single invertible transformation: $g_i = g^i$, $i \in \mathbb{N}$. For $g \in End(V)$ we say that an eigenvalue λ_1 of g is *dominant* if it has algebraic multiplicity one and

$$|\lambda_1| > |\lambda_k|$$

for all eigenvalues λ_k of g different from λ_1.

DEFINITION 1.132. An endomorphism g is called *proximal* if it has a dominant eigenvalue; an automorphism $g \in GL(V)$ is *very proximal* if both g and g^{-1} are proximal elements of $GL(V)$.

For a proximal endomorphism g we let $\tilde{A}_g \subset V$ denote the (one-dimensional) eigenspace corresponding to the dominant eigenvalue λ_1 and let $\tilde{E}_g \subset V$ denote the sum of the rest of the generalized eigenspaces of g. We project $\tilde{A}_g$ and $\tilde{E}_g$, respectively, to a point A_g and a hyperplane E_g in the projective space $P(V)$. We will refer to A_g as the *attractive point* and E_g the *exceptional hyperplane* for the action of a proximal projective transformation g on $\mathbb{P}(V)$. (The reason for the terminology will become clear from the next lemma.)

It is clear that proximality depends only on the projectivization of g.

We now work out limits of sequences $g^i \in End(P(V))$, when g is proximal. We already know that the sequence (g^i) of projective transformations subconverges to a quasiprojective transformation, the issue is to compute the rank, range and the kernel of the limit.

LEMMA 1.133. *If $g \in End(V)$ is a proximal endomorphism of $P(V)$, then each convergent subsequence in the sequence (g^k) of projective transformations converges to a rank 1 (quasiconstant) quasiprojective transformation $\hat{g}$. The image $\mathrm{Im}_{\hat{g}}$ of $\hat{g}$ equals $A_g := P(\tilde{A}_g)$ and the kernel $\mathrm{Ker}_{\hat{g}}$ of $\hat{g}$ equals $E_g := P(\tilde{E}_g)$.*

PROOF. We normalize g so that $\lambda_1 = 1$; hence, all eigevalues of g restricted to $\tilde{E}_g$ have absolute value < 1. Clearly, the restriction of g to $\tilde{A}_g$ is the identity, while, by Part 3 of Exercise 1.128, the restriction of the sequence g^i to $\tilde{E}_g$ converges to the zero linear map. Lemma follows. $\square$

COROLLARY 1.134. *Given a proximal endomorphism $g \in End(P(V))$, for every for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that for every $i \geqslant N$, the projective transformation $g^i \in End(P(V))$ maps the complement of the ε-neighborhood of the hyperplane $E_g \subset P(V)$ inside the ball*

$$B(A_g, \varepsilon)$$

of radius ε and center A_g.

We will be using quasiprojective transformations and proximal elements of $GL(V)$ in the proof of the Tits' Alternative, Section 13.4.

1.16. Kernels and distance functions

A *kernel* on a set X is a symmetric map $\psi : X \times X \to \mathbb{R}_+$ such that $\psi(x, x) = 0$. (Symmetry of ψ means that $\psi(x, y) = \psi(y, x)$ for all x, y in X.) Fix $p \in X$ and define the associated *Gromov kernel*

$$k_p(x, y) := \frac{1}{2} \left(\psi(x, p) + \psi(p, y) - \psi(x, y) \right),$$

cf. section 9.3 for the definition of the *Gromov product* in metric spaces. Clearly,

$$\forall x \in X, \quad k_p(x,x) = \psi(x,p).$$

DEFINITION 1.135. 1. A kernel ψ is *positive semidefinite* if for every natural number n, every subset $\{x_1, \ldots, x_n\} \subset X$ and every vector $\lambda \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \geqslant 0\,. \tag{1.15}$$

2. A kernel ψ is *conditionally negative semidefinite* if for every $n \in \mathbb{N}$, every subset $\{x_1, \ldots, x_n\} \subset X$ and every vector $\lambda \in \mathbb{R}^n$ with $\sum_{i=1}^n \lambda_i = 0$, the following holds:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \leqslant 0\,. \tag{1.16}$$

This is not a particularly transparent definition. A better way to think about this definition is in terms of the vector space $V = V(X)$ of consisting of functions with finite support $X \to \mathbb{R}$. Then each kernel ψ on X defines a symmetric bilinear form on V (denoted Ψ):

$$\Psi(f,g) = \sum_{x,y \in X} \psi(x,y) f(x) g(y).$$

With this notation, the left hand side of (1.15) becomes simply $\Psi(f,f)$, where

$$\lambda_i := f(x_i), \quad \mathrm{Supp}(f) \subset \{x_1, \ldots, x_n\} \subset X.$$

Thus, a kernel is positive semidefinite if and only if Ψ is a positive semidefinite bilinear form. Similarly, ψ is conditionally negative semidefinite if and only if the restriction of $-\Psi$ to the subspace V_0 consisting of functions with zero average, is a positive semidefinite bilinear form.

NOTATION 1.136. We will use the lower case letters to denote kernels and the corresponding upper case letters to denote the associated bilinear forms on V.

Below is yet another interpretation of the conditionally negative semidefinite kernels. For a subset $\{x_1, \ldots, x_n\} \subset X$ define the symmetric matrix M with the entries

$$m_{ij} = -\psi(x_i, x_j), \quad 1 \leqslant i, j \leqslant n.$$

For $\lambda = (\lambda_1, \ldots, \lambda_n)$, the left hand-side of the inequality (1.16) equals

$$q(\lambda) = \lambda^T M \lambda,$$

a symmetric bilinear form on $\mathbb{R}^n$. Then, the condition (1.16) means that q is positive semi-definite on the hyperplane

$$\sum_{i=1}^{n} \lambda_i = 0$$

in $\mathbb{R}^n$. Suppose, for a moment, that this form is actually positive-definite, Since $\psi(x_i, x_j) \geqslant 0$, it follows that the form q on $\mathbb{R}^n$ has signature $(n-1, 1)$. The standard basis vectors $e_1, \ldots, e_n$ in $\mathbb{R}^n$ are null-vectors for q; the condition $m_{ij} \leqslant 0$ amounts to the requirement that these vectors belong to the same, say, positive, light cone.

The following theorem gives yet another interpretation of conditionally negative semidefinite kernels in terms of embedding in Hilbert spaces. It was first proven by J. Schoenberg in [**Sch38**] in the case of finite sets, but the same proof works for infinite sets as well.

THEOREM 1.137. *A kernel ψ on X is conditionally negative semidefinite if and only if there exists a map $F : X \to \mathcal{H}$ to a Hilbert space such that*

$$\psi(x, y) = \|F(x) - F(y)\|^2.$$

Here $\|\cdot\|$ denotes the norm on $\mathcal{H}$. Furthermore, if G is a group acting on X preserving the kernel ψ then the map F is equivariant with respect to a homomorphism $G \to \mathrm{Isom}(\mathcal{H})$.

PROOF. 1. Suppose that the map F exists. Then, for every $p = x_0 \in X$, the associated Gromov kernel $k_p(x, y)$ equals

$$k_p(x, y) = \langle F(x), F(y) \rangle ,$$

and, hence, for every finite subset $\{x_0, x_1, \ldots, x_n\} \subset X$, the corresponding matrix with the entries $k_p(x_i, x_j)$ is the Gramm matrix of the set

$$\{y_i := F(x_i) - F(x_0) : i = 1, \ldots, n\} \subset \mathcal{H}.$$

Hence, this matrix is positive semidefinite. Accordingly, Gromov kernel determines a positive semidefinite bilinear form on the vector space $V = V(X)$.

We will verify that ψ is conditionally negative semidefinite by considering subsets X_0 in X of the form $\{x_0, x_1, \ldots, x_n\}$. (Since the point x_0 was arbitrary, this will suffice.)

Let $f : X_0 \to \mathbb{R}$ be such that

$$\sum_{i=0}^{n} f(x_i) = 0. \tag{1.17}$$

Thus,

$$f(x_0) := -\sum_{i=1}^{n} f(x_i).$$

Set $y_i := F(x_i), i = 0, \ldots, n$. Since the kernel K is positive semidefinite, we have

$$\sum_{i,j=1}^{n} \left(|y_0 - y_i|^2 + |y_0 - y_j|^2 - |y_i - y_j|^2\right) f(x_i) f(x_j) = \tag{1.18}$$

$$2 \sum_{i,j=1}^{n} k_p(x_i, x_j) f(x_i) f(x_j) \geqslant 0.$$

The left hand side of this equation equals

$$2 \left(\sum_{i=1}^{n} f(x_i) \right) \cdot \left(\sum_{j=1}^{n} |y_0 - y_j|^2 f(x_j) \right) -$$

$$\sum_{i,j=1}^{n} |y_i - y_j|^2 f(x_i) f(x_j).$$

Since $f(x_0) := -\sum_{i=1}^{n} f(x_i)$, we can rewrite this expression as

$$-f(x_0)^2|y_0 - y_0|^2 - 2\left(\sum_{j=1}^{n} |y_0 - y_j|^2 f(x_0)f(x_j)\right) - \sum_{i,j=1}^{n} |y_i - y_j|^2 f(x_i)f(x_j) =$$

$$\sum_{i,j=0}^{n} |y_i - y_j|^2 f(x_i)f(x_j) = \sum_{i,j=0}^{n} \psi(x_i, x_j)f(x_i)f(x_j).$$

Taking into account the inequality (1.18), we conclude that

$$\sum_{i,j=0}^{n} \psi(x_i, x_j)f(x_i)f(x_j) \leqslant 0. \tag{1.19}$$

In other words, the kernel ψ on X is conditionally negative semidefinite.

2. Suppose that ψ is conditionally negative semidefinite. Fix $p \in X$ and define the Gromov kernel

$$k_p(x, y) := \frac{1}{2}\left(\psi(x, p) + \psi(p, y) - \psi(x, y)\right).$$

The key to the proof is:

LEMMA 1.138. *$k_p(x, y)$ is a positive semidefinite kernel on X.*

PROOF. Consider a subset $X_0 = \{x_1, \ldots, x_n\} \subset X$ and a function $f : X_0 \to \mathbb{R}$.

a. We first consider the case when $p \notin X_0$. Then we set $x_0 := p$ and extend the function f to p by

$$f(x_0) := -\sum_{i=1}^{n} f(x_i).$$

The resulting function $f : \{x_0, \ldots, x_n\} \to \mathbb{R}$ satisfies (1.17) and, hence,

$$\sum_{i,j=0}^{n} \psi(x_i, x_j)f(x_i)f(x_j) \leqslant 0.$$

The same argument as in the first part of the proof of Theorem 1.137 (run in the reverse) then shows that

$$\sum_{i,j=1}^{n} k_p(x_i, x_j)f(x_i)f(x_j) \geqslant 0.$$

Thus, k_p is positive semidefinite on functions whose support is disjoint from $\{p\}$.

b. Suppose that $p \in X_0$, $f(p) = c \neq 0$. We define a new function $g(x) := f(x) - c\delta_p$. Here δ_p is the characteristic function of the subset $\{p\} \subset X$. Then $p \notin \mathrm{Supp}(g)$ and, hence, by the Case (a),

$$K_p(g, g) \geqslant 0.$$

On the other hand,

$$K_p(f, f) = F(g, g) + 2cK(g, \delta_p) + c^2 K(\delta_p, \delta_p) = F(g, g),$$

since the other two terms vanish (as $k_p(x, p) = 0$ for every $x \in X$). Thus, K_p is positive semidefinite. ☐

Now, consider the vector space $V = V(X)$ equipped with the positive semidefinite bilinear form $\langle f, g\rangle = K(f, g)$. Define the Hilbert space $\mathcal{H}$ as the metric completion of

$$V/\{f \in V : \langle f, f\rangle = 0\}.$$

Then we have a natural map $F : X \to \mathcal{H}$ which sends $x \in X$ to the projection of the δ-function δ_x (the indicator function $\mathbf{1}_x$); we obtain:

$$\langle F(x), F(y)\rangle = k_p(x, y).$$

Let us verify now that

$$\langle F(x) - F(y), F(x) - F(y)\rangle = \psi(x, y). \tag{1.20}$$

The left hand side of this expression equals

$$\langle F(x), F(x)\rangle + \langle F(y), F(y)\rangle - 2k_p(x, y) = \psi(x, p) + \psi(y, p) - 2k_p(x, y).$$

Then, the equality (1.20) follows from the definition of the Gromov kernel k. The part of the theorem dealing with G-invariant kernels is clear from the construction. □

Below we list several elementary properties of positive semidefinite and conditionally negative semidefinite kernels.

LEMMA 1.139. *Each kernel of the form $\psi(x, y) = f(x)f(y)$ is positive semidefinite.*

PROOF. This follows from the computation:

$$\sum_{i,j=1}^{n} \lambda_i\lambda_j\psi(x_i, x_j) = \left(\sum_{i=1}^{n} \lambda_i f(x_i)\right)\left(\sum_{j=1}^{n} \lambda_j f(x_j)\right) = \left(\sum_{i=1}^{n} \lambda_i f(x_i)\right)^2 \geqslant 0. \quad \square$$

Before proving the next lemma we will need the notion of *Hadamard product* of two matrices: If $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times m$ matrices, then their *Hadamard product*, denoted by $A \circ B$ is the matrix with the entried $(a_{ij}b_{ij})$. The fact that we will need about Hadamard product is known as the *Schur Product Theorem*:

- If A, B are positive semidefinite $n \times n$ matrices, then their entrywise product (the *Hadamard product*) $A \circ B$ is again positive semidefinite.

A proof of this theorem reduces to two calculations: For each (row) vector $v \in \mathbb{R}^n$

$$v^T (A \circ B)v = Tr(A\,\mathrm{diag}(v)B\,\mathrm{diag}(v))$$

(where $\mathrm{diag}(v)$ is the diagonal matrix with the diagonal entries equal to $v_i, i = 1, \ldots, n$). Then for the matrix $M = B^{1/2}\,\mathrm{diag}(v)A^{1/2}$ (note that square roots exist since A and B are positive semidefinite) we have:

$$v^T (A \circ B)v = Tr(A^{1/2}A^{1/2}\,\mathrm{diag}(v)B^{1/2}B^{1/2}\,\mathrm{diag}(v)) =$$

$$Tr(A^{1/2}\,\mathrm{diag}(v)B^{1/2}B^{1/2}\,\mathrm{diag}(v)A^{1/2}) = Tr(M^T M) \geqslant 0.$$

LEMMA 1.140. *Sums and products of positive semidefinite kernels are again positive semidefinite. The set of positive semidefinite kernels is closed in the space of all kernels with respect to the topology of pointwise convergence.*

PROOF. The only statement which is not immediate from the definitions is that product of positive semidefinite kernels $\theta(x,y) = \varphi(x,y)\psi(x,y)$ is again positive semidefinite. In order to prove so it suffices to consider the case $x, y \in X = \{x_1, \ldots, x_n\}$. Let $A = (a_{ij})$, $a_{ij} = \varphi(x_i, x_j)$, and $B = (b_{ij})$, $b_{ij} = \psi(x_i, x_j)$ denote the Gramm matrices of the kernels φ and ψ. Then the product kernel θ is given by the matrix

$$C = A \circ B.$$

Since A and B are positive semidefinite, so is C and, hence, θ. $\square$

COROLLARY 1.141. *If a kernel $\psi(x,y)$ is positive semidefinite, so is the kernel $\exp(\psi(x,y))$.*

PROOF. This follows from the previous lemma since

$$\exp(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!}. \quad \square.$$

THEOREM 1.142 (J. Schoenberg, [**Sch38**]). *If $\psi(x,y)$ is a conditionally negative semidefinite kernel then for each $s > 0$ the function $\varphi(x,y) = \exp(-s\psi(x,y))$ is a positive semidefinite kernel.*

PROOF. If X is empty, there is nothing to prove, therefore, fix $p \in X$ and consider the kernel

$$k(x,y) = \psi(x,p) + \psi(y,p) - \psi(x,y)$$

(twice the Gromov kernel). This kernel is positive semidefinite according to Lemma 1.138. We have:

$$\exp(-s\psi(x,y)) = \exp(s\ k(x,y)) \exp(-s\psi(x,p)) \exp(-s\psi(y,p)).$$

The first term on the right hand side is positive semidefinite according to Corollary 1.141. The product of the other two terms is positive semidefinite by Lemma 1.139. Therefore, Lemma 1.140 implies that $\varphi(x,y)$ is positive semidefinite. $\square$

Note that Schoenberg uses Theorem 1.142 to prove in [**Sch38**] the following neat result: For every conditionally negative semidefinite kernel $\psi : X \times X \to \mathbb{R}_+$ and every $0 < \alpha \leqslant 1$, the power ψ^α is also a conditionally negative semidefinite kernel. In other words, if a metric space (X, dist) embeds isometrically into a Hilbert space, so does every metric space

$$(X, \mathrm{dist}^\alpha), \quad 0 < \alpha \leqslant 1.$$

EXERCISE 1.143. Use this theorem of Schoenberg (and other results from this section) to prove the following. Suppose that (Z, μ) is a space with a measure μ. Then for each $p \in [1,2]$, the Banach space $L^p(Z, \mu)$ embeds linearly and isometrically into a Hilbert space. Furthermore, if $G \curvearrowright (Z, \mu)$ is a group preserving the measure μ, then such isometric embedding can be chosen to be equivariant with respect to a representation $G \to \mathrm{Isom}(\mathcal{H})$.

CHAPTER 2

Differential geometry

In this book we will use some elementary Differential and Riemannian geometry, basics of which are reviewed in this chapter. All the manifolds that we consider are second countable.

2.1. Smooth manifolds

We expect the reader to know basics of differential topology, that can be found, for instance, in [**GP10**], [**Hir76**], [**War83**]. Below is only a brief review.

Unless stated otherwise, all maps between smooth manifolds, vector fields and differential forms are assumed to be infinitely differentiable.

We will use the notation$\Lambda^k(M)$ for the space of differential k-forms on M. Every vector field X on M defines the *contraction operator*

$$i_X : \Lambda^{\ell+1}(M) \to \Lambda^{\ell}(M), \quad i_X(\omega)(X_1, \ldots, X_\ell) = \omega(X, X_1, \ldots, X_\ell).$$

The *Lie derivative* along the vector field X is defined as

$$L_X : \Lambda^k(M) \to \Lambda^k(M),$$
$$L_X(\omega) = i_X d\omega + d(i_X\omega).$$

For a smooth n–dimensional manifold M, a *k–dimensional submanifold* in M is a subset $N \subset M$ with the property that every point $p \in N$ is contained in the domain U of a chart $\varphi : U \to \mathbb{R}^n$ such that $\varphi(U \cap N) = \varphi(U) \cap \mathbb{R}^k$.

If $k = n$ then, by the inverse function theorem, N is an open subset in M; in this case N is also called an *open submanifold* in M. (The same is true in the topological category, but the proof is harder and requires Brouwer's Invariance of Domain Theorem, see e.g. [**Hat02**], Theorem 2B.3.)

Suppose that $U \subset \mathbb{R}^n$ is an open subset. A *piecewise-smooth function* $f : U \to \mathbb{R}^m$ is a continuous function such that for every $x \in U$ there exists a neighborhood V of x in U, a diffeomorphism $\phi : V \to V' \subset \mathbb{R}^n$, a triangulation T of V', so that the composition

$$f \circ \phi^{-1} : (V', T) \to \mathbb{R}^m$$

is smooth on each simplex. Note that composition $g \circ f$ is again piecewise-smooth, provided that g is smooth; however, composition of piecewise-smooth maps need not be piecewise-smooth.

One then defines *piecewise smooth k–dimensional submanifolds* N of a smooth manifold M. Such N is a topological submanifold which is locally the image of $\mathbb{R}^k$ in $\mathbb{R}^n$ under a piecewise-smooth homeomorphism $\mathbb{R}^n \to \mathbb{R}^n$. We refer the reader to [**Thu97**] for the detailed discussion of piecewise-smooth manifolds.

If $k = n - 1$ we also sometimes call a submanifold a (*piecewise smooth*) *hypersurface.*

Below we review two alternative ways of defining submanifolds. Consider a smooth map $f : M \to N$ of a m-dimensional manifold $M = M^m$ to an n-dimensional manifold $N = N^n$. The map $f : M \to N$ is called an *immersion* if for every $p \in M$, the linear map $df_p : T_pM \to T_{f(p)}N$ is injective. If, moreover, f defines a homeomorphism from M to $f(M)$ with the subspace topology, then f is called a *smooth embedding*.

EXERCISE 2.1. Construct an injective immersion $\mathbb{R} \to \mathbb{R}^2$ which is not a smooth embedding.

If N is a submanifold in M then the inclusion map $i : N \to M$ is a smooth embedding. This, in fact, provides an alternative definition for k-dimensional submanifolds: They are images of smooth embeddings with k–dimensional manifolds (see Corollary 2.4). Images of immersions provide a large class of subsets, called *immersed submanifolds*.

A smooth map $f : M^k \to N^n$ is called a *submersion* if for every $p \in M$, the linear map df_p is surjective. The following theorem can be found for instance, in [**GP10**], [**Hir76**], [**War83**].

THEOREM 2.2. (1) *If $f : M^m \to N^n$ is an immersion, then for every $p \in M$ and $q = f(p)$ there exists a chart $\varphi : U \to \mathbb{R}^m$ of M with $p \in U$, and a chart $\psi : V \to \mathbb{R}^n$ of N with $q \in V$ such that the composition*

$$\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$$

is of the form

$$\overline{f}(x_1, \dots, x_m) = (x_1, \dots, x_m, \underbrace{0, \dots, 0}_{n-m \text{ times}}).$$

(2) *If $f : M^m \to N^n$ is a submersion, then for every $p \in M$ and $q = f(p)$ there exists a chart $\varphi : U \to \mathbb{R}^m$ of M with $p \in U$, and a chart $\psi : V \to \mathbb{R}^n$ of N with $q \in V$ such that the composition*

$$\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$$

is of the form

$$\overline{f}(x_1, \dots, x_n, \dots, x_m) = (x_1, \dots, x_n).$$

(3) *The* constant rank theorem *is a combination of (1) and (2). Suppose that the derivative of $f : M^m \to N^n$ has constant rank r. Then then for every $p \in M$ and $q = f(p)$ there exists a chart $\varphi : U \to \mathbb{R}^m$ of M with $p \in U$, and a chart $\psi : V \to \mathbb{R}^n$ of N with $q \in V$ such that the composition*

$$\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$$

is of the form

$$\overline{f}(x_1, \dots, x_m) = (x_1, \dots, x_r, 0, \dots, 0).$$

In particular, $f(U)$ is a submanifold of dimension r in N.

EXERCISE 2.3. Prove Theorem 2.2. *Hint.* Use the Inverse Function Theorem and the Implicit Function Theorem from Vector Calculus.

COROLLARY 2.4. (1) *If $f : M^m \to N^n$ is a smooth embedding then $f(M^m)$ is a m-dimensional submanifold of N^n.*

(2) *If* $f : M^m \to N^n$ *is a submersion then for every* $x \in N^n$ *the fiber* $f^{-1}(x)$ *is a submanifold of codimension* n.

EXERCISE 2.5. Every submersion $f : M \to N$ is an *open map*, i.e., the image of an open subset in M is an open subset in N.

Let $f : M^m \to N^n$ be a smooth map and $y \in N$ is a point such that for some $x \in f^{-1}(y)$, the map $df_x : T_xM \to T_yN, y = f(x)$, is not surjective. Then the point $y \in N$ is called a *singular value* of f. A point $y \in N$ which is not a singular value of f is called a *regular value* of f. Thus, for every regular value $y \in N$ of f, the preimage $f^{-1}(y)$ is either empty or a smooth submanifold of dimension $m - n$.

THEOREM 2.6 (Sard's theorem). *Almost every point* $y \in N$ *is a regular value of* f.

2.2. Smooth partition of unity

DEFINITION 2.7. Let M be a smooth manifold and $\mathcal{U} = \{B_i : i \in I\}$ a locally finite covering of M by open subsets diffeomorphic to Euclidean balls. A collection of smooth functions $\{\eta_i : i \in I\}$ on M is called a *smooth partition of unity* for the cover $\mathcal{U}$ if the following conditions hold:

(1) $\sum_i \eta_i \equiv 1$.
(2) $0 \leqslant \eta_i \leqslant 1, \quad \forall i \in I$.
(3) $\mathrm{Supp}(\eta_i) \subset \overline{B_i}, \quad \forall i \in I$.

LEMMA 2.8. *Every open cover* $\mathcal{U}$ *as above admits a smooth partition of unity.*

2.3. Riemannian metrics

A *Riemannian metric* (also known as the *metric tensor*) on a smooth n-dimensional manifold M, is a positive definite inner product $\langle \cdot, \cdot \rangle_p$ defined on the tangent spaces T_pM of M; this inner product is required to depend smoothly on the point $p \in M$. We will suppress the subscript p in this notation; we let $\|\cdot\|$ denote the norm on T_pM determined by the Riemannian metric. The subspace of TM consisting of unit tangent vectors is a submanifold denoted UM and called the *unit tangent bundle*: UM is a smooth submanifold of TM and the restriction of the projection $TM \to M$ is a bundle, whose fibers are $n-1$-dimensional spheres.

The Riemannian metric is usually denoted $g = g_x = g(x), x \in M$ or ds^2. We will use the notation dx^2 to denote the Euclidean Riemannian metric on $\mathbb{R}^n$:

$$dx^2 := dx_1^2 + \ldots + dx_n^2.$$

Here and in what follows we use the convention that for tangent vectors u, v,

$$dx_i dx_j(u, v) = u_i v_j$$

and dx_i^2 stands for $dx_i dx_i$. A Riemannian metric on an open subset $\Omega \subset \mathbb{R}^n$ is determined by its *Gramm matrix* $A_x, x \in \Omega$, where A_x is a positive-definite symmetric matrix depending smoothly on x:

$$\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle_x = g_{ij}(x),$$

the ij-th entry of the matrix A_x.

A *Riemannian manifold* is a smooth manifold equipped with a Riemannian metric.

Two Riemannian metrics g, h on a manifold M are said to be *conformal* to each other, if $h_x = \lambda(x) g_x$, where $\lambda(x)$ is a smooth positive function on M, called the *conformal factor*. In matrix notation, we just multiply the matrix A_x of g_x by a scalar function. Such modification of Riemannian metrics does not change the angles between tangent vectors. A Riemannian metric g_x on a domain Ω in $\mathbb{R}^n$ is called *conformally-Euclidean* if it is conformal to dx^2, i.e., it is given by

$$\lambda(x)dx^2 = \lambda(x)(dx_1^2 + \ldots + dx_n^2).$$

Thus, the square of the norm of a vector $v \in T_x\Omega$ with respect to g_x is given by

$$\lambda(x) \sum_{i=1}^{n} v_i^2.$$

Given an immersion $f : M^m \to N^n$ and a Riemannian metric g on N, one defines the *pull-back* Riemannian metric $f^*(g)$ by

$$\langle v, w \rangle_p = \langle df(v), df(w) \rangle_q, p \in M, q = f(p) \in N,$$

where in the right-hand side we use the inner product defined by g and in the left-hand side the one defined by $f^*(g)$. It is useful to rewrite this definition in terms of symmetric matrices, when M, N are open subsets of $\mathbb{R}^n$. Let A_y be the matrix-function defining g. Then $f^*(g)$ is given by the matrix-function B_x, where

$$y = f(x), \quad B_x = (D_x f) A_y \ (D_x f)^T$$

and $D_x f$ is the Jacobian matrix of f at the point x.

Let us compute how pull-back works in "calculus terms" (this is useful for explicit computations of the pull-back metrics $f^*(g)$), when $g(y)$ is a Riemannian metric on an open subset U in $\mathbb{R}^n$. Suppose that

$$g(y) = \sum_{i,j} g_{ij}(y) dy_i dy_j$$

and $f = (f_1, \ldots, f_n)$ is a diffeomorphism $V \subset \mathbb{R}^n \to U$. Then

$$f^*(g) = h,$$

$$h(x) = \sum_{i,j} g_{ij}(f(x)) df_i df_j.$$

Here for a function $\phi : \mathbb{R}^n \to \mathbb{R}$, e.g., $\phi(x) = f_i(x)$,

$$d\phi = \sum_{k=1}^{n} d_k \phi = \sum_{k=1}^{n} \frac{\partial \phi}{\partial x_k} dx_k,$$

and, thus,

$$df_i df_j = \sum_{k,l=1}^{n} \frac{\partial f_i}{\partial x_k} \frac{\partial f_j}{\partial x_l} dx_k dx_l.$$

A special case of the above is when N is a submanifold in a Riemannian manifold M. One can define a Riemannian metric on N either by using the inclusion map and the pull-back metric, or by considering, for every $p \in N$, the subspace T_pN of T_pM, and restricting the inner product $\langle \cdot, \cdot \rangle_p$ to it. Both procedures define the same Riemannian metric on N.

Measurable Riemannian metrics. The same definition makes sense if the inner product depends only measurably on the point $p \in M$, equivalently, the matrix-function A_x is only measurable. This generalization of Riemannian metrics will be used in our discussion of quasiconformal groups, Chapter 21, section 21.6.

Gradient and divergence. A Riemannian metric g on M defines isomorphisms between tangent and cotangent spaces of M:

$$T_p(M) \to T_p^*(M),$$

where each $v \in T_p(M)$ corresponds to the linear functional

$$v^* \in T_p(M), \quad v^*(w) = \langle v, w \rangle .$$

In particular, one defines the *gradient vector field* ∇u of a function $u : M \to \mathbb{R}$ by dualizing the 1-form du.

Suppose now that M is n-dimensional. For a vector field X on M, the *divergence* $\operatorname{div} X$ is a function on M, which, for every n-form ω, satisfies

$$\operatorname{div} X\omega = L_X(\omega),$$

where L_X is the Lie derivative along X.

In local coordinates, divergence and gradient are given by the formulae:

$$\operatorname{div} X = \sum_{i=1}^{n} \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left(\sqrt{\det(g)} X^i \right),$$

and

$$(\nabla u)^i = \sum_{j=1}^{n} g^{ij} \frac{\partial u}{\partial x_j}.$$

Here

$$X = (X^1, \ldots, X^n),$$

and $(g^{ij}) = (g_{ij})^{-1}$ is the inverse matrix of the metric tensor g.

Length and distance. Given a Riemannian metric on M, one defines the *length* of a path $\mathfrak{p} : [a, b] \to M$ by

$$\text{length}(\mathfrak{p}) = \int_a^b \|\mathfrak{p}'(t)\| dt. \tag{2.1}$$

By abusing the notation, we will frequently denote $\text{length}(\mathfrak{p})$ by $\text{length}(\mathfrak{p}([a, b]))$.

Then, provided that M is connected, one defines the Riemannian *distance function*

$$\operatorname{dist}(p, q) = \inf_{\mathfrak{p}} \text{length}(\mathfrak{p}),$$

where the infimum is taken over all paths in M connecting p to q.

A smooth map $f : (M, g) \to (N, h)$ of Riemannian manifolds is called a *Riemannian isometry* if $f^*(h) = g$. In most cases, such maps do not preserve the Riemannian distances. This leads to a somewhat unfortunate terminological confusion, since the same name *isometry* is used to define maps between metric spaces which preserve the distance functions. Of course, if a Riemannian isometry $f : (M, g) \to (N, h)$ is also a diffeomorphism, then it preserves the Riemannian distance function and, hence, f is an isometry of Riemannian metric spaces.

A *Riemannian geodesic segment* is a path $\mathfrak{p} : [a, b] \subset \mathbb{R} \to M$ which is a local length-minimizer, i.e.:

There exists $c \geqslant 0$ so that for all t_1, t_2 in J sufficiently close to each other,

$$\operatorname{dist}(\mathfrak{p}(t_1), \mathfrak{p}(t_2)) = \text{length}(\mathfrak{p}([t_1, t_2])) = c|t_1 - t_2|.$$

If $c = 1$, we say that $\mathfrak{p}$ has *unit speed*. Thus, a unit speed geodesic is a locally-distance preserving map from an interval to (M, g). This definition extends to infinite geodesics in M, which are maps $\mathfrak{p} : J \to M$, defined on intervals $J \subset M$, whose restrictions to each finite interval are finite geodesics. A Riemannian metric is said to be *complete* if every geodesic segment extends to a complete geodesic $\gamma : \mathbb{R} \to M$. According to the Hopf–Rinow theorem, a Riemannian metric on a connected manifold is complete if and only if the associated distance function is complete.

A smooth map $f : (M, g) \to (N, h)$ is called *totally-geodesic* if it maps geodesics in (M, g) to geodesics in (N, h). If, in addition, $f^*(h) = g$, then such f is locally distance-preserving.

Injectivity and convexity radii. For every complete Riemannian manifold M and a point $p \in M$, there exists the *exponential map*

$$\exp_p : T_pM \to M,$$

which sends every vector $v \in T_pM$ to the point $\gamma_v(1)$, where $\gamma_v(t)$ is the unique geodesic in M with $\gamma(0) = p$ and $\gamma'(0) = v$. If $S(0, r) \subset T_pM$, $B(0, r) \subset T_pM$ are the round sphere and round ball of radius r, then

$$\exp(S(0, r)), \quad \exp(B(0, r)) \subset M$$

are the *geodesic r-sphere* and the geodesic r-ball in M centered at p.

The *injectivity radius $InjRad(p)$* of M at the point $p \in M$ is the supremum of the numbers r so that $\exp_p |B(0, r)$ is a diffeomorphism to its image. The *radius of convexity $ConRad(p)$* is the supremum of r's so that $r \leqslant InRad(p)$ and $C = \exp_p(B(0, r))$ is a convex subset of M, i.e., every $x, y \in C$ are connected by a (distance–realizing) geodesic segment entirely contained in C. It is a basic fact of Riemannian geometry that for every $p \in M$,

$$ConRad(p) > 0,$$

see e.g. [**dC92**].

2.4. Riemannian volume

For every n-dimensional Riemannian manifold (M, g) one defines the *volume element* (or *volume density*) denoted $\mathrm{d}V$ (or $\mathrm{d}A$ if M is 2-dimensional). Given n vectors $v_1, \ldots, v_n \in T_pM$, $\mathrm{d}V(v_1 \wedge \ldots \wedge v_n)$ is the volume of the parallelepiped in T_pM spanned by these vectors. This volume is nothing but $\sqrt{|\det(G(v_1, \ldots, v_n))|}$, where $G(v_1, \ldots, v_n)$ is the Gramm matrix with the entries $\langle v_i, v_j \rangle$. If $ds^2 = \rho^2(x)dx^2$, is a conformally-Euclidean metric on an open subset of $\mathbb{R}^n$, $\rho > 0$, then the volume density of ds^2 is given by

$$\rho^n(x)dx_1 \ldots dx_n.$$

Thus, every Riemannian manifold has a canonical measure, given by the integral of its volume form

$$mes(E) = \int_A \mathrm{d}V.$$

THEOREM 2.9 (Generalized Rademacher's theorem). *Let $f : M \to N$ be a Lipschitz map of Riemannian manifolds. Then f is differentiable almost everywhere.*

EXERCISE 2.10. Deduce Theorem 2.9 from Theorem 1.75 and the fact that M is second countable.

We now define *volumes of maps and submanifolds*. The simplest and the most familiar notion of volume of maps comes from the vector calculus. Let Ω be a bounded region in $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}^n$ be a smooth map. Then the *geometric* volume of f is defined as

$$Vol(f) := \int_\Omega |J_f(x)| dx_1 \dots dx_n, \tag{2.2}$$

where J_f is the Jacobian determinant of f. Note that we are integrating here a non-negative quantity, hence, the geometric volume of a map is always non-negative. If f were 1-1 and $J_f(x) > 0$ for every x, then, of course,

$$Vol(f) = \int_\Omega J_f(x) dx_1 \dots dx_n = Vol(f(\Omega)).$$

More generally, if $f : \Omega \to \mathbb{R}^m$ (now, m need not be equal to n), then

$$Vol(f) = \int_\Omega \sqrt{|\det(G_f)|},$$

where G_f is the Gramm matrix with the entries $\left\langle \frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial x_j} \right\rangle$, where brackets denote the usual inner product in $\mathbb{R}^m$. In case f is 1-1, the reader will recognize in this formula the familiar expression for the volume of an immersed submanifold $\Sigma = f(\Omega)$ in $\mathbb{R}^m$,

$$Vol(f) = \int_\Sigma dS.$$

The Gramm matrix above makes sense also for maps whose target is an m-dimensional Riemannian manifold (M, g), with partial derivatives replaced with vectors $df(X_i)$ in M, where X_i are coordinate vector fields in Ω:

$$X_i = \frac{\partial}{\partial x_i}, i = 1, \dots, n.$$

Furthermore, one can take the domain of the map f to be an arbitrary smooth manifold N (possibly with boundary). The definition of volume still makes sense and is independent of the choice of local charts on N used to define the integral: This independence is a corollary of the change of variables formula in the integral in $\mathbb{R}^n$. More precisely, consider charts $\varphi_\alpha : U_\alpha \to V_\alpha \subset N$, so that $\{V_\alpha\}_{\alpha \in J}$ is a locally-finite open covering of N. Let $\{\eta_\alpha\}$ be a partition of unity on N corresponding to this covering. Then for $\zeta_\alpha = \eta_\alpha \circ \varphi_\alpha$, $f_\alpha = f \circ \varphi_\alpha$,

$$Vol(f) = \sum_{\alpha \in J} \int_{U_\alpha} \zeta_\alpha \sqrt{|\det(G_{f_\alpha})|} dx_1 \dots dx_n$$

In particular, if f is 1-1 and $\Sigma = f(N)$, then

$$Vol(f) = Vol(\Sigma).$$

REMARK 2.11. The formula for $Vol(f)$ makes sense when $f : N \to M$ is merely Lipschitz, in view of Theorem 2.9.

Thus, one can define the volume of an immersed submanifold, as well as that of a piecewise smooth submanifold; in the latter case we subdivide a piecewise-smooth submanifold in a union of images of simplices under smooth maps.

By abuse of language, sometimes, when we consider an open submanifold N in M, so that boundary ∂N of N a submanifold of codimension 1, while we denote the volume of N by $Vol(N)$, we shall call the volume of ∂N the *area*, and denote it by $Area\,(\partial N)$.

EXERCISE 2.12. (1) Suppose that $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ is a smooth map so that $|d_x f(u)| \leqslant 1$ for every unit vector u and every $x \in \Omega$. Show that $|J_f(x)| \leqslant 1$ for every x and, in particular,

$$Vol(f(\Omega)) = |\int_\Omega J_f dx_1 \dots dx_n| \leqslant Vol(f) \leqslant Vol(\Omega).$$

Hint: Use the fact that under the linear map $A = d_x f$, the image of every r-ball is contained in an r-ball.

(2) Prove the same thing if the map f is merely 1-Lipschitz.

More general versions of the above exercises are the following.

EXERCISE 2.13. Let (M, g) and (N, h) be n-dimensional Riemannian manifolds.

(1) Let $f : M \to N$ be a smooth map such that for every $x \in M$, the norm of the linear map

$$df_x : \left(T_x M, \langle\cdot,\cdot\rangle_g\right) \to (T_f(x)N, \langle\cdot,\cdot\rangle_h)$$

is at most L.

Prove that $|J_f(x)| \leqslant L^n$ for every x and that for every open subset U of M

$$Vol(f(\Omega)) \leqslant L^n Vol(\Omega).$$

(2) Prove the same statement for an L–Lipschitz map $f : M \to N$.

A consequence of Theorem 2.2 is the following.

THEOREM 2.14. *Consider a compact Riemannian manifold M^m, a submersion $f : M^m \to N^n$. For every $x \in N$ set $M_x := f^{-1}(x)$. Then, for every $p \in N$ and every $\epsilon > 0$ there exists an open neighborhood W of p such that for every $x \in W$,*

$$1 - \epsilon \leqslant \frac{Vol(M_x)}{Vol(M_p)} \leqslant 1 + \epsilon.$$

PROOF. First note that, by compactness of M_p, for every neighborhood U of M_p there exists a neighborhood W of p such that $f^{-1}(W) \subset U$.

According to Theorem 2.2, (2), for every $x \in M_p$ there exists a chart of M, $\varphi_x : U_x \to \tilde{U}_x$, with U_x containing x, and a chart of N, $\psi_x : V_x \to \tilde{V}_x$ with V_x containing p, such that $\psi_x \circ f \circ \varphi_x^{-1}$ is a restriction of the projection to the first n coordinates. Without loss of generality we may assume that $\tilde{U}_x$ is an open cube in $\mathbb{R}^m$. Therefore, $\overline{V}_x$ is also a cube in $\mathbb{R}^n$, and $\tilde{U}_x = \tilde{V}_x \times \tilde{Z}_x$, where $\tilde{Z}_x$ is an open subset in $\mathbb{R}^{m-n}$.

Since M_p is compact, it can be covered by finitely many such domains of charts $U_1, \dots, U_k$. Let $V_1, \dots, V_k$ be the corresponding domains of charts containing p. For the open neighborhood $U = \bigcup_{i=1}^k U_i$ of M_p consider an open neighborhood W of p, contained in $\bigcap_{i=1}^k V_i$, such that $f^{-1}(W) \subseteq U$.

For every $x \in W$, $M_x = \bigcup_{l=1}^{k} (U_i \cap M_x)$. Fix $l \in \{1, \ldots, k\}$. Let $(g_{ij}(y))_{1 \leqslant i,j \leqslant n}$ be the matrix-valued function on $\tilde{U}_l$, defining the pull-back by φ_l of the Riemannian metric on M.

Since g_{ij} is continuous, there exists a neighborhood $\tilde{W}_l$ of $\tilde{p} = \psi_l(p)$ such that for every $\tilde{x} \in W_l$ and for every $\tilde{t} \in \tilde{Z}_l$ we have,

$$(1-\epsilon)^2 \leqslant \frac{\det \left[g_{ij}(\tilde{x},\tilde{t})\right]_{n+1 \leqslant i,j \leqslant m}}{\det \left[g_{ij}(\tilde{p},\tilde{t})\right]_{n+1 \leqslant i,j \leqslant m}} \leqslant (1+\epsilon)^2 .$$

Recall that the volumes of $M_x \cap U_i$ and of $M_p \cap U_l$ are obtained by integrating respectively

$$(\det \left[g_{ij}(\tilde{x},\tilde{t})\right]_{n+1 \leqslant i,j \leqslant k})^{1/2}$$

and

$$(\det \left[g_{ij}(\tilde{p},\tilde{t})\right]_{n+1 \leqslant i,j \leqslant k})^{1/2}$$

on Z_l. The volumes of M_x and M_p are obtained by combining this with a partition of unity.

It follows that for $x \in \bigcap_{i=1}^{k} \psi_i^{-1}(\overline{W}_l)$,

$$1-\epsilon \leqslant \frac{Vol(M_x)}{Vol(M_p)} \leqslant 1+\epsilon .$$

□

Finally, we recall an important formula for volume computations:

<u>Theorem</u> 2.15 (Coarea formula, see e.g. Theorem 6.3 in **[Cha06]** and 3.2.22 in **[Fed69]**). *Let $f : M \to (0,\infty)$ be a smooth function on a Riemannian manifold M. For almost every $t \in (0,\infty)$, the level set $\mathcal{H}_t := f^{-1}(t)$ is a smooth hypersurface in M; let $\mathrm{d}A_t$ be the Riemannian area density induced on $\mathcal{H}_t$ and $\mathrm{d}V$ be the Riemannian volume density of M. Then, for every function $g \in L^1(M)$,*

$$\int_M g\,|\nabla f| \mathrm{d}V = \int_0^\infty \mathrm{d}t \int_{\mathcal{H}_t} g\,\mathrm{d}A_t .$$

2.5. Volume growth and isoperimetric functions. Cheeger constant

In this section we present several basic notions, initially introduced in Riemannian geometry and later adapted and used in group theory and in combinatorics. These notions and their *coarse analogues* will appear frequently in this book.

Volume growth. Given a Riemannian manifold M and a basepoint $x_0 \in M$, the *(volume) growth function* is defined as

$$\mathfrak{G}_{M,x_0}(r) := Vol\, B(x_0, r),$$

the volume of the metric ball of radius r and center at x_0 in M.

Remarks 2.16. (1) For two different points x_0, y_0, we have

$$\mathfrak{G}_{M,x_0}(r) \leqslant \mathfrak{G}_{M,y_0}(r+d), \text{ where } d = \mathrm{dist}(x_0,y_0) .$$

(2) Suppose that the action of the isometry group of M is *cobounded* on M, i.e., there exists a constant κ such that the orbit of $B(x_0, \kappa)$ under the group $\mathrm{Isom}(M)$, is the entire manifold M. (For instance, this is the case if M is a regular covering space of a compact Riemannian manifold.) Then, for every two basepoints x_0, y_0

$$\mathfrak{G}_{M,x_0}(r) \leqslant \mathfrak{G}_{M,y_0}(r+\kappa)\,.$$

Thus, in this case, the growth rate of the function $\mathfrak{G}$ does not depend on the choice of the basepoint.

We refer the reader to Section 5.7 for the detailed discussion of volume growth and its relation to *group growth.*

Exercise 2.17. Assume again that the action $\mathrm{Isom}(M) \curvearrowright M$ is cobounded, the constant κ is as above, and that M is complete.

(1) Prove that the growth function is *almost sub-multiplicative*, that is:

$$\mathfrak{G}_{M,x_0}\left((r+t)\kappa\right) \leqslant \mathfrak{G}_{M,x_0}(r\kappa)\mathfrak{G}_{M,x_0}((t+1)\kappa)\,.$$

(2) Prove that the growth function of M is at most exponential, that is, there exists $a > 1$ such that

$$\mathfrak{G}_{M,x_0}(x) \leqslant a^x\,, \text{ for every } x \geqslant 0\,.$$

Isoperimetric inequalities and isoperimetric functions. Isoperimetric problems in geometry go back to the antiquity: (Dido's problem) *Which region of the given perimeter in the Euclidean plane $\mathbb{R}^2$ has the least area?* The answer "the round disk" is intuitively obvious, but, surprisingly, hard to prove. This classical problem explain the terminology *isoperimetric* below.

In general, isoperimetric problem in Riemannian geometry have the following *minimax* form.

Consider a complete connected n-dimensional Riemannian manifold M (which may or may not be closed). Fix a number $k \in \mathbb{N}$ and all consider closed k-dimensional submanifolds $Z \subset M$ (or maps $Z \to M$ of closed k-manifolds to M or, more generally, k-cycles in M). Assume, now, depending on the context, that each $Z \subset M$ bounds a $k+1$-dimensional submanifold, or a $k+1$-chain B, or that the map $Z \to M$ extends to a map $B \to M$, where B is a compact manifold with boundary equal to Z. The the latter case, one typically assumes that $Z = \mathbb{S}^k$ and B is the $k+1$-ball. To unify the notation, we will simply say that $Z = \partial B$, even in the case of maps $Z \to M$.

Next, among all these B's (or their maps), one looks for the one of the least $k+1$-volume. (The minimum may not exist, in which case one takes the infimum.) This least volume is the *filling volume of* Z. Lastly, among all Z's with $Vol_k(Z) \leqslant L$, one looks for the ones which have the largest filling volume (again, taking the supremum in general). This defines the *isoperimetric function* of M:

$$IP_{M,k}(L) = \sup_{Z, Vol_k(Z)\leqslant L}\ \inf_{B,\partial B = Z} Vol_{k+1}(B). \tag{2.3}$$

In each setting (submanifolds, maps, cycles), we get a different isoperimetric function, of course.

We will be primarily interested in two cases (Z having codimension and dimension 1 respectively):

1. $k = n-1$, Z is a (smooth) closed hypersurface in M.

2. $k = 1$, $Z = \mathbb{S}^1$, where we consider Lipschitz maps $Z \to M$ and their extensions $B = \mathbb{D}^2 \to M$ ("filling disks").

Perhaps surprisingly, asymptotic behavior of isoperimetric functions in these two cases goes long way towards determining the asymptotic geometry of M. Suppose, for instance, that M is a regular cover of a compact Riemannian manifold, with the group G of covering transformations. Then the dichotomy *linear/superlinear* for both isoperimetric functions serves as a major demarkation line in the world of finitely generated groups:

1. The condition $IP_{M,1}(L) \approx L$ (linear growth of the filling area) yields the class of *Gromov-hyperbolic groups* G. This linearity condition can be regarded as *asymptoically negative sectional curvature* of the manifold M (and the group G).

2. The condition $IP_{M^n,n-1}(L) \approx L$ yields the class of *nonamenable groups* G.

Here we are using the notation $\approx$ introduced in the Definition 1.3.

The following Riemannian geometry theorem illustrates the power of this linear/nonlinear dichotomy:

THEOREM 2.18. *Suppose that M is a Riemannian manifold which is the universal cover of a compact Riemannian manifold. Then*

$$IP_{M,1}(L) \approx L \Rightarrow IP_{M,k}(L) \approx L,$$

for all $k \geqslant 2$. Here for $k \geqslant 2$ one can equally use either the homological filling or filling of maps of spheres by maps of disks. For the former, one needs to assume that $H_i(M) = 0, i \leqslant k$ and for the latter one requires that $\pi_i(M) = 0, i \leqslant k$.

As far as we know, this theorem does not have a "purely Riemannian" proof: One first verifies that the group $G = \pi_1(M)$ is Gromov-hyperbolic (Theorem 9.172), then proves that all such groups have linear isoperimetric functions of in all degrees [**Lan00, Min01**] and, then uses the approximate equality of isoperimetric functions of M and of G (cf. Theorem 5.98).

Below we discuss the "codimension 1" isoperimetric function in more detail. If M is connected and noncompact, each closed hypersurface in M bounds exactly one compact submanifold, which leads to

DEFINITION 2.19. Suppose that $F : \mathbb{R}_+ \to \mathbb{R}_+$ is a function and M is a (connected) noncompact n-dimensional Riemannian manifold. Then M is said to satisfy the *isoperimetric inequality* of the form

$$Vol(\Omega) \leqslant F\left(Area(\partial\Omega)\right),$$

if this inequality holds for all open submanifolds $\Omega \subset M$ with compact closure and smooth boundary.

EXERCISE 2.20. The above definition is equivalent to the inequality

$$IP_{n-1,M}(L) \leqslant F(L)$$

for every $L > 0$. (Note: Hypersurfaces in the definition of $IP_{n-1,M}$ need not be connected.)

For instance, if M is the Euclidean plane, then

$$4\pi \mathrm{Ar}(c) \leqslant \ell^2(c), \tag{2.4}$$

for every loop c (with equality realized precisely in the case when c is a round circle). Thus,

$$IP_{\mathbb{R}^2}(\ell) = \frac{\ell^2}{4\pi}.$$

The Cheeger constant. As the main dichotomy in the case of codimension 1 isoperimetric inequality is linear/nonlinear, it makes sense to look at the ratio between areas of hypersurfaces in M and volumes of domains in M which they bound. If M is compact, connected and the hypersurface is connected, then there are exactly two such domains. In line with the definition of the isoperimetric function, we will be choosing the domain with the least volume. This motivates:

DEFINITION 2.21. The *Cheeger (isoperimetric) constant* $h(M)$ (or *isoperimetric ratio*) of M is the infimum of the ratios

$$\frac{Area(\partial\Omega)}{\min\left[Vol(\Omega)\,,\,Vol(M\setminus\Omega)\right]},$$

where Ω varies over all open nonempty submanifolds with compact closure and smooth boundary.

In particular, if $h(M) \geqslant \kappa > 0$, then the following isoperimetric inequality holds in M:

$$Vol(\Omega) \leqslant \frac{1}{\kappa} Area(\partial\Omega).$$

Cheeger constant was defined by Cheeger for compact manifolds in [**Che70**]. Further details can be found for instance in P. Buser's book [**Bus10**]. Note that when M is a Riemannian manifold of infinite volume, one may replace the denominator in the ratio defining the Cheeger constant by $Vol(\Omega)$.

Assume now that M is the universal cover of a compact Riemannian manifold N. A natural question to ask is to what extent the growth function and the Cheeger constant of M depend on the choice of the Riemannian metric on N. The first question, in a way, was one of the origins of the geometric group theory.

V.A. Efremovich [**Efr53**] noted that two growth functions corresponding to two different choices of metrics on N are asymptotically equal (see Definition 1.4) and, moreover, that their asymptotic equivalence class is determined by the fundamental group of N only. See Proposition 5.78 for a slightly more general statement.

A similar phenomenon occurs with the Cheeger constant: Positivity of $h(M)$ does not depend on the metric on N, it depends only on a certain property of $\pi_1(N)$, namely, the non-amenability, see Remark 16.14. This was proved much later by R. Brooks [**Bro81a, Bro82a**]. Brooks' argument has a global analytic flavor, as it uses the connection established by Cheeger [**Che70**] between positivity of the isoperimetric constant and positivity of spectrum of the Laplace–Beltrami operator on M. Note that, even though in his paper Cheeger only considers compact manifolds, the same argument works for universal covers of compact manifolds. This result was highly influential in global analysis on manifolds and harmonic analysis on graphs and manifolds.

2.6. Curvature

Instead of defining the Riemannian curvature tensor, we will only describe some properties of Riemannian curvature. First, if (M, g) is a 2-dimensional Riemannian manifold, one defines the *Gaussian curvature* of (M, g), which is a smooth function $K : M \to \mathbb{R}$, whose values are denoted $K(p)$ and K_p.

More generally, for an n-dimensional Riemannian manifold (M, g), one defines the *sectional curvature*, which is a function $\Lambda^2 M \to \mathbb{R}$, denoted $K_p(u, v) = K_{p,g}(u, v)$:

$$K_p(u, v) = \frac{\langle R(u, v)u, v\rangle}{|u \wedge v|^2},$$

provided that $u, v \in T_pM$ are linearly independent. Here R is the Riemannian curvature tensor and $|u \wedge v|$ is the area of the parallelogram in T_pM spanned by the vectors u, v. Sectional curvature depends only on the 2-plane P in T_pM spanned by u and v. The curvature tensor $R(u, v)w$ does not change if we replace the metric g with a conformal metric $h = ag$, where $a > 0$ is a constant. Thus,

$$K_{p,h}(u, v) = a^{-2} K_{p,g}(u, v).$$

Totally geodesic Riemannian isometric immersions $f : (M, g) \to (N, h)$ preserve sectional curvature:

$$K_p(u, v) = K_q(df(u), df(v)), \quad q = f(p).$$

In particular, sectional curvature is invariant under Riemannian isometries of equidimensional Riemannian manifolds. In the case when M is 2-dimensional, $K_p(u, v) = K_p$, is the Gaussian curvature of M.

Gauss-Bonnet formula. Our next goal is to connect areas of triangles to curvature.

<u>Theorem</u> 2.22 (Gauss-Bonnet formula). *Let (M, g) be a Riemannian surface with the Gaussian curvature $K(p), p \in M$ and the area form dA. Then for every 2-dimensional triangle $\blacktriangle \subset M$ with geodesic edges and vertex angles α, β, γ,*

$$\int_{\blacktriangle} K(p)dA = (\alpha + \beta + \gamma) - \pi.$$

In particular, if $K(p)$ is constant equal κ, we get

$$-\kappa Area(\blacktriangle) = \pi - (\alpha + \beta + \gamma).$$

The quantity $\pi - (\alpha + \beta + \gamma)$ is called the *angle deficit* of the triangle Δ.

Curvature and volume. Below we describe the relation of uniform lower and upper bounds on the sectional curvature and the growth of volumes of balls, that will be used in the sequel. The references for these results are [**BC01**, Section 11.10], [**CGT82**], [**Gro86**], [**Gö0**] and [**GHL04**], Theorem 3.101, p. 140.

We will use the following notation: For $\kappa \in \mathbb{R}$, we let $A_\kappa(r)$ and $V_\kappa(r)$ denote the area of the sphere, respectively the volume of the ball of radius r, in the n–dimensional space of constant sectional curvature κ. We will also denote by $A(x, r)$ the area of the geodesic sphere of radius r and center x in a given Riemannian manifold M. Likewise, $V(x, r)$ will denote the volume of the geodesic ball centered at x and of radius r in M.

<u>Theorem</u> 2.23 (Bishop–Gromov–Günther). *Let M be a complete n–dimensional Riemannian manifold.*

(1) *Assume that the sectional curvature on M is at least a. Then, for every point $x \in M$:*
   - $A(x,r) \leqslant A_a(r)$ *and* $V(x,r) \leqslant V_a(r)$.
   - *The functions* $r \mapsto \frac{A(x,r)}{A_a(r)}$ *and* $r \mapsto \frac{V(x,r)}{V_a(r)}$ *are non-increasing.*

(2) *Assume that the sectional curvature on M is at most b. Then, for every $x \in M$ with injectivity radius $\rho_x = InjRad_M(x)$:*
   - *For all $r \in (0, \rho_x)$, we have $A(x,r) \geqslant A_b(r)$ and $V(x,r) \geqslant V_b(r)$.*
   - *The functions* $r \mapsto \frac{A(x,r)}{A_b(r)}$ *and* $r \mapsto \frac{V(x,r)}{V_b(r)}$ *are non-decreasing on the interval* $(0, \rho_x)$.

The results (1) in the theorem above are also true if the Ricci curvature of M is at least $(n-1)a$.

Exercise 2.24. Use this inequality to show that every n-dimensional Riemannian manifold M of nonnegative Ricci curvature has at most polynomial growth:

$$\mathfrak{G}_M(r) \precsim r^n.$$

Theorem 2.23 follows from infinitesimal versions of the above inequalities (see Theorems 3.6 and 3.8 in [**Cha06**]). A consequence of the infinitesimal version of Theorem 2.23, (1), is the following theorem which will be useful in the proof of the quasiisometric invariance of positivity of the Cheeger constant:

<u>Theorem</u> 2.25 (Buser's inequality [**Bus82**], [**Cha06**], Theorem 6.8). *Let M be a complete n–dimensional manifold with sectional curvature at least a. Then there exists a positive constant λ depending on n, a and $r > 0$, such that the following holds. Given a hypersurface $\mathcal{H} \subset M$ and a ball $B(x,r) \subset M$ such that $B(x,r) \setminus \mathcal{H}$ is the union of two open subsets Ω_1, Ω_2 separated by $\mathcal{H}$, we have:*

$$\min\left[Vol(\Omega_1)\,,\, Vol(\Omega_2)\right] \leqslant \lambda Area\left[\mathcal{H} \cap B(x,r)\right].$$

2.7. Riemannian manifolds of bounded geometry

Definition 2.26. We say that a Riemannian manifold M has *bounded geometry* if it is connected, complete, has uniform upper and lower bounds for the sectional curvature:

$$a \leqslant K_p(u,v) \leqslant b$$

(for all $p \in M, u, v \in T_p(M)$) and a uniform lower bound for the injectivity radius:

$$InjRad(x) > \epsilon > 0.$$

Probably the correct terminology should be "uniformly locally bounded geometry", but we prefer shortness to an accurate description. The numbers a, b, ϵ in this definition are called *geometric bounds* on M. For instance, every compact connected Riemannian manifold M has bounded geometry, every covering space of M (with pull-back Riemannian metric) also has bounded geometry. More generally, if M is connected, complete and the action of the isometry group on M is cobounded, then M has bounded geometry.

Exercise 2.27. Every noncompact manifold of bounded geometry has infinite volume.

Remark 2.28. One frequently encounters weaker notions of bounded geometry for Riemannian manifold, e.g.:

1. There exists $L \geqslant 1$ and $R > 0$ such that every ball of radius R in M is L-bi-Lipschitz equivalent to the ball of radius R in $\mathbb{R}^n$. (This notion is used, for instance, by Gromov in [**Gro93**], §0.5.A_3).

2. The Ricci curvature of M has a uniform lower bound ([**Cha06**], [**Cha01**]).

For the purposes of this book, the restricted condition in Definition 2.26 suffices.

The following theorem connects Gromov's notion of bounded geometry with the one used in this book:

Theorem 2.29 (See e.g. Theorem 1.14, [**Att94**]). *Let M be a Riemannian manifold of bounded geometry with geometric bounds a, b, ϵ. Then for every $x \in M$ and $0 < r < \epsilon/2$, the exponential map*

$$\exp_x : B(0, r) \to B(x, r) \subset M$$

is an L-bi-Lipschitz diffeomorphism, where $L = L(a, b, \epsilon)$.

This theorem also allows one to refine the notion of partition of unity in the context of Riemannian manifolds of bounded geometry:

Lemma 2.30. *Let M be a Riemannian manifold of bounded geometry and let $\mathcal{U} = \{B_i = B(x_i, r_i) : i \in I\}$ be a locally finite covering of M by metric balls so that $InjRad_M(x_i) > 2r_i$ for every i and*

$$B\left(x_i, \frac{3}{4}r_i\right) \cap B\left(x_j, \frac{3}{4}r_j\right) = \emptyset, \ \forall i \neq j.$$

Then $\mathcal{U}$ admits a smooth partition of unity $\{\eta_i : i \in I\}$ which, in addition, satisfies the following properties:

1. $\eta_i \equiv 1$ on every ball $B(x_i, \frac{r_i}{2})$.

2. Every smooth functions η_i is L–Lipschitz for some L independent of i.

In what follows we keep the notation $V_\kappa(r)$ from Theorem 2.23 for the volume of a ball of radius r in the n–dimensional space of constant sectional curvature κ.

Lemma 2.31. *Let M be complete n–dimensional Riemannian manifold with bounded geometry, let $a \leqslant b$ and $\rho > 0$ be such that the sectional curvature of M varies in the interval $[a, b]$ and that at every point of M the injectivity radius is larger than ρ. Then:*

(1) *For every $\delta > 0$, every δ–separated set in M is ϕ-uniformly discrete, with $\phi(r) = \frac{V_a(r+\lambda)}{V_b(\lambda)}$, where λ is the minimum of $\frac{\delta}{2}$ and ρ.*

(2) *For every $2\rho > \delta > 0$ and every maximal δ–separated set N in M, the multiplicity of the covering $\{B(x, \delta) \mid x \in N\}$ is at most $\frac{V_a\left(\frac{3\delta}{2}\right)}{V_b\left(\frac{\delta}{2}\right)}$.*

Proof. (1) Let S be a δ–separated subset in M.

According to Theorem 2.23, for every point $x \in S$ and radius $r > 0$ we have:

$$V_a(r + \lambda) \geqslant Vol\,[B_M(x, r + \lambda)] \geqslant \text{card}\,\left[\overline{B}(x, r) \cap S\right]\, V_b(\lambda)\,.$$

This inequality implies that

$$\text{card}\left[\overline{B}(x,r)\cap S\right] \leqslant \frac{V_a(r+\lambda)}{V_b(\lambda)},$$

whence, S with the induced metric is ϕ-uniformly discrete, with the required ϕ.

(2) Let F be a subset in N such that the intersection

$$\bigcap_{x\in F} B(x,\delta)$$

is non-empty. Let y be a point in this intersection. Then the ball $B\left(y, \frac{3\delta}{2}\right)$ contains the disjoint union $\bigsqcup_{x\in F} B\left(x, \frac{\delta}{2}\right)$, whence

$$V_a\left(\frac{3\delta}{2}\right) \geqslant Vol\left[B_M\left(y, \frac{3\delta}{2}\right)\right] \geqslant \text{card }[F]\, V_b\left(\frac{\delta}{2}\right).$$

□

2.8. Metric simplicial complexes of bounded geometry and systolic inequalities

In this section we describe a discretization of manifolds of bounded geometry via *metric simplicial complexes*. Another method of approximating of Riemannian manifolds by simplicial complexes will be described in §5.3, cf. Theorem 5.50.

Let X be a simplicial complex and d a path-metric on X. Then (X,d) is said to be a *metric simplicial complex* if the restriction of d to each simplex is isometric to a Euclidean simplex. The main example of a metric simplicial complex is a generalization of a graph with the standard metric described below.

Let X be a connected simplicial complex. As usual, we will often conflate X and its geometric realization. Metrize each k-simplex of X to be isometric to the standard k-simplex Δ^k in the Euclidean space:

$$\Delta^k = (\mathbb{R}_+)^{k+1} \cap \{x_0 + \ldots + x_k = 1\}.$$

Thus, for each m-simplex Δ^m and its face Δ^k, the inclusion $\Delta^k \to \Delta^m$ is an isometric embedding. This allows us to define a length-metric on X so that each simplex is isometrically embedded in X, similarly to the definition of the standard metric on a graph and the Riemannian distance function. Namely, a *piecewise-linear (PL) path* $\mathfrak{p}$ in X is a path $\mathfrak{p} : [a,b] \to X$, whose domain can be subdivided in finitely many intervals $[a_i, a_{i+1}]$ such that each restriction

$$\mathfrak{p}|_{[a_i,a_{i+1}]}$$

is a piecewise-linear path whose image is contained in a single simplex of X. Lengths of such paths are defined using the Euclidean metric on simplices of X. Then

$$d(x,y) = \inf_{\mathfrak{p}} \text{length}(\mathfrak{p}),$$

where the infimum is taken over all PL paths in X connecting x to y. The metric d is then a path-metric; we call this metric the *standard metric* on X.

Exercise 2.32. Verify that the standard metric is complete and that X is a geodesic metric space.

Definition 2.33. A metric simplicial complex X has *bounded geometry* if it is connected and if there exist $L \geqslant 1$ and $N < \infty$ such that:

- every vertex of X is incident to at most N edges;
- the length of every edge is in the interval $[L^{-1}, L]$.

In particular, the set of vertices of X with the induced metric is a uniformly discrete metric space.

Thus, a metric simplicial complex of bounded geometry is necessarily finite-dimensional.

EXAMPLE 2.34.
- If Y is a finite connected metric simplicial complex, then its universal cover (with the pull-back path metric) has bounded geometry.
- A connected simplicial complex (with the standard metric) has bounded geometry if and only if there is a uniform bound on the valency of the vertices in its 1-skeleton.

Metric simplicial complexes of bounded geometry appear in the context of Riemannian manifolds with bounded geometry.

DEFINITION 2.35. Let M be a Riemannian manifold. A *uniform triangulation* $\mathcal{T}$ of M is a metric simplicial complex X of bounded geometry together with a bi-Lipschitz homeomorphism $f : X \to M$.

Every smooth manifold admits a triangulation (see [**Cai61**] for an especially simple proof); however, a general Riemannian manifold M will not have a uniform triangulation. An easy sufficient condition for uniformity of (any) triangulation of M is compactness of M. Lifting a finite triangulation $\mathcal{T}$ of a compact Riemannian manifold M to its Riemannian covering $M' \to M$ results in a uniform triangulation $\mathcal{T}'$ of M'.

Proofs of the following theorem are outlined in [**Att94**, Theorem 1.14] and [**ECH$^+$92**, Theorem 10.3.1]; a detailed proof in the case of hyperbolic manifolds can be found in [**Bre09**].

THEOREM 2.36. *Every Riemannian manifold of bounded geometry admits a uniform triangulation. Furthermore, there exists a function $L = L(m, a, b, \epsilon)$ with the following property. Let M be an m-dimensional Riemannian manifold of bounded geometry with the geometric bounds a, b, ϵ. Then M admits a uniform triangulation $\mathcal{T}$, which is a simplicial complex X equipped with the standard metric together with an L-bi-Lipschitz homeomorphism $f : X \to M$, such that geometric bounds on X depend only on m, a, b and ϵ.*

However, given lack of a detailed proof, this theorem should be currently treated as a conjecture. Nevertheless, a *homotopy form* of this theorem is not all that hard and its proof can be found for instance in [**GP88**]:

THEOREM 2.37. *Under the hypothesis of Theorem 2.36, there exists a simplicial complex Y of bounded geometry (with the standard metric) and a pair of L-Lipschitz maps*

$$f : M \to Y, \quad \bar{f} : Y \to M$$

which form a homotopy-equivalence between M and Y. Furthermore, the homotopies

$$H : M \times [0, 1] \to M, \quad H : Y \times [0, 1] \to Y$$

between $\bar{f} \circ f$ and id_M, and $f \circ \bar{f}$ and id_Y respectively are also L-Lipschitz maps. Here L and geometric bounds on Y depend only on m, a, b and ϵ.

The simplicial complex Y appears as the nerve of a cover $\mathcal{B}$ of M by suitably chosen convex metric balls. The map $f : M \to Y$ comes from a partition of unity associated with the cover $\mathcal{B}$.

The main application of uniform triangulations (or, their homotopy analogues) in this book comes in the form of *systolic inequalities* which we describe below.

Let M be a Riemannian manifold. The *k-systole $sys_k(M)$* of M is defined as the infimum of volumes of homologically nontrivial k-cycles in M. In the following proof, by abusing the terminology, we will conflate singular k-chains

$$S = \sum_{i=1}^{N} a_i \sigma_i$$

(where $a_i \in \mathbb{Z} \setminus \{0\}$ and σ_i's are singular simplices) and their *support sets* in X, i.e., unions of images of the singular k-simplices σ_i.

THEOREM 2.38. *Every Riemannian manifold M of bounded geometry has positive k-systoles for all k.*

PROOF. Given M we take either a uniform triangulation $\mathcal{T} = (X, f)$ of M (if it exists) or its homotopy analogue. Since the map f is a bi-Lipschitz homotopy equivalence, it suffices to prove positivity of k-systoles for X. The key to the proof is the following *Deformation Theorem* of Federer and Flemming, which first appeared in their work on the Plateau Problem [**FF60**, §5]. Another proof of this fundamental fact can be found in Federer's book [**Fed69**, 4.2.9]; an especially readable proof is given in [**ECH$^+$92**, Theorem 10.3.3]. Suppose that Δ^n is the standard n-simplex. For each interior point $x \in \Delta^n$ we define the *radial projection* $p_x : \Delta^n \setminus \{x\} \to \partial\Delta^n$. We will need:

THEOREM 2.39 (Deformation Theorem). *Suppose that S is a singular k-chain in Δ, $k < n$. Then for almost every point $x \in \Delta^n$, the k-volume of the chain $p_x(S)$ in $\partial\Delta$ does not exceed $C\mathrm{Vol}_k(S)$, where the constant C depends only on the dimension n of the simplex.*

We will refer to the projections p_x satisfying the conclusion of this theorem as *Federer-Flemming projections.*

Suppose now that S is a singular k-cycle in a bounded geometry D-dimensional simplicial complex X. In each n-simplex Δ^n in X whose dimension is greater than k, we apply a Federer–Flemming projection p_x to S. By combining these projections, we obtain a chain S_1 in $X^{(n-1)}$, which is homologous to S and whose volume is at most $C\mathrm{V}ol_k(S)$. After repeating the process at most $D-k$ times, we obtain a k-cycle S' in the k-skeleton $X^{(k)}$ (homologous to S); the volume of S' is at most $C^{D-k}\mathrm{V}ol_k(S)$. Let V_k denote the volume of the standard k-simplex. If $\mathrm{V}ol_k(S')$ is less than V_k, then S' cannot cover any k-simplex in X. Therefore, for each k-simplex Δ^k in X we apply a radial projection p_x to S' from any point x which does not belong to S' (at this stage, we no longer care about the volume of the image). The result is a k-cycle T in the $k-1$-skeleton $X^{(k-1)}$ of X, which is still homologous to S. However, $k-1$-dimensional simplicial complexes have zero kth homology groups, which means that T (and, hence, S) is homologically trivial. Therefore, assuming that $S \in Z_k(M)$ was homologically nontrivial, we obtain a lower bound on it volume:

$$\mathrm{V}ol_k(S) \geqslant L^{-k} C^{k-D} V_k.$$

□

2.9. Harmonic functions

For the detailed discussion of the material in this section we refer the reader to **[Li12]** and **[SY94]**.

Let M be a Riemannian manifold. Given a smooth function $f : M \to \mathbb{R}$, we define the *energy* of f as the integral

$$E(f) = \int_M |df|^2 \mathrm{d}V = \int_M |\nabla f|^2 \mathrm{d}V.$$

Here the gradient vector field ∇f is obtained by dualizing the differential 1-form df using the Riemannian metric on M. Note that energy is defined even if f only belongs to the Sobolev space $W^{1,2}_{loc}(M)$ of functions differentiable a.e. on M with locally square-integrable partial derivatives.

<u>THEOREM</u> 2.40 (Lower semicontinuity of the energy functional). *Let (f_i) be a sequence of functions in $W^{1,2}_{loc}(M)$ which converges (in $W^{1,2}_{loc}(M)$) to a function f. Then*

$$E(f) \leqslant \liminf_{i\to\infty} E(f_i).$$

DEFINITION 2.41. A function $h \in W^{1,2}_{loc}$ is called *harmonic* if it is *locally energy-minimizing*: For every point $p \in M$ and a small metric ball $B = B(p, r) \subset M$,

$$E(h\big|_B) \leqslant E(u), \quad \forall u : \overline{B} \to \mathbb{R}, u\big|_{\partial B} = h\big|_{\partial B}.$$

Equivalently, for every relatively compact open subset $\Omega \subset M$ with smooth boundary

$$E(h\big|_B) \leqslant E(u), \quad \forall u : \overline{\Omega} \to \mathbb{R}, u\big|_{\partial\Omega} = h\big|_{\partial\Omega}.$$

It turns out that harmonic functions h on M are automatically smooth and, moreover, satisfy the equation $\Delta h = 0$, where Δ is the *Laplace–Beltrami operator* on M:

$$\Delta u = \operatorname{div} \nabla u.$$

In local coordinates (assuming that M is n-dimensional)

$$\Delta u = \sum_{i,j=1}^{n} \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left(g^{ij} \sqrt{|g|} \frac{\partial u}{\partial x_j} \right).$$

In terms of the Levi–Civita connection ∇ on M,

$$\Delta(u) = Trace(H(u)), \quad H(u)(X, Y) = \nabla_X \nabla_Y (u) - \nabla_{\nabla_X Y}(u),$$

where X, Y are vector fields on M. In local coordinates, setting

$$H_{ij} = H(u) \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right),$$

we have

$$Trace(H) = \sum_{i,j=1}^{n} g^{ij} H_{ij}.$$

If $M = \mathbb{R}^n$ with the flat metric, then Δ is the usual Laplace operator:

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} u.$$

Exercise 2.42. Work out the formula for Δu in the case of a conformally-Euclidean metric g on an open subset of $\mathbb{R}^n$. Conclude that harmonicity with respect to g is equivalent to harmonicity with respect to the flat metric.

Lastly, if we use the normal (geodesic) coordinates on a Riemannian manifold then

$$\Delta u(p) = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} u(0).$$

A function u on M is called *subharmonic* if

$$\Delta u \geqslant 0.$$

Example 2.43. If $n = 1$ and $M = \mathbb{R}$ then a function is harmonic if and only if it is linear, and is subharmonic if and only if it is convex.

Exercise 2.44. Suppose that $h : M \to \mathbb{R}$ is a harmonic function and $f : \mathbb{R} \to \mathbb{R}$ is a smooth convex function. Then the composition $u = f \circ h$ is subharmonic. Hint: Verify that $\Delta u(p) \geqslant 0$ for every $p \in M$ using normal coordinates on M defined via the exponential map $\exp_p : T_pM \to M$. This reduces the problem to a Euclidean computation.

Theorem 2.45 (Maximum Principle). *Suppose that M is connected, $\Omega \subset M$ is a relatively compact subset with smooth boundary and $h : M \to \mathbb{R}$ is a harmonic function. Then $h\big|_{\overline{\Omega}}$ attains maximum on the boundary of Ω and, moreover, if $h\big|_{\Omega}$ attains its maximum at a point of Ω, then h is constant.*

Corollary 2.46. *Let $h_i : M \to \mathbb{R}, i = 1, 2$ be two harmonic functions such that $h_1 \leqslant h_2$. Then either $h_1 = h_2$ or $h_1 < h_2$.*

Proof. The difference $h = h_1 - h_2 \leqslant 0$ is also a harmonic function on M. Suppose that the subset $A = \{h_1(x) = h_2(x)\}$ is nonempty. Then for every relatively compact subset with smooth boundary $\Omega \subset M$ and $A \cap \Omega \neq \emptyset$, the maximum of $h\big|_{\Omega}$ is attained on

$$A \cap \Omega.$$

Therefore, $h\big|_{\Omega}$ is identically zero. Taking an exhaustion of M by subsets Ω as above, we conclude that h vanishes on the entire M. ☐

Theorem 2.47 (Li-Schoen's Mean Value Inequality for subharmonic functions). *Suppose that Ricci curvature of the Riemannian n-manifold M is bounded below by a constant r. Then there exists a function $C(n, r, R)$ such that for every nonegative subharmonic function $u : M \to \mathbb{R}$, and normal ball $B(p, R)$, we have*

$$u^2(p) \leqslant C(n, r, R) \int_{B(p,R)} u^2 \mathrm{d}V.$$

As a corollary, one obtains a similar mean value inequality for harmonic functions (without any positivity assumption):

Corollary 2.48. *Suppose that M, p and R satisfy the hypothesis of the previous theorem. Then for every harmonic function $h : M \to \mathbb{R}$ we have*

$$h^2(p) \leqslant \sqrt{C(n, r, R)} \int_{B(p,R)} h^2 \mathrm{d}V.$$

Proof. The composition $u = h^2$ of h with the convex function $x \mapsto x^2$ is subharmonic. Therefore,

$$u^2(p) \leqslant C(n, r, R) \int_{B(p,R)} u^2 \mathrm{d}V.$$

Thus,

$$h^4(p) \leqslant C(n, r, R) \int_{B(p,R)} u^2 \mathrm{d}V \leqslant C(n, r, R) \left(\int_{B(p,R)} u \mathrm{d}V \right)^2,$$

which implies the inequality

$$h^2(p) \leqslant \sqrt{C(n, r, R)} \int_{B(p,R)} h^2 \mathrm{d}V. \quad \square$$

Theorem 2.49 (Yau's gradient estimate). *Suppose that M^n is a complete n-dimensional Riemannian manifold with Ricci curvature $\geqslant a$. Then for every positive harmonic function h on M, every $x \in M$ with $InjRad(x) \geqslant \epsilon$,*

$$|\nabla h(x)| \leqslant C(\epsilon, n) h(x).$$

The following two theorem are a part of the so called *elliptic regularity theory* for solutions of 2nd order elliptic PDEs, see e.g. [**GT83**].

Theorem 2.50 (Derivative bounds). *For every harmonic function h on a manifold of bounded geometry, there exists $L(r)$ such that for every $x \in M$ and every harmonic function $h : M \to \mathbb{R}$, whose restriction to the ball $B(x, r)$ takes values in $[0, 1]$, we have*

$$|\nabla |\nabla h(x)|^2| \le L(r),$$

as long as $\nabla h(x) \neq 0$.

Note that similar estimates hold for higher-order derivatives of harmonic functions; we will only need a bound on the 2nd derivatives.

Theorem 2.51 (Compactness Property). *Suppose that (f_i) is a sequence of harmonic functions on M so that there exists $p \in M$ for which the sequence $(f_i(p))$ is bounded. Then the family of functions (f_i) is precompact in $W^{1,2}_{loc}(M)$. Furthermore, every limit of a subsequence in (f_i) is a harmonic function.*

We will use these properties of harmonic functions in Chapter 19, in the proof of Stallings Theorem on ends of groups via harmonic functions. Since in the proof it suffices to work with 2-dimensional Riemannian manifolds (Riemann surfaces), the properties of harmonic functions we are using follow from more elementary properties of harmonic functions of one complex variable (real parts of holomorphic functions). For instance, the upper bounds on the 1st and 2nd derivatives and Compactness Property follow from Cauchy's integral formula; the maximum principle for harmonic functions follows from the maximum principle for holomorphic functions. Similarly, Corollary 2.48 follows from [Poisson's Integral Formula].

2.10. Spectral interpretation of the Cheeger constant

Let M be a complete connected Riemannian manifold of infinite volume. Then the vector space $V = L^2(M) \cap C^\infty(M)$ contains no nonzero constant functions. We let Δ_M denote the restriction of the Laplace-Beltrami operator to the space V and let $\lambda_1(M)$ be the lowest eigenvalue of Δ_M. The number $\lambda_1(M)$ is also known as the *spectral gap* of the manifold M. The eigenvalue $\lambda_1(M)$ can be computed as

$$(2.5)\quad \inf\left\{\frac{\int_M |\nabla f|^2}{\int_M f^2} \mid f : M \to \mathbb{R} \text{ is smooth, nonzero, with compact support}\right\},$$

see **[CY75]** or Chapter I of **[SY94]**. J. Cheeger proved in **[Che70]** that

$$\lambda_1(M) \geqslant \frac{1}{4}h^2(M),$$

where $h(M)$ is the Cheeger constant of M. Even though Cheeger's original result was formulated for compact manifolds, his argument works for noncompact manifolds as well, see **[SY94]**. Cheeger's inequality is complemented by the following inequality due to P. Buser (see **[Bus82]**, or **[SY94]**) which holds for all complete Riemannian manifolds whose Ricci curvature is bounded below by some $a \in \mathbb{R}$:

$$\lambda_1(M) \leqslant \alpha h(M) + \beta h^2(M),$$

for some $\alpha = \alpha(a), \beta = \beta(a)$. Combined, Cheeger and Buser inequalities imply

THEOREM 2.52. $h(M) = 0 \iff \lambda_1(M) = 0$.

2.11. Comparison geometry

In the setting of general metric spaces it is still possible to define a notion of (upper and lower bound for the) sectional curvature, which, moreover, coincide with the standard ones for Riemannian manifolds. This is done by comparing geodesic triangles in a metric space to geodesic triangles in a *model space of constant curvature*. In what follows, we only discuss the metric definition of upper bound for the sectional curvature, the lower bound case is similar but will not be used in this book.

2.11.1. Alexandrov curvature and $CAT(\kappa)$ spaces. For a real number $\kappa \in \mathbb{R}$, we denote by X_κ the *model surface of constant curvature* κ. If $\kappa = 0$ then X_κ is the Euclidean plane. If $\kappa < 0$ then X_κ will be discussed in detail in Chapter 8, it is the *upper half-plane with the rescaled hyperbolic metric:*

$$X_\kappa = \left(\mathbf{U}^2, |\kappa|^{-1}\frac{dx^2+dy^2}{y^2}\right).$$

If $\kappa > 0$ then X_κ is the 2–dimensional sphere $S\left(0, \frac{1}{\sqrt{\kappa}}\right)$ in $\mathbb{R}^3$ with the Riemannian metric induced from $\mathbb{R}^3$.

Let X be a geodesic metric space, and let Δ be a geodesic triangle in X. Given $\kappa > 0$ we say that Δ is κ*–compatible* if its perimeter is at most $\frac{2\pi}{\sqrt{\kappa}}$. By default, every triangle is κ–compatible for $\kappa \leqslant 0$.

We will prove later on (see §8.11) the following:

LEMMA 2.53. *Let $\kappa \in \mathbb{R}$ and let $a \leqslant b \leqslant c$ be three numbers such that $c \leqslant a+b$ and $a+b+c < \frac{2\pi}{\sqrt{\kappa}}$ if $\kappa > 0$. Then there exists a geodesic triangle in X_κ with side-lengths a, b and c, and this triangle is unique up to congruence.*

Therefore, for every $\kappa \in \mathbb{R}$ and every κ–compatible triangle $\Delta = \Delta(A, B, C) \subset X$ with vertices $A, B, C \in X$ and lengths a, b, c of the opposite sides, there exists a triangle (unique, up to congruence)

$$\tilde{\Delta}(\tilde{A}, \tilde{B}, \tilde{C}) \subset X_\kappa$$

with the side-lengths a, b, c. The triangle $\tilde{\Delta}(\tilde{A}, \tilde{B}, \tilde{C})$ is called the *κ–comparison triangle* or a *κ–Alexandrov triangle.*

For every point P on, say, the side AB of Δ, we define the *κ–comparison point* $\tilde{P} \in \tilde{A}\tilde{B}$, such that

$$d(A, P) = d(\tilde{A}, \tilde{P}).$$

Definition 2.54. We say that *the triangle Δ is $CAT(\kappa)$* if it is κ–compatible and for every pair of points P and Q on the triangle, their κ–comparison points $\tilde{P}, \tilde{Q}$ satisfy

$$\mathrm{dist}_{X_\kappa}\left(\tilde{P}, \tilde{Q}\right) \geqslant \mathrm{dist}_X\left(P, Q\right) .$$

Definition 2.55.
(1) A $CAT(\kappa)$*–domain* in X is an open convex set $U \subset X$, and such that all the geodesic triangles entirely contained in U are $CAT(\kappa)$.
(2) The space X has *Alexandrov curvature at most κ* if it is covered by $CAT(\kappa)$–domains.

Note that a $CAT(\kappa)$–domain U for $\kappa > 0$ must have diameter strictly less than $\frac{\pi}{\sqrt{\kappa}}$. Otherwise, one can construct geodesic triangles in U with two equal edges and the third reduced to a point, with perimeter $\geqslant \frac{2\pi}{\sqrt{\kappa}}$.

The point of Definition 2.55 is that it applies to non-Riemannian metric spaces where such notions as tangent vectors, Riemannian metric, curvature tensor cannot be defined, while one can still talk about curvature being bounded from above by κ.

Proposition 2.56. *Let X be a Riemannian manifold. Its Alexandrov curvature is at most κ if and only if its sectional curvature in every point is $\leqslant \kappa$.*

Proof. The "if" implication follows from the Rauch-Toponogov comparison theorem (see [**dC92**, Proposition 2.5]). For the "only if" implication we refer to [**Rin61**] or to [**GHL04**, Chapter III]. □

Definition 2.57. A metric space X is called a $CAT(\kappa)$*-space* if the entire X is a $CAT(\kappa)$-domain. We will use the definition only for $\kappa \leqslant 0$. A metric space X is said to be a $CAT(-\infty)$*-space* if X is a $CAT(\kappa)$-space for every κ.

Note that for the moment we do not assume X to be metrically complete. This is because there are naturally occurring incomplete $CAT(0)$ spaces, called *Euclidean buildings*, which, nevertheless, are *geodesically complete* (every geodesic segment is contained in a complete geodesic).

Clearly, every Hilbert space is $CAT(0)$.

Exercise 2.58. Let X be a simplicial tree with a path-metric d. Show that (X, d) is $CAT(-\infty)$.

This exercise leads to the following definition:

Definition 2.59. A geodesic metric space X such that for every geodesic triangle in X with the sides xy, yz, zx, the side xy is contained in the union $yz \cup zx$, is called a *real tree*.

Exercise 2.60. 1. Show that a geodesic metric space X is a real tree if and only if X is $CAT(-\infty)$.

2. Consider the following metric space: Take the union of the x-axis in $\mathbb{R}^2$ and all vertical lines $\{x = q\}$, where q's are rational numbers. Equip X with the path-metric induced from $\mathbb{R}^2$. Show that X is an real tree.

We note that real trees are also called $\mathbb{R}$-*trees* or *metric trees* in the literature. A real tree is called *complete* if it is complete as a metric space. While the simplest examples of real trees are given by simplicial trees equipped with their standard path-metrics, we will see in Chapter 9 that other real trees also arise naturally in the geometric group theory. We refer to Section 9.2 for further discussion of real trees.

Exercise 2.61. Let Γ be a connected metric graph with the path-metric. Show that Γ is a $CAT(1)$ if and only if Γ contains no circuits of length $< 2\pi$.

More interesting examples come from polygonal complexes. Their origins lie in two areas of mathematics, going back to 1940s and 1950s:

- The *small cancellation theory*, which is an area of the combinatorial group theory.
- Alexandrov's theory of spaces of curvature bounded from above.

Suppose that X is a connected almost regular 2-dimensional cell complex. We equip X with a path-metric where each 2-face is isometric to a constant curvature κ 2-dimensional polygon with unit edges. This defines structure of a metric graph on the link $Lk(v)$ of each vertex v of X, where each corner c of each 2-face F determines an edge of $Lk(v)$ whose length is the angle of F at c. We refer the reader to [**BH99, Bal95**] for proofs of the following theorem:

Theorem 2.62. *The metric space X has Alexandrov curvature $\leqslant \kappa$ if and only if each connected component of the link $Lk(v)$ of each vertex v of X is a $CAT(1)$ space.*

To make this theorem more concrete, we assume that each 2-dimensional face of X has n edges and for each vertex $v \in X$ the combinatorial length of the shortest circuit in the link $Lk(v)$ is at least m. Then Theorem 2.62 implies:

Corollary 2.63. *1. Suppose that $\kappa = 0$, $n \geqslant 3$ and $m \geqslant 6$, or $n \geqslant 4$ and $m \geqslant 4$, or $n \geqslant 6$ and $m \geqslant 3$. Then X has Alexandrov curvature $\leqslant 0$.*

2. Suppose that $\kappa = -1$, $n \geqslant 3$ and $m \geqslant 7$, or $n \geqslant 4$ and $m \geqslant 5$, or $n \geqslant 6$ and $m \geqslant 4$, or $n \geqslant 7$ and $m \geqslant 3$. Then X has Alexandrov curvature $\leqslant -1$.

In the case of spaces of non-positive curvature one can connect local and global curvature bounds:

Theorem 2.64 (Cartan-Hadamard Theorem). *If X is a simply connected complete metric space with Alexandrov curvature at most κ for some $\kappa \leqslant 0$, then X is a $CAT(\kappa)$–space.*

We refer the reader to [**Bal95**] and [**BH99**] for proofs of this theorem, and a detailed discussion of $CAT(\kappa)$–spaces, with $\kappa \leqslant 0$.

Definition 2.65. Simply-connected complete Riemannian manifolds of sectional curvature $\leqslant 0$ are called *Hadamard manifolds.* Thus, every Hadamard manifold is a $CAT(0)$ space.

An important property of $CAT(0)$-spaces is *convexity of the distance function.* Suppose that X is a geodesic metric space. A function $F : X \times X \to \mathbb{R}$ is said to be *convex* if for every pair of geodesics $\alpha(s), \beta(s)$ in X (which are parameterized with constant, but not necessarily unit, speed), the function

$$f(s) = F(\alpha(s), \beta(s))$$

is a convex function of one variable. Thus, the distance function dist of X is convex, whenever for every pair of geodesics a_0a_1 and b_0b_1 in X, the points $a_s \in a_0a_1$ and $b_s \in b_0b_1$ such that $\text{dist}(a_0, a_s) = s\text{dist}(a_0, a_1)$ and $\text{dist}(b_0, b_s) = s\text{dist}(b_0, b_1)$ satisfy

$$\text{dist}(a_s, b_s) \leqslant (1-s)\text{dist}(a_0, b_0) + s\text{dist}(a_1, b_1)\,. \tag{2.6}$$

Note that in the case of a normed vector space X, a function $f : X \times X \to \mathbb{R}$ is convex if and only if the epi-graph

$$\{(x, y, t) \in X^2 \times \mathbb{R} : f(x, y) \geqslant t\}$$

is convex.

Proposition 2.66. *A geodesic metric space X is $CAT(0)$ if and only if the distance on X is convex.*

Proof. Suppose first that X is $CAT(0)$. Consider two geodesics a_0b_0 and a_1b_1 in X. On the geodesic a_0b_1 consider the point c_s such that $\text{dist}(a_0, c_s) = s\text{dist}(a_0, b_1)$. The fact that the triangle with edges a_0a_1, a_0b_1 and a_1b_1 is $CAT(0)$ and the Thales theorem in $\mathbb{R}^2$, imply that $\text{dist}(a_s, c_s) \leqslant s\text{dist}(a_1, b_1)$. The same argument applied to the triangle with edges a_0b_1, a_0b_0, b_0b_1, implies that $\text{dist}(c_s, b_s) \leqslant (1-s)\text{dist}(a_0, b_0)$. The inequality (2.6) follows from

$$\text{dist}(a_s, b_s) \leqslant \text{dist}(a_s, c_s) + \text{dist}(c_s, b_s)\,.$$

Conversely, assume that (2.6) is satisfied.

In the special case when $a_0 = a_1$, this implies the comparison property in Definition 2.54 when one of the two points P, Q is a vertex of the triangle. When $a_0 = b_0$, (2.6) again implies the comparison property where

$$\frac{\text{dist}(A, P)}{\text{dist}(A, B)} = \frac{\text{dist}(B, Q)}{\text{dist}(B, C)}\,.$$

We now consider the general case of two points $P \in AB$ and $Q \in BC$ such that $\frac{\text{dist}(A,P)}{\text{dist}(A,B)} = s$ and $\frac{\text{dist}(B,Q)}{\text{dist}(B,C)} = t$ with $s < t$. Consider $B' \in AB$ such that $\text{dist}(A, B') = \frac{s}{t}\text{dist}(A, B)$. Then, according to the above, $\text{dist}(B', C) \leqslant \text{dist}(\widetilde{B'}, \widetilde{C})$, and

$$\text{dist}(P, Q) \leqslant t\text{dist}(B', C) \leqslant t\text{dist}(\widetilde{B'}, \widetilde{C}) = \text{dist}(\widetilde{P}, \widetilde{Q})\,.$$

□

Corollary 2.67. *Every $CAT(0)$-space X is uniquely geodesic, i.e., for any two points $p, q \in X$, the (arc-length parameterized) geodesic from p to q is unique.*

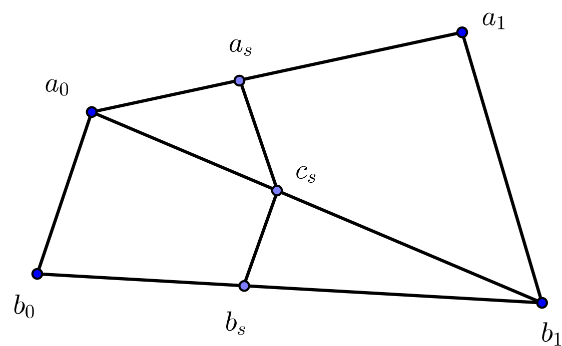


FIGURE 2.1. Argument for convexity of the distance function.

PROOF. It suffices to apply the inequality (2.6) to a geodesic bigon, that is, in the special case when $a_0 = b_0$ and $a_1 = b_1$. $\square$

2.11.2. Cartan's fixed point theorem. Let X be a metric space and $A \subset X$ be a subset. Define the function

$$\rho(x) = \rho_A(x) = \sup_{a \in A} d^2(x, a).$$

PROPOSITION 2.68. *Let X be a complete CAT(0) space. Then for every bounded subset $A \subset X$, the function $\rho = \rho_A$ attains unique minimum in X.*

PROOF. Consider a sequence (x_n) in X such that

$$\lim_{n\to\infty} \rho(x_n) = r = \inf_{x \in X} \rho(x).$$

We claim that the sequence (x_n) is Cauchy. Given $\epsilon > 0$ let $x = x_i, x' = x_j$ be points in this sequence such that

$$r \leqslant \rho(x) < r + \epsilon, \quad r \leqslant \rho(x') < r + \epsilon.$$

Let p be the midpoint of $xx' \subset X$; hence, $r \leqslant \rho(p)$. Let $a \in A$ be such that

$$\rho(p) - \epsilon < d^2(p, a).$$

Consider the Euclidean comparison triangle $\tilde{T} = T(\tilde{x}, \tilde{x}', \tilde{a})$ for the triangle $T(x, x', a)$. In the Euclidean plane we have (by the parallelogram identity (1.6)):

$$d^2(\tilde{x}, \tilde{x}') + 4\, d^2(\tilde{a}, \tilde{p}) = 2\left(d^2(\tilde{a}, \tilde{x}) + d^2(\tilde{a}, \tilde{x}')\right).$$

Applying the comparison inequality for the triangles T and $\tilde{T}$, we obtain:

$$d(a, p) \leqslant d(\tilde{a}, \tilde{p}).$$

Thus:

$$d(x, x')^2 + 4(r - \epsilon) < d^2(x, x') + 4d^2(a, p) \leqslant 2\left(d^2(a, x) + d^2(a, x')\right) <$$

$$2(\rho(x) + \rho(x')) < 4r + 4\epsilon.$$

It follows that

$$d(x, x')^2 < 8\epsilon$$

and, therefore, the sequence (x_n) is Cauchy. By completeness of X, the function ρ attains minimum in X; the same Cauchy argument implies that the point of minimum is unique. □

As a corollary, we obtain a fixed-point theorem for isometric group actions on complete $CAT(0)$ spaces, which was first proven by E. Cartan in the context of Riemannian manifolds of nonpositive curvature and then extended by J. Tits to geodesic metric spaces with convex distance function:

THEOREM 2.69. *Let X be a complete $CAT(0)$ metric space and $G \subset \mathrm{Isom}(X)$ be a subgroup which has bounded orbits: One (equivalently every) subset of the form*

$$G \cdot x = \{g(x) : g \in G\}$$

is bounded. Then G fixes a point in X.

PROOF. Let A denote a (bounded) orbit of G in X and let ρ_A be the corresponding function on X. Then, by uniqueness of the minimum point m of ρ_A, the group G will fix m. □

COROLLARY 2.70. *1. Every finite group action on a complete $CAT(0)$ space has a fixed point. For instance, every action of a finite group on a complete real tree or on a Hilbert space fixes a point.*

2. If G is a compact group acting isometrically and continuously on a Hilbert space $\mathcal{H}$, then G fixes a point in $\mathcal{H}$.

EXERCISE 2.71. Prove that this corollary holds for all real trees T, not necessarily complete ones. Hint: For a finite subset $F \subset T$ consider its *span* T_F , i.e., the union of all geodesic segments connecting points of F. Show that T_F is isometric to a complete metric tree and is G-invariant if F was. In fact, T_F is isometric to a finite metric simplicial complex (which, as a simplicial complex, is isomorphic to a finite simplicial tree).

DEFINITION 2.72. A group G is said to have the *Property FA* if for every isometric action $G \curvearrowright T$ on a complete real tree T, G fixes a point in T.

Thus, all finite groups have the Property FA.

2.11.3. Ideal boundary, horoballs and horospheres. In this section we discuss the notion of the *ideal boundary* of a metric space. This is a particularly useful concept when the metric space is $CAT(0)$, and it generalizes the concept introduced for non-positively curved simply connected Riemannian manifolds by P. Eberlein and B. O'Neill in [**EO73**, Section 1].

Let X be a geodesic metric space. Two geodesic rays ρ_1 and ρ_2 in X are called *asymptotic* if they are at finite Hausdorff distance; equivalently if the function $t \mapsto \mathrm{dist}(\rho_1(t), \rho_2(t))$ is bounded on $[0, \infty)$.

Clearly, being asymptotic is an equivalence relation on the set of geodesic rays in X.

DEFINITION 2.73. The *ideal boundary* of a metric space X is the collection of equivalence classes of geodesic rays. It is usually denoted either by $\partial_\infty X$ or by $X(\infty)$.

An equivalence class $\xi \in \partial_\infty X$ is called an *ideal point* or *point at infinity* of X, and the fact that a geodesic ray ρ is contained in this class is sometimes expressed by the equality $\rho(\infty) = \xi$. When a geodesic ray ρ represents an equivalence class $\xi \in \partial_\infty X$, the ray ρ is said to be *asymptotic to* ξ.

The space of geodesic rays in X has a natural compact-open topology, or, equivalently, topology of uniform convergence on compacts (recall that we regard geodesic rays as maps from $[0, \infty)$ to X). Thus, we topologize $\partial_\infty X$ by giving it the quotient topology τ.

EXERCISE 2.74. Every isometry $g : X \to X$ induces a homeomorphism $g_\infty : \partial_\infty X \to \partial_\infty X$.

This exercise explains why we consider rays emanating from different points of X: Otherwise, most isometries of X would not act on $\partial_\infty X$.

Convention. From now on, in this section, we assume that X is a complete $CAT(0)$ metric space.

LEMMA 2.75. *If X is locally compact then for every point $x \in X$ and every point $\xi \in \partial_\infty X$ there exists a unique geodesic ray ρ with $\rho(0) = x$ and $\rho(\infty) = \xi$. We will also use the notation $x\xi$ for the ray ρ.*

PROOF. Let $r : [0, \infty) \to X$ be a geodesic ray with $r(\infty) = \xi$. For every $n \in \mathbb{N}$, according to Corollary 2.67, there exists a unique geodesic $\mathfrak{g}_n$ joining x and $r(n)$. The convexity of the distance function implies that every $\mathfrak{g}_n$ is at Hausdorff distance $\text{dist}(x, r(0))$ from the segment of r between $r(0)$ and $r(n)$.

By the Arzela-Ascoli Theorem, a subsequence $\mathfrak{g}_{n_k}$ of geodesic segments converges in the compact-open topology to a geodesic ray ρ with $\rho(0) = x$. Moreover, ρ is at Hausdorff distance $\text{dist}(x, r(0))$ from r.

Assume that ρ_1 and ρ_2 are two asymptotic geodesic rays with $\rho_1(0) = \rho_2(0) = x$. Let M be such that $\text{dist}(\rho_1(t), \rho_2(t)) \leqslant M$, for every $t \geqslant 0$. Consider $t \in [0, \infty)$, and $\varepsilon > 0$ arbitrarily small. Convexity of the distance function implies that

$$\text{dist}(\rho_1(t), \rho_2(t)) \leqslant \varepsilon \text{dist}(\rho_1(t/\varepsilon), \rho_2(t/\varepsilon)) \leqslant \varepsilon M .$$

It follows that $\text{dist}(\rho_1(t), \rho_2(t)) = 0$ and, hence, $\rho_1 = \rho_2$. □

In particular, for a fixed point $p \in X$ one can identify the set $\overline{X} := X \sqcup \partial_\infty X$ with the set of geodesic segments and rays with initial point p. In what follows, we will equip $\overline{X}$ with the topology induced from the compact-open topology on the space of these segments and rays.

EXERCISE 2.76.
(1) Prove that the embedding $X \hookrightarrow \overline{X}$ is a homeomorphism to its image.
(2) Prove that the topology on $\overline{X}$ is independent of the chosen basepoint p. In other words, given p and q two points in X, the map $[p, x] \mapsto [q, x]$ (with $x \in \overline{X}$) is a homeomorphism.

(3) In the special case when X is a Hadamard manifold, show that for every point $p \in X$, the ideal boundary $\partial_\infty X$ is homeomorphic to the unit sphere S in the tangent space T_pM *via* the map

$$v \in S \subset T_pM \to \exp_p(\mathbb{R}_+ v) \in \partial_\infty X.$$

An immediate consequence of the Arzela–Ascoli Theorem is that $\overline{X}$ is compact, provided that X is locally compact.

Consider a geodesic ray $r : [0, \infty) \to X$, and an arbitrary point $x \in X$. The function $t \mapsto \text{dist}(x, r(t)) - t$ is decreasing (due to the triangle inequality) and bounded from below by $-\text{dist}(x, r(0))$. Therefore, there exists a limit

$$b_r(x) := \lim_{t\to\infty} [\text{dist}(x, r(t)) - t] \ . \tag{2.7}$$

Definition 2.77. The function $b_r : X \to \mathbb{R}$ thus defined, is called the *Busemann function for the ray r.*

For a proof of the next result see e.g. **[Bal95]**, Chapter 2, Proposition 2.5.

Theorem 2.78. *If r_1 and r_2 are two asymptotic rays then $b_{r_1} - b_{r_2}$ is a constant function.*

In particular, it follows that the collections of sublevel sets and the level sets of a Busemann function do not depend on the ray r, but only on the point at infinity that r represents.

Exercise 2.79. Show that b_r is linear with slope -1 along the ray r. In particular,

$$\lim_{t\to\infty} b_r(t) = -\infty.$$

Definition 2.80. A sublevel set of a Busemann function, $b_r^{-1}(-\infty, a]$ is called a *(closed) horoball with center* $\xi = r(\infty)$; we denote such horoballs as $\overline{B}(\xi)$ or $\overline{B}(r)$. A level set $b_r^{-1}(a)$ of a Busemann function is called a *horosphere with center* ξ, it is denoted $\Sigma(\xi)$. In the case when X is 2-dimensional, horospheres are called *horocycles*. Lastly, an open sublevel set $b_r^{-1}(-\infty, a)$ is called an *open horoball with center* $\xi = r(\infty)$, and denoted $B(\xi)$ or $B(r)$.

Informally, one can think informally of horoballs $B(\xi)$ and horospheres $\Sigma(\xi)$ as metric balls and metric spheres of infinite radii in X, centered at ξ, whose radii are determined by the choice of the Busemann function b_r (which is determined only up to a constant) and by the choice of the value a of b_r.

Lemma 2.81. *Let r be a geodesic ray and let B be the open horoball $b_r^{-1}(-\infty, 0)$. Then $B = \bigcup_{t\geqslant 0} B(r(t), t)$.*

Proof. Indeed, if for a point x,

$$b_r(x) = \lim_{t\to\infty} [\text{dist}(x, r(t)) - t] < 0,$$

then, for some sufficiently large t, $\text{dist}(x, r(t)) - t < 0$. Whence $x \in B(r(t), t)$.

Conversely, suppose that $x \in X$ is such that for some $s \geqslant 0$,

$$\text{dist}(x, r(s)) - s = \delta_s < 0.$$

Then, because the function $t \mapsto \operatorname{dist}(x, r(t)) - t$ is decreasing, it follows that for every $t \geqslant s$,

$$\operatorname{dist}(x, r(t)) - t \leqslant \delta_s\,.$$

Whence, $b_r(x) \leqslant \delta_s < 0$. ☐

LEMMA 2.82. *Let X be a $CAT(0)$ space. Then every Busemann function on X is convex and 1-Lipschitz.*

PROOF. Recall that the distance function on any metric space is 1-Lipschitz. Since Busemann functions are limits of normalized distance functions, it follows that Busemann functions are 1-Lipschitz as well. (This part does not require the $CAT(0)$ assumption.) Similarly, since the distance function is convex, Busemann functions are also convex as limits of normalized distance functions. ☐

Furthermore, if X is a Hadamard manifold, then every Busemann function b_r is smooth, with gradient of constant norm 1, see [**BGS85**].

LEMMA 2.83. *Assume that X is a complete $CAT(0)$ space. Then:*

- *Open and closed horoballs in X are convex sets.*
- *A closed horoball is the closure of an open horoball.*

PROOF. The first property follows immediately from the convexity of Busemann functions. Let $f = b_r$ be a Busemann function. Consider the closed horoball

$$\overline{B} = \{x : f(x) \leqslant t\}.$$

Since this horoball is a closed subset of X, it contains the closure of the open horoball

$$B = \{x : f(x) < t\}.$$

Suppose now that $f(x) = t$. Since $\lim_{s\to\infty} f(s) = -\infty$, there exists s such that $f(r(s)) < t$. Convexity of f implies that for $z = r(s)$,

$$f(y) < f(x) = t, \quad \forall y \in xz \setminus \{x\}.$$

Therefore, x belongs to the closure of the open horoball B, which implies that $\overline{B}$ is the closure of B. ☐

EXERCISE 2.84. 1. Suppose that X is the Euclidean space $\mathbb{R}^n$, r is the geodesic ray in X with $r(0) = 0$ and $r'(0) = u$, where u is a unit vector. Show that

$$b_r(x) = -x \cdot u.$$

In particular, closed (resp. open) horoballs in X are closed (resp. open) half-spaces, while horospheres are hyperplanes.

2. Construct an example of a proper $CAT(0)$ space and an open horoball $B \subset X$, $B \neq X$, so that B is not equal to the interior of the closed horoball $\overline{B}$. Can this happen in the case of Hadamard manifolds?

CHAPTER 3

Groups and their actions

This chapter covers some basic group–theoretic material as well as group actions on topological and metric spaces. We also briefly discuss Lie groups, group cohomology and its relation to the structural theory of groups. For detailed treatment of the basic group theory we refer to [**Hal76**] and [**LS77**].

Notation and terminology. With very few exceptions, in a group G we use the multiplication sign $\cdot$ to denote its binary operation. We denote its identity element either by e or by 1. We denote the inverse of an element $g \in G$ by g^{-1}. Given a subset S in G we denote by S^{-1} the subset $\{g^{-1} \mid g \in S\}$. Note that for abelian groups the neutral element is usually denoted 0, the inverse of x by $-x$ and the binary operation by $+$. We will use the notation

$$[x, y] = xyx^{-1}y^{-1}$$

for the *commutator* of elements x, y of a group G.

A surjective homomorphism is called an *epimorphism*, while an injective homomorphism is called a *monomorphism.* If two groups G and G' are isomorphic we write $G \simeq G'$. An isomorphism of groups $\varphi : G \to G$ is also called an *automorphism.* In what follows, we denote by $\mathrm{Aut}(G)$ the group of automorphisms of G.

We use the notation $H < G$ or $H \leqslant G$ to denote that H is a subgroup in G. Given a subgroup H in G:

- the *order* $|H|$ of H is its cardinality;
- the *index* of H in G, denoted $|G : H|$, is the common cardinality of the quotients G/H and $H\backslash G$.

The *order* of an element g in a group $(G, \cdot)$ is the order of the subgroup $\langle g \rangle$ of G generated by g. In other words, the order of g is the minimal positive integer n such that $g^n = 1$. If no such integer exists, then g is said to be *of infinite order.* In this case, $\langle g \rangle$ is isomorphic to $\mathbb{Z}$.

For every positive integer m we denote by $\mathbb{Z}_m$ *the cyclic group of order* m, $\mathbb{Z}/m\mathbb{Z}$. Given $x, y \in G$ we let x^y denote the conjugation of x by y, i.e., yxy^{-1}.

3.1. Subgroups

Given two subsets A, B in a group G we denote by AB the subset

$$\{ab \ : \ a \in A,\, b \in B\} \subset G.$$

Similarly, we will use the notation

$$A^{-1} = \{a^{-1} : a \in A\}.$$

A *normal subgroup* K in G is a subgroup such that for every $g \in G$, $gKg^{-1} = K$ (equivalently $gK = Kg$). We use the notation $K \lhd G$ to denote that K is a normal subgroup in G. When H and K are subgroups of G and either H or K is a normal subgroup of G, the subset $HK \subset G$ becomes a subgroup of G.

A subgroup K of a group G is called *characteristic* if for every automorphism $\phi : G \to G$, $\phi(K) = K$. Note that every characteristic subgroup is normal (since conjugation is an automorphism). But not every normal subgroup is characteristic:

EXAMPLE 3.1. Let G be the group $(\mathbb{Z}^2, +)$. Since G is abelian, every subgroup is normal. But, for instance, the subgroup $\mathbb{Z} \times \{0\}$ is not invariant under the automorphism $\phi : \mathbb{Z}^2 \to \mathbb{Z}^2$, $\phi(m, n) = (n, m)$.

DEFINITION 3.2. The *center* $Z(G)$ of a group G is defined as the subgroup consisting of elements $h \in G$ so that $[h, g] = 1$ for each $g \in G$.

It is easy to see that the center is a characteristic subgroup of G.

DEFINITION 3.3. A *subnormal descending series* in a group G is a series

$$G = N_0 \rhd N_1 \rhd \cdots \rhd N_n \rhd \cdots$$

such that N_{i+1} is a normal subgroup in N_i for every $i \geqslant 0$.

If all N_i's are normal subgroups of G, then the series is called *normal*.

A subnormal series of a group is called a *refinement* of another subnormal series if the terms of the latter series all occur as terms in the former series.

The following is a basic result in group theory:

LEMMA 3.4. *If G is a group, $N \lhd G$, and $A \lhd B < G$, then BN/AN is isomorphic to $B/A(B \cap N)$.*

DEFINITION 3.5. Two subnormal series

$$G = A_0 \rhd A_1 \rhd \ldots \rhd A_n = \{1\} \text{ and } G = B_0 \rhd B_1 \rhd \ldots \rhd B_m = \{1\}$$

are called *isomorphic* if $n = m$ and there exists a bijection between the sets of partial quotients $\{A_i/A_{i+1} \mid i = 1, \ldots, n-1\}$ and $\{B_i/B_{i+1} \mid i = 1, \ldots, n-1\}$ such that the corresponding quotients are isomorphic.

LEMMA 3.6. *Any two finite subnormal series*

$$G = H_0 \geqslant H_1 \geqslant \ldots \geqslant H_n = \{1\} \text{ and } G = K_0 \geqslant K_1 \geqslant \ldots \geqslant K_m = \{1\}$$

possess isomorphic refinements.

PROOF. Define $H_{ij} = (K_j \cap H_i)H_{i+1}$. The following is a subnormal series

$$H_{i0} = H_i \geqslant H_{i1} \geqslant \ldots \geqslant H_{im} = H_{i+1} \,.$$

When inserting all these in the series of H_i one obtains the required refinement.

Likewise, define $K_{rs} = (H_s \cap K_r)K_{r+1}$ and by inserting the series

$$K_{r0} = K_r \geqslant K_{r1} \geqslant \ldots \geqslant K_{rn} = K_r$$

in the series of K_r, we define its refinement.

According to Lemma 3.4

$$H_{ij}/H_{ij+1} = (K_j \cap H_i)H_{i+1}/(K_{j+1} \cap H_i)H_{i+1} \simeq K_j \cap H_i/(K_{j+1} \cap H_i)(K_j \cap H_{i+1}) \,.$$

Similarly, one proves that $K_{ji}/K_{ji+1} \simeq K_j \cap H_i/(K_{j+1} \cap H_i)(K_j \cap H_{i+1})$. $\square$

Definition 3.7. A group G is a *torsion group* if all its elements have finite order.

A group G is said to be *without torsion* (or *torsion-free*) if all its non-trivial elements have infinite order.

Note that the subset $\mathrm{Tor}\, G = \{g \in G \mid g \text{ of finite order}\}$ of the group G, sometimes called the *torsion* of G, is in general not a subgroup.

Definition 3.8. A group G is said to have property * *virtually* if some finite index subgroup H of G has the property *.

For instance, a group is *virtually torsion-free* if it contains a torsion-free subgroup of finite index, a group is *virtually abelian* if it contains an abelian subgroup of finite index and a *virtually free group* is a group which contains a free subgroup of finite index.

Remark 3.9. Note that this terminology widely used in group theory is not entirely consistent with the notion of *virtually isomorphic groups*, which involves not only taking finite index subgroups but also quotients by finite normal subgroups.

The following properties of finite index subgroups will be useful.

Lemma 3.10. *If $N \lhd H$ and $H \lhd G$, N of finite index in H and H finitely generated, then N contains a finite index subgroup K which is normal in G.*

Proof. By hypothesis, the quotient group $F = H/N$ is finite. For an arbitrary $g \in G$ the conjugation by g is an automorphism of H, hence H/gNg^{-1} is isomorphic to F. A homomorphism $H \to F$ is completely determined by the images in F of elements of a finite generating set of H. Therefore there are finitely many such homomorphisms, and finitely many possible kernels of them. Thus, the set of subgroups gNg^{-1}, $g \in G$, forms a finite list $N, N_1, .., N_k$. The subgroup $K = \bigcap_{g\in G} gNg^{-1} = N \cap N_1 \cap \cdots \cap N_k$ is normal in G and has finite index in N, since each of the subgroups $N_1, \ldots, N_k$ has finite index in H. □

Proposition 3.11. *Let G be a finitely generated group. Then:*

(1) *For every $n \in \mathbb{N}$ there exist finitely many subgroups of index n in G.*

(2) *Every finite index subgroup H in G contains a subgroup K which is finite index and characteristic in G.*

Proof. (1) Let $H \leqslant G$ be a subgroup of index n. We list the left cosets of H:

$$H = g_1 \cdot H, g_2 \cdot H, \ldots, g_n \cdot H,$$

and label these cosets by the numbers $\{1, \ldots, n\}$. The action by left multiplication of G on the set of left cosets of H defines a homomorphism $\phi : G \to S_n$ such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$ and H is the inverse image under ϕ of the stabilizer of 1 in S_n. Note that there are $(n-1)!$ ways of labeling the left cosets, each defining a different homomorphism with these properties.

Conversely, if $\phi : G \to S_n$ is such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$, then $G/\phi^{-1}(\mathrm{Stab}\,(1))$ has cardinality n.

Since the group G is finitely generated, a homomorphism $\phi : G \to S_n$ is determined by the image of a generating finite set of G, hence there are finitely many distinct such homomorphisms. The number of subgroups of index n in H is equal to the number η_n of homomorphisms $\phi : G \to S_n$ such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$, divided by $(n-1)!$.

(2) Let H be a subgroup of index n. For every automorphism $\varphi : G \to G$, $\varphi(H)$ is a subgroup of index n. According to (1) the set $\{\varphi(H) \mid \varphi \in \mathrm{Aut}\,(G)\}$ is finite, equal $\{H, H_1, \ldots, H_k\}$. It follows that

$$K = \bigcap_{\varphi \in \mathrm{Aut}\,(G)} \varphi(H) = H \cap H_1 \cap \ldots \cap H_k.$$

Then K is a characteristic subgroup of finite index in H hence in G. □

Exercise 3.12. Does the conclusion of Proposition 3.11 still hold for groups which are not finitely generated?

Let S be a subset in a group G, and let $H \leqslant G$ be a subgroup. The following are equivalent:

(1) H is the smallest subgroup of G containing S;

(2) $H = \bigcap_{S \subset G_1 \leqslant G} G_1$;

(3) $H = \left\{ s_1 s_2 \cdots s_n \ : \ n \in \mathbb{N}, s_i \in S \text{ or } s_i^{-1} \in S \text{ for every } i \in \{1, 2, \ldots, n\} \right\}$.

The subgroup H satisfying any of the above is denoted $H = \langle S \rangle$ and is said to be *generated by* S. The subset $S \subset H$ is called a *generating set* of H. The elements in S are called *generators* of H.

When S consists of a single element x, $\langle S \rangle$ is usually written as $\langle x \rangle$; it is the cyclic subgroup consisting of powers of x.

We say that a normal subgroup $K \lhd G$ is *normally generated* by a set $R \subset K$ if K is the smallest normal subgroup of G which contains R, i.e.,

$$K = \bigcap_{R \subset N \lhd G} N\,.$$

We will use the notation

$$K = \langle\langle R \rangle\rangle$$

for this subgroup. The subgroup K is also called the *normal closure* or the *conjugate closure* of R in G. Other notations for K which appear in the literature are R^G and $\langle R \rangle^G$.

3.2. Virtual isomorphisms of groups and commensurators

In this section we consider a weakening of the notion of a group isomorphism to the one of a *virtual isomorphism*. This turns out to be the right *algebraic* concept in the context of geometric group theory.

Definition 3.13. (1) Two groups G_1 and G_2 are called *co-embeddable* if there exist injective group homomorphisms $G_1 \to G_2$ and $G_2 \to G_1$.

(2) The groups G_1 and G_2 are *commensurable* if there exist finite index subgroups $H_i \leqslant G_i$, $i = 1, 2$, such that H_1 is isomorphic to H_2.

An isomorphism $\varphi : H_1 \to H_2$ is called an *abstract commensurator* of G_1 and G_2.

(3) We say that two groups G_1 and G_2 are *virtually isomorphic* (abbreviated as VI) if there exist finite index subgroups $H_i \subset G_i$ and finite normal subgroups $F_i \triangleleft H_i$, $i = 1, 2$, so that the quotients H_1/F_1 and H_2/F_2 are isomorphic.

An isomorphism $\varphi : H_1/F_1 \to H_2/F_2$ is called a *virtual isomorphism* of G_1 and G_2. When $G_1 = G_2$, φ is called *virtual automorphism*.

EXAMPLE 3.14. All countable free groups are co-embeddable. However, a free group of infinite rank is not virtually isomorphic to a free group of infinite rank.

PROPOSITION 3.15. *All the relations in Definition 3.13 are equivalence relation between groups.*

PROOF. The fact that co-embeddability is an equivalence relation is immediate. It suffices to prove that virtual isomorphism is an equivalence relation. The only non-obvious property is transitivity. We need:

LEMMA 3.16. *Let F_1, F_2 be normal finite subgroups of a group G. Then their normal closure $F = \langle\langle F_1, F_2\rangle\rangle$ (i.e., the smallest normal subgroup of G containing F_1 and F_2) is again finite.*

PROOF. Let $f_1 : G \to G_1 = G/F_1$, $f_2 : G_1 \to G_1/f_1(F_2)$ be the quotient maps. Since the kernel of each f_1, f_2 is finite, it follows that the kernel of $f = f_2 \circ f_1$ is finite as well. On the other hand, the kernel of f is clearly the subgroup F. □

Suppose now that G_1 is VI to G_2 and G_2 is VI to G_3. Then we have

$$F_i \triangleleft H_i < G_i, |G_i : H_i| < \infty, |F_i| < \infty, \quad i = 1, 2, 3,$$

and

$$F'_2 \triangleleft H'_2 < G_2, |G_2 : H'_2| < \infty, |F'_2| < \infty,$$

so that

$$H_1/F_1 \cong H_2/F_2, \quad H'_2/F'_2 \cong H_3/F_3.$$

The subgroup $H''_2 := H_2 \cap H'_2$ has finite index in G_2. By the above lemma, the normal closure in H''_2

$$K_2 := \langle\langle F_2 \cap H''_2, F'_2 \cap H''_2\rangle\rangle$$

is finite. We have quotient maps

$$f_i : H''_2 \to C_i = f_i(H''_2) \leqslant H_i/F_i, i = 1, 3,$$

with finite kernels and cokernels. The subgroups $E_i := f_i(K_2)$, are finite and normal in C_i, $i = 1, 3$. We let $H'_i, F'_i \subset H_i$ denote the preimages of C_i and E_i under the quotient maps $H_i \to H_i/F_i$, $i = 1, 3$. Then $|F'_i| < \infty, |G_i : H'_i| < \infty, i = 1, 3$. Lastly,

$$H'_i/F'_i \cong C_i/E_i \cong H''_2/K_2, i = 1, 3.$$

Therefore, G_1, G_3 are virtually isomorphic. □

Given a group G, we define $VI(G)$ as the set of equivalence classes of virtual automorphisms of G with respect to the following equivalence relation. Two virtual automorphisms of G, $\varphi : H_1/F_1 \to H_2/F_2$ and $\psi : H'_1/F'_1 \to H'_2/F'_2$, are *equivalent* if for $i = 1, 2$, there exist $\widetilde{H}_i$, a finite index subgroup of $H_i \cap H'_i$, and $\widetilde{F}_i$, a normal subgroup in $\widetilde{H}_i$ containing the intersections $\widetilde{H}_i \cap F_i$ and $\widetilde{H}_i \cap F'_i$, such that φ and ψ induce the same automorphism from $\widetilde{H}_1/\widetilde{F}_1$ to $\widetilde{H}_2/\widetilde{F}_2$.

Lemma 3.16 implies that the composition induces a binary operation on $VI(G)$, and that $VI(G)$ with this operation becomes a group, called *the group of virtual automorphisms of* G.

Let $Comm(G)$ be the set of equivalence classes of virtual automorphisms of G with respect to the equivalence relation defined as above, with the normal subgroups F_i and F_i' trivial. As in the case of $VI(G)$, the set $Comm(G)$, endowed with the binary operation defined by the composition, becomes a group, called the *abstract commensurator of the group* G.

Let Γ be a subgroup of a group G. The *commensurator of* Γ *in* G, denoted by $\mathrm{Comm}_G(\Gamma)$, is the set of elements g in G such that the conjugation by g defines an abstract commensurator of Γ: $g\Gamma g^{-1} \cap \Gamma$ has finite index in both Γ and $g\Gamma g^{-1}$.

EXERCISE 3.17. Show that $Comm_G(\Gamma)$ is a subgroup of G.

EXERCISE 3.18. Show that for $G = SL(n, \mathbb{R})$ and $\Gamma = SL(n, \mathbb{Z})$, $Comm_G(\Gamma)$ contains $SL(n, \mathbb{Q})$.

3.3. Commutators and the commutator subgroup

Recall that the commutator of two elements x, y of a group G is defined as $[x, y] = xyx^{-1}y^{-1}$. Thus:

- two elements x, y commute, i.e., $xy = yx$, if and only if $[x, y] = 1$.
- $xy = [x, y]yx$.

Thus, the commutator $[x, y]$ 'measures the degree of non-commutation' of the elements h and k. In Lemma 11.30 we will prove some further properties of commutators.

Let H, K be two subgroups of G. We denote by $[H, K]$ the subgroup of G generated by all commutators $[h, k]$ with $h \in H$, $k \in K$.

DEFINITION 3.19. The *commutator subgroup (or* derived subgroup*) of* G is the subgroup $G' = [G, G]$. As above, we may say that the commutator subgroup G' of G 'measures the degree of non-commutativity' of the group G.

A group G is *abelian* if every two elements of G *commute*, i.e., $ab = ba$ for all $a, b \in G$.

EXERCISE 3.20. Suppose that S is a generating set of G. Then G is abelian if and only if $[a, b] = 1$ for all $a, b \in S$.

PROPOSITION 3.21.
(1) *G' is a characteristic subgroup of G;*
(2) *G is* abelian *if and only if $G' = \{1\}$;*
(3) *$G_{ab} = G/G'$ is an abelian group (called the* abelianization *of G);*
(4) *if $\varphi : G \to A$ is a homomorphism to an abelian group A, then φ factors through the abelianization: Given the quotient map $p : G \to G_{ab}$, there exists a homomorphism $\overline{\varphi} : G_{ab} \to A$ such that $\varphi = \overline{\varphi} \circ p$.*

PROOF. (1) The set $S = \{[x, y] \mid x, y \in G\}$ is a generating set of G' and for every automorphism $\psi : G \to G$, $\psi(S) = S$.

Part (2) follows from the equivalence $xy = yx \Leftrightarrow [x, y] = 1$, and (3) is an immediate consequence of (2).

Part (4) follows from the fact that $\varphi(S) = \{1\}$. □

Recall that the *finite dihedral group* of order $2n$, denoted by D_{2n} or $I_2(n)$, is the group of symmetries of the regular Euclidean n-gon, i.e., the group of isometries of the unit circle $\mathbb{S}^1 \subset \mathbb{C}$ generated by the rotation $r(z) = e^{\frac{2\pi i}{n}} z$ and the reflection $s(z) = \bar{z}$. Likewise, the *infinite dihedral group* D_∞ is the group of isometries of $\mathbb{Z}$ (with the metric induced from $\mathbb{R}$); the group D_∞ is generated by the translation $t(x) = x + 1$ and the symmetry $s(x) = -x$.

EXERCISE 3.22. Find the commutator subgroup and the abelianization for the finite dihedral group D_{2n} and for the infinite dihedral group D_∞.

EXERCISE 3.23. Let S_n (the symmetric group on n symbols) be the group of permutations of the set $\{1, 2, \ldots, n\}$, and $A_n \subset S_n$ be the alternating subgroup, consisting of even permutations.

(1) Prove that for every $n \notin \{2, 4\}$ the group A_n is generated by the set of cycles of length 3.
(2) Prove that if $n \geqslant 3$, then for every cycle σ of length 3 there exists $\rho \in S_n$ such that $\sigma^2 = \rho\sigma\rho^{-1}$.
(3) Use (1) and (2) to find the commutator subgroup and the abelianization for A_n and for S_n.
(4) Find the commutator subgroup and the abelianization for the group H of permutations of $\mathbb{Z}$ defined in Example 4.8.

Note that it is not necessarily true that the commutator subgroup G' of G consists entirely of commutators $\{[x, y] : x, y \in G\}$. However, occasionally, every element of the derived subgroup is indeed a single commutator. For instance, every element of the alternating group $A_n < S_n$ is the commutator in S_n, see [**Ore51**].

This leads to an interesting invariant (of geometric flavor) called the *commutator norm* (or *commutator length*) $\ell_c(g)$ of $g \in G'$, which is the least number k so that g can be expressed as a product

$$g = [x_1, y_1] \cdots [x_k, y_k],$$

as well as the *stable commutator norm* of g:

$$\limsup_{n\to\infty} \frac{\ell_c(g^n)}{n}.$$

See [**Bav91, Cal08, Cal09**] for further details. For instance, if G is the free group on two generators (see Definition 4.20), then every nontrivial element of G' has stable commutator norm greater than 1.

3.4. Semi-direct products and short exact sequences

Let $G_i, i \in I$, be a collection of groups. The *direct product* of these groups, denoted

$$G = \prod_{i\in I} G_i$$

is the Cartesian product of sets G_i with the group operation given by

$$(a_i) \cdot (b_i) = (a_i b_i).$$

Note that each group G_i is the quotient of G by the (normal) subgroup

$$\prod_{j\in I\setminus\{i\}} G_j.$$

A group G is said to *split* as a direct product of its normal subgroups $N_i \lhd G, i = 1, \dots, k$, if one of the following equivalent statements holds:

- $G = N_1 \cdots N_k$ and $N_i \cap N_j = \{1\}$ for all $i \neq j$;
- for every element g of G there exists a unique k-tuple

$$(n_1, \dots, n_k), n_i \in N_i, i = 1, \dots, k$$

  such that $g = n_1 \cdots n_k$.

Then, G is isomorphic to the direct product $N_1 \times \ldots \times N_k$. Thus, finite direct products G can be defined either *extrinsically*, using groups N_i as quotients of G, or *intrinsically*, using normal subgroups N_i of G.

Similarly, one defines *semidirect products* of two groups, by taking the above *intrinsic* definition and relaxing the normality assumption:

Definition 3.24. (1) (*with the ambient group as given data*) A group G is said to *split as a semidirect product of two subgroups* N and H, which is denoted by $G = N \rtimes H$, if and only if N is a *normal subgroup* of G, H is a *subgroup* of G, and one of the following equivalent statements holds:

- $G = NH$ and $N \cap H = \{1\}$;
- $G = HN$ and $N \cap H = \{1\}$;
- for every element g of G there exists a unique $n \in N$ and $h \in H$ such that $g = nh$;
- for every element g of G there exists a unique $n \in N$ and $h \in H$ such that $g = hn$;
- there exists *a retraction* $G \to H$, i.e., a homomorphism which restricts to the identity on H, and whose kernel is N.

Observe that the map $\varphi : H \to \mathrm{Aut}\,(N)$ defined by $\varphi(h)(n) = hnh^{-1}$, is a group homomorphism.

(2) (*with the quotient groups as given data*) Given any two groups N and H (not necessarily subgroups of the same group) and a group homomorphism $\varphi : H \to \mathrm{Aut}\,(N)$, one can define a new group $G = N \rtimes_\varphi H$ which is a semidirect product of a copy of N and a copy of H in the above sense, defined as follows. As a set, $N \rtimes_\varphi H$ is defined as the cartesian product $N \times H$. The binary operation $*$ on G is defined by

$$(n_1, h_1) * (n_2, h_2) = (n_1\varphi(h_1)(n_2), h_1h_2)\,,\ \forall n_1, n_2 \in N \text{ and } h_1, h_2 \in H\,.$$

The group $G = N \rtimes_\varphi H$ is called the *semidirect product of* N *and* H *with respect to* φ.

Remarks 3.25. (1) If a group G is the semidirect product of a normal subgroup N with a subgroup H in the sense of (1), then G is isomorphic to $N \rtimes_\varphi H$ defined as in (2), where

$$\varphi(h)(n) = hnh^{-1}\,.$$

(2) The group $N \rtimes_\varphi H$ defined in (2) is a semidirect product of the normal subgroup $N_1 = N \times \{1\}$ and the subgroup $H = \{1\} \times H$ in the sense of (1).

(3) If both N and H are normal subgroups in (1), then G is a direct product of N and H.

If φ is the trivial homomorphism, sending every element of H to the identity automorphism of N, then $N \rtimes_\phi H$ is the direct product $N \times H$.

Here is yet another way to define semidirect products. An *exact sequence* is a sequence of groups and group homomorphisms

$$\dots G_{n-1} \overset{\varphi_{n-1}}{\longrightarrow} G_n \overset{\varphi_n}{\longrightarrow} G_{n+1} \dots$$

such that $\mathrm{Im}\,\varphi_{n-1} = \mathrm{Ker}\,\varphi_n$ for every n. A *short exact sequence* is an exact sequence of the form:

$$\{1\} \longrightarrow N \overset{\varphi}{\longrightarrow} G \overset{\psi}{\longrightarrow} H \longrightarrow \{1\}\,. \tag{3.1}$$

In other words, φ is an isomorphism from N to a normal subgroup $N' \lhd G$ and ψ descends to an isomorphism $G/N' \simeq H$.

Definition 3.26. A short exact sequence *splits* if there exists a homomorphism $\sigma : H \to G$ (called a *section*) such that

$$\psi \circ \sigma = Id.$$

When the sequence splits we shall sometimes write it as

$$1 \to N \to G \overset{\curvearrowleft}{\to} H \to 1.$$

Then, every split exact sequence determines a decomposition of G as the semidirect product $\varphi(N) \rtimes \sigma(H)$. Conversely, every semidirect product decomposition $G = N \rtimes H$ defines a split exact sequence, where φ is the identity embedding and $\psi : G \to H$ is the retraction.

Examples 3.27. (1) The dihedral group D_{2n} is isomorphic to $\mathbb{Z}_n \rtimes_\varphi \mathbb{Z}_2$, where $\varphi(1)(k) = n - k$.

(2) The infinite dihedral group D_∞ is isomorphic to $\mathbb{Z} \rtimes_\varphi \mathbb{Z}_2$, where $\varphi(1)(k) = -k$.

(3) The permutation group S_n is the semidirect product of A_n and $\mathbb{Z}_2 = \{\mathrm{id}, (12)\}$.

(4) The group $(\mathrm{Aff}(\mathbb{R})\,, \circ)$ of affine maps $f : \mathbb{R} \to \mathbb{R}, f(x) = ax + b\,$, with $a \in \mathbb{R}^*$ and $b \in \mathbb{R}$ is a semidirect product $\mathbb{R} \rtimes_\varphi \mathbb{R}^*$, where $\varphi(a)(x) = ax$.

Proposition 3.28. (1) *Every isometry ϕ of $\mathbb{R}^n$ is of the form $\phi(x) = Ax + b$, where $b \in \mathbb{R}^n$ and $A \in O(n)$.*

(2) *The group $\mathrm{Isom}(\mathbb{R}^n)$ splits as the semidirect product $\mathbb{R}^n \rtimes O(n)$, with the obvious action of the orthogonal $O(n)$ on $\mathbb{R}^n$.*

Sketch of proof of (1). For every vector $a \in \mathbb{R}^n$ we denote by T_a the translation of vector a, $x \mapsto x + a$.

If $\phi(0) = b$, then the isometry $\psi = T_{-b} \circ \phi$ fixes the origin 0. Thus, it suffices to prove that an isometry fixing the origin is an element of $O(n)$. Indeed:

- an isometry of $\mathbb{R}^n$ preserves straight lines, because these are bi-infinite geodesics;
- an isometry is a homogeneous map, i.e., $\psi(\lambda v) = \lambda\psi(v)$; this is due to the fact that (for $0 < \lambda \leqslant 1$) $w = \lambda v$ is the unique point in $\mathbb{R}^n$ satisfying
$$d(0, w) + d(w, v) = d(0, v).$$
- an isometry map is an additive map, i.e., $\psi(a + b) = \psi(a) + \psi(b)$ because an isometry preserves parallelograms.

Thus, ψ is a linear transformation of $\mathbb{R}^n$, $\psi(x) = Ax$ for some matrix A. Orthogonality of the matrix A follows from the fact that the image of an orthonormal basis under ψ is again an orthonormal basis. □

EXERCISE 3.29. 1. Prove statement (2) of Proposition 3.28. Note that $\mathbb{R}^n$ is identified with the group of translations of the n-dimensional affine space *via* the map $b \mapsto T_b$.

2. Suppose that G is a subgroup of $\mathrm{Isom}(\mathbb{R}^n)$. Is it true that G is isomorphic to the semidirect product $T \rtimes Q$, where $T = G \cap \mathbb{R}^n$ and Q is the projection of G to $O(n)$?

In sections 3.9.5 and 3.9.6 we discuss semidirect products and short exact sequences in more detail.

3.5. Direct sums and wreath products

Let X be a non-empty set, and let $\mathcal{G} = \{G_x \mid x \in X\}$ be a collection of groups indexed by X. Consider the set of maps $Map_f(X, \mathcal{G})$ with finite support, i.e.,

$$Map_f(X, \mathcal{G}) := \{f : X \to \bigsqcup_{x\in X} G_x \; ; \; f(x) \in G_x \, , \, f(x) \neq 1_{G_x}$$
$$\text{for only finitely many } x \in X\} \, .$$

DEFINITION 3.30. The *direct sum* $\bigoplus_{x\in X} G_x$ is defined as $Map_f(X, \mathcal{G})$, endowed with the pointwise multiplication of functions:
$$(f \cdot g)(x) = f(x) \cdot g(x) \, , \, \forall x \in X.$$

Clearly, if A_x are abelian groups, then $\bigoplus_{x\in X} A_x$ is abelian.

When $G_x = G$ is the same group for all $x \in X$, the direct sum is the set of maps
$$Map_f(X, G) := \{f : X \to G \mid f(x) \neq 1_G \text{ for only finitely many } x \in X\} \, ,$$
and we denote it either by $\bigoplus_{x\in X} G$ or by $G^{\oplus X}$.

If, in this latter case, the set X is itself a group H, then there is a natural action of H on the direct sum, defined by
$$\varphi : H \to \mathrm{Aut}\left(\bigoplus_{h\in H} G\right) \, , \, \varphi(h)f(x) = f(h^{-1}x) \, , \, \forall x \in H \, .$$
Thus, we define the semidirect product
$$\left(\bigoplus_{h\in H} G\right) \rtimes_\varphi H \, . \tag{3.2}$$

Definition 3.31. The semidirect product (3.2) is called *the wreath product of G with H*, and it is denoted by $G \wr H$. The wreath product $G = \mathbb{Z}_2 \wr \mathbb{Z}$ is called the *lamplighter group*.

This useful construction is a source of many interesting examples in group theory, for instance, we will see in §12.5 how it is used to prove failure of QI rigidity of the class of virtually solvable groups.

3.6. Geometry of group actions

3.6.1. Group actions. Let G be a group or a semigroup and X be a set. An *action of G on X on the left* is a map

$$\mu : G \times X \to X, \quad \mu(g, a) = g(a),$$

so that

(1) $\mu(1, x) = x$;
(2) $\mu(g_1 g_2, x) = \mu(g_1, \mu(g_2, x))$ for all $g_1, g_2 \in G$ and $x \in X$.

Remark 3.32. If G is a group, then the two properties above imply that

$$\mu(g, \mu(g^{-1}, x)) = x$$

for all $g \in G$ and $x \in X$.

An *action of G on X on the right* is a map

$$\mu : X \times G \to X, \quad \mu(a, g) = (a)g,$$

so that

(1) $\mu(x, 1) = x$;
(2) $\mu(x, g_1 g_2) = \mu(\mu(x, g_1), g_2)$ for all $g_1, g_2 \in G$ and $x \in X$.

Note that the difference between an action on the left and an action on the right is the order in which the elements of a product act.

If not specified, an action of a group G on a set X is always to the left, and it is often denoted $G \curvearrowright X$. Every left action amounts to a homomorphism from G to the group $Bij(X)$ of bijections of X. An action is called *effective* or *faithful* if this homomorphism is injective. Given an action $\mu : G \times X \to X$ we will use the notation $g(x)$ for $\mu(g, x)$.

If X is a metric space, an *isometric action* is an action so that $\mu(g, \cdot)$ is an isometry of X for each $g \in G$. In other words, an isometric action is a group homomorphism

$$G \to \mathrm{Isom}(X).$$

A group action $G \curvearrowright X$ on a set X is called *free* if for every $x \in X$, the *stabilizer of x in G*,

$$G_x = \{g \in G : g(x) = x\}$$

is $\{1\}$.

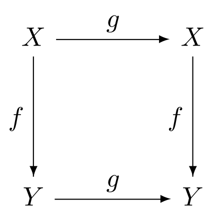
Given an action $\mu : G \curvearrowright X$, a map $f : X \to Y$ is called *G–invariant* if

$$f(\mu(g, x)) = f(x), \quad \forall g \in G, x \in X.$$

Given two actions $\mu : G \curvearrowright X$ and $\nu : G \curvearrowright Y$, a map $f : X \to Y$ is called *G–equivariant* if

$$f(\mu(g, x)) = \nu(g, f(x)), \quad \forall g \in G, x \in X.$$

In other words, for each $g \in G$ we have a commutative diagram,



A *topological group* is a group G equipped with the structure of a topological space, so that the group operations (multiplication and inversion) are continuous maps. If G is a group without specified topology, we will always assume that G is *discrete*, i.e., is given the discrete topology. When referring to homomorphisms or isomorphism of topological groups, we will always mean continuous homomorphisms and homeomorphic isomorphisms.

The usual algebraic concepts have *local analogues* for topological groups. One says that a map $\phi : G_2 \to G_1$ is a *local embedding* of topological groups if it is continuous on its domain, which is an (open) neighborhood U of $1 \in G_2$, $\phi(1) = 1$ and

$$\phi(g_1 g_2) = \phi(g_1)\phi(g_2),$$

whenever all three elements $g_1, g_2, g_1 g_2$ belong to U. This allows one to talk about *locally isomorphic* topological groups. A topological group G is called *locally compact*, if it admits a basis of topology at $1 \in G$ consisting of relatively compact neighborhoods. A topological group is σ-compact if it is the union of countablly many compact subsets.

LEMMA 3.33. *Each open subgroup $H \leq G$ is also closed.*

PROOF. The complement $G \setminus H$ equals the union

$$\bigcup_{g \notin H} gH$$

of open subsets. Therefore, H is closed. □

A topological group G is said to be *compactly generated* if it there exists a compact subset $K \subset G$ generating G. Every compactly generated group is σ-compact. The converse is not true in general: For locally compact σ-compact spaces (even compactly generated groups) second countability may not hold. Nevertheless, for every locally compact σ-compact group G there exists a compact normal subgroup N such that G/N is second countable [**Com84**, Theorem 3.7].

THEOREM 3.34 (See [**Str74**]). *Every locally compact second countable Hausdorff group has a proper left-invariant metric.*

THEOREM 3.35 (Cartan–Iwasawa–Mal'cev). *Every connected locally compact group contains unique, up to conjugation, maximal compact subgroup.*

We will us this theorem in the context of Lie groups discussed in the next section.

If G is a topological group and X is a topological space, a *continuous action* of G on X is a continuous map μ satisfying the above *action* axioms.

A continuous action $\mu : G \curvearrowright X$ is called *proper* if for every compact subsets $K_1, K_2 \subset X$, the set

$$G_{K_1,K_2} = \{g \in G : g(K_1) \cap K_2 \neq \emptyset\} \subset G$$

is compact. If G has discrete topology, a proper action is called *properly discontinuous* action, as G_{K_1,K_2} is finite.

EXERCISE 3.36. Suppose that X is locally compact and $G \curvearrowright X$ is proper. Show that the quotient X/G is Hausdorff.

Recall that a topological space X is called *Baire* if it satisfies the Baire property: *Countable intersections of open dense subsets of X are dense in X.* According to Baire theorem, each complete metric space is Baire.

LEMMA 3.37. *Suppose that $G \times X \to X$ is a continuous transitive action of a σ-compact group on a Hausdorff Baire space X. Then for each $x \in X$ the orbit map $G \to X, g \mapsto g(x)$ descends to a homeomorphism $\phi : G/G_x \to X$.*

PROOF. We let $p : G \to G/G_x$ denote the quotient map. The map ϕ is defined by $\phi(p(g)) = g(x)$. We leave it to the reader to verify that ϕ is a continuous bijection, equivariant with respect to the G-action on G/G_x and X. Since X is Hausdorff, for each compact subset $K \subset G$ the restriction of the map ϕ to $p(K)$ is a homeomorphism to its image, which is necessarily closed. Since G is σ-compact, there exists a countable collection $K_i, i \in I$, of compact subsets of G, whose union equals G. Since the orbit map $G \to X$ is surjective and X is Baire, there exists a compact subset $K \subset G$ whose image has nonempty interior in X. Therefore, the restriction of ϕ^{-1} to the interior U of $\phi(p(K))$ is continuous. Since ϕ is G-equivariant and the G-orbit of U is the entire X, we conclude that $\phi^{-1} : X \to G/G_x$ is continuous. □

A topological action $G \curvearrowright X$ is called *cocompact* if there exists a compact $C \subset X$ so that

$$G \cdot C := \bigcup_{g \in G} gC = X.$$

EXERCISE 3.38. If the action $G \curvearrowright X$ is cocompact, then the quotient space X/G (equipped with the quotient topology) is compact.

The following is a converse to the above exercise:

LEMMA 3.39. *Suppose that X is a locally compact space and the action $G \curvearrowright X$ is such that the quotient space X/G is compact. Then G acts cocompactly on X.*

PROOF. Let $p : X \to Y = X/G$ be the quotient. For every $x \in X$ choose a relatively compact (open) neighborhood $U_x \subset X$ of x. Then the collection

$$\{p(U_x)\}_{x \in X}$$

is an open cover of Y. Since Y is compact, this open cover has a finite subcover

$$\{p(U_{x_i}) : i = 1, \ldots, n\}$$

The union

$$C := \bigcup_{i=1}^{n} \mathrm{cl}(U_{x_i})$$

is compact in X and projects onto Y. Hence, $G \cdot C = X$. □

In the context of non-proper metric spaces, the concept of a cocompact group action is replaced with the one of a *cobounded action.* An isometric action $G \curvearrowright X$ is called *cobounded* if there exists $D < \infty$ such that for some point $x \in X$,

$$\bigcup_{g \in G} g(B(x, D)) = X.$$

Equivalently, given any pair of points $x, y \in X$, there exists $g \in G$ such that $\mathrm{dist}(g(x), y) \leqslant 2D$. Clearly, if X is proper, the action $G \curvearrowright X$ is cobounded if and only if it is cocompact. We call a metric space X *quasihomogeneous* if the action $\mathrm{Isom}(X) \curvearrowright X$ is cobounded.

Similarly, we have to modify the notion of a properly discontinuous action: An isometric action $G \curvearrowright X$ on a metric space is called *properly discontinuous* if for every bounded subset $B \subset X$, the set

$$G_{B,B} = \{g \in G : g(B) \cap B \neq \emptyset\}$$

is finite. Assigning two different meaning to the same notation of course, creates ambiguity, the way out of this conundrum is to think of the concept of proper discontinuity applied to different categories of actions: Topological and isometric. In the former case we use compact subsets, in the latter case we use bounded subsets. For proper metric spaces, both definitions, of course, are equivalent.

Example 3.40. Let G be an infinite discrete group equipped with the discrete metric, taking only the values 0 and 1. Then the action $G \curvearrowright G$ is properly discontinuous as a topological action, but is not properly discontinuous as an isometric action.

3.6.2. Linear actions.

3.6.3. Several facts from representation theory. In this section, V will denote a finite-dimensional vector space over a field $\mathbb{K}$ whose algebraic closure will be denoted $\bar{\mathbb{K}}$. We let $End(V)$ denote the algebra of (linear) endomorphisms of V and $GL(V)$ the group of invertible endomorphisms of V. Linear actions of groups G on V are called *representations* of G on V.

Lemma 3.41.

$$\tau : End(V) \times End(V) \to \mathbb{K}, \tau(A, B) = \mathrm{tr}(AB^T)$$

is a nondegenerate bilinear form on $End(V)$, regarded as a vector space over $\mathbb{K}$.

Proof. Representing A and B by their matrix entries $(a_{ij}), (b_{kl})$, we obtain:

$$\mathrm{tr}(AB^T) = \sum_{i,j} a_{ij} b_{ij}.$$

Therefore, if for some i, j, $a_{ij} \neq 0$, we take B such that $b_{kl} = 0$ for all $(k, l) \neq (i, j)$ and $b_{ij} = 1$. Then $\mathrm{tr}(AB^T) = a_{ij} \neq 0$. □

If V is a vector space and $A \subset End(V)$ is a subsemigroup, then A is said to *act irreducibly* on V if V contains no proper subspaces $V' \subset V$ such that $aV' \subset V'$ for all $a \in A$. An action is said to be *absolutely irreducible* iff the corresponding action on the vector space $V \otimes_{\mathbb{K}} \bar{\mathbb{K}}$ is irreducible.

A proof of the following theorem can be found, for instance, in [**Lan02**, Chapter XVII, §3, Corollary 3.3]:

THEOREM 3.42 (Burnside's theorem). *If $A \subset End(V)$ is a subalgebra which acts irreducibly on V, then $A = End(V)$. In particular, if $G \subset End(V)$ is a subsemigroup acting irreducibly, then G spans $End(V)$ as a vector space.*

LEMMA 3.43. *If a linear action of a group G on V is absolutely irreducible, then so is the action G on $W = \Lambda^k V$.*

PROOF. Since G spans $End(V)$, the action of G on W is absolutely reducible iff the action of $End(V)$ is. However, all exterior product representations of $End(V)$ are absolutely irreducible; this is a special case of irreducibility of Weyl modules, see e.g. [**FH94**, Theorem 6.3, Part 4]. □

EXERCISE 3.44. Suppose that $\mathbb{K} \subset \mathbb{L}$ is a field extension and the linear action $G \curvearrowright V$ is absolutely irreducible. Show that the action of G on $V \otimes_{\mathbb{K}} \mathbb{L}$ (regarded as a vector space over $\mathbb{L}$) is also absolutely irreducible. Give example of an irreducible representation which is not absolutely irreducible.

3.6.4. Lie groups. References for this section are [**FH94, Hel01, OV90, War83**].

A *Lie algebra* is a vector space $\mathfrak{g}$ over a field F, equipped with a binary operation $[\cdot,\cdot] : \mathfrak{g}^2 \to \mathfrak{g}$, called the *Lie bracket*, which satisfies the following axioms:

1. The Lie bracket is bilinear:

$$[\lambda x, y] = \lambda\mu[x, y], \quad [x + y, z] = [x, z] + [x, z]$$

for all $\lambda \in F$, $x, y, z \in \mathfrak{g}$.

2. The Lie bracket is anti-symmetric:

$$[x, y] = -[x, y].$$

3. The Lie bracket satisfies the *Jacobi identity*:

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0.$$

In this book we will consider only finite-dimensional real and complex Lie algebras, i.e., we will assume that $F = \mathbb{R}$ or $F = \mathbb{C}$ and $\mathfrak{g}$ is finite-dimensional as a vector space.

EXAMPLE 3.45. Lie algebra $\mathfrak{g}$ which is the vector space of $n \times n$ matrices $Mat_n(F)$ with coefficients in the field F, with the Lie bracket given by the commutator

$$[X, Y] = XY - YX.$$

An *ideal* in a Lie algebra $\mathfrak{g}$ is a vector subspace $J \subset \mathfrak{g}$ such that for every $x \in \mathfrak{g}, y \in J$ we have:

$$[x, y] \in J.$$

For instance, the subspace J consisting of scalar multiplies of the identity matrix $I \in Mat_n(F)$ is an ideal in $\mathfrak{g} = Mat_n(F)$. A Lie algebra $\mathfrak{g}$ is called *simple* if it is not 1-dimensional and every ideal in $\mathfrak{g}$ is either 0 or the entire $\mathfrak{g}$.

If $\mathfrak{g}_1, \ldots, \mathfrak{g}_m$ are Lie algebras, their *direct sum* is the direct sum of the vector spaces

$$\mathfrak{g}_1 \oplus \ldots \oplus \mathfrak{g}_m$$

with the Lie bracket, given by

$$[x_1 + \ldots x_m, y_1 + \ldots + y_m] = \sum_{i=1}^{m} [x_i, y_i]$$

for $x_i, y_i \in \mathfrak{g}_i$, $i = 1, \ldots m$. A Lie algebra $\mathfrak{g}$ is called *semisimple* if it is isomorphic to the direct sum of finitely many simple Lie algebras.

A *Lie group* is a group G which has the structure of a smooth manifold, so that the binary operation $G \times G \to G$ and inversion $g \mapsto g^{-1}, G \to G$ are smooth maps. Actually, every Lie group G can be made into a real analytic manifold with real analytic group operations. We will mostly use the notation e for the neutral element of G. We will assume that G is a real n-dimensional manifold, although we will sometimes also consider *complex Lie groups.* A *homomorphism* of Lie groups is a group homomorphism which is also a smooth map.

EXAMPLE 3.46. Groups $GL(n, \mathbb{R}), SL(n, \mathbb{R}), O(n), O(p, q)$ are (real) Lie groups. Every countable discrete group (a topological group with discrete topology) is a Lie group. (Recall that we require our manifolds to be second countable. If we were to drop this requirement, then any discrete group becomes a Lie group.)

Here $O(p, q)$ is the group of linear isometries of the quadratic form

$$x_1^2 + \ldots x_p^2 - x_{p+1}^2 - \ldots - x_{p+q}^2$$

of the signature (p, q). The most important, for us, case is the group $O(n, 1) \cong O(1, n)$. The group $PO(n, 1) = O(n, 1)/\pm I$ is the group of isometries of the hyperbolic n-space.

EXERCISE 3.47. Show that the group $PO(n, 1)$ embeds in $O(n, 1)$ as the subgroup stabilizing the *future light cone*

$$x_1^2 + \ldots + x_n^2 - x_{n+1}^2 > 0, \quad x_{n+1} > 0.$$

The tangent space $V = T_e G$ of a Lie group G at the identity element $e \in G$ has the structure of a Lie algebra, called the *Lie algebra* $\mathfrak{g}$ of the group G.

EXAMPLE 3.48. 1. The Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ of $SL(n, \mathbb{C})$ consists of trace-free $n \times n$ complex matrices. The Lie bracket operation on $\mathfrak{sl}(n, \mathbb{C})$ is given by

$$[A, B] = AB - BA.$$

2. The Lie algebra of the unitary subgroup $U(n) < GL(n, \mathbb{C})$ equals the space of skew-hermitian matrices

$$\mathfrak{u}(n) = \{A \in Mat_n(\mathbb{C}) : A = -A^*\}.$$

3. The Lie algebra of the orthogonal subgroup $O(n) < GL(n, \mathbb{R})$ equals the space of skew-symmetric matrices

$$\mathfrak{o}(n) = \{A \in Mat_n(\mathbb{R}) : A = -A^T\}.$$

EXERCISE 3.49. $\mathfrak{u}(n) \oplus i\mathfrak{u}(n) = Mat_n(\mathbb{C})$, is the Lie algebra of the group $GL(n, \mathbb{C})$.

Every Lie group G has a left-invariant Riemannian metric, i.e., a Riemannian metric invariant under the left multiplication

$$L_g : G \to G, \quad L_g(x) = gx$$

by elements of G. Indeed, pick a positive-definite inner product $\langle \cdot, \cdot \rangle_e$ on T_eG. The map $L_g : G \to G$ is a diffeomorphism and the action of G on itself *via* left multiplication is simply-transitive. We define the inner product $\langle \cdot, \cdot \rangle_g$ on T_gG as the image of $\langle \cdot, \cdot \rangle_e$ under the derivative $DL_g : T_eG \to T_gG$. Similarly, if G is a compact Lie group, then it admits a bi-invariant Riemannian metric, i.e., a Riemannian metric invariant under both left and right multiplication. Namely, if $\langle \cdot, \cdot \rangle$ is a left-invariant Riemannian metric on G, define the right-invariant metric by the formula:

$$\langle u, v \rangle' := \int_G \langle DR_g(u), DR_g(v) \rangle \, dVol(g),$$

where $dVol$ is the volume form of the Riemannian metric $\langle \cdot, \cdot \rangle$ and R_g is the right multiplication by g:

$$R_g : G \to G, \quad R_g(x) = xg.$$

Every Lie group G acts on itself *via* inner automorphisms

$$\rho(g)(x) = gxg^{-1}.$$

This action is smooth and the identity element $e \in G$ is fixed by the entire group G. Therefore, G acts linearly on the tangent space $V = T_eG$ at the identity $e \in G$. The action of G on V is called *the adjoint representation of the group* G and denoted by Ad. Thus, one obtains a homomorphism

$$Ad : G \to GL(V).$$

LEMMA 3.50. *For every connected Lie group G the kernel of $Ad : G \to GL(V)$ is contained in the center of G.*

PROOF. There is a local diffeomorphism

$$\exp : V \to G$$

called *the exponential map* of the group G, sending $0 \in V$ to $e \in G$. In the case when $G = GL(n, \mathbb{R})$ this map is the ordinary matrix exponential map. The map exp satisfies the identity

$$g \exp(v) g^{-1} = \exp(\mathrm{Ad}(g)v), \quad \forall v \in V, g \in G.$$

Thus, if $\mathrm{Ad}(g) = Id$, then g commutes with every element of G of the form $\exp(v), v \in V$. The set of such elements is open in G. Now, if we are willing to use a real analytic structure on G, then it would immediately follow that g belongs to the center of G. Below is an alternative argument. Let $g \in \mathrm{Ker}(Ad)$. The centralizer $Z(g)$ of g in G is given by the equation

$$Z(g) = \{h \in G : [h, g] = 1\}.$$

Since the commutator is a continuous map, $Z(g)$ is a closed subgroup of G. Moreover, as we observed above, this subgroup has nonempty interior in G (containing e). Since $Z(g)$ acts transitively on itself by, say, left multiplication, $Z(g)$ is open in G. As G is connected, we conclude that $Z(g) = G$. Therefore the kernel of Ad is contained in the center of G. ☐

A Lie group G is called *simple* if G contains no connected proper normal subgroups. Equivalently, a Lie group G is simple if its Lie algebra $\mathfrak{g}$ is simple.

Example 3.51. The group $SL(2,\mathbb{R})$ is simple, but its center is isomorphic to $\mathbb{Z}_2$.

Thus, a simple Lie group need not be simple as an abstract group. A Lie group G is *semisimple* if its Lie algebra is semisimple.

Examples of semisimple Lie groups are $SL(n,\mathbb{R})$, $SO(p,q)$, $SO(n,\mathbb{C})$.

Below are several deep structural theorems about Lie groups:

Theorem 3.52 (S. Lie). *1. For every finite-dimensional real Lie algebra $\mathfrak{g}$ there exists a unique, up to an isomorphism, simply-connected Lie group G whose Lie algebra is isomorphic to $\mathfrak{g}$. 2. Two Lie groups are locally isomorphic if and only if their Lie algebras are isomorphic.*

Theorem 3.53 (E. Cartan). *Every closed subgroup H of a Lie group G has structure of a Lie group so that the inclusion $H \hookrightarrow G$ is an embedding of smooth manifolds.*

The next theorem is a corollary of the *Peter-Weyl theorem*, see e.g., [**OV90**, Theorem 10, page 245]:

Theorem 3.54. *Next, every compact Lie group is linear, i.e., embeds in $GL(V)$ for some finite-dimensional real vector space V.*

While there are nonlinear (connected) Lie groups, e.g., the universal cover of $SL(2,\mathbb{R})$, each Lie group is *locally* linear.

Theorem 3.55 (I. D. Ado, [**Ado36**]). *Every finite-dimensional Lie algebra $\mathfrak{g}$ over a field F of characteristic zero (e.g., over the real numbers) admits a faithful finite-dimensional representation. In particular, every Lie group locally embeds in $GL(V)$ for some finite-dimensional real vector space V.*

We refer the reader to [**FH94**, Theorem E.4] for a proof. Note that if a Lie group G has discrete center, then the adjoint representation of G is a local embedding of G in $GL(\mathfrak{g})$, where $\mathfrak{g}$ is the Lie algebra of G. The difficulty is in the case of groups with nondiscrete center.

3.6.5. Haar measure and lattices.

Definition 3.56. A (left) Haar measure on a topological group G is a countably additive, nontrivial measure μ on Borel subsets of G satisfying:

(1) $\mu(gE) = \mu(E)$ for every $g \in G$ and every Borel subset $E \subset G$.
(2) $\mu(K)$ is finite for every compact $K \subset G$.
(3) Every Borel subset $E \subset G$ is *outer regular*:

$$\mu(E) = \inf\{\mu(U) : E \subset U,\ U \text{ is open in } G\}.$$

(4) Every open set $E \subset G$ is *inner regular*:

$$\mu(E) = \sup\{\mu(U) : U \subset E,\ U \text{ is open in } G\}.$$

By Haar's Theorem, see [**Bou63**], every locally compact topological group G admits a Haar measure and this measure is unique up to scaling. Similarly, one defines right-invariant Haar measures. In general, left and right Haar measures are not the same, but they are for some important classes of groups:

Definition 3.57. A locally compact group G is *unimodular* if left and right Haar measures are constant multiples of each other.

Important examples of Haar measures come from Riemannian geometry. Let X be a homogeneous Riemannian manifold, G is the isometry group. Then X has a natural measure ω defined by the volume form of the Riemannian metric on X. We have the natural surjective map $G \to X$ given by the orbit map $g \mapsto g(o)$, where $o \in X$ is a base-point. The fibers of this map are stabilizers G_x of points $x \in X$. Arzela-Ascoli theorem implies that each subgroup G_x is compact. Transitivity of the action $G \curvearrowright X$ implies that all the subgroups G_x are conjugate. Setting $K = G_o$, we obtain the identification $X = G/K$. Now, let μ be the pull-back of ω under the projection map $G \to X$. By construction, μ is left-invariant (since the metric on X is G–invariant).

DEFINITION 3.58. Let G be a topological group with finitely many connected components and μ a Haar measure on G. A *lattice* in G is a discrete subgroup $\Gamma < G$ so that the quotient $Q = \Gamma\backslash G$ admits a finite G–invariant measure (for the action to the right of G on Q) induced by the Haar measure. A lattice Γ is called *uniform* if the quotient Q is compact.

If G is a Lie group, then the measure above can also be obtained by taking a Riemannian metric on G which is left-invariant under G and right-invariant under K, the maximal compact subgroup of G. Note that when G is unimodular, the volume form thus obtained is also right-invariant under G.

Thus, if one considers the quotient $X := G/K$, then X has a Riemannian metric which is (left) invariant under G. Hence, Γ is a lattice if and only if Γ acts on X properly discontinuously so that $vol(\Gamma \setminus X)$ is finite. Note that the action of Γ on X is *a priori* not free.

THEOREM 3.59. *A locally compact second countable group G is unimodular, provided that it contains a lattice.*

PROOF. For arbitrary $g \in G$ consider the push-forward $\nu = R_g(\mu)$ of the (left) Haar measure μ on G; here R_g is the right multiplication by g:

$$\nu(E) = \mu(Eg).$$

Then ν is also a left Haar measure on G. By the uniqueness of Haar measure, $\nu = c\mu$ for some constant $c > 0$.

LEMMA 3.60. *Every discrete subgroup $\Gamma < G$ admits a* measurable fundamental set*, i.e., a measurable subset of $D \subset G$ such that*

$$\bigcup_{\gamma\in\Gamma} \gamma D = G, \quad \mu(\gamma D \cap D) = 0, \quad \forall \gamma \in \Gamma \setminus \{e\}.$$

PROOF. Since $\Gamma < G$ is discrete, there exists an open neighborhood V of $e \in G$ such that $\Gamma \cap V = \{e\}$. Since G is a topological group, there exists another open neighborhood U of $e \in G$, such that $UU^{-1} \subset V$. Then for $\gamma \in \Gamma$ we have

$$\gamma u = u', u \in U, u' \in U \Rightarrow \gamma = u'u^{-1} \in U \Rightarrow \gamma = e.$$

In other words, Γ-images of U are pairwise disjoint. Since G is a second countable, there exists a countable subset

$$E = \{g_i \in G : i \in \mathbb{N}\}$$

such that

$$G = \bigcup_i U g_i.$$

Clearly, each set

$$W_n := Ug_n \setminus \bigcup_{i<n} \Gamma U g_i$$

is measurable, and so is their union

$$D = \bigcup_{n=1}^{\infty} W_n.$$

Let us verify that D is a measurable fundamental set. First, note that for every $x \in G$ there exists the least n such that $x \in Ug_n$. Therefore,

$$G = \bigcup_{n=1}^{\infty} \left(Ug_n \setminus \bigcup_{i<n} Ug_i \right).$$

Next,

$$\Gamma \cdot D = \bigcup_{n=1}^{\infty} \left(\Gamma U g_n \setminus \bigcup_{i<n} \Gamma U g_i \right) =$$

$$\Gamma \cdot \bigcup_{n=1}^{\infty} \left(Ug_n \setminus \bigcup_{i<n} Ug_i \right) \supset \bigcup_{n=1}^{\infty} \left(Ug_n \setminus \bigcup_{i<n} Ug_i \right) = G.$$

Therefore, $\Gamma \cdot D = G$. Next, suppose that

$$x \in \gamma D \cap D.$$

Then, for some n, m

$$x \in W_n \cap \gamma W_m.$$

If $m < n$, then

$$\gamma W_m \subset \Gamma \bigcup_{i<n} U g_i,$$

which is disjoint from W_n, a contradiction. Thus, $W_n \cap \gamma W_m = \emptyset$ for $m < n$ and all $\gamma \in \Gamma$. If $n < m$, then

$$W_n \cap \gamma W_m = \gamma^{-1} \left(\gamma W_n \cap W_m \right) = \emptyset.$$

Thus, $n = m$, which implies that

$$Ug_n \cap \gamma U g_n \neq \emptyset \Rightarrow U \cap \gamma U \neq \emptyset \Rightarrow \gamma = e.$$

Therefore, for all $\gamma \in \Gamma \setminus \{e\}$, $\gamma D \cap D = \emptyset$. □

We can now prove Theorem 3.59. Let $\Gamma < G$ be a lattice and let $D \subset G$ be its measurable fundamental set. Then

$$0 < \mu(D) = \mu(\Gamma \backslash G) < \infty$$

since Γ is a lattice. For every $g \in G$, Dg is again a measurable fundamental set for Γ and, thus, $\mu(D) = \mu(Dg)$. Hence,

$$\mu(D) = \mu(Dg) = c\nu(D).$$

It follows that $c = e$. Thus, μ is also a right Haar measure. □

3.6.6. Geometric actions. Suppose now that X is a metric space. We will equip the group of isometries $\mathrm{Isom}(X)$ of X with the *compact-open topology*, equivalent to the topology of uniform convergence on compact sets. A subgroup $G \leqslant \mathrm{Isom}(X)$ is called *discrete* if it is discrete with respect to the subset topology.

EXERCISE 3.61. Suppose that X is proper. Show that the following are equivalent for a subgroup $G \leqslant \mathrm{Isom}(X)$:

a. G is discrete.

b. The action $G \curvearrowright X$ is properly discontinuous.

c. For every $x \in X$ and an infinite sequence $g_i \in G$,

$$\lim_{i\to\infty} d(x, g_i(x)) = \infty.$$

Hint: Use Arzela–Ascoli theorem.

DEFINITION 3.62. A *geometric action* of a group G on a metric space X is an isometric properly discontinuous cobounded action $G \curvearrowright X$.

For instance, if X is a homogeneous Riemannian manifold with the isometry group G and $\Gamma < G$ is a uniform lattice, then Γ acts geometrically on X. Note that every geometric action on a proper metric space is cocompact.

LEMMA 3.63. *Suppose that a group G acts geometrically on a proper metric space X. Then $G\backslash X$ has a metric defined by*

$$\mathrm{dist}(\bar{a}, \bar{b}) = \inf\{\mathrm{dist}(p,q)\ ;\ p \in Ga\,,\ q \in Gb\} = \inf\{\mathrm{dist}(a,q)\ ;\ q \in Gb\}\,, \tag{3.3}$$

where $\bar{a} = Ga$ and $\bar{b} = Gb$.

Moreover, this metric induces the quotient topology of $G\backslash X$.

PROOF. The infimum in (3.3) is attained, i.e., there exists $g \in G$ such that

$$\mathrm{dist}(\bar{a}, \bar{b}) = \mathrm{dist}(a, gb).$$

Indeed, take $g_0 \in G$ arbitrary, and let $R = \mathrm{dist}(a, gb)$. Then

$$\mathrm{dist}(\bar{a}, \bar{b}) = \inf\{\mathrm{dist}(a,q)\ ;\ q \in Gb \cap \overline{B}(a,R)\}.$$

Now, for every $gb \in \overline{B}(a,R)$,

$$gg_0^{-1}\overline{B}(a,R) \cap \overline{B}(a,R) \neq \emptyset.$$

Since G acts properly discontinuously on X, this implies that the set $Gb \cap \overline{B}(a,R)$ is finite, hence the last infimum is over a finite set, and it is attained. We leave it to the reader to verify that dist is the Hausdorff distance between the orbits $G \cdot a$ and $G \cdot b$. Clearly the projection $X \to G\backslash X$ is a contraction. One can easily check that the topology induced by the metric dist on $G\backslash X$ coincides with the quotient topology. $\square$

3.7. Zariski topology and algebraic groups

The proof of Tits' theorem relies in part on some basic results from theory of affine algebraic groups. We recall some terminology and results needed in the argument. For a more thorough presentation, see [**Hum75**] and [**OV90**].

The proof of the following general lemma is straightforward, and left as an exercise to the reader.

LEMMA 3.64. *For every commutative ring A the following two statements are equivalent:*

(1) *every ideal in A is finitely generated;*

(2) *the set of ideals satisfies the ascending chain condition (ACC), that is, every ascending chain of ideals*

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \cdots$$

stabilizes, i.e., there exists an integer N such that $I_n = I_N$ for every $n \geqslant N$.

DEFINITION 3.65. A commutative ring is called *noetherian*, if it satisfies one (hence, both) statements in Lemma 3.64.

Note that a field seen as a ring is always noetherian. Other examples of noetherian rings come from the following theorem:

THEOREM 3.66 (Hilbert's ideal basis theorem, see [**Hum75**]). *If A is a noetherian ring then the ring of multivariable polynomials $A[X_1, \ldots, X_n]$ is also noetherian.*

From now on, we fix a field $\mathbb{K}$.

DEFINITION 3.67. An *affine algebraic set* in $\mathbb{K}^n$ is a subset Z in $\mathbb{K}^n$ that is the solution set of a system of multivariable polynomial equations $p_j = 0$, $\forall j \in J$, with coefficients in $\mathbb{K}$:

$$Z = \{(x_1, \ldots, x_n) \in \mathbb{K}^n \; ; \; p_j(x_1, \ldots, x_n) = 0, j \in J\}.$$

We will frequently say "algebraic subset" or "affine variety" when referring to an affine algebraic set.

For instance, algebraic subsets in the affine line (the 1-dimensional vector space V over $\mathbb{K}$) are finite subsets and the entire of V, since every nonzero polynomial in one variable has at most finitely many zeroes.

There is a one-to-one map associating to every algebraic subset in $\mathbb{K}^n$ an ideal in $\mathbb{K}[X_1, \ldots, X_n]$:

$$Z \mapsto I_Z = \{p \in \mathbb{K}[X_1, \ldots, X_n] \; ; \; p\big|_Z \equiv 0\}.$$

Note that I_Z is the kernel of the homomorphism $p \mapsto p|_Z$ from $\mathbb{K}[X_1, \ldots, X_n]$ to the ring of functions on Z. Thus, the ring $\mathbb{K}[X_1, \ldots, X_n]/I_Z$ may be seen as a ring of functions on Z; this quotient ring is called the *coordinate ring of Z* or the *ring of polynomials on Z*, and denoted $\mathbb{K}[Z]$.

Theorem 3.66 and Lemma 3.64 imply the following.

LEMMA 3.68. (1) *The set of algebraic subsets of $\mathbb{K}^n$ satisfies the descending chain condition (DCC): every descending chain of algebraic subsets*

$$Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_i \supseteq \cdots$$

stabilizes, i.e., for some integer $N \geqslant 1$, $Z_i = Z_N$ for every $i \geqslant N$.

(2) *Every algebraic set is defined by finitely many equations.*

DEFINITION 3.69. A *morphism* between two affine varieties Y in $\mathbb{K}^n$ and Z in $\mathbb{K}^m$ is a map of the form $\varphi : Y \to Z$, $\varphi = (\varphi_1, \ldots, \varphi_m)$, such that each φ_i is in $\mathbb{K}[Y]$, $i \in \{1, 2, ..., m\}$.

Note that every morphism is induced by a morphism $\tilde{\varphi} : \mathbb{K}^n \to \mathbb{K}^m$, $\tilde{\varphi} = (\tilde{\varphi}_1, \ldots, \tilde{\varphi}_m)$, with $\tilde{\varphi}_i : \mathbb{K}^n \to \mathbb{K}$ a polynomial function for every $i \in \{1, 2, \ldots, m\}$.

An *isomorphism* between two affine varieties Y and Z is an invertible map $\varphi : Y \to Z$ such that both φ and φ^{-1} are morphisms. When $Y = Z$, an isomorphism is called an *automorphism*.

EXERCISE 3.70. 1. If $f : Y \to Z$ is a morphism of affine varieties and $W \subset Z$ is a subvariety, then $f^{-1}(W)$ is a subvariety in Y. In particular, every linear automorphism of $V = \mathbb{K}^n$ sends subvarieties to subvarieties and, hence, the notion of a subvariety is independent of the choice of a basis in V.

2. Show that the projection map $f : \mathbb{C}^2 \to \mathbb{C}$, $f(x, y) = x$, does not map subvarieties to subvarieties.

Let V be an n-dimensional vector space over a field $\mathbb{K}$. The *Zariski topology* on V is the topology having as closed sets all the algebraic subsets in V. It is clear that the intersection of algebraic subsets is again an algebraic subset. Let $Z = Z_1 \cup \ldots \cup Z_\ell$ be a finite union of algebraic subsets, Z_i defined by the ideal I_{Z_i}. Then Z is defined by the ideal

$$I_Z = I_{Z_1} \cdot \ldots \cdot I_{Z_\ell}$$

generated by the products

$$\prod_{i=1}^{\ell} p_i$$

of elements $p_i \in I_{Z_i}$.

The induced topology on a subvariety $Z \subseteq V$ is also called the Zariski topology. Note that this topology can also be defined directly using polynomial functions in $\mathbb{K}[Z]$. According to Exercise 3.70, morphisms between affine varieties are continuous with respect to the Zariski topologies.

The *Zariski closure* of a subset $E \subset V$ can also be defined by means of the set P_E of all polynomials which vanish on E, i.e., it coincides with

$$\{x \in V \mid p(x) = 0, \forall p \in P_E\}.$$

A subset $Y \subset Z$ in an affine variety is called *Zariski-dense* if its Zariski closure is the entire of Z.

Lemma 3.68, Part (1), implies that closed subsets in Zariski topology satisfy the descending chain condition (DCC).

DEFINITION 3.71. A topological space such that the closed sets satisfy the DCC is called *noetherian*.

LEMMA 3.72. *Every subspace of a noetherian topological space (with the subspace topology) is noetherian.*

Proof. Let X be a space with topology $\mathcal{T}$ such that $(X, \mathcal{T})$ is noetherian, and let Y be an arbitrary subset in X. Consider a descending chain of closed subsets in Y:

$$Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_n \supseteq \dots$$

Every Z_i is equal to $Y \cap C_i$ for some closed set C_i in X. We leave it to the reader to check that C_i can be taken equal to the closure $\overline{Z}_i$ of Z_i in X.

The descending chain of closed subsets in X,

$$\overline{Z}_1 \supseteq \overline{Z}_2 \supseteq \cdots \supseteq \overline{Z}_n \supseteq \dots$$

stabilizes, hence, so does the chain of the subsets Z_i. □

Proposition 3.73. *Every noetherian topological space X is compact.*

Proof. Compactness of X is equivalent to the condition that for every family $\{Z_i \ : \ i \in I\}$ of closed subsets in X, if $\bigcap_{i\in I} Z_i = \emptyset$, then there exists a finite subset J of I such that $\bigcap_{j\in J} Z_j = \emptyset$. Assume that all finite intersections of a family as above are non-empty. Then we construct inductively a descending sequence of closed sets that never stabilizes. The initial step consists of picking an arbitrary set Z_{i_1}, with $i_1 \in I$. At the nth step we have a non-empty intersection $Z_{i_1} \cap Z_{i_2} \cap \dots \cap Z_{i_n}$; hence, there exists $Z_{i_{n+1}}$ with $i_{n+1} \in I$ such that $Z_{i_1} \cap Z_{i_2} \cap \dots \cap Z_{i_n} \cap Z_{i_{n+1}}$ is a non-empty proper closed subset of $Z_{i_1} \cap Z_{i_2} \cap \dots \cap Z_{i_n}$. □

We now discuss a strong version of connectedness, relevant in the setting of noetherian spaces.

Lemma 3.74. *For a topological space X the following properties are equivalent:*

(1) *every open non-empty subset of X is dense in X;*

(2) *any two open non-empty subsets have non-empty intersection;*

(3) *X cannot be written as a finite union of proper closed subsets.*

We leave the proof of this lemma as an exercise to the reader.

Definition 3.75. A topological space is called *irreducible* if it is non-empty and one of (hence all) the properties in Lemma 3.74 is (are) satisfied. A subset of a topological space is *irreducible* if, when endowed with the subset topology, it is an irreducible space.

Exercise 3.76. (1) Prove that $\mathbb{K}^n$ with Zariski topology is irreducible, provided that the field $\mathbb{K}$ is infinite.

(2) Prove that an algebraic variety Z is irreducible if and only if $\mathbb{K}[Z]$ does not contain zero divisors.

The following properties are straightforward and their proof is left as an exercise to the reader.

Lemma 3.77. (1) *The image of an irreducible space under a continuous map is irreducible.*

(2) *The cartesian product of two irreducible spaces is an irreducible space, when endowed with the product topology.*

Note that the Zariski topology on $\mathbb{K}^{n+m} = \mathbb{K}^n \times \mathbb{K}^m$ is *not* the product topology (unless $nm = 0$ or $\mathbb{K}$ is finite). Hence, irreducibility of products of irreducible varieties cannot be derived from Lemma 3.77.

Lemma 3.78. *Let V_1, V_2 be finite-dimensional vector spaces over $\mathbb{K}$ and let $Z_i \subset V_i, i = 1, 2,$ be irreducible subvarieties. Then the product $Z := Z_1 \times Z_2 \subset V = V_1 \times V_2$ is an irreducible subvariety in the vector space V.*

Proof. Let $Z = W_1 \cup W_2$ be a union of two proper subvarieties. For every $z \in Z_1$ the product $\{z\} \times Z_2$ is isomorphic to Z_2 (*via* the projection to the second factor) and, hence, is irreducible. On the other hand,

$$\{z\} \times Z_2 = ((\{z\} \times Z_2) \cap W_1) \cup ((\{z\} \times Z_2) \cap W_2)$$

is a union of two subvarieties. Thus, for every $z \in Z_1$, one of these subvarieties has to be the entire $\{z\} \times Z_2$. In other words, either $\{z\} \times Z_2 \subset W_1$ or $\{z\} \times Z_2 \subset W_2$. We then partition Z_1 in two subsets A_1, A_2:

$$A_i = \{z \in Z_1 : \{z\} \times Z_2 \subset W_i\}, i = 1, 2.$$

Since each W_1, W_2 is a proper subvariety, both A_1, A_2 are proper subsets of Z_1. We will now prove that both A_1, A_2 are subvarieties in Z_1. We will consider the case of A_1 since the other case is obtained by relabeling. Let $f_1, \ldots, f_k$ denote generators of the ideal of W_1. We will think of each f_i as a function of two variables $f = f(X_1, X_2)$, where X_k stands for the tuple of coordinates in $V_k, k = 1, 2$. Then

$$A_1 = \{z \in Z_2 : f_i(z, z_2) = 0, \forall z \in Z_1, i = 1, \ldots, k\}.$$

However, for every fixed $z \in Z_1$, the function $f_i(z, \cdot)$ is a polynomial function $f_{i,z}$ on Z_2. Therefore, A_1 is the solution set of the system of polynomial equations on Z_1:

$$\{f_{i,z} = 0 : i = 1, \ldots, k, z \in Z_1\}.$$

It follows that A_1 is a subvariety, which contradicts irreducibility of Z_2. □

Lemma 3.79. *Let $(X, \mathcal{T})$ be a topological space.*

1. *A subset Y of X is irreducible if and only if its closure $\overline{Y}$ in X is irreducible.*
2. *If Y is irreducible and $Y \subseteq A \subseteq \overline{Y}$ then A is irreducible.*
3. *Every irreducible subset Y of X is contained in a maximal irreducible subset.*
4. *The maximal irreducible subsets of X are closed and they cover X.*

Proof. (1) For every open subset $U \subset X$, $U \cap Y \neq \emptyset$ if and only if $U \cap \overline{Y} \neq \emptyset$. This and Lemma 3.74, (2), imply the equivalence.

(2) Follows directly from (1).

(3) The family $\mathcal{I}_Y$ of irreducible subsets containing Y has the property that every ascending chain has a maximal element, which is the union. It can be easily verified that the union is again irreducible, using Lemma 3.74, (2). It follows from Zorn's Lemma that $\mathcal{I}_Y$ contains a maximal element.

(4) Singletons are irreducible and cover X. Now, the statement follows from (1) and (3), since . □

Theorem 3.80. *A noetherian topological space X is a union of finitely many distinct maximal irreducible subsets $X_1, X_2, \ldots, X_n$, such that for every i, X_i is not contained in $\bigcup_{j\neq i} X_j$. Moreover, every maximal irreducible subset in X coincides with one of the subsets $X_1, X_2, \ldots, X_n$. This decomposition of X is unique up to a renumbering of the X_i's.*

Proof. Let $\mathcal{F}$ be the collection of closed subsets of X that cannot be written as a finite union of maximal irreducible subsets. Assume that $\mathcal{F}$ is non-empty. Since X is noetherian, $\mathcal{F}$ satisfies the DCC, hence by Zorn's Lemma it contains a minimal element Y. As Y is not irreducible, it can be decomposed as $Y = Y_1 \cup Y_2$, where Y_i are closed and, by the minimality of Y, both Y_i decompose as finite unions of irreducible subsets (maximal in Y_i). According to Lemma 3.79, (3), Y itself can be written as union of finitely many maximal irreducible subsets, a contradiction. It follows that $\mathcal{F}$ is empty.

If $X_i \subseteq \bigcup_{j\neq i} X_j$ then $X_i = \bigcup_{j\neq i} (X_j \cap X_i)$. As X_i is irreducible it follows that $X_i \subseteq X_j$ for some $j \neq i$, hence, by maximality, $X_i = X_j$, contradicting the fact that we took only distinct maximal irreducible subsets. A similar argument is used to prove that every maximal irreducible subset of X must coincide with one of the sets X_i.

Now assume that X can be also written as a union of distinct maximal irreducible subsets $Y_1, Y_2, \ldots, Y_m$ such that for every i, Y_i is not contained in $\bigcup_{j\neq i} Y_j$. For every $i \in \{1, 2, ..., m\}$ there exists a unique $j_i \in \{1, 2, \ldots, n\}$ such that $Y_i = X_{j_i}$. The map $i \mapsto j_i$ is injective, and if some $k \in \{1, 2, \ldots, n\}$ is not in the image of this map then it follows that

$$X_k \subseteq \bigcup_{i=1}^{m} Y_i \subseteq \bigcup_{j\neq k} X_j,$$

a contradiction. □

Definition 3.81. The subsets X_i defined in Theorem 3.80 are called *the irreducible components* of X. In other words, irreducible components of X are maximal irreducible subsets of X.

Note that we can equip every Zariski–open subset U of a (finite-dimensional) vector space V with the Zariski topology, which is the subset topology with respect to the Zariski topology on V. Then U is also Noetherian. We will be using the Zariski topology primarily in the context of the group $GL(V)$, which we identify with the Zariski open subset of $V \otimes V^*$, the space of $n \times n$ matrices with nonzero determinant.

Definition 3.82. An *algebraic subgroup* of $GL(V)$ is a Zariski-closed subgroup of $GL(V)$.

Given an algebraic subgroup G of $GL(V)$, the binary operation $G \times G \to G$, $(g, h) \mapsto gh$ is a morphism. The inversion map $g \mapsto g^{-1}$, as well as the left-multiplication and right-multiplication maps $g \mapsto ag$ and $g \mapsto ga$, by a fixed element $a \in G$, are automorphisms of the variety G.

Example 3.83.

(1) The subgroup $SL(V)$ of $GL(V)$ is algebraic, defined by the equation $\det(g) = 1$.

(2) The group $GL(n,\mathbb{K})$ can be identified with an algebraic subgroup of the group $SL(n+1,\mathbb{K})$ by mapping every matrix $A \in GL(n,\mathbb{K})$ to the matrix

$$\begin{pmatrix} A & 0 \\ 0 & \frac{1}{\det(A)} \end{pmatrix}.$$

Therefore, in what follows, it will not matter if we consider algebraic subgroups of $GL(n,\mathbb{K})$ or of $SL(n,\mathbb{K})$.

(3) The group $O(V)$ is an algebraic subgroup, as it is given by the system of equations $M^T M = \mathrm{Id}_V$.

(4) More generally, given an arbitrary quadratic form q on V, its stabilizer $O(q)$ is obviously algebraic. A special instance of this is the *symplectic group* $Sp(2k,\mathbb{K})$, preserving the form with the following matrix (given with respect to the standard basis in $V = \mathbb{K}^{2n}$)

$$J = \begin{pmatrix} 0 & K \\ -K & 0 \end{pmatrix}, \text{ where } K = \begin{pmatrix} 0 & \dots & 1 \\ 0 & \iddots & 0 \\ 1 & \dots & 0 \end{pmatrix}.$$

LEMMA 3.84. *If Γ is a subgroup of an algebraic group G, then its Zariski closure $\overline{\Gamma}$ in G is also a subgroup.*

PROOF. We let $\mu : G \times G \to G, \lambda : G \to G$ denote the multiplication and inversion maps respectively. Both maps are continuous if we equip $G \times G, G$ with their respective Zariski topologies. Therefore, for each subset $E \subset G$,

$$\lambda(\overline{E}) \subset \overline{\lambda(E)},$$

which implies that

$$\lambda(\overline{\Gamma}) \subset \overline{\Gamma}.$$

Hence, $\overline{\Gamma}$ is closed under the inversion. Similarly, for each $g \in \Gamma$,

$$\mu(g, \overline{\Gamma}) \subset \overline{\Gamma}$$

and, thus, for each $h \in \overline{\Gamma}$,

$$\mu(\overline{\Gamma}, h) \subset \overline{\Gamma}.$$

It follows that $\mu(\overline{\Gamma}, \overline{\Gamma}) \subset \overline{\Gamma}$. □

The vector space $V = \mathbb{K}^n$ also has the *standard* or *classical* topology, given by the suitable norm on V. We use the terminology *classical topology* for the induced topology on subsets of V. Classical topology, of course, is stronger than Zariski topology.

THEOREM 3.85 (See for instance Chapter 3, §2, in [**OV90**]). (1) *An algebraic subgroup of $GL(n,\mathbb{C})$ is irreducible in the Zariski topology if and only if it is connected in the classical topology.*

(2) *A connected (in classical topology) algebraic subgroup of $GL(n,\mathbb{R})$ is irreducible in the Zariski topology.*

We will not need this theorem; the following proposition will suffice for our purposes:

PROPOSITION 3.86. *Let G be an algebraic subgroup in $GL(V)$.*

(1) *Only one irreducible component of* G *contains the identity element. This component is called the* identity component *and is denoted by* G_0.
(2) *The subset* G_0 *is a normal subgroup of finite index in* G *whose cosets are the irreducible components of* G.

PROOF. (1) Let $X_1, ..., X_k$ be irreducible components of G containing the identity. According to Lemma 3.78, the product set $X_1 \times \ldots \times X_k$ is irreducible. Since the product map is a morphism, the subset $X_1 \cdots X_k \subset G$ is irreducible as well; hence by Lemma 3.79, (3), and by Theorem 3.80 this subset is contained in some X_j. The fact that every X_i with $i \in \{1, ..., k\}$ is contained in $X_1 \cdots X_k$, hence in X_j, implies that $k = 1$.

(2) Since the inversion map $g \mapsto g^{-1}$ is an algebraic automorphism of G (but not a group automorphism, of course) it follows that G_0 is stable with respect to the inversion. Hence for every $g \in G_0$, gG_0 contains the identity element, and is an irreducible component. It follows that $gG_0 = G_0$. Likewise, for every $x \in G$, xG_0x^{-1} is an irreducible component containing the identity element, hence it equals G_0. The cosets of G_0 (left or right) are images of G_0 under automorphisms, therefore also irreducible components. Thus, there can only be finitely many of them. □

REMARK 3.87. Proposition 3.86, (2), implies that for algebraic groups the irreducible components are disjoint. This is not true in general for algebraic varieties, consider, for instance, the subvariety $\{xy = 0\} \subset \mathbb{K}^2$.

We now relate Lie groups and algebraic groups. For $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, each algebraic subgroup $G < GL(n, \mathbb{F})$ is necessarily closed in the classical topology, hence, is a (complex, resp. real) Lie subgroup of $GL(n, \mathbb{F})$ by Theorem3.53. Below is a simpler argument which does not rely upon Cartan's theorem.

THEOREM 3.88. *Each algebraic subgroup* $G < GL(n, \mathbb{F})$ *is a Lie subgroup.*

PROOF. Since G is a subgroup, it is a (real or complex) submanifold in $GL(n, \mathbb{F})$ iff it contains a nonempty open subset which is a submanifold in $GL(n, \mathbb{F})$. The subgroup G is the zero-set of a polynomial map $p : GL(n, \mathbb{F}) \to \mathbb{F}^k$. Let r denote the maximum of ranks of the derivative dp of p on G, it is attained on an open nonempty subset U of G. Let V denote the subset of $GL(n, \mathbb{F})$ where dp has rank r. By the constant rank theorem (Theorem 2.2), V is a smooth submanifold of $GL(n, \mathbb{F})$. Applying the constant rank theorem to the restriction $p : V \to \mathbb{F}^k$, we conclude that $U = G \cap V$ is a smooth submanifold in V and, hence, in $GL(n, \mathbb{F})$. □

3.8. Group actions on complexes

3.8.1. ***G*–complexes.** Let G be a (discrete) topological group and let X be a cell complex, defined via disjoint unions of balls U_n and attaching maps e^n, see section 1.7.2. We say that X is a *G-complex*, or that we have a *cellular* action $G \curvearrowright X$, if $G \times X \to X$ is a topological action and for every n we have a G-action $G \curvearrowright U_n$ such that the attaching map

$$e^n : \partial U_n \to X^{(n-1)}$$

is G-equivariant.

DEFINITION 3.89. A cellular action $G \curvearrowright X$ is said to be *without inversions* if whenever $g \in G$ preserves a cell s in X, it fixes this cell pointwise.

A topological action $G \curvearrowright X$ on the geometric realization of a simplicial complex is called *simplicial* if it sends simplices to simplices and is affine on each simplex. As with cellular actions, a simplicial complex equipped with a simplicial group action is called a *simplicial G-complex*.

Equivalently, one describes simplicial G-complexes as follows. Let $G \curvearrowright V(X)$ be a (set-theoretic) action of G on the vertex set of a simplicial complex X, which sends simplices to simplices. Then this action defines a simplicial action of G on X. (Use the canonical affine extension of $G \curvearrowright V(X)$ to the geometric realization of each simplex.)

The following is immediate from the definition of X'', since barycentric subdivisions are canonical:

LEMMA 3.90. *Let X be an almost regular cell complex and $G \curvearrowright X$ be an action without inversions. Then $G \curvearrowright X$ induces a simplicial action without inversions $G \curvearrowright X''$.*

LEMMA 3.91. *Let X be a simplicial complex and $G \curvearrowright X$ be a free simplicial action. Then this action is properly discontinuous on X (in the weak topology).*

PROOF. Let K be a compact in X. Then K is contained in a finite union of simplices $\sigma_1, \ldots, \sigma_k$ in X. Let $F \subset G$ be the subset consisting of elements $g \in G$ so that $gK \cap K \neq \emptyset$. Then, assuming that F is infinite, it contains distinct elements g, h such that $g(\sigma) = h(\sigma)$ for some $\sigma \in \{\sigma_1, \ldots, \sigma_n\}$. Then $f := h^{-1}g(\sigma) = \sigma$. Since the action $G \curvearrowright X$ is linear on each simplex, f fixes a point in σ. This contradicts the assumption that the action of G on X is free. $\square$

3.8.2. Borel construction. Every group G admits a *classifying space* $E(G)$, which is a contractible cell complex admitting a free cellular action $G \curvearrowright E(G)$. The space $E(G)$ is far from being unique, we will use the one obtained by *Milnor's Construction*, see for instance [**Hat02**, Section 1.B]. A benefit of this construction is that $E(G)$ is a simplicial complex and the construction of $G \curvearrowright E(G)$ is canonical. Simplices in $E(G)$ are ordered tuples of elements of g: $[g_0, \ldots, g_n]$ is an n-simplex with the obvious inclusions. To verify contractibility of $E = E(G)$, note that each $i+1$-skeleton E^{i+1} contains the cone over the i-skeleton E^i, consisting of simplices of the form

$$[1, g_0, \ldots, g_n], g_0, \ldots, g_n \in G.$$

(The point $[1, \ldots, 1] \in E^{i+1}$ is the tip of this cone.) Therefore, the straight-line homotopy to $[1, \ldots, 1]$ gives the required contraction.

The group G acts on $E(G)$ by the left multiplication

$$g \times [g_0, \ldots, g_n] \mapsto [gg_0, \ldots, gg_n].$$

Clearly, this action is free and, moreover, each simplex has trivial stabilizer. The action $G \curvearrowright E(G)$ has two obvious properties that we will be using:

1. If the group G is finite, then each skeleton $E(G)^i$ is compact.
2. If $G_1 \leqslant G_2$, then there exists an equivariant embedding $E(G_1) \hookrightarrow E(G_2)$.

We will use only these properties and not the actual construction of $E(G)$ and the action $G \curvearrowright E(G)$.

Suppose now that X is a cell complex and $G \curvearrowright X$ is a cellular action without inversions. Our next goal is to replace X with a new cell complex $\widehat{X}$ which admits a homotopy-equivalence $p : \widehat{X} \to X$ so that the action $G \curvearrowright X$ lifts (*via* p) to a free cellular action $G \curvearrowright \widehat{X}$. The action $G \curvearrowright \widehat{X}$ is called the *Borel Construction*. We first explain the construction when when X is a simplicial complex since the idea is much clearer in this case.

For each simplex $\sigma \in X$ consider its (pointwise) stabilizer $G_\sigma \leqslant G$. Clearly, if $\sigma_1 \subset \sigma_2$, then

$$G_{\sigma_2} \leqslant G_{\sigma_1}.$$

For each simplex σ define $\widehat{X}_\sigma := \sigma \times E(G_\sigma)$. The group G_σ acts naturally on $\widehat{X}_\sigma$. Whenever $\sigma_1 \subset \mathrm{Supp}(\sigma_2)$ we have the natural embedding $E(G_{\sigma_1}) \hookrightarrow E(G_{\sigma_2})$ and, hence, embeddings

$$\widehat{X}_{\sigma_1} = \sigma_1 \times E(G_{\sigma_1}) \supset \sigma_1 \times E(G_{\sigma_2}) \subset \widehat{X}_{\sigma_2}.$$

Henceforth, we glue $\widehat{X}_{\sigma_2}$ to $\widehat{X}_{\sigma_1}$ by identifying the two copies of the product subcomplex $\sigma_1 \times E(G_{\sigma_2})$. Let $\widehat{X}$ denote the regular cell complex resulting from these identifications.

For general cell complexes we have to modify the above construction. Define the *support* $\mathrm{Supp}(\sigma)$ of an n-cell σ in X to be the smallest subcomplex in X whose underlying space contains the image of $\mathbb{S}^{n-1}$ under the attaching map of σ. Since G acts on X without inversions, for every $\sigma_1 \subset \mathrm{Supp}(\sigma_2)$,

$$G_{\sigma_2} \leqslant G_{\sigma_1},$$

where G_σ is the stabilizer of σ in G. As before, for each n-dimensional cell σ define $\widehat{X}_\sigma := D^n \times E(G_\sigma)$. The group G_σ acts on $\widehat{X}_\sigma$ preserving the product structure and fixing D^n pointwise. Whenever $\sigma_1 \subset \mathrm{Supp}(\sigma_2)$ we have the natural embedding $E(G_{\sigma_1}) \hookrightarrow E(G_{\sigma_2})$ and hence inclusions

$$\widehat{X}_{\sigma_1} = \sigma_1 \times E(G_{\sigma_1}) \supset \sigma_1 \times E(G_{\sigma_2}) \subset \mathrm{Supp}(\sigma_2) \times E(G_{\sigma_2}).$$

At the same time, we have the attaching map $e_{\sigma_2} : \partial D^n \to \mathrm{Supp}(\sigma_2)$ and, thus the attaching map

$$\widehat{e}_{\sigma_2} := e_{\sigma_2} \times Id : \partial D^n \times E(G_{\sigma_2}) \to \mathrm{Supp}(\sigma_2) \times E(G_{\sigma_2}).$$

Here n is the dimension of the cell σ_2. We now define $\widehat{X}$ by induction on skeleta of X. We begin with $\widehat{X}_0$ obtained by replacing each 0-cell σ in X with $\widehat{X}_\sigma$. Assume that $\widehat{X}_{n-1}$ is constructed by gluing spaces $\widehat{X}_\tau$, where τ's are cells in $X^{(n-1)}$. For each n-cell σ the attaching map $\widehat{e}_\sigma$ defined above will yield an attaching map

$$\partial D^n \times E(G_\sigma) \to \widehat{X}_{n-1}.$$

We then glue the spaces $\widehat{X}_\sigma$ to $\widehat{X}_{n-1}$ *via* these attaching maps. We have a natural projection $p : \widehat{X} \to X$ which corresponds to the projection

$$\widehat{X}_\sigma := D^n \times E(G_\sigma) \to D^n$$

for each n-cell σ in X. Since each D^n is contractible, it follows that p restricts to a homotopy-equivalence

$$\widehat{X}_n \to X^{(n)}$$

for every n. Naturality of the construction ensures that the action $G \curvearrowright X$ lifts to an action $G \curvearrowright \widehat{X}$; it is clear from the construction that for each cell σ, the stabilizer of $\widehat{X}_\sigma$ in G is G_σ. Since G_σ acts freely on $E(G_\sigma)$, it follows that the action $G \curvearrowright \widehat{X}$ is free. Suppose now that $G \curvearrowright X$ is properly discontinuous. Then, G_σ is finite for each σ and, thus $\widehat{X}_\sigma$ has finite i-skeleton for each i. Moreover, if X/G were compact, then the action of G on each i-skeleton of $\widehat{X}$ is compact as well.

The construction of the complex $\widehat{X}$ and the action $G \curvearrowright \widehat{X}$ is called the *Borel construction*. One application of the Borel construction is the following

LEMMA 3.92. *Suppose that $G \curvearrowright X$ is a cocompact properly discontinuous action. Then there exists a properly discontinuous, cellular, free action $G \curvearrowright \widehat{X}$ which is cocompact on each skeleton and so that X is homotopy-equivalent to $\widehat{X}$.*

3.8.3. Groups of finite type. If G is a group admitting a free properly discontinuous cocompact action on a graph Γ, then G is finitely generated, as, by the covering theory, $G \cong \pi_1(\Gamma/G)/p_*(\pi_1(\Gamma))$, where $p : \Gamma \to \Gamma/G$ is the covering map. Groups of *finite type* $\mathbf{F}_n$ are higher-dimensional generalizations of this example.

DEFINITION 3.93. A group G is said to have *type* $\mathbf{F}_n$, $1 \leqslant n \leqslant \infty$, if it admits a free properly discontinuous cellular action on an $n-1$-connected n-dimensional cell complex Y, which is cocompact on each skeleton. A group G has the *type* $\mathbf{F}$ if there exists a finite $K(G, 1)$ complex.

Clearly,

$$\mathbf{F} \subset \mathbf{F}_\infty \subset \ldots \mathbf{F}_n \subset \mathbf{F}_{n-1} \subset \ldots \subset \mathbf{F}_1.$$

Note that we allow the complex Y in the definition of groups of finite type to be infinite-dimensional. Thus, a group has type $\mathbf{F}_\infty$ if and only if it admits a free properly discontinuous cellular action on a contractible cell complex Y such that $Y^{(i)}/G$ is finite for each i.

Observe also that a group G is finitely generated if and only it it has type $\mathbf{F}_1$. In Lemma 4.89 we will prove that a group G is finitely presented if and only if it has type $\mathbf{F}_2$.

EXAMPLE 3.94 (See [**Bie76b**]). Let $\mathbb{F}_2$ be free group on 2 generators a, b. Consider the group $G = \mathbb{F}_2^n$ which is the n-fold direct product of $\mathbb{F}_2$. Define the homomorphism $\phi : G \to \mathbb{Z}$ which sends all generators a_i, b_i of G to the generator $1 \in \mathbb{Z}$. Let $K := Ker(\phi)$. Then K is of type $\mathbf{F}_{n-1}$ but not of type $\mathbf{F}_n$.

In view of Lemma 3.92, we obtain:

COROLLARY 3.95. *A group G has type $\mathbf{F}_n$ if and only if it admits a properly discontinuous cocompact cellular action on an $n-1$-connected n-dimensional cell complex X, which is cocompact on each skeleton.*

PROOF. One direction is obvious. Suppose, therefore, that we have an action $G \curvearrowright X$ as above. We apply Borel construction to this action and obtain a free properly discontinuous action $G \curvearrowright \widehat{X}$ which is cocompact on each skeleton of $\widehat{X}$. If $n = \infty$, we let $Y := \widehat{X}$. Otherwise, we let Y denote the n-skeleton of $\widehat{X}$. Recall that the inclusion $Y \hookrightarrow \widehat{X}$ induces monomorphisms of all homotopy groups π_j, $j \leqslant n-1$. Since X is $n-1$–connected, the same holds for $\widehat{X}$ and hence Y. ☐

COROLLARY 3.96. *Every finite group has type $\mathbf{F}_\infty$.*

PROOF. Start with the action of G on a complex X which is a point and then apply the above corollary. ☐

3.9. Cohomology

The purpose of this section is to introduce cohomology of groups and to give explicit formulae for cocycles and coboundaries in small degrees. We refer the reader to [**Bro82b**, Chapter III, Section 1] for the more thorough discussion. We will also connect group cohomology to two group-theoretic constructions: *Semidirect products* and *coextensions*.

3.9.1. Group rings and modules. Suppose that R is a commutative ring with unit element 1. The *R-ring* RG of a group G is the set of formal sums $\sum_{g\in G} m_g\, g$, where m_g are elements of R which are equal to zero for all but finitely many values of g. The most important examples for us will be the integer group ring $\mathbb{Z}G$ and the rational group ring $\mathbb{Q}G$. So far, RG is just a set, but it becomes a ring once endowed with the two operations:

- addition:
$$\sum_{g\in G} m_g\, g + \sum_{g\in G} n_g\, g = \sum_{g\in G} (m_g + n_g)\, g$$
- multiplication defined by the convolution of maps to $\mathbb{Z}$, that is
$$\sum_{a\in G} m_a\, a + \sum_{b\in G} n_b\, b = \sum_{g\in G} \left(\sum_{ab=g} m_a n_b \right) g\,.$$

According to a Theorem of G. Higman [**Hig40**], every integer group ring is an integral domain. Both R and G embed as subsets of RG by identifying every $m \in \mathbb{Z}$ with $m1_G$ and every $g \in G$ with $1g$. Every group homomorphism $\varphi : G \to H$ induces a homomorphism between group rings, which by abuse of notation we shall denote also by φ. In particular, the trivial homomorphism $o : G \to \{1\}$ induces a retraction $o : \mathbb{Z}G \to R$, called the *augmentation*. If the homomorphism $\varphi : G \to H$ is an isomorphism, then so is the homomorphism between group rings. This implies that an action of a group G on another group H (by automorphisms) extends to an action of G on the group ring $\mathbb{Z}H$ (by automorphisms).

Let L be a ring and M be an abelian group. We say that M is a (left) L-module if we are given a map
$$(\ell, m) \mapsto \ell \cdot m, L \times M \to M,$$
which is additive in both variables and so that
$$(\ell_1 \star \ell_2) \cdot m = \ell_1 \cdot (\ell_2 \cdot m), \tag{3.4}$$
where $\star$ denotes the multiplication operation in L.

Similarly, M is a *right* L-module if we are given an additive (in both variables) map
$$(m, \ell) \mapsto m \cdot \ell, M \times L \to M,$$
so that
$$m \cdot (\ell_1 \star \ell_2) = (m \cdot \ell_1) \cdot \ell_2. \tag{3.5}$$

Lastly, M is an *L-bimodule* if M has structure of both left and right L-module.

In the case when R is a field F (say, $R = \mathbb{Q}$), a left RG-module is an F-vector space endowed with a linear G-action. In the case $R = \mathbb{Z}$, we will refer to (left) $\mathbb{Z}G$-modules simply as G-modules.

3.9.2. Group cohomology. Let G be a group and let M, N be left $\mathbb{Z}G$-modules; then $Hom_G(M, N)$ denotes the $\mathbb{Z}$-submodule of G-invariant elements in the $\mathbb{Z}$-module $Hom(M, N)$, where G acts on homomorphisms (of abelian groups) $u : M \to N$ by the formula:

$$(gu)(m) = g \cdot u(g^{-1}m).$$

If C_* is a chain complex of abelian groups and A is a G-module, then $Hom_G(C_*, A) \subset Hom(C_*, A)$ is the chain subcomplex formed by submodules $Hom_G(C_k, A)$ in $Hom(C_k, A)$. The *standard chain complex* $C_* = C_*(G)$ of G with coefficients in A is defined as follows:

$C_k(G) = \mathbb{Z} \otimes \prod_{i=0}^{k} G$, is the $\mathbb{Z}G$-module freely generated by $(k+1)$-tuples $(g_0, \ldots, g_k)$ of elements of G with the G-action given by

$$g \cdot (g_0, \ldots, g_k) = (gg_0, \ldots, gg_k).$$

The reader should think of each tuple as spanning a k-simplex. The boundary operator on this chain complex is the natural one:

$$\partial_k(g_0, \ldots, g_k) = \sum_{i=0}^{k} (-1)^i (g_0, \ldots, \hat{g}_i, \ldots g_k),$$

where $\hat{g}_i$ means that we omit this entry in the $k+1$-tuple. The dual cochain complex C^* is defined by:

$$C^k = Hom(C_k, A), \quad \delta_k(f)((g_0, \ldots, g_{k+1})) = f(\partial_{k+1}(g_0, \ldots, g_{k+1})), f \in C^k.$$

Thus, C_* and C^* are just the simplicial chain and cochain complexes of the simplicial complex defining the Milnor's classifying space EG of the group G (see Section 3.8.2), with which the reader is probably familiar with from a basic algebraic topology course.

Suppose for a moment that A is a trivial G-module. Then, for $BG = (EG)/G$, the simplicial cochain complex $C^*(BG, A)$ is naturally isomorphic to the subcomplex of G-invariant cochains in $C^*(G, A)$, i.e., the subcomplex

$$(C^*(G, A))^{RG} = Hom_{RG}(C_*, A).$$

If A is a nontrivial RG-module, then $Hom_{RG}(C_*, A)$ is still isomorphic to a certain natural cochain complex based on the simplicial complex $C_*(BG)$ (a cochain complex with twisted coefficients, or coefficients in a certain sheaf), but the definition is more involved and we will omit it.

Definition 3.97. Define the subspaces of i-cocycles and i-coboundaries in $Hom_G(C_i, A)$ as

$$Z^i(G, A) := \mathrm{Ker}(\delta_i), \quad B^i(G, A) := \mathrm{Im}(\delta_{i-1}),$$

respectively. The cohomology groups of G with coefficients in the G-module A are defined as

$$H^*(G, A) := H_*(Hom_G(C_*, A)).$$

In other words,

$$H^i(G, A) = Z^i(G, A)/B^i(G, A).$$

In particular, if A is a trivial G-module, then $H^*(G, A) = H^*(BG, A)$.

DEFINITION 3.98. The *(integer) cohomological dimension* of a group G, is defined as

$$cd(G) = \sup\{q \in \mathbb{Z} : \exists A, \text{a } \mathbb{Z}G\text{-module, such that } H^q(G, A) \neq 0\}.$$

Note that the definition of cohomological dimension we gave is, in fact, a theorem rather than the standard definition. We refer the reader to **[Bro82b]** for the usual definition of cohomological dimension in terms of projective resolutions.

EXAMPLE 3.99. 1. Suppose that G admits a finite $K(G, 1)$ CW complex X. Then $cd(G) \leqslant \dim(X)$.

2. If G is a nontrivial finite group, then $cd(G) = \infty$.

REMARK 3.100. Analogously to the integer cohomological dimension, one defines the *rational* cohomological dimension $cd_{\mathbb{Q}}(G)$ as the supremum of degrees q such that $H^q(G, A) \neq 0$ for some G-module A which is also a vector space over $\mathbb{Q}$ (on which G acts linearly).[1] One advantage $cd_{\mathbb{Q}}$ has over the integer cohomological dimension is that (unlike the latter) the former is an invariant under virtual isomorphisms of groups (see **[Bro82b]**) and, for finitely generated groups, is invariant under quasiisometries, see Theorem 6.62.

So far, all definitions looked very natural. Our next step is to reduce the number of variables in the definition of cochains by one using the fact that cochains in $Hom_G(C_k, A)$ are G-invariant. The drawback of this reduction, as we will see, will be lack of naturality, but the advantage will be new formulae for cohomology groups which are useful in some applications.

By G-invariance, for $f \in Hom_G(C_k, A)$ we have:

$$f(g_0, \ldots, g_k) = g_0 \cdot f(1, g_0^{-1}g_1, \ldots, g_0^{-1}g_k).$$

In other words, it suffices to restrict cochains to the set of $(k+1)$-tuples where the first entry is $1 \in G$. Every such tuple has the form

$$(1, g_1, g_1g_2, \ldots, g_1 \cdots g_k)$$

(we will see below why). The latter is commonly denoted

$$[g_1|g_2|\ldots|g_k].$$

Note that, computing the value of the coboundary,

$$\delta_{k-1}f(1, g_1, g_1g_2, \ldots, g_1 \cdots g_k) = \delta_{k-1}f([g_1|g_2|\ldots|g_k])$$

we get

$$\delta_{k-1}f(1, g_1, g_1g_2, \ldots, g_1 \cdots g_k) =$$

$$f(g_1, \ldots, g_1 \cdots g_k) - f(1, g_1g_2, \ldots, g_1 \cdots g_k) + f(1, g_1, g_1g_2g_3, \ldots, g_1 \cdots g_k) - \ldots =$$

$$g_1 \cdot f(1, g_2, \ldots, g_2 \cdots g_k) - f([g_1g_2|g_3|\ldots|g_k]) + f([g_1|g_2g_3|g_4|\ldots|g_k]) - \ldots =$$

$$g_1 \cdot f([g_2|\ldots|g_k]) - f([g_1g_2|g_3|\ldots|g_k]) + f([g_1|g_2g_3|g_4|\ldots|g_k]) - \ldots$$

[1] Even more generally, given an arbitrary commutative ring R one defines the appropriate group cohomology using the group ring RG instead of $\mathbb{Z}G$, and the cohomological dimension $cd_R(G)$, see **[Bro82b]**.

Thus,

$$\delta_{k-1}f([g_1|g_2|\dots|g_k]) = g_1 \cdot f([g_2|\dots|g_k]) - f([g_1g_2|g_3|\dots|g_k]) +$$

$$f([g_1|g_2g_3|g_4|\dots|g_k]) - \dots$$

Then, we let $\bar{C}^k$ ($k \geq 1$) denote the abelian group of functions f sending k-tuples $[g_1|\dots|g_k]$ of elements of G to elements of A; we equip these groups with the above coboundary homomorphisms δ_k. For $k = 0$, we have to use the empty symbol $[\ ]$, $f([\ ]) = a \in A$, so that such functions f are identified with elements of A. Thus, $\bar{C}_0 = A$ and the above formula for δ_0 reads as:

$$\delta_0 : a \mapsto c_a, \quad c_a([g]) = g \cdot a - a.$$

The resulting chain complex $(\bar{C}_*, \delta_*)$ is called the *inhomogeneous bar complex* of G with coefficients in A. We now compute the coboundary maps δ_k for this complex for small values of k:

(1) $\delta_0 : a \mapsto f_a, \quad f_a([g]) = g \cdot a - a.$
(2) $\delta_1(f)([g_1, g_2]) = g_1 \cdot f([g_2]) - f([g_1g_2]) + f([g_1]).$
(3) $\delta_2(f)([g_1|g_2|g_3]) = g_1 \cdot f([g_2|g_3]) - f([g_1g_2|g_3]) + f([g_1|g_2g_3]) - f([g_1|g_2]).$

Therefore, spaces of coboundaries and cocycles for $(\bar{C}_*, \delta_*)$ in small degrees are (we now drop the bar notation for simplicity):

(1) $B^1(G, A) = \{f_a : G \to A, \forall a \in A | f_a(g) = g \cdot a - a\}.$
(2) $Z^1(G, A) = \{f : G \to A | f(g_1g_2) = f(g_1) + g_1 \cdot f(g_2)\}.$
(3) $B^2(G, A) = \{h : G \times G \to A | \exists f : G \to A, h(g_1, g_2) = f(g_1) - f(g_1g_2) + g_1 \cdot f(g_2)\}.$
(4) $Z^2(G, A) = \{f : G \times G \to A | g_1 \cdot f(g_2, g_3) - f(g_1, g_2) = f(g_1g_2, g_3) - f(g_1, g_2g_3)\}.$

Let us look at the definition of $Z^1(G, A)$ more closely. In addition to the left action of G on A, we define a *trivial right action* of G on A: $a \cdot g = a$. Then a function $f : G \to A$ is a 1-cocycle if and only if

$$f(g_1g_2) = f(g_1) \cdot g_2 + g_1 \cdot f(g_2).$$

The reader will immediately recognize here the Leibnitz formula for the derivative of the product. Hence, 1-cocycles $f \in Z^1(G, A)$ are called *derivations* of G with values in A. The 1-coboundaries are called *principal derivations* or *inner derivations.* If A is trivial as a left G-module, then, of course, all principal derivations are zero and derivations are just homomorphisms $G \to A$.

Nonabelian derivations. The notions of derivation and principal derivation can be extended to the case when the target group is nonabelian; we will use the notation N for the target group with the binary operation $\star$ and $g \cdot n$ for the action of G on N by automorphisms, i.e.,

$$g \cdot n = \varphi(g)(n), \qquad \text{where } \varphi : G \to Aut(N) \text{ is a homomorphism.}$$

DEFINITION 3.101. A function $d : G \to N$ is called a *derivation* if

$$d(g_1g_2) = d(g_1) \star g_1 \cdot d(g_2), \quad \forall g_1, g_2 \in G.$$

A derivation is called *principal* if it is of the form $d = d_n$, where

$$d_n(g) = n^{-1} \star (g \cdot n).$$

The space of derivations is denoted $Der(G,N)$ and the subspace of principal derivations is denoted $Prin(G,N)$ or, simply, $P(G,N)$.

EXERCISE 3.102. Verify that every principal derivation is indeed a derivation.

EXERCISE 3.103. Verify that every derivation d satisfies

- $d(1) = 1$;
- $d(g^{-1}) = g^{-1} \cdot [d(g)]^{-1}$.

We will use derivations in the context of free solvable groups in section 11.6. In section 3.9.5 we will discuss derivations in the context of semidirect products, while in section 3.9.6 we explain how 2nd cohomology group $H^2(G,A)$ can be used to describe central coextensions.

Nonabelian cohomology. We would like to define the 1-st cohomology $H^1(G,N)$, where the group N is nonabelian and we have an action of G on N. The problem is that neither $Der(G,N)$ nor $Prin(G,N)$ is a group, so taking quotient $Der(G,N)/Prin(G,N)$ makes no sense. Nevertheless, we can think of the formula

$$f \mapsto f + d_a, a \in A,$$

in the abelian case (defining action of $Prin(G,A)$ on $Der(G,A)$) as the *left* action of the group A on $Der(G,A)$:

$$a(f) = f', \quad f'(g) = -a + f(g) + (g \cdot a).$$

The latter generalizes in the nonabelian case, as the group N acts to the left on $Der(G,N)$ by

$$n(f) = f', \quad f'(g) = n^{-1} \star f(g) \star (g \cdot n).$$

Then, one defines $H^1(G,N)$ as the quotient

$$N \backslash Der(G,N).$$

EXAMPLE 3.104. 1. Suppose that G-action on N is trivial. Then $Der(G,N) = Hom(G,N)$ and N acts on homomorphisms $f : G \to N$ by postcomposition with inner automorphisms. Thus, $H^1(G,N)$ in this case is

$$N \backslash Hom(G,N),$$

the set of conjugacy classes of homomorphisms $G \to N$.

2. Suppose that $G \cong \mathbb{Z} = \langle 1 \rangle$ and the action φ of $\mathbb{Z}$ on N is arbitrary. We have $\eta := \varphi(1) \in Aut(N)$. Then $H^1(G,N)$ is the set of *twisted conjugacy classes of elements of* N: Two elements $m_1, m_2 \in N$ are said to be in the same η-twisted conjugacy class if there exists $n \in N$ so that

$$m_2 = n^{-1} \star m_1 \star \eta(n).$$

Indeed, every derivation $d \in Der(\mathbb{Z},N)$ is determined by the image $m = d(1) \in N$. Then two derivations d_i so that $m_i = d_i(1)$ $(i = 1,2)$ are in the same N-orbit if m_1, m_2 are in the same η-twisted conjugacy class.

3.9.3. Bounded cohomology of groups. *An isometric Banach* $\mathbb{Z}G$*-module* V if a Banach space equipped with an isometric action of the group on G. Using $C_*(G)$, which is the bar-complex of G, one defines the *bounded cochain complex*

$$C_b^*(G,V) = Hom_{G,b}(C_*,V),$$

where $C_b^k(G,V)$ consists of G-equivariant bounded maps $G^{k+1} \to V$, with the usual coboundary operator. Accordingly, one defines the *bounded cohomology groups* of G with coefficients in V:

$$H_b^*(G,V) := H_*(C_b^*(G,V)).$$

Alternatively, one can use the subcomplex of bounded functions $\bar{C}_b^*(G,V)$ in the inhomogeneous bar-complex of the group G and obtain

$$H_b^k(G,V) \cong Z_b^k(G,V)/B_b^k(G,V),$$

where the spaces of cocycles and coboundaries in the right hand-side refer to the bounded elements of the groups of inhomogeneous cocycles and coboundaries.

The same definitions go through if instead of the entire V one uses a $\mathbb{Z}G$-submodule $A \subset V$; then one defines the bounded cohomology groups $H_b^k(G,A)$ via maps $G^{k+1} \to A$.

We now consider the special case, when V (and, hence, A) is a trivial G-module. (The most important cases are, of course, $V = \mathbb{R}$ and $A = \mathbb{Z}$, $A = \mathbb{R}$.) Then for a classifying space $Y = BG$ of G one defines the subcomplex $C_b^*(Y,A)$ of the cochain complex $C^*(Y,A)$. The homology of this subcomplex is the *bounded cohomology* $H_b^*(Y,A)$ of Y with coefficients in A.

EXERCISE 3.105. Verify that $H_b^*(Y,A) \cong H_b^*(G,A)$.

Note that the above isomorphism holds even if Y is not a $K(G,1)$ but merely has G as its fundamental group, see [**Bro81b**].

It is instructive to identify elements of $Z_b^2(G,A)$, where A is a subgroup of $\mathbb{R}$, which appear as ordinary coboundaries: For $f \in C^1(G,A)$, i.e., $f : G \to A$,

$$\delta_1(f)([g_1,g_2]) = f(g_2) - f(g_1g_2) + f(g_1)$$

is a bounded 2-cocycle if and only if there exists a constant D so that for all $g_1, g_2 \in G$,

$$|f(g_1) + f(g_2) - f(g_1g_2)| \leqslant D. \tag{3.6}$$

In other words, such f is "almost a homomorphism $f : G \to A$", with the error $\leq D$ in the definition of a homomorphism.

DEFINITION 3.106. A map $f : G \to \mathbb{R}$ is called a *quasimorphism* if it satisfies the inequality (3.6) for all $g_1, g_2 \in G$ and a fixed constant D.

Quasi-morphisms appear frequently in geometric group theory; they were first used by R. Brooks in [**Bro81b**], who was proving that, while for the free group F_n of rank $n \geq 2$, $H^2(F_n,\mathbb{R}) = 0$, nevertheless, the vector space $H_b^2(F_n,\mathbb{R})$ is infinite-dimensional. Namely, he constructed an infinite-dimensional space of equivalence classes of quasimorphisms $F_n \to \mathbb{R}$, where

$$f_1 \sim f_2 \iff \|f_1 - f_2\| < \infty.$$

Taking coboundaries of these quasimorphisms shows that $H_b^2(F_n,\mathbb{R})$ has infinite dimension.

Many interesting groups do not admit nontrivial homomorphisms of $\mathbb{R}$ but admit unbounded quasimorphisms. For instance, a hyperbolic Coxeter group G does not admit nontrivial homomorphisms to $\mathbb{R}$. However, if G is a nonelementary hyperbolic group, it has infinite-dimensional space of equivalence classes of quasimorphisms, see [**EF97a**] for details. We refer the reader to Monod's paper [**Mon06**] for a survey of applications of bounded cohomology of groups, as well as Calegari's book [**Cal09**] for the in-depth discussion of quasimorphisms defined by the *commutator norm*.

We will encounter elements of the group $H_b^2(G, \mathbb{Z})$ in Section 9.19 when discussing central coextensions of hyperbolic groups, as we will be proving subjectivity of the homomorphism $H_b^2(G, \mathbb{Z}) \to H^2(G, \mathbb{Z})$.

Analogously to the bounded cohomology, one defines ℓ_p-cohomology and ℓ_p-homology groups, we refer the reader to [**AG99, BP03, Pan95**] for the detailed discussion.

3.9.4. Ring derivations. Our next goal is to extend the notion of derivation in the context of (noncommutative) rings. Typical rings that the reader should have in mind are *integer group rings*.

Definition 3.107. Let M be an L-bimodule. A *derivation* (with respect to this bimodule structure) is a map $d : L \to M$ such that:

(1) $d(\ell_1 + \ell_2) = d(\ell_1) + d(\ell_2)$,
(2) $d(\ell_1 \star \ell_2) = d(\ell_1) \cdot \ell_2 + \ell_1 \cdot d(\ell_2)$.

The space of derivations is an abelian group, which will be denoted $Der(L, M)$.

Below is the key example of a bimodule that we will be using in the context of derivations. Let G, H be groups, $\varphi : G \to Bij(H)$ is an action of G on H by set-theoretic automorphisms. We let $L := \mathbb{Z}G, M := \mathbb{Z}H$ be the integer group rings, where we regard the ring M as an abelian group and ignore its multiplicative structure.

Every action $\varphi : G \curvearrowright H$ determines the left L-module structure on M by:

$$(\sum_i a_i g_i) \cdot (\sum_j b_j h_j) := \sum_{i,j} a_i b_j g_i \cdot h_j, \quad a_i \in \mathbb{Z}, b_j \in \mathbb{Z},$$

where $g \cdot h = \varphi(g)(h)$ for $g \in G, h \in H$. We define the structure of a *right L-module* on M by:

$$(m, \ell) \mapsto mo(\ell) = o(\ell)m, \quad o(\ell) \in \mathbb{Z},$$

where $o : L \to \mathbb{Z}$ is the augmentation of $\mathbb{Z}G = L$.

Derivations with respect for the above group ring bimodules will be called *group ring derivations*.

Exercise 3.108. Verify the following properties of group ring derivations:

(P_1) $d(1_G) = 0$, whence $d(m) = 0$ for every $m \in \mathbb{Z}$.

(P_2) $d(g^{-1}) = -g^{-1} \cdot d(g)$.

(P_3) $d(g_1 \cdots g_m) = \sum_{i=1}^m (g_1 \cdots g_{i-1}) \cdot d(g_i) o(g_{i+1} \cdots g_m)$.

(P_4) Every derivation $d \in Der(\mathbb{Z}G, \mathbb{Z}H)$ is uniquely determined by its values $d(x)$ on the generators x of G.

Fox Calculus. We now consider the special case when $G = H = F_X$, is the free group on the generating set X. In this context, the theory of derivations was developed by R. H. Fox in [**Fox53**].

LEMMA 3.109. *Every map $d : X \to M = \mathbb{Z}G$ extends to a group ring derivation $d \in Der(\mathbb{Z}G, M)$.*

PROOF. We set

$$d(x^{-1}) = -x^{-1} \cdot d(x), \quad \forall x \in X$$

and $d(1) = 0$. We then extend d inductively to the free group G by

$$d(yu) = d(y) + y \cdot d(u),$$

where $y = x \in X$ or $y = x^{-1}$ and yu is a reduced word in the alphabet $X \cup X^{-1}$. Lastly, we extend d by additivity to the rest of the ring $L = \mathbb{Z}G$. In order to verify that d is a derivation, we need to check only that

$$d(uv) = d(u) + u \cdot d(v),$$

where $u, v \in F_X$. The verification is a straightforward induction on the length of the reduced word u and is left to the reader. □

DEFINITION 3.110. To each generator $x_i \in X$ we associate a derivation ∂_i, called the *Fox derivative*, defined by $\partial_i x_j = \delta_{ij} \in \{0, 1\}$, which is regarded as the subset of $\mathbb{Z}\cdot 1_G \subset \mathbb{Z}G$. The maps ∂_i then extend to derivations $\partial_i \in Der(\mathbb{Z}F_X, \mathbb{Z}F_X)$ as in Lemma 3.109. In particular,

$$\partial_i(x_i^{-1}) = -x_i^{-1}.$$

Importance of the derivations ∂_i comes from:

PROPOSITION 3.111. *Suppose that $G = F_r$ is the free group of rank $r < \infty$. Then every derivation $d \in Der(\mathbb{Z}G, \mathbb{Z}G)$ can be written as a sum*

$$d = \sum_{i=1}^{r} k_i \partial_i \,, \quad \text{where } k_i = d(x_i) \in \mathbb{Z}\,.$$

Furthermore, $Der(\mathbb{Z}G, \mathbb{Z}G)$ is a free abelian group with the basis $\partial_i, i = 1, \ldots, r$.

PROOF. The first assertion immediately follows from Exercise 3.108 (part (P_4)), and from the fact that both sides of the equation evaluated on x_j equal k_j. Thus, the derivations $\partial_i, i = 1, \ldots, k$, generate $Der(\mathbb{Z}G, \mathbb{Z}G)$. Independence of these generators follows from the fact that $\partial_i x_j = \delta_{ij}$. □

3.9.5. Derivations and split extensions. Components of homomorphisms to semidirect products.

DEFINITION 3.112. Let G and L be two groups and let N, H be subgroups in G.

(1) Assume that $G = N \times H$. Every group homomorphism $F : L \to G$ splits as a product of two homomorphisms $F = (f_1, f_2)$, $f_1 : L \to N$ and $f_2 : L \to H$, called the *components* of F.
(2) Assume now that G is a semidirect product $N \rtimes H$. Then every homomorphism $F : L \to G$ determines (and is determined by) a pair (d, f), where

- $f : L \to H$ is a homomorphism (the composition of F and the retraction $G \to H$);
- a map $d = d_F : L \to N$, called *derivation* associated with F. The derivation d is determined by the formula

$$F(\ell) = d(\ell)f(\ell).$$

EXERCISE 3.113. Show that d is indeed a derivation in the sense of Section 3.9.2.

EXERCISE 3.114. Verify that for every derivation d and a homomorphism $f : L \to H$ there exists a homomorphism $F : L \to G$ with the components d, f.

Extensions and coextensions.

DEFINITION 3.115. Given a short exact sequence

$$\{1\} \longrightarrow N \longrightarrow G \longrightarrow H \longrightarrow \{1\},$$

we call the group G an *extension of N by H* or a *coextension of H by N*.[2]

Given two classes of groups $\mathcal{A}$ and $\mathcal{B}$, the groups that can be obtained as extensions of N by H with $N \in \mathcal{A}$ and $H \in \mathcal{B}$, are called $\mathcal{A}$-by-$\mathcal{B}$ *groups* (e.g. abelian-by-finite, nilpotent-by-free etc.).

Two extensions defined by the short exact sequences

$$\{1\} \longrightarrow N_i \xrightarrow{\varphi_i} G_i \xrightarrow{\psi_i} H_i \longrightarrow \{1\}$$

$(i = 1, 2)$ are *equivalent* if there exist isomorphisms

$$f_1 : N_1 \to N_2, \quad f_2 : G_1 \to G_2, \quad f_3 : H_1 \to H_2$$

that determine a commutative diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N_1 & \longrightarrow & G_1 & \longrightarrow & H_1 & \longrightarrow & 1 \\ & & \downarrow {\scriptstyle f_1} & & \downarrow {\scriptstyle f_2} & & \downarrow {\scriptstyle f_3} & & \\ 1 & \longrightarrow & N_2 & \longrightarrow & G_2 & \longrightarrow & H_2 & \longrightarrow & 1 \end{array}$$

We now use the notion of an isomorphism of exact sequences to reinterpret the notion of a split extension.

PROPOSITION 3.116. *Consider a short exact sequence*

$$1 \to N \xrightarrow{\iota} G \xrightarrow{\pi} Q \to 1\,. \tag{3.7}$$

The following are equivalent:

(1) *the sequence splits;*

(2) *there exists a subgroup H in G such that the projection π restricted to H becomes an isomorphism.*

(3) *the extension G is equivalent to an extension corresponding to a semidirect product $N \rtimes Q$;*

[2] Our terminology is a bit nonstandard, as both constructions are called *extensions* in the literature. We settled on the coextension terminology following the paper [**MN82**] where it was used for semigroups.

(4) *there exists a subgroup H in G such that $G = N \rtimes H$.*

Proof. It is clear that (2) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2): Let $\sigma : Q \to \sigma(H) \subset G$ be a section. The equality $\pi \circ \sigma = \mathrm{id}_Q$ implies that π restricted to H is both surjective and injective.

The implication (2) $\Rightarrow$ (3) is obvious.

(3) $\Rightarrow$ (2): Assume that there exists H such that $\pi|_H$ is an isomorphism. The fact that it is surjective implies that $G = NH$. The fact that it is injective implies that $H \cap N = \{1\}$.

(2) $\Rightarrow$ (4): Since π restricted to H is surjective, it follows that for every $g \in G$ there exists $h \in H$ such that $\pi(g) = \pi(h)$, hence $gh^{-1} \in \mathrm{Ker}\,\pi = \mathrm{Im}\,\iota$.

Assume that $g \in G$ can be written as $g = \iota(n_1)h_1 = \iota(n_2)h_2$, with $n_1, n_2 \in N$ and $h_1, h_2 \in H$. Then $\pi(h_1) = \pi(h_2)$, which, by the hypothesis that π restricted to H is an isomorphism, implies $h_1 = h_2$, whence $\iota(n_1) = \iota(n_2)$ and $n_1 = n_2$ by the injectivity of ι.

(4) $\Rightarrow$ (2): The existence of the decomposition for every $g \in G$ implies that π restricted to H is surjective.

The uniqueness of the decomposition implies that $H \cap \mathrm{Im}\,\iota = \{1\}$, whence π restricted to H is injective. □

Remark 3.117. Every sequence with free nonabelian group Q splits, see Lemma 4.24.

Examples 3.118. (1) For $n \geqslant 1$, the short exact sequence

$$1 \longrightarrow (2\mathbb{Z})^n \longrightarrow \mathbb{Z}^n \longrightarrow \mathbb{Z}_2^n \longrightarrow 1$$

does not split.

(2) Let F_n be a free group of rank $n \geqslant 2$ (see Definition 4.20) and let F_n' be its commutator subgroup (see Definition 3.19). Note that the abelianization of F_n as defined in Proposition 3.21, (3), is $\mathbb{Z}^n$. The short exact sequence

$$1 \longrightarrow F_n' \longrightarrow F_n \longrightarrow \mathbb{Z}^n \longrightarrow 1$$

does not split.

From now on, we restrict to the case of exact sequences

$$1 \to A \xrightarrow{\iota} G \xrightarrow{\pi} Q \to 1, \tag{3.8}$$

where A is an abelian group. Recall that the set of derivations $Der(Q, A)$ has a natural structure of an abelian group.

Remarks 3.119. (1) The short exact sequence (3.8) uniquely defines an action of Q on A. Indeed, G acts on A by conjugation and, since the kernel of this action contains A, it defines an action of Q on A. In what follows we shall denote this action by $(q, a) \mapsto q \cdot a$, and by φ the homomorphism $Q \to \mathrm{Aut}(A)$ defined by this action.

(2) If the short exact sequence (3.8) splits, the group G is isomorphic to $A \rtimes_\varphi Q$.

Classification of splittings.

Below we discuss classification of all splittings of short exact sequences (3.8) which do split. We use the additive notation for the binary operation on A. We begin with few observations. From now on, we fix a section σ_0 and, hence, a semidirect product decomposition $G = A \rtimes Q$. Note that every splitting of a short exact sequence (3.8), is determined by a section $\sigma : Q \to G$. Furthermore, every section $\sigma : Q \to G$ is determined by its components (d_σ, π) with respect to the semidirect product decomposition given by σ_0 (see Remark 3.112). Since π is fixed, a section σ is uniquely determined by its derivation d_σ. Conversely, every derivation $d \in Der(Q, A)$ determines a section σ, so that $d = d_\sigma$. Thus, the set of sections of (3.8) is in bijective correspondence with the abelian group of derivations $Der(Q, A)$.

Our next goal is to discuss the equivalence relation between different sections (and derivations). We say that an automorphism $\alpha \in Aut(G)$ is a *shearing* (with respect to the semidirect product decomposition $G = A \rtimes Q$) if $\alpha(A) = A, \alpha|A = Id$ and α projects to the identity on Q. Examples of shearing automorphisms are *principal shearing automorphisms*, which are given by conjugations by elements $a \in A$. It is clear that shearing automorphisms act on splittings of the short exact sequence (3.8).

Exercise 3.120. The group of shearing automorphisms of G is isomorphic to the abelian group $Der(Q, A)$: Every derivation $d \in Der(Q, A)$ determines a shearing automorphism $\alpha = \alpha_d$ of G by the formula

$$\alpha(a \star q) = (a + d(q)) \star q,$$

which gives the bijective correspondence.

In view of this exercise, the classification of splittings modulo shearing automorphisms yields a very boring answer: All sections are equivalent under the group of shearing transformations. A finer classification of splittings is given by the following definition. Two splittings σ_1, σ_2 are said to be A-*conjugate* if they differ by a principal shearing automorphism: There exists $a \in A$ such that

$$\sigma_2(q) = a\sigma_1(q)a^{-1}, \forall q \in Q.$$

If d_1, d_2 are the derivations corresponding to the sections σ_1, σ_2, then

$$(d_2(q), q) = (a, 1)(d_1(q), q)(-a, 1) \Leftrightarrow d_2(q) = d_1(q) - [q \cdot a - a]\,.$$

In other words, d_1, d_2 differ by the principal derivation corresponding to $a \in A$.

Thus, we proved the following

Proposition 3.121. *A-conjugacy classes of splittings of the short exact sequence* (3.8) *are in bijective correspondence with the quotient*

$$Der(Q, A)/Prin(Q, A),$$

where $Prin(Q, A)$ is the subgroup of principal derivations.

Note that $Der(Q, A) \cong Z^1(Q, A)$, $Prin(Q, A) = B^1(Q, A)$ and the quotient $Der(Q, A)/Prin(Q, A)$ is $H^1(Q, A)$, the first cohomology group of Q with coefficients in the $\mathbb{Z}Q$–module A.

Below is another application of $H^1(Q, A)$. Let L be a group and let

$$F : L \to G = A \rtimes Q$$

be a homomorphism. The group G, of course, acts on the homomorphisms F by postcomposition with inner automorphisms. Two homomorphisms are said to be *conjugate* if they belong to the same orbit of this G-action.

LEMMA 3.122. *1. A homomorphism $F : L \to G$ is conjugate to a homomorphism with the image in Q if and only if the derivation d_F of F is principal.*

2. Furthermore, suppose that $F_i : L \to G$ are homomorphisms with components $(d_i, \pi), i = 1, 2$. Then F_1 and F_2 are A-conjugate if and only if

$$[d_1] = [d_2] \in H^1(L, A).$$

PROOF. Let $g = qa \in G, a \in A, q \in Q$. If $(qa)F(\ell)(qa)^{-1} \in Q$, then $aF(\ell)a^{-1} \in Q$. Thus, for (1) it suffices to consider A-conjugation of homomorphisms $F : L \to G$. Hence, (2) $\Rightarrow$ (1). To prove (2) we note that the composition of F with an inner automorphism defined by $a \in A$ has the derivation equal to $d_F - d_a$, where d_a is the principal derivation determined by a. □

3.9.6. Central coextensions and second cohomology. We restrict ourselves to the case of central coextensions (a similar result holds for general extensions with abelian kernels, see e.g. **[Bro82b]**). In this case, A is trivial as a G-module and, hence, $H^*(G, A) \cong H^k(K(G, 1), A)$. This cohomology group can be also computed as $H^k(Y, A)$, where $G = \pi_1(Y)$ and Y is a $k+1$-connected cell complex.

Let G be a group and A an abelian group. A *central coextension* of G by A is a short exact sequence

$$1 \to A \xrightarrow{\iota} \tilde{G} \xrightarrow{r} G \to 1$$

where $\iota(A)$ is contained in the center of $\tilde{G}$. Choose a set-theoretic section

$$s : G \to \tilde{G}, s(1) = 1, r \circ s = Id.$$

Then, the group $\tilde{G}$ is be identified (as a set) with the direct product $A \times G$. With this identification, the group operation on $\tilde{G}$ has the form

$$(a, g) \cdot (b, h) = (a + b + f(g, h), gh),$$

where $f(1, 1) = 0 \in A$. Here the function $f : G \times G \to A$ measures the failure of s to be a homomorphism:

$$f(g, h) = s(g)s(h)\left(s(gh)\right)^{-1}.$$

Not every function $f : G \times G \to A$ corresponds to a central extension:

EXERCISE 3.123. A function f gives rise to a central coextension if and only if it satisfies the *cocycle identity*:

$$f(g, h) + f(gh, k) = f(h, k) + f(g, hk).$$

In other words, the set of such functions is the abelian group of cocycles $Z^2(G, A)$, see §3.9.2. We will refer to f simply as a *cocycle*.

Two central coextensions are said to be *equivalent* if there exist an isomorphism τ making the following diagram commutative:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & A & \longrightarrow & \tilde{G}_1 & \longrightarrow & G & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle id} & & \\
1 & \longrightarrow & A & \longrightarrow & \tilde{G}_2 & \longrightarrow & G & \longrightarrow & 1
\end{array}$$

EXERCISE 3.124. A coextension is trivial, meaning equivalent to the product $A \times G$, if and only if the central coextension splits.

We will use the notation $\mathbb{E}(G, A)$ to denote the set of equivalence classes of coextensions. In the language of cocycles, $r_1 \sim r_2$ if and only if

$$f_1 - f_2 = \delta c,$$

where $c : G \to A$, and

$$\delta c(g, h) = c(g) + c(h) - c(gh)$$

is the coboundary, $\delta c \in B^2(G, A)$. Recall that

$$H^2(G, A) = Z^2(G, A)/B^2(G, A)$$

is the 2-nd cohomology group of G with coefficients in A.

The set $\mathbb{E}(G, A)$ has natural structure of an abelian group, where the sum of two coextensions

$$A \to G_i \xrightarrow{r_i} G$$

is defined by

$$G_3 = \{(g_1, g_2) \in G_1 \times G_2 | r_1(g_1) = r_2(g_2)\} \xrightarrow{r} G,$$

$$r(g_1, g_2) = r_1(g_1) = r_2(g_2).$$

The kernel of this coextension is the subgroup A embedded diagonally in $G_1 \times G_2$. In the language of cocycles $f : G \times G \to A$, the sum of coextensions corresponds to the sum of cocycles and the trivial element is represented by the cocycle $f = 0$.

To summarize:

THEOREM 3.125 (See Chapter IV in [**Bro82b**].). *There exists an isomorphism of abelian groups*

$$H^2(K(G, 1), A) \cong H^2(G, A) \to \mathbb{E}(G, A).$$

The conclusion, thus, is that a group G with nontrivial 2-nd cohomology group $H^2(G, A)$ admits nontrivial central coextensions with the kernel A. How does one construct groups with nontrivial $H^2(G, A)$? Suppose that G admits a finite 2-dimensional $K(G, 1)$ complex Y, such that $\chi(G) := \chi(Y) \geqslant 2$. Then for $A \cong \mathbb{Z}$, we have

$$\chi(G) = 1 - b_1(Y) + b_2(Y) \geqslant 2 \Rightarrow b_2(Y) > 0.$$

The universal coefficients theorem shows that, for such groups G, if A is an abelian group which admits an epimorphism to $\mathbb{Z}$, then $H^2(G, A) \neq 0$.

Pull-backs of central coextensions. We fix an abelian group A and consider behavior of the groups $\mathbb{E}(G, A)$ under group homomorphisms $f : G_1 \to G_2$.

Lemma 3.126. *Every homomorphism $f : G_1 \to G_2$ induces a homomorphism*

$$f^* : \mathbb{E}(G_2, A) \to \mathbb{E}(G_1, A).$$

Moreover, f lifts to a homomorphism of the corresponding central extensions $\tilde{G}_1 \to \tilde{G}_2$.

Proof. Given a central coextension e_2:

$$0 \to A \to \tilde{\mathcal{G}}_2 \overset{p_2}{\to} G_2 \to 1,$$

we define a group $\tilde{G}_1$ as the *fiber product*:

$$\tilde{G}_1 := \{(g_1, \tilde{g}_2) \in G_1 \times \tilde{G}_2 : f(g_1) = p_2(\tilde{g}_2)\}.$$

The reader will verify that $\tilde{G}_1$ is a subgroup of the direct product $G_1 \times \tilde{G}_2$. The subgroup $A < 1 \times \tilde{G}_2$ is contained in the center of the product group. The subgroup $\tilde{G}_1 < G_1 \times \tilde{G}_2$ admits two projections: The projection to the first factor, G_1, which we denote p_1 and the projection to the second factor $\tilde{G}_2$, which we denote $\tilde{f}$. Let us identify the kernel of the homomorphism p_1:

$$p_1(g_1, \tilde{g}_2) = 1 \iff p_2(\tilde{g}_2) = 1 \iff \tilde{g}_2 \in A.$$

Therefore, the kernel of p_1 is naturally isomorphic to the group A. Hence, we obtain a central coextension $e_1 = f^*(e_2)$:

$$0 \to A \to \tilde{G}_1 \to G_1 \to 1 \tag{3.9}$$

and a homomorphism $\tilde{f} : \tilde{G}_1 \to \tilde{G}_2$, such that the following digram is commutative:

$$\begin{array}{ccc} \tilde{G}_1 & \xrightarrow{\tilde{f}} & \tilde{G}_2 \\ \downarrow{\scriptstyle p_1} & & \downarrow{\scriptstyle p_2} \\ G_1 & \xrightarrow{\tilde{f}} & G_1 \end{array} .$$

Thus, f indeed determines a natural map $f^* : \mathbb{E}(G_2, A) \to \mathbb{E}(G_1, A)$. We leave it to the reader to verify that f^* is a homomorphism. ☐

Exercise 3.127. 1. If f is surjective, so is $\tilde{f}$.

2. If s_2 is a set-theoretic section of p_2, then

$$s_1(g_1) = (g_1, s_2 f(g_1))$$

is a set-theoretic section of p_1.

3. Use Part 2 to verify commutativity of the diagram:

$$\begin{array}{ccc} H^2(G_2, A) & \xrightarrow{H^2(f)} & H^2(G_2, A) \\ \downarrow & & \downarrow \\ \mathbb{E}(G_2, A) & \xrightarrow{f^*} & \mathbb{E}(G_2, A) \end{array} .$$

Here the vertical arrows are the isomorphisms given by Theorem 3.125 and $H^2(f)$ is the homomorphism of 2nd cohomology groups induced by $f : G_1 \to G_2$. On the level of cocycles, the homomorphism $H^2(f)$ is given by

$$\omega_2 \in Z^2(G_2, A) \mapsto \omega_1 \in Z^2(G_1, A),$$
$$\omega_1(x, y) = \omega_2(f(x), f(y)).$$

Let us also identify the kernel of the homomorphism $\tilde{f}$. Suppose that s_1, s_2 are the sections as in Part 2 of Exercise 3.127. Assume that the section s_2 is normalized: $s_2(1) = 1$. Then for each $k \in K = \mathrm{Ker}(f)$, $s_1(k) = (k, 1)$, i.e., the restriction of s_1 to K is a homomorphism (even though, $s_1 : G_1 \to \tilde{G}_1$ is not, in general). Since $\tilde{f}$ is the restriction of the projection to the second factor, we conclude that

$$\tilde{K} = \mathrm{Ker}(\tilde{f}) = s_1(K).$$

In particular, kernels of f and $\tilde{f}$ are isomorphic.

Suppose for a moment, that the central coextension (3.9) splits, i.e., there exists a homomorphism $s : G_1 \to \tilde{G}_1$ right-inverse to p_1. Then the homomorphisms $s\big|_K$ and $s_1\big|_K$ differ by a homomorphism $\varphi : K \to A$:

$$s_1(k) = s(k)\varphi(k),$$

where we identify A with the subgroup $1 \times A < \tilde{G}_1$. Since the subgroup A is contained in the center of $\tilde{G}_1$, we obtain:

$$\varphi(gkg^{-1}) = \varphi(k)$$

for all $k \in K, g \in G_1$. In other words, the action of G_1 by conjugation on K fixes the homomorphism φ.

CHAPTER 4

Finitely generated and finitely presented groups

4.1. Finitely generated groups

A group which has a finite generating set is called *finitely generated*.

DEFINITION 4.1. The *rank* of a finitely generated group G, denoted rank (G), is the minimal number o generators of G.

REMARK 4.2. In French, the terminology for finitely generated groups is *groupe de type fini*. On the other hand, in English, being a *group of finite type* is a much stronger requirement than finite generation (typically, this means that the group has type $\mathbf{F}_\infty$).

EXERCISE 4.3. Show that every finitely generated group is countable.

EXAMPLES 4.4.
(1) The group $(\mathbb{Z}, +)$ is finitely generated by both $\{1\}$ and $\{-1\}$. Also, any set $\{p, q\}$ of coprime integers generates $\mathbb{Z}$.
(2) The group $(\mathbb{Q}, +)$ is not finitely generated.

EXERCISE 4.5. Prove that the transposition (12) and the cycle $(12 \ldots n)$ generate the permutation group S_n.

REMARKS 4.6.
(1) Every quotient $\bar{G}$ of a finitely generated group G is finitely generated; we can take as generators of $\bar{G}$ the images of the generators of G.
(2) If N is a normal subgroup of G, and both N and G/N are finitely generated, then G is finitely generated. Indeed, take a finite generating set $\{n_1, \ldots, n_k\}$ for N, and a finite generating set $\{g_1 N, \ldots, g_m N\}$ for G/N. Then
$$\{g_i, n_j \ : \ 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant k\}$$
is a finite generating set for G.

We will see in examples below that if N is a normal subgroup in a group G and G is finitely generated, it *does not* necessarily follow that N is finitely generated (not even if G is a semidirect product of N and G/N).

EXAMPLE 4.7. Let G be the wreath product $\mathbb{Z} \wr \mathbb{Z} \cong N \ltimes \mathbb{Z}$, where N is the (countably) infinite direct sum of copies of $\mathbb{Z}$. Then G is 2-generated (take a generator of the quotient group $\mathbb{Z}$ and a generator of one of the direct summands of N). On the other hand, the subgroup N is not finitely generated.

EXAMPLE 4.8. Let H be the group of permutations of $\mathbb{Z}$ generated by the transposition $t = (01)$ and the translation map $s(i) = i + 1$. Let H_i be the group of permutations of $\mathbb{Z}$ supported on $[-i,i] = \{-i, -i+1, \dots, 0, 1, \dots, i-1, i\}$, and let H_ω be the group of finitely supported permutations of $\mathbb{Z}$ (i.e., the group of bijections $f : \mathbb{Z} \to \mathbb{Z}$ such that f is the identity outside a finite subset of $\mathbb{Z}$),

$$H_\omega = \bigcup_{i=0}^{\infty} H_i \, .$$

Then H_ω is a normal subgroup in H and $H/H_\omega \simeq \mathbb{Z}$, while H_ω is not finitely generated.

Indeed from the relation $s^k t s^{-k} = (k\, k+1)\, ,\, k \in \mathbb{Z}\, ,$ it immediately follows that H_ω is a subgroup in H. It is, likewise, easy to see that $s^k H_i s^{-k} \subset H_{i+k}$, whence $s^k H_\omega s^{-k} \subset H_\omega$ for every $k \in \mathbb{Z}$.

If $g_1, \dots, g_k$ is a finite set generating H_ω, then there exists an $i \in \mathbb{N}$ so that all g_j's are in H_i, hence $H_\omega = H_i$. On the other hand, clearly, H_i is a proper subgroup of H_ω.

Exercise 4.9. 1. Let F be a non-abelian free group (see Definition 4.20). Let $\varphi : F \to \mathbb{Z}$ be any non-trivial homomorphism. Prove that the kernel of φ is not finitely generated.

2. Let F be a free group of finite rank with free generators $x_1, \dots, x_n$; set $G := F \times F$. Then G has the generating set

$$\{(x_i, 1), (1, x_j) : 1 \le i, j \le n\}.$$

Define homomorphism $\phi : G \to \mathbb{Z}$ sending every generator of G to $1 \in \mathbb{Z}$. Show that the kernel K of ϕ is finitely generated. Hint: Use the elements (x_i, x_j^{-1}), $(x_i x_j^{-1}, 1)$, $(1, x_i x_j^{-1})$, $1 \le i, j \le n$, of the subgroup K.

We will see later that a *finite index* subgroup of a finitely generated group is always finitely generated (Lemma 4.86 or Theorem 5.35). The next lemma shows that extensions of finitely generated groups are again finitely generated:

Lemma 4.10. *Suppose that we have a short exact sequence of groups*

$$1 \to G_1 \xrightarrow{i} G_2 \xrightarrow{\pi} G_3 \to 1$$

such that the groups G_1, G_3 are finitely generated. Then G_2 is also finitely generated.

Proof. Let S_1, S_3 be finite generating sets of G_1, G_3. For each $\bar{s} \in S_3$ pick $s \in \pi^{-1} S_3$. We claim that

$$S_2 := i(S_1) \cup \{s | \bar{s} \in S_3\}$$

is a generating set of G_2. Indeed, each $g \in G_2$ projects to $\pi(g)$, which is a product

$$\bar{s}_1^{\pm 1} \cdots \bar{s}_k^{\pm 1}, \quad s_i \in S_3.$$

Therefore, by normality of $i(G_1)$ in G_2, the element g itself has the form

$$h \cdot s_1^{\pm 1} \cdots s_k^{\pm 1}, \quad h \in i(G_1).$$

Since h is a product of the elements $s \in S_1$ (and their inverses), the claim follows. □

A similar proof applies to *wreath products*. Recall that the *wreath product* $A \wr C$ of groups A and C is the semidirect product

$$(\oplus_C A) \rtimes C$$

where C acts on the direct sum by precompositions: $f(x) \mapsto f(xc^{-1})$. Thus, elements of wreath products $A \wr C$ are pairs (f, c), where $f : C \to A$ is a function with finite support and $c \in C$. The product structure on this set is given by the formula

$$(f_1(x), c_1) \cdot (f_2(x), c_2)) = (f_1(xc_2^{-1})f_2(x), c_1c_2).$$

Here and below we use multiplicative notation when dealing with wreath products. For each $q \in A$ we define the function $\delta_a : C \to A$ is the function which sends $1 \in C$ to $a \in A$ and sends all other elements of C to $1 \in A$.

LEMMA 4.11. *If $a_i, i \in I, c_j, j \in J$ are generators of A and C, respectively, the elements $(1, c_j), j \in J$ and $(\delta_{a_i}, 1), i \in I$, generate $G_A := A \wr C$. In particular, if A and C are finitely generated, so is $A \wr C$.*

PROOF. It is enough to show that each $(f, 1) \in G_A$ is a product of the elements $(1, c_j), (\delta_{a_i}, 1)$. Since the maps δ_a generate

$$\oplus_C A,$$

it suffices to prove this claim for each δ_a. If

$$a = a_{i_1} \dots a_{i_k},$$

then, clearly,

$$(\delta_a, 1) = (\delta_{a_{i_1}}, 1) \dots (\delta_{a_{i_k}}, 1).$$

Lemma follows. □

Below we describe a finite generating set for the group $GL(n, \mathbb{Z})$. In the proof we use the *elementary matrices* $N_{i,j} = I_n + E_{i,j}$ $(i \neq j)$; here I_n is the identity $n \times n$ matrix and the matrix $E_{i,j}$ has a unique non-zero entry 1 in the intersection of the i–th row and the j–th column.

PROPOSITION 4.12. *The group $GL(n, \mathbb{Z})$ is generated by*

$$s_1 = \begin{pmatrix} 0 & \dots & & \dots & 0 & 1 \\ 1 & \ddots & & & \vdots & 0 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0 \end{pmatrix}, s_2 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix},$$

$$s_3 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, s_4 = \begin{pmatrix} -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

PROOF. *Step 1.* The permutation group S_n acts (effectively) on $\mathbb{Z}^n$ by permuting the basis vectors; we, thus, obtain a monomorphism $\varphi : S_n \to GL(n, \mathbb{Z})$, so that $\varphi(12 \dots n) = s_1$, $\varphi(12) = s_2$. Consider now the corresponding action of S_n on $n \times n$ matrices. Multiplication of a matrix by s_1 on the left permutes rows cyclically, multiplication to the right does the same with columns. Multiplication by s_2 on the left swaps the first two rows, multiplication to the right does the same with columns. Therefore, by multiplying an elementary matrix A by appropriate products of s_1, s_1^{-1} and s_2 on the left and on the right, we obtain the matrix s_3. In view of Exercise 4.5, the permutation $(12\dots n)$ and the transposition (12) generate the permutation group S_n. Thus, every elementary matrix N_{ij} is a product of s_1, s_1^{-1}, s_2 and s_3.

Let d_j denote the diagonal matrix with the diagonal entries $(1,\dots,1,-1,1,\dots 1)$, where -1 occurs in j-th place. Thus, $d_1 = s_4$. The same argument as above, shows that for every d_j and $s = (1j) \in S_n$, $sd_js = d_1$. Thus, all diagonal matrices d_j belong to the subgroup generated by s_1, s_2 and s_4.

Step 2. Now, let g be an arbitrary element in $GL(n,\mathbb{Z})$. Let $a_1,\dots,a_n$ be the entries of the first column of g. We will prove that there exists an element p in $\langle s_1, s_2, s_3, s_4\rangle \subset GL(n,\mathbb{Z})$, such that pg has the entries $1, 0, \dots, 0$ in its first column. We argue by induction on $k = C_1(g) = |a_1| + \cdots + |a_n|$. Note that $k \geqslant 1$. If $k = 1$, then $(a_1,\dots,a_n)$ is a permutation of $(\pm 1, 0, \dots, 0)$; hence, it suffices to take p in $\langle s_1\,,\, s_2, s_4\rangle$ permuting the rows so as to obtain $1, 0, \dots, 0$ in the first column.

Assume that the statement is true for all integers $1 \leqslant i < k$; we will prove it for k. After to permuting rows and multiplying by $d_1 = s_4$ and d_2, we may assume that $a_1 > a_2 > 0$. Then $N_{1,2}d_2g$ has the following entries in the first column: $a_1 - a_2, -a_2, a_3, \dots a_n$. Therefore, $C_1\left(N_{1,2}d_2g\right) < C_1(g)$. By the induction assumption, there exists an element p of $\langle s_1, s_2, s_3, s_4\rangle$ such that $pN_{1,2}d_2g$ has the entries of its first column equal to $1, 0, \dots, 0$. This proves the claim.

Step 3. We leave it to the reader to check that for every pair of matrices $A, B \in GL(n-1,\mathbb{R})$ and row vectors $L = (l_1,\dots,l_{n-1})$ and $M = (m_1,\dots,m_{n-1})$

$$\begin{pmatrix} 1 & L \\ 0 & A \end{pmatrix} \cdot \begin{pmatrix} 1 & M \\ 0 & B \end{pmatrix} = \begin{pmatrix} 1 & M + LB \\ 0 & AB \end{pmatrix}.$$

Therefore, the set of matrices

$$\left\{ \begin{pmatrix} 1 & L \\ 0 & A \end{pmatrix} \;;\; A \in GL(n-1,\mathbb{Z})\,,\, L \in \mathbb{Z}^{n-1} \right\}$$

is a subgroup of $GL(n,\mathbb{Z})$ isomorphic to $\mathbb{Z}^{n-1} \rtimes GL(n-1,\mathbb{Z})$.

Using this, an induction on n and Step 2, one shows that there exists an element p in $\langle s_1, s_2, s_3, s_4\rangle$ such that pg is upper triangular and with entries on the diagonal equal to 1. It, therefore, suffices to prove that every integer upper triangular matrix as above is in $\langle s_1, s_2, s_3, s_4\rangle$. This can be done for instance by repeating the argument in Step 2 with multiplications on the right. □

The wreath product (see Definition 3.31) is a useful construction of a finitely generated group from two finitely generated groups:

Exercise 4.13. Let G and H be groups, and S and X be their respective generating sets. Prove that $G \wr H$ is generated by

$$\{(f_s, 1_H) \mid s \in S\} \cup \{(f_1, x) \mid x \in X\},$$

where $f_s : H \to G$ is defined by

$$f_s(1_H) = s\,,\; f_s(h) = 1_G\,,\; \forall h \neq 1_H.$$

In particular, if G and H are finitely generated then so is $G \wr H$.

Exercise 4.14. Let G be a finitely generated group and let S be an infinite set of generators of G. Show that there exists a finite subset F of S so that G is generated by F.

Exercise 4.15. An element g of the group G is a *non-generator* if for every generating set S of G, the complement $S \setminus \{g\}$ is still a generating set of G.

(a) Prove that the set of non-generators forms a subgroup of G. This subgroup is called the *Frattini subgroup*.
(b) Compute the Frattini subgroup of $(\mathbb{Z}, +)$.
(c) Compute the Frattini subgroup of $(\mathbb{Z}^n, +)$. (*Hint:* You may use the fact that $Aut(\mathbb{Z}^n)$ is $GL(n, \mathbb{Z})$, and that the $GL(n, \mathbb{Z})$–orbit of e_1 is the set of vectors $(k_1, \dots, k_n)$ in $\mathbb{Z}^n$ such that $\gcd(k_1, \dots, k_n) = 1$.)

Definition 4.16. *A group G is said to have* bounded generation property *(or is* boundedly generated*) if there exists a finite subset $\{t_1, \dots, t_m\} \subset G$ such that every $g \in G$ can be written as*

$$g = t_1^{k_1} t_2^{k_2} \cdots t_m^{k_m} ,$$

where $k_1, k_2, \dots, k_m$ are integers.

Clearly, all finitely generated abelian groups have the bounded generation property, and so are all finite groups. On the other hand, the nonabelian free groups, which we will introduce in the next section, obviously, do not have the bounded generation property. For other examples of boundedly generated groups see Proposition 11.73. We also note that Alexey Muranov [**Mur05**] constructed examples of infinite boundedly generated *simple* groups.

4.2. Free groups

Let X be a set. Its elements are called *letters* or *symbols*. We define the set of *inverse letters* (or *inverse symbols*) $X^{-1} = \{a^{-1} \mid a \in X\}$. We will think of $X \cup X^{-1}$ as an *alphabet*.

A *word* in $X \cup X^{-1}$ is a finite (possibly empty) string of letters in $X \cup X^{-1}$, i.e., an expression of the form

$$a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_k}^{\epsilon_k},$$

where $a_i \in X, \epsilon_i = \pm 1$; here $x^1 = x$ for every $x \in X$. We will use the notation 1 for the *empty word* (the one which has no letters).

Convention 4.17. Sometimes, by abusing the terminology, we will refer to words in $X \cup X^{-1}$ as *words in X*.

Denote by X^* the set of words in the alphabet $X \cup X^{-1}$, where the empty word, denoted by 1, is included. For instance,

$$a_1 a_2 a_1^{-1} a_2 a_2 a_1 \in X^*.$$

The *length* of a word w is the number of letters in this word. The length of the empty word is 0.

A word $w \in X^*$ is *reduced* if it contains no pair of consecutive letters of the form aa^{-1} or $a^{-1}a$. The *reduction* of a word $w \in X^*$ is the deletion of all pairs of consecutive letters of the form aa^{-1} or $a^{-1}a$.

For instance, the words

$$1, a_2a_1, a_1a_2a_1^{-1}$$

are reduced, while

$$a_2a_1a_1^{-1}a_3$$

is not reduced.

More generally, a word w is *cyclically reduced* if it is reduced and, in addition, the first and the last letters of w are not inverses of each other. Equivalently, conjugating w by an element of $X \cup X^{-1}$:

$$w' = awa^{-1}, \quad a \in X \cup X^{-1}$$

results in a word w' whose reduction has length $\gg$ the length of w.

We define an equivalence relation on X^* by $w \sim w'$ if w can be obtained from w' by a finite sequence of reductions and their inverses, i.e., the relation $\sim$ on X^* is generated by

$$ua_ia_i^{-1}v \sim uv, \quad ua_i^{-1}a_iv \sim uv$$

where $u, v \in X^*$.

Proposition 4.18. *Any word $w \in X^*$ is equivalent to a unique reduced word.*

Proof. *Existence.* We prove the statement by induction on the length of a word. For words of length 0 and 1 the statement is clearly true. Assume that it is true for words of length n and consider a word of length $n+1$, $w = a_1 \cdots a_n a_{n+1}$, where $a_i \in X \cup X^{-1}$. According to the induction hypothesis, there exists a reduced word $u = b_1 \cdots b_k$ with $b_j \in X \cup X^{-1}$ such that $a_2 \cdots a_{n+1} \sim u$. Then $w \sim a_1u$. If $a_1 \neq b_1^{-1}$ then a_1u is reduced. If $a_1 = b_1^{-1}$ then $a_1u \sim b_2 \cdots b_k$ and the latter word is reduced.

Uniqueness. Let $F(X)$ be the set of reduced words in $X \cup X^{-1}$. For every $a \in X \cup X^{-1}$ we define a map $L_a : F(X) \to F(X)$ by

$$L_a(b_1 \cdots b_k) = \begin{cases} ab_1 \cdots b_k & \text{if} \quad a \neq b_1^{-1}, \\ b_2 \cdots b_k & \text{if} \quad a = b_1^{-1}. \end{cases}$$

For every word $w = a_1 \cdots a_n$ define $L_w = L_{a_1} \circ \cdots \circ L_{a_n}$. For the empty word 1 define $L_1 = \mathrm{id}$. It is easy to check that $L_a \circ L_{a^{-1}} = \mathrm{id}$ for every $a \in X \cup X^{-1}$, and to deduce from it that $v \sim w$ implies $L_v = L_w$.

We prove by induction on the length that if w is reduced then $w = L_w(1)$. The statement clearly holds for w of length 0 and 1. Assume that it is true for reduced words of length n and let w be a reduced word of length $n+1$. Then $w = au$, where $a \in X \cup X^{-1}$ and u is a reduced word that does not begin with a^{-1}, i.e., such that $L_a(u) = au$. Then $L_w(1) = L_a \circ L_u(1) = L_a(u) = au = w$.

In order to prove uniqueness it suffices to prove that if $v \sim w$ and v, w are reduced then $v = w$. Since $v \sim w$ it follows that $L_v = L_w$, hence $L_v(1) = L_w(1)$, that is $v = w$. □

Exercise 4.19. Give a geometric proof of this proposition using identification of $w \in X^*$ with the set of edge-paths $\mathfrak{p}_w$ in a regular tree T of valence $2|X|$, which start at a fixed vertex v_0. The reduced path $\mathfrak{p}^*$ in T corresponding to the reduction w^* of w is the unique geodesic in T connecting v_0 to the terminal point of $\mathfrak{p}$. Uniqueness of w^* then translates to the fact that a tree contains no circuits.

Let $F(X)$ be the set of reduced words in $X \cup X^{-1}$. Proposition 4.18 implies that $X^*/\sim$ can be identified with $F(X)$.

DEFINITION 4.20. The *free group over* X is the set $F(X)$ endowed with the product $*$ defined by: $w * w'$ is the unique reduced word equivalent to the word ww'. The unit is the empty word.

The cardinality of X is called the *rank* of the free group $F(X)$.

We note that, at the moment, we have two, a priori distinct, notions of rank for (finitely generated) free groups: One is the least number of generators and the second is the cardinality of the set X. We will see, however, that the two numbers are the same.

The set $F = F(X)$ with the product defined in Definition 4.20 is indeed a group. The inverse of a reduced word

$$w = a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_k}^{\epsilon_k}$$

is given by

$$w^{-1} = a_{i_k}^{-\epsilon_k} a_{i_{k-1}}^{-\epsilon_{k-1}} \cdots a_{i_1}^{-\epsilon_1}.$$

It is clear that the product ww^{-1} projects to the empty word 1 in F.

EXERCISE 4.21. A free group of rank at least 2 is not abelian. Thus, *free non-abelian* means 'free of rank at least 2.'

The *free semigroup* $F^s(X)$ *with the generating set* X is defined in the fashion similar to $F(X)$, except that we only allow the words in the alphabet X (and not in X^{-1}), in particular the reduction is not needed.

PROPOSITION 4.22 (Universal property of free groups). *A map* $\varphi : X \to G$ *from the set* X *to a group* G *can be extended to a homomorphism* $\Phi : F(X) \to G$ *and this extension is unique.*

PROOF. *Existence.* The map φ can be extended to a map on $X \cup X^{-1}$ (which we denote also φ) by $\varphi(a^{-1}) = \varphi(a)^{-1}$.

For every reduced word $w = a_1 \cdots a_n$ in $F = F(X)$ define

$$\Phi(a_1 \cdots a_n) = \varphi(a_1) \cdots \varphi(a_n).$$

Set $\Phi(1_F) := 1_G$, the identity element of G. We leave it to the reader to check that Φ is a homomorphism.

Uniqueness. Let $\Psi : F(X) \to G$ be a homomorphism such that $\Psi(x) = \varphi(x)$ for every $x \in X$. Then for every reduced word $w = a_1 \cdots a_n$ in $F(X)$,

$$\Psi(w) = \Psi(a_1) \cdots \Psi(a_n) = \varphi(a_1) \cdots \varphi(a_n) = \Phi(w).$$

□

COROLLARY 4.23. *Every group is the quotient of a free group.*

PROOF. Apply Proposition 4.22 to the group G and a generating set X of G (e.g., $X = G$). □

LEMMA 4.24. *Every short exact sequence* $1 \to N \to G \overset{r}{\to} F(X) \to 1$ *splits. In particular,* G *contains a subgroup isomorphic to* $F(X)$.

Proof. Indeed, for each $x \in X$ consider choose an element $t_x \in G$ projecting to x; the map $x \mapsto t_x$ extends to a group homomorphism $s : F(X) \to G$. Composition $r \circ s$ is the identity homomorphism $F(X) \to F(X)$ (since it is the identity on the generating set X). Therefore, the homomorphism s is a splitting of the exact sequence. Since $r \circ s = Id$, s is a monomorphism. □

Corollary 4.25. *Every short exact sequence* $1 \to N \to G \to \mathbb{Z} \to 1$ *splits.*

4.3. Presentations of groups

Let G be a group and S a generating set of G. According to Proposition 4.22, the inclusion map $i : S \to G$ extends uniquely to an epimorphism $\pi_S : F(S) \to G$. The elements of $\mathrm{Ker}(\pi_S)$ are called *relators* (or *relations*) of the group G with the generating set S.

N.B. In the above, by an abuse of language we used the symbol s to designate two different objects: s is a letter in $F(S)$, as well as an element in the group G.

If $R = \{r_i \mid i \in I\} \subset F(S)$ is such that $\mathrm{Ker}(\pi_S)$ is normally generated by R (i.e., $\langle\langle R \rangle\rangle = \mathrm{Ker}(\pi_S)$) then we say that the ordered pair (S, R), usually denoted $\langle S|R\rangle$, is *a presentation of* G. The elements $r \in R$ are called *defining relators* (or *defining relations*) of the presentation $\langle S|R\rangle$.

A group G is said to be *finitely presented* if it admits a finite presentation, i.e., a presentation with finitely many generators and relators.

By abuse of language we also say that the generators $s \in S$ and the *relations* $r = 1$, $r \in R$, constitute a presentation of the group G. Sometimes we will write presentations in the form

$$\langle s_i, i \in I | r_j = 1, j \in J\rangle$$

where

$$S = \{x_i\}_{i \in I}, \quad R = \{r_j\}_{j \in J}.$$

If both S and R are finite, then the pair S, R is called *a finite presentation of* G. A group G is called *finitely presented* if it admits a finite presentation. Sometimes it is difficult, and even algorithmically impossible, to find a finite presentation of a finitely presented group, see [**BW11**].

Conversely, given an alphabet S and a set R of (reduced) words in the alphabet S, we can form the quotient

$$G := F(S)/\langle\langle R \rangle\rangle .$$

Then $\langle S|R\rangle$ is a presentation of G. By abusing notation, we will often write

$$G = \langle S|R\rangle ,$$

if G is a group with the presentation $\langle S|R\rangle$. If w is a word in the generating set S, we will use $[w]$ to denote its projection to the group G. An alternative notation for the equality

$$[v] = [w]$$

is

$$v \equiv_G w.$$

Note that the significance of a presentation of a group is the following:

- every element in G can be written as a finite product $x_1 \cdots x_n$ with
$$x_i \in S \cup S^{-1} = \{s^{\pm 1} : s \in S\}$$
i.e., as a word in the alphabet $S \cup S^{-1}$;
- a word $w = x_1 \cdots x_n$ in the alphabet $S \cup S^{-1}$ is *equal to the identity in* G, $w \equiv_G 1$, *if and only if* in $F(S)$ the word w is the product of finitely many conjugates of the words $r_i \in R$, i.e.,
$$w = \prod_{i=1}^{m} r_i^{u_i}$$
for some $m \in \mathbb{N}$, $u_i \in F(S)$ and $r_i \in R$.

Below are few examples of group presentations:

EXAMPLES 4.26. (1) $\langle a_1, \dots, a_n \mid [a_i, a_j], 1 \leqslant i, j \leqslant n \rangle$ is a finite presentation of $\mathbb{Z}^n$;

(2) $\langle x, y \mid x^n, y^2, yxyx \rangle$ is a presentation of the finite dihedral group D_{2n} ;

(3) $\langle x, y \mid x^2, y^3, [x, y] \rangle$ is a presentation of the cyclic group $\mathbb{Z}_6$.

Let $\langle S|R \rangle$ be a presentation of a group G. Let H be a group and $\psi : X \to H$ be a map which "preserves the relators", i.e., $\psi(r) = 1$ for every $r \in R$. Then:

LEMMA 4.27. *The map ψ extends to a group homomorphism $\psi : G \to H$.*

PROOF. By the universal property of free groups, the map ψ extends to a homomorphism $\tilde{\psi} : F(X) \to H$. We need to show that $\langle\langle R \rangle\rangle$ is contained in $\mathrm{Ker}(\tilde{\psi})$. However, $\langle\langle R \rangle\rangle$ consists of products of elements of the form grg^{-1}, where $g \in F, r \in R$. Since $\tilde{\psi}(grg^{-1}) = 1$, the claim follows. □

EXERCISE 4.28. The group $\bigoplus_{x \in X} \mathbb{Z}_2$ has presentation
$$\langle x \in X | x^2, [x, y], \forall x, y \in X \rangle .$$

PROPOSITION 4.29 (Finite presentability is independent of the generating set). *Assume that a group G has finite presentation $\langle S \mid R \rangle$, and let $\langle X \mid T \rangle$ be an arbitrary presentation of G, such that X is finite. Then there exists a finite subset $T_0 \subset T$ such that $\langle X \mid T_0 \rangle$ is a presentation of G.*

PROOF. Every element $s \in S$ can be written as a word $a_s(X)$ in X. The map $i_{SX} : S \to F(X)$, $i_{SX}(s) = a_s(X)$ extends to a unique homomorphism $p : F(S) \to F(X)$. Moreover, since $\pi_X \circ i_{SX}$ is an inclusion map of S into $F(S)$, and both π_S and $\pi_X \circ p$ are homomorphisms from $F(S)$ to G extending the map $S \to G$, by the uniqueness of the extension we have that
$$\pi_S = \pi_X \circ p.$$
This implies that $\mathrm{Ker}(\pi_X)$ contains $p(r)$ for every $r \in R$.

Likewise, every $x \in X$ can be written as a word $b_x(S)$ in S, and this defines a map $i_{XS} : X \to F(S)$, $i_{XS}(x) = b_x(S)$, which extends to a homomorphism $q : F(X) \to F(S)$. A similar argument shows that $\pi_S \circ q = \pi_X$.

For every $x \in X$,
$$\pi_X(p(q(x))) = \pi_S(q(x)) = \pi_X(x).$$

This implies that for every $x \in X$, $x^{-1}p(q(x))$ is in $\operatorname{Ker}(\pi_X)$. Let N be the normal subgroup of $F(X)$ normally generated by

$$\{p(r) \mid r \in R\} \cup \{x^{-1}p(q(x)) \mid x \in X\}\,.$$

We have that $N \leqslant \operatorname{Ker}(\pi_X)$. Therefore, there is a natural projection

$$\mathrm{proj} : F(X)/N \to F(X)/\operatorname{Ker}(\pi_X).$$

Let $\bar{p} : F(S) \to F(X)/N$ be the homomorphism induced by p. Since $\bar{p}(r) = 1$ for all $r \in R$, it follows that $\bar{p}(\operatorname{Ker}\pi_S) = 1$, hence, $\bar{p}$ induces a homomorphism

$$\varphi : F(S)/\operatorname{Ker}(\pi_S) \to F(X)/N.$$

We next observe that the homomorphism φ is onto. Indeed, $F(X)/N$ is generated by elements of the form $xN = p(q(x))N$, and the latter is the image under φ of $q(x)\operatorname{Ker}(\pi_S)$.

Consider the homomorphism

$$\mathrm{proj} \circ \varphi : F(S)/\operatorname{Ker}(\pi_S) \to F(X)/\operatorname{Ker}(\pi_X)$$

Both the domain and the target groups are isomorphic to G. Each element x of the generating set X is sent by the isomorphism $G \to F(S)/\operatorname{Ker}(\pi_S)$ to $q(x)\operatorname{Ker}(\pi_S)$. The same element x is sent by the isomorphism $G \to F(X)/\operatorname{Ker}(\pi_X)$ to $x\operatorname{Ker}(\pi_X)$. Note that

$$\mathrm{proj} \circ \varphi\,(q(x)\operatorname{Ker}(\pi_S)) = \mathrm{proj}(xN) = x\operatorname{Ker}(\pi_X).$$

This means that, modulo the two isomorphisms mentioned above, the map $\mathrm{proj} \circ \varphi$ is id_G. This implies that φ is injective, hence, a bijection. Therefore, proj is also a bijection. This happens if and only if $N = \operatorname{Ker}(\pi_X)$. In particular, $\operatorname{Ker}(\pi_X)$ is normally generated by the finite set of relators

$$\Re = \{p(r) \mid r \in R\} \cup \{x^{-1}p(q(x)) \mid x \in X\}.$$

Since $\Re = \langle\langle T \rangle\rangle$, every relator $\rho \in \Re$ can be written as a product

$$\prod_{i \in I_\rho} t_i^{v_i}$$

with $v_i \in F(X), t_i \in T$ and I_ρ finite. It follows that $\operatorname{Ker}(\pi_X)$ is normally generated by the finite subset

$$T_0 = \bigcup_{\rho \in \Re} \{t_i \mid i \in I_\rho\}$$

of T. □

Proposition 4.29 can be reformulated as follows: If G is finitely presented, X is finite and

$$1 \to N \to F(X) \to G \to 1$$

is a short exact sequence, then N is normally generated by finitely many elements $n_1, \ldots, n_k$. This can be generalized to an arbitrary short exact sequence:

Lemma 4.30. *Consider a short exact sequence*

$$1 \to N \to K \xrightarrow{\pi} G \to 1\,, \quad \textit{with } K \textit{ finitely generated.} \tag{4.1}$$

If G is finitely presented, then N is normally generated by finitely many elements $n_1, \ldots, n_k \in N$.

Proof. Let S be a finite generating set of K; then $\overline{S} = \pi(S)$ is a finite generating set of G. Since G is finitely presented, by Proposition 4.29 there exist finitely many words $r_1, \ldots, r_k$ in S such that

$$\left\langle \overline{S} \mid r_1(\overline{S}), \ldots, r_k(\overline{S}) \right\rangle$$

is a presentation of G.

Define $n_j = r_j(S)$, an element of N by the assumption.

Let n be an arbitrary element in N and $w(S)$ a word in S such that $n = w(S)$ in K. Then $w(\overline{S}) = \pi(n) = 1$, whence in $F(S)$ the word $w(S)$ is a product of finitely many conjugates of $r_1, \ldots, r_k$. When projecting such a relation *via* $F(S) \to K$ we obtain that n is a product of finitely many conjugates of $n_1, \ldots, n_k$. □

Proposition 4.31. *Suppose that N a normal subgroup of a group G. If both N and G/N are finitely presented then G is also finitely presented.*

Proof. Let X be a finite generating set of N and let Y be a finite subset of G such that $\overline{Y} = \{yN \mid y \in Y\}$ is a generating set of G/N. Let $\langle X \mid r_1, \ldots, r_k \rangle$ be a finite presentation of N and let $\left\langle \overline{Y} \mid \rho_1, \ldots, \rho_m \right\rangle$ be a finite presentation of G/N. The group G is generated by $S = X \cup Y$ and this set of generators satisfies a list of relations of the following form:

$$r_i(X) = 1\,,\ 1 \leqslant i \leqslant k\,,\ \rho_j(Y) = u_j(X)\,,\ 1 \leqslant j \leqslant m\,, \tag{4.2}$$

$$x^y = v_{xy}(X)\,,\ x^{y^{-1}} = w_{xy}(X) \tag{4.3}$$

for some words u_j, v_{xy}, w_{xy} in S.

We claim that this is a complete set of defining relators of G.

All the relations above can be rewritten as $t(X, Y) = 1$ for a finite set T of words t in S. Let K be the normal subgroup of $F(S)$ normally generated by T.

The epimorphism $\pi_S : F(S) \to G$ defines an epimorphism $\varphi : F(S)/K \to G$. Let wK be an element in $\mathrm{Ker}(\varphi)$, where w is a word in S. Due to the set of relations (4.3), there exist a word $w_1(X)$ in X and a word $w_2(Y)$ in Y, such that $wK = w_1(X)w_2(Y)K$.

Applying the projection $\pi : G \to G/N$, we see that $\pi(\varphi(wK)) = 1$, i.e., $\pi(\varphi(w_2(Y)K)) = 1$. This implies that $w_2(Y)$ is a product of finitely many conjugates of $\rho_i(Y)$, hence $w_2(Y)K$ is a product of finitely many conjugates of $u_j(X)K$, by the second set of relations in (4.2). This and the relations (4.3) imply that $w_1(X)w_2(Y)K = v(X)K$ for some word $v(X)$ in X. Then the image $\varphi(wK) = \varphi(v(X)K)$ is in N; therefore, $v(X)$ is a product of finitely many conjugates of relators $r_i(X)$. This implies that $v(X)K = K$.

We have thus obtained that $\mathrm{Ker}(\varphi)$ is trivial, hence φ is an isomorphism, equivalently that $K = \mathrm{Ker}(\pi_S)$. This implies that $\mathrm{Ker}(\pi_S)$ is normally generated by the finite set of relators listed in (4.2) and (4.3). □

We continue with a list of finite presentations of some important groups:

Examples 4.32. (1) *Surface groups*:

$$\Pi_n = \langle a_1, b_1, \ldots, a_n, b_n | [a_1, b_1] \cdots [a_n, b_n] \rangle\,,$$

is the fundamental group of the closed connected oriented surface of genus n, see e.g. **[Hat02, Mas91]**.

(2) *Right–angled Artin groups (RAAGs).* Let $\mathcal{G}$ be a finite graph with the vertex set $V = \{x_1, \dots, x_n\}$ and the edge set E consisting of the edges $\{[x_i, x_j]\}_{i,j}$. Define the *right–angled Artin group* by

$$A_{\mathcal{G}} := \langle V | [x_i, x_j], \text{whenever } [x_i, x_j] \in E \rangle .$$

Here we commit a useful abuse of notation: In the first instance $[x_i, x_j]$ denotes the commutator and in the second instance it denotes the edge of $\mathcal{G}$ connecting x_i to x_j.

EXERCISE 4.33. a. If $\mathcal{G}$ contains no edges then $A_{\mathcal{G}}$ is a free group on n generators.

b. If $\mathcal{G}$ is the complete graph on n vertices then

$$A_{\mathcal{G}} \cong \mathbb{Z}^n.$$

(3) *Coxeter groups.* Let $\mathcal{G}$ be a finite simple graph. Let V and E denote be the vertex and the edge set of $\mathcal{G}$ respectively. Put a label $m(e) \in \mathbb{N} \setminus \{1\}$ on each edge $e = [x_i, x_j]$ of $\mathcal{G}$. Call the pair

$$\Gamma := (\mathcal{G}, m : E \to \mathbb{N} \setminus \{1\})$$

a *Coxeter graph.* Then Γ defines the *Coxeter group* C_Γ:

$$C_\Gamma := \left\langle x_i \in V | x_i^2, (x_i x_j)^{m(e)}, \text{ whenever there exists an edge } e = [x_i, x_j] \right\rangle .$$

See **[Dav08]** for the detailed discussion of Coxeter groups.

(4) *Artin groups.* Let Γ be a Coxeter graph. Define

$$A_\Gamma := \left\langle x_i \in V | \underbrace{x_i x_j \cdots}_{m(e) \text{ terms}} = \left(\underbrace{x_j x_i \cdots}_{m(e) \text{ terms}} \right), \text{ whenever } e = [x_i, x_j] \in E \right\rangle .$$

Then A_Γ is a right-angled Artin group if and only if $m(e) = 2$ for every $e \in E$. In general, C_Γ is the quotient of A_Γ by the subgroup normally generated by the set

$$\{x_i^2 : x_i \in V\}.$$

(5) *Shephard groups:* Let Γ be a Coxeter graph. Label vertices of Γ with natural numbers $n_x, x \in V(\Gamma)$. Now, take a group, a Shepherd group, S_Γ to be generated by vertices $x \in V$, subject to Artin relators and, in addition, relators

$$x^{n_x}, \quad x \in V.$$

Note that, in the case $n_x = 2$ for all $x \in V$, the group which we obtain is the Coxeter group C_Γ. Shephard groups (and von Dyck groups below) are *complex analogues* of Coxeter groups.

(6) *Generalized von Dyck groups:* Let Γ be a labeled graph as in the previous example. Define a group D_Γ to be generated by vertices $x \in V$, subject to the relators

$$x^{n_x}, \quad x \in V;$$
$$(xy)^{m(e)}, e = [x, y] \in E.$$

If Γ consists of a single edge, then D_Γ is called a *von Dyck group.* Every von Dyck group D_Γ is an index 2 subgroup in the Coxeter group C_Δ, where Δ is the triangle with edge-labels p, q, r, which are the vertex-edge labels of Γ.

(7) *Integer Heisenberg group:*

$$H_{2n+1}(\mathbb{Z}) := \langle x_1, \ldots, x_n, y_1, \ldots, y_n, z \mid$$

$$[x_i, z] = 1, [y_j, z] = 1, [x_i, x_j] = 1, [y_i, y_j] = 1, [x_i, y_j] = z^{\delta_{ij}}, 1 \leqslant i, j \leqslant n \rangle .$$

(8) *Baumslag–Solitar groups:*

$$BS(m, n) = \langle a, b | ab^m a^{-1} = b^n \rangle ,$$

where m, n are nonzero integers.

EXERCISE 4.34. Show that $H_{2n+1}(\mathbb{Z})$ is isomorphic to the group appearing in Example 11.36, (3).

The classes of groups described so far were defined combinatorially, in terms of their presentations. Below are several important classes of finitely presented groups which are defined *geometrically*:

(1) $CAT(-1)$ groups: Groups G which act geometrically on $CAT(-1)$ metric spaces.
(2) $CAT(0)$ groups: Groups G which act geometrically on $CAT(0)$ metric spaces.
(3) Automatic groups: We refer the reader to [**ECH$^+$92**] for the definition.
(4) Hyperbolic and relatively hyperbolic groups, which will be defined in Chapter 9.
(5) Semihyperbolic groups, see [**JA95**].

An important feature of finitely presented groups is provided by the following theorem, see e.g. [**Hat02**]:

THEOREM 4.35. *Every finitely generated group is the fundamental group of a smooth closed manifold of dimension* 4.

Laws in groups.

DEFINITION 4.36. An *identity* (or *law*) is a non-trivial reduced word

$$w = w(x_1, \ldots, x_n)$$

in the letters $x_1, \ldots, x_n$ and their inverses. A group G is said to *satisfy the identity (law)* $w = 1$ if this equality is satisfied in G whenever $x_1, \ldots, x_n$ are replaced by arbitrary elements in G. In other words, for the group

$$Q = \langle x_1, \ldots, x_n | w \rangle ,$$

the pull-back map

$$Hom(Q, G) \longrightarrow Hom(F_n, G)$$

is surjective.

EXAMPLES 4.37 (Groups satisfying a law). (1) Abelian groups. Here the law is

$$w(x_1, x_2) = x_1 x_2 x_1^{-1} x_2^{-1} .$$

(2) Solvable groups, see section 11.6, equation (11.10).

(3) Free Burnside groups. The *free Burnside group*

$$B(n,m) = \left\langle x_1, \ldots, x_n \mid w^n \text{ for every word } w \text{ in } x_1^{\pm 1}, \ldots, x_n^{\pm 1} \right\rangle .$$

It is known that these groups are infinite for sufficiently large m (see [**Ady79**], [**Ol′91a**], [**Iva94**], [**Lys96**], [**DG**] and references therein).

Note that free nonabelian groups (and, hence, groups containing them) do not satisfy any law.

4.4. The rank of a free group determines the group. Subgroups

PROPOSITION 4.38. *Two free groups $F(X)$ and $F(Y)$ are isomorphic if and only if X and Y have the same cardinality.*

PROOF. A bijection $\varphi : X \to Y$ extends to an isomorphism $\Phi : F(X) \to F(Y)$ by Proposition 4.22. Therefore, two free groups $F(X)$ and $F(Y)$ are isomorphic if X and Y have the same cardinality.

Conversely, let $\Phi : F(X) \to F(Y)$ be an isomorphism. Take $N(X) \leqslant F(X)$, the subgroup generated by the subset $\{g^2 : g \in F(X)\}$; clearly, N is normal in $F(X)$. Then, $\Phi(N(X)) = N(Y)$ is the normal subgroup generated by $\{h^2\,; h \in F(Y)\}$. It follows that Φ induces an isomorphism $\Psi : F(X)/N(X) \to F(Y)/N(Y)$.

LEMMA 4.39. *The quotient $\bar{F} := F/N$ is isomorphic to $A = \mathbb{Z}_2^{\oplus X}$, where $F = F(X)$.*

PROOF. Recall that A has the presentation

$$\left\langle x \in X | x^2, [x,y], \forall x, y \in X \right\rangle ,$$

see Exercise 4.28. We now prove the assertion of the lemma. Consider the map $\eta : F \to A$ sending the generators of F to the obvious generators of A. Thus, $\pi(g) = \pi(g^{-1})$ for all $g \in F$. We conclude that for all $g, h \in X$,

$$1 = \pi((hg)^2) = \pi([g,h]),$$

and, therefore, $\bar{F}$ is abelian.

Since A satisfies the law $a^2 = 1$ for all $a \in A$, it is clear that $\eta = \varphi \circ \pi$, where $\pi : F \to \bar{F}$ is the quotient map. We next construct the inverse ψ to ϕ. We define ψ on the generators $x \in X$ of A: $\psi(x) = \bar{x} = \pi(x)$. We need to show that ψ preserves the relators of A (as in Lemma 4.27): Since $\bar{F}$ is abelian, $[\psi(x), \psi(y)] = 1$ for all $x, y \in X$. Moreover, $\psi(x)^2 = 1$ since $\bar{F}$ also satisfies the law $g^2 = 1$. It is clear that ϕ, ψ are inverses to each other. □

Thus, $F(X)/N(X)$ is isomorphic to $\mathbb{Z}_2^{\oplus X}$, while $F(Y)/N(Y)$ is isomorphic to $\mathbb{Z}_2^{\oplus Y}$. It follows that $\mathbb{Z}_2^{\oplus X} \cong \mathbb{Z}_2^{\oplus Y}$ as $\mathbb{Z}_2$–vector spaces. Therefore, X and Y have the same cardinality, by uniqueness of the dimension of vector spaces. □

REMARK 4.40. Proposition 4.38 implies that for every cardinal number n there exists, up to isomorphism, exactly one free group of rank n. We denote this group by F_n.

Recall that the *rank* of a finitely generated group G is the least number of generators of G. In other words,

$$\text{rank}\,(G) = \min\{r : \exists \text{ an epimorphism } F_r \to G\}.$$

Corollary 4.41. *For each finite n, the number n is the least cardinality of a generating set of F_n. In other words,* $\mathrm{rank}\,(F_n) = n$.

Proof. If this theorem fails, there exists a epimorphism

$$h : F(X) \to F(Y), \quad |X| = m < |Y| = n.$$

This epimorphism projects to an epimorphism of the abelian quotients

$$\bar{h} : A = F(X)/N(X) \to B = F(Y)/N(Y).$$

However, A and B are vector spaces over $\mathbb{Z}_2$ of dimensions m and n respectively. This contradicts the assumption that $m < n$. □

Theorem 4.42 (Nielsen–Schreier). *Any subgroup of a free group is a free group.*

This theorem will be proven in Corollary 4.80 using topological methods; see also [**LS77**, Proposition 2.11].

4.5. Free constructions: Amalgams of groups and graphs of groups

4.5.1. Amalgams. Amalgams (amalgamated free products and HNN extensions) allow one to build more complicated groups starting with a given pair of groups or a group and a pair of its subgroups which are isomorphic to each other.

Amalgamated free products. As a warm-up we first define the *free product* of groups $G_1 = \langle X_1 | R_1 \rangle, G_2 = \langle X_2 | R_2 \rangle$ by the presentation:

$$G_1 \star G_2 = \langle G_1, G_2 | \quad \rangle,$$

which is a shorthand for the presentation:

$$\langle X_1 \sqcup X_2 | R_1 \sqcup R_2 \rangle.$$

For instance, the free group of rank 2 is isomorphic to $\mathbb{Z} \star \mathbb{Z}$.

More generally, suppose that we are given subgroups $H_i \leqslant G_i$ $(i = 1, 2)$ and an isomorphism

$$\phi : H_1 \to H_2 \ .$$

Define the *amalgamated free product*

$$G_1 \star_{H_1 \cong H_2} G_2 = \left\langle G_1, G_2 | \phi(h) h^{-1}, h \in H_1 \right\rangle.$$

In other words, in addition to the relators in G_1, G_2 we identify $\phi(h)$ with h for each $h \in H_1$. A common shorthand for the amalgamated free product is

$$G_1 \star_H G_2,$$

where $H \cong H_1 \cong H_2$ (the embeddings of H into G_1 and G_2 are suppressed in this notation).

HNN extensions. This construction is named after G. Higman, B. Neumann and H. Neumann who first introduced it in [**HNN49**]. It is a variation on the amalgamated free product where $G_1 = G_2$. Namely, suppose that we are given a group G, its subgroup H and a monomorphism $\phi : H \to G$. Then the HNN extension of G *via* ϕ is defined as

$$G \star_{H,\phi} = \left\langle G, t | t h t^{-1} = \phi(h), \forall h \in H \right\rangle.$$

A common shorthand for the HNN extension is

$$G \star_H$$

where the monomorphism ϕ is suppressed in this notation.

EXERCISE 4.43. Suppose that H is the trivial subgroup. Then

$$G\star_H \cong G \star \mathbb{Z}.$$

EXERCISE 4.44. Let $G = \langle S|R\rangle$, where R is a single relator which contains each letter $x \in X$ exactly twice (possibly as x^{-1}). Show that G is isomorphic to the free product of the fundamental group of a closed surface and a free group. Give an example where the free factor is nontrivial.

More generally, one defines *simultaneous HNN extension* of G along a collection of isomorphic subgroups: Suppose that we are given a collection of subgroups $H_j, j \in J$ of G and isomorphic embeddings $\phi_j : H_j \to G$. Then define the group

$$G\star_{\phi_j:H_j\to G, j\in J} = \left\langle G, t_j, j \in J | t_j h t_j^{-1} = \phi_j(h), \forall h \in H_j, j \in J \right\rangle.$$

4.5.2. Graphs of groups. In this section, graphs are no longer assumed to be simplicial, but are assumed to connected. The notion of graphs of groups is a very useful generalization of both the amalgamated free product and the HNN extension.

Suppose that Γ is a graph. Assign to each vertex v of Γ a *vertex group* G_v; assign to each edge e of Γ an *edge group* G_e. We orient each edge e so that its head is e_+ and the tail is e_- (this allows for the possibility that $e_+ = e_-$). Suppose, furthermore, that for each edge e we are given *monomorphisms*

$$\phi_{e_+} : G_e \to G_{e_+}, \phi_{e_-} : G_e \to G_{e_-}.$$

REMARK 4.45. More generally, one can allow non-injective homomorphisms

$$G_e \to G_{e_+}, \quad G_e \to G_{e_-},$$

but we will not consider them here, see [**Mas91**].

The oriented graph Γ together with the collection of vertex and edge groups and the monomorphisms $\phi_{e_\pm}$ is called a *graph of groups* $\mathcal{G}$ based on the graph Γ.

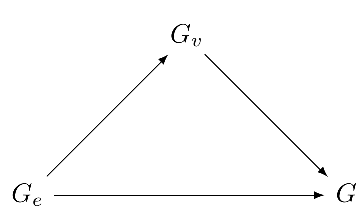
Our next goal is to convert (connected) graphs of groups $\mathcal{G}$ into groups. We first do this in the case when Γ is simply-connected, i.e., is a tree.

DEFINITION 4.46. Suppose that Γ is a tree. The *fundamental group* $\pi(\mathcal{G}) = \pi_1(\mathcal{G})$ of a graph of groups based on a tree Γ is a group G satisfying the following:

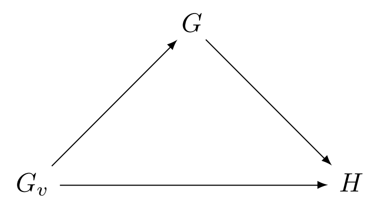
1. There is a collection of *compatible homomorphisms*

$$G_v \to G, G_e \to G, v \in V(\Gamma), e \in E(\Gamma),$$

i.e., that whenever $v = e_\pm$, we have the commutative diagram



2. The group G is *universal* with respect to the above property, i.e., given any group H and a collection of compatible homomorphisms $G_v \to H, G_e \to H$, there exists a unique homomorphism $G \to H$ such that we have commutative diagrams



for all $v \in V(\Gamma)$.

Note that the above definition easily implies that $G = \pi(\mathcal{G})$ is unique (up to an isomorphism). For the existence of $\pi(\mathcal{G})$ see [**Ser80**] and the discussion below. It is also a nontrivial (but not a very difficult to prove) fact that the homomorphisms $G_v \to G$ are injective.

Suppose now that Γ is connected but not simply-connected. We then let $T \subset \Gamma$ be a maximal subtree and $\mathcal{T} \subset \mathcal{G}$ be the corresponding subgraph of groups. Set $G_T := \pi(\mathcal{T})$. For each edge $e = [v, w] \in E(\Gamma)$ which is not in T, we have embeddings $\psi_{e_\pm}$ obtained by composing $\phi_{e_\pm} : G_e \to G_v, G_w$ with embeddings $G_v, G_w \to G_T$. Thus, for each edge e which is not in T, we have two isomorphisms $G_e \to G_e^\pm < G_T$ and, accordingly, we obtain isomorphisms $G_e^- \to G_e^+$. Lastly, using these isomorphisms, define the simultaneous HNN extension G of G_T. Lastly, set $\pi(\mathcal{G}) = G$.

Whenever $G \cong \pi(\mathcal{G})$, we will say that $\mathcal{G}$ determines a *graph of groups decomposition* of G. The decomposition $\mathcal{G}$ is called *trivial* if there is a vertex v so that the natural homomorphism $G_v \to G$ is onto.

EXAMPLE 4.47. 1. Suppose that the graph Γ consists of a single edge e whose head e_+ is the vertex called 2 and the tail e_- is the vertex called 1. Assume that $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_2$. Then

$$\pi(\mathcal{G}) \cong G_1 \star_{H_1 \cong H_2} G_2.$$

2. Suppose that the graph Γ is a monogon, consisting of an edge e connecting the vertex called 1 to itself. Suppose, furthermore, $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_1$. Then

$$\pi(\mathcal{G}) \cong G_1 \star_{H_1 \cong H_2} .$$

Once this example is understood, one can show that for every graph of groups $\mathcal{G}$, the group $\pi_1(\mathcal{G})$ exists by describing this group in terms of generators and relators in the manner similar to the definition of the amalgamated free product and the HNN extension. In the next section we will see how to construct $\pi_1(\mathcal{G})$ using topology.

4.5.3. Converting graphs of groups into amalgams. Suppose that $\mathcal{G}$ is a graph of groups and $G = \pi_1(\mathcal{G})$. Our goal is to convert $\mathcal{G}$ into an amalgam decomposition of G. There are two cases to consider:

1. Suppose that the graph Γ underlying $\mathcal{G}$ contains a oriented edge $e = [v_1, v_2]$ so that e separates Γ in the sense that the graph Γ' obtained form Γ by removing e (and keeping v_1, v_2) is a disjoint union of connected subgraphs $\Gamma_1 \sqcup \Gamma_2$, where $v_i \in V(\Gamma_i)$. Let $\mathcal{G}_i$ denote the subgraph in the graph of groups $\mathcal{G}$, corresponding to $\Gamma_i, i = 1, 2$. Then set

$$G_i := \pi_1(\mathcal{G}_i), \quad i = 1, 2, \quad G_3 := G_e.$$

We have composition of embeddings $G_e \to G_{v_i} \to G_i \to G$. Then the universal property of $\pi_1(\mathcal{G}_i)$ and $\pi_1(\mathcal{G})$ implies that $G \cong G_1 \star_{G_3} G_2$: One simply verifies that G satisfies the universal property for the amalgam $G_1 \star_{G_3} G_2$.

2. Suppose that Γ contains an oriented edge $e = [v_1, v_2]$ such that e does not separate Γ. Let $\Gamma_1 := \Gamma'$, where Γ' is obtained from Γ by removing the edge e as in the Case 1. Set $G_1 := \pi_1(\mathcal{G}_1)$ as before. Then the embeddings

$$G_e \to G_{v_i}, i = 1, 2$$

induce embeddings $G_e \to G_i$ with the images H_1, H_2 respectively. Similarly to the Case 1, we obtain

$$G \cong G_1 \star_{G_e} = G_1 \star_{H_1 \cong H_2}$$

where the isomorphism $H_1 \to H_2$ is given by the composition

$$H_1 \to G_e \to H_2.$$

Clearly, $\mathcal{G}$ is trivial if and only if the corresponding amalgam $G_1 \star_{G_3} G_2$ or $G_1 \star_{G_e}$ is trivial.

4.5.4. Topological interpretation of graphs of groups. Let $\mathcal{G}$ be a graph of groups. Suppose that for all vertices and edges $v \in V(\Gamma)$ and $e \in E(\Gamma)$ we are given connected cell complexes M_v, M_e with the fundamental groups G_v, G_e respectively. For each edge $e = [v, w]$ assume that we are given a continuous map $f_{e_\pm} : M_e \to M_{e_\pm}$ which induces the monomorphism $\phi_{e_\pm}$. This collection of spaces and maps is called a *graph of spaces*

$$\mathcal{G}_M := \{M_v, M_e, f_{e_\pm} : M_e \to M_{e_\pm} : v \in V(\Gamma), e \in E(\Gamma)\}.$$

In order to construct $\mathcal{G}_M$ starting from $\mathcal{G}$, recall that each group G admits a cell complex $K(G, 1)$ whose fundamental group is G and whose universal cover is contractible, see §3.8.2. Given a group homomorphism $\phi : H \to G$, there exists a continuous map, unique up to homotopy,

$$f : K(H, 1) \to K(G, 1)$$

which induces the homomorphism ϕ. Then one can take $M_v := K(G_v, 1)$, $M_e := K(G_e, 1)$, etc.

To simplify the picture (although this is not the general case), the reader can think of each M_v as a manifold with several boundary components which are homeomorphic to $M_{e_1}, M_{e_2}, \ldots$, where e_j are the edges having v as their *head* or *tail*. Then assume that the maps $f_{e_\pm}$ are homeomorphisms onto the respective boundary components.

For each edge e we form the product $M_e \times [0, 1]$ and then form the double mapping cylinders for the maps $f_{e_\pm}$, i.e., identify points of $M_e \times \{0\}$ and $M_e \times \{1\}$ with their images under f_{e_-} and f_{e_+} respectively. Let M denote the resulting cell complex. It then follows from the Seifert–Van Kampen theorem [**Mas91**] that

THEOREM 4.48. *The group $\pi_1(M)$ is isomorphic to $\pi(\mathcal{G})$.*

This theorem allows one to think of the graphs of groups and their fundamental groups *topologically* rather than *algebraically*.

EXERCISE 4.49. Use the above interpretation to show that for each vertex $v \in V(\Gamma)$ the canonical homomorphism $G_v \to \pi(\mathcal{G})$ is injective.

EXAMPLE 4.50. The group $F(X)$ is isomorphic to $\pi_1(\vee_{x \in X} \mathbb{S}^1)$.

4.5.5. Constructing finite index subgroups. In this section we use the topological interpretation of graphs of groups in order to construct finite index subgroups. The main result (Theorem 4.52) will be used in the proof of quasi-isometric rigidity of virtually free groups in Chapter 18.

Let $\mathcal{G}$ be a finite graph of groups. Suppose that we are given a *compatible* collection of finite index subgroups $G'_v < G_v, G'_e < G_e$ for each vertex and edge group of $\mathcal{G}$, i..e, a collection of subgroups such that whenever $v = e_\pm$, we have

$$G_v \cap \phi_{e_\pm}(G'_e) = G'_v \cap \phi_{e_\pm}(G_e).$$

We refer to this equality as the *compatibility condition.*

THEOREM 4.51. *For every compatible collection of finite index subgroups as above, there exists a finite-index subgroup $G' < G$ such that*

$$G' \cap G_v = G'_v, \quad G' \cap G_e = G'_e$$

for every vertex v and edge e. Furthermore, $G' = \pi_1(\mathcal{G}')$, where $\mathcal{G}'$ is another finite graph of groups, for which there exists a morphism of graphs of groups

$$p : \mathcal{G}' \to \mathcal{G}$$

inducing the inclusion $G' \hookrightarrow G$.

PROOF. This theorem is proven by John Hempel in **[Hem87]** (Theorem 2.2). Our proof mostly follows his arguments.

Let Γ denote the graph underlying $\mathcal{G}$. For each vertex group G_v (resp. edge group G_e) of $\mathcal{G}$ we let X_v (resp. X_e) denote a classifying space of this group. Then, as in §4.5.4, we convert the graph of groups $\mathcal{G}$ into a graph of spaces X, with vertex spaces X_v and edge spaces X_e. We will use the notation

$$f_{e_\pm} : X_e \to X_{e_\pm}$$

for the attaching maps inducing the monomorphisms $\phi_{e_\pm}$. It will be convenient to assume that distinct attaching maps have disjoint images.

We will construct the subgroup G' as the fundamental group of another graph of spaces X' which admits a finite cover $p : X' \to X$, such that $G' = p_*(\pi_1(X'))$. The group inclusions $G'_v \to G_v, G'_e \to G_v$ are induced by finite covers of spaces

$$X'_v \to X_v, \quad X'_e \to X_e.$$

We now assemble the spaces X'_v, X'_e into a finite connected graph of spaces X'. We let d_v, d_e denote the degrees of these covers, i.e.,

$$d_v = |G_v : G'_v| \ .$$

Set

$$d = \prod_{v \in V(\Gamma)} d_v \ .$$

Now, for each $v \in V(\Gamma)$ we let $\tilde{X}_v$ denote the disjoint union of d/d_v copies of X'_v.

We will use the notation X'_{v_i} for components of $\tilde{X}_v$.

Our next goal is to describe how to connect components X'_{v_i} to each other via copies of the double mapping cones for the maps $X'_e \to X'_v$. We then observe that by the definition of $\tilde{X}_v$ and the compatibility assumption, for each edge e with $e_+ = v, e_- = w$, the number of components preimages of $f_{e_+}(X_e)$ in $\tilde{X}_v$ equals the number of components of preimages of $f_{e_-}(X_e)$ in $\tilde{X}_w$. We therefore, match these subsets of $\tilde{X}_v, \tilde{X}_w$ in pairs. For every such pair, we connect the corresponding vertices v_i, w_j by an edge e_{ij}. This defines a new graph $\tilde{\Gamma}'$ whose vertices are v_i's and edges are e_{ij}'s, where v runs through the vertex set of Γ. The graph $\tilde{\Gamma}'$ is, a priori, disconnected, we pick a connected component Γ' of this graph.

We then construct a graph of spaces X' based on Γ' as follows. For each vertex v_i, we take, of course, X'_{v_i} as the associated vertex space. For every edge e_{ij} define

$$X'_{e_{ij}} = X'_e$$

where e_{ij} corresponds to the matching of preimages of $f_{e_\pm}(X_e)$. Accordingly, we let the map

$$f_{e_{ij+}} : X'_{e_{ij}} \to X'_{v_i}$$

be the lift of the attaching map $f_{e_+} : X_e \to X_v$. Note that these lifts exist by the compatibility assumption. We do the same for the vertex e_{ij-}. As the result, we obtain a connected graph of spaces.

We leave it to the reader to verify that the covering maps

$$X'_{v_i} \to X_v, \quad X'_{e_{ij}} \to X_e$$

assemble to a covering map $X' \to X$. This covering map is finite-to-one by the construction. It induces an embedding $G' = \pi_1(X') \to G = \pi_1(X)$. Again, by construction, this embedding satisfies the requirements of the theorem. $\square$

As an application we obtain:

THEOREM 4.52. *Let $\mathcal{G}$ be a finite graph of finite groups. Then its fundamental group $G = \pi_1(\mathcal{G})$ is virtually free.*

PROOF. For each vertex group G_v of $\mathcal{G}$ we let $G'_v < G_v$ be the trivial subgroup; we make the same choice for the edge groups. Let $G' < G$ denote the finite index subgroup and the morphism

$$\mathcal{G}' \to \mathcal{G}$$

given by Theorem 4.51. By construction, $\mathcal{G}'$ has trivial vertex groups. Hence, for the underlying graph Γ' of $\mathcal{G}'$ we obtain

$$G' = \pi_1(\Gamma')$$

which is free. $\square$

4.5.6. Graphs of groups and group actions on trees. An *action* of a group G on a tree T is an action $G \curvearrowright T$ such that each element of G acts as an automorphism of T, i.e., such action is a homomorphism $G \to Aut(T)$. A tree T with the prescribed action $G \curvearrowright T$ is called a G–tree. An action $G \curvearrowright T$ is said to be *without inversions* if whenever $g \in G$ preserves an edge e of T, it fixes e pointwise. The action is called *bounded* (or *trivial*) if there is a vertex $v \in T$ fixed by the entire group G.

REMARK 4.53. Later on, in Chapter 9, we will encounter more complicated (non-simplicial) *real* trees and group actions on such trees.

Our next goal is to explain the relation between the graph of groups decompositions of G and actions of G on simplicial trees without inversions.

Suppose that $G \cong \pi(\mathcal{G})$ is a graph of groups decomposition of G. We associate with $\mathcal{G}$ a graph of spaces $M = M_{\mathcal{G}}$ as in §4.5.4. Let X denote the universal cover of the corresponding cell complex M. Then X is the disjoint union of the copies of the universal covers $\tilde{M}_v, \tilde{M}_e \times (0,1)$ of the complexes M_v and $M_e \times (0,1)$. We will refer to this partitioning of X as the *tiling* of X. In other words, X has the structure of a graph of spaces, where each vertex/edge space is homeomorphic to $\tilde{M}_v, v \in V(\Gamma)$, $\tilde{M}_e \times [0,1], e \in E(\Gamma)$. Let T denote the graph corresponding to X: Each copy of $\tilde{M}_v$ determines a vertex in T and each copy of $\tilde{M}_e \times [0,1]$ determines an edge in T.

EXAMPLE 4.54. Suppose that Γ consists of two vertices 1 and 2 and the edge $[1,2]$ connecting them, M_1 and M_2 are surfaces of genus 1 with a single boundary component each. Let M_e be the circle. We assume that the maps $f_{e_\pm}$ are homeomorphisms of this circle to the boundary circles of M_1, M_2. Then, M is a surface of genus 2. The graph T is sketched in Figure 4.1.

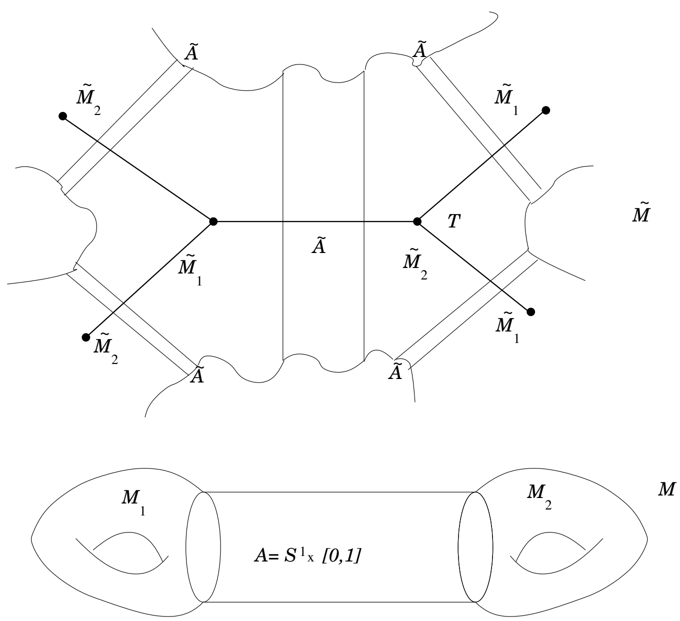


FIGURE 4.1. *U*niversal cover of the genus 2 surface.

The Mayer–Vietoris theorem, applied to the above tiling of X, implies that $0 = H_1(X,\mathbb{Z}) \cong H_1(T,\mathbb{Z})$. Therefore, $T = T(\mathcal{G})$ is a tree. The group $G = \pi_1(M)$ acts on X by deck-transformations, preserving the tiling. Thus, we obtain the induced action $G \curvearrowright T$. If $g \in G$ preserves some $\tilde{M}_e \times (0,1)$, then g comes from the fundamental group of M_e. Therefore, such g also preserves the orientation on the segment $[0,1]$. Hence, the action $G \curvearrowright T$ is without inversions. Observe that the stabilizer of each $\tilde{M}_v$ in G is conjugate in G to $\pi_1(M_v) = G_v$. Moreover, $T/G = \Gamma$.

EXAMPLE 4.55. Let $G = BS(n,m)$ be the Baumslag-Solitar group described in Example 4.32, (8). The group G clearly has the structure of a graph of groups since it is isomorphic to the HNN extension of $\mathbb{Z}$,

$$\mathbb{Z}\star_{H_1 \cong H_2}$$

where the subgroups $H_1, H_2 \subset \mathbb{Z}$ have the indices n and m respectively. In order to construct the cell complex $K(G,1)$, take the circle $\mathbb{S}^1 = M_v$, the cylinder $\mathbb{S}^1 \times [0,1]$

and attach the ends to this cylinder to M_v by the maps of the degrees p and q respectively. Now, consider the associated G–tree T. Its vertices have valence $n+m$: Each vertex v has m incoming and n outgoing edges so that for each outgoing edge e we have $v = e_-$ and for each incoming edge we have $v = e_+$. The vertex stabilizer $G_v \cong \mathbb{Z}$ permutes (transitively) incoming and outgoing edges among each other. The stabilizer of each outgoing edge is the subgroup H_1 and the stabilizer of each incoming edge is the subgroup H_2. Thus, the action of $\mathbb{Z}$ on the set of incoming edges is *via* the group $\mathbb{Z}/m$ and on the set of outgoing edges *via* the group $\mathbb{Z}/n$.

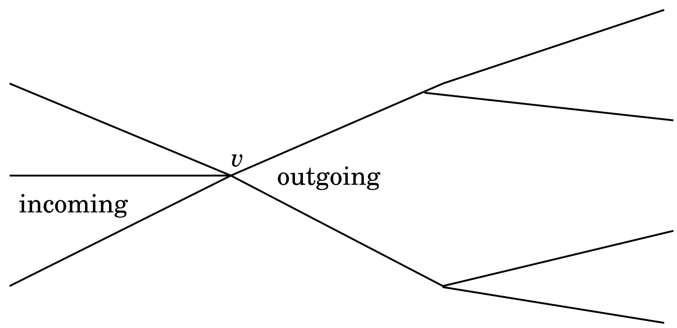


FIGURE 4.2. Tree for the group $BS(2,3)$.

LEMMA 4.56. *The action $G \curvearrowright T$ is bounded if and only if the graph of groups decomposition of G is trivial.*

PROOF. Suppose that G fixes a vertex $\tilde{v} \in T$. Then $\pi_1(M_v) = G_v = G$, where $v \in \Gamma$ is the projection of $\tilde{v}$. Hence, the decomposition of G is trivial. Conversely, suppose that G_v maps onto G. Let $\tilde{v} \in T$ be the vertex which projects to v. Then $\pi_1(M_v)$ is the entire $\pi_1(M)$ and, hence, G preserves $\tilde{M}_{\tilde{v}}$. Therefore, the group G fixes $\tilde{v}$. □

Conversely, each action of G on a simplicial tree T yields a realization of G as the fundamental group of a graph of groups $\mathcal{G}$, such that $T = T(\mathcal{G})$. Here is the construction of $\mathcal{G}$. Furthermore, an *unbounded* action leads to a *nontrivial* graph of groups.

If the action $G \curvearrowright T$ has inversions, we replace T with its barycentric subdivision T'. Then G acts on T' without inversions. If the action $G \curvearrowright T$ is unbounded, so is $G \curvearrowright T'$. Thus, from now on, we assume that G acts on T without inversions. Then, the quotient T/G is a graph Γ: $V(\Gamma) = V(T)/G$ and $E(\Gamma) = E(T)/G$. For every vertex $\tilde{v}$ and edge $\tilde{e}$ of T we let $G_{\tilde{v}}$ and $G_{\tilde{e}}$ be their respective stabilizes in G. Clearly, whenever $\tilde{e} = [\tilde{v}, \tilde{w}]$, we get the embedding

$$G_{\tilde{e}} \to G_{\tilde{v}}.$$

If $g \in G$ maps oriented the edge $\tilde{e} = [\tilde{v}, \tilde{w}]$ to an oriented edge $\tilde{e}' = [\tilde{v}', \tilde{w}']$, we obtain isomorphisms

$$G_{\tilde{v}} \to G_{\tilde{v}'}, \quad G_{\tilde{w}} \to G_{\tilde{w}'}, \quad G_{\tilde{e}} \to G_{\tilde{e}'}$$

induced by conjugation *via* g and the following diagram is commutative:

$$\begin{array}{ccc} G_{\tilde{e}} & \longrightarrow & G_{\tilde{v}} \\ \downarrow & & \downarrow \\ G_{\tilde{e}'} & \longrightarrow & G_{\tilde{v}'} \end{array}$$

We set $G_v := G_{\tilde{v}}$, $G_e := G_{\tilde{e}}$, where v and e are the projections of $\tilde{v}$ and edge $\tilde{e}$ to Γ. For every edge e of Γ oriented as $e = [v, w]$, we define the monomorphism $G_e \to G_v$ as follows. By applying an appropriate element $g \in G$ as above, we can assume that $\tilde{e} = [\tilde{v}, \tilde{w}]$. We then define the embedding $G_e \to G_v$ to make the diagram

$$\begin{array}{ccc} G_{\tilde{e}} & \longrightarrow & G_{\tilde{v}} \\ \downarrow & & \downarrow \\ G_{e} & \longrightarrow & G_{v} \end{array}$$

commutative. The result is a graph of groups $\mathcal{G}$. We leave it to the reader to verify that the functor $(G \curvearrowright T) \to \mathcal{G}$ described above is the inverse of the functor $\mathcal{G} \to (G \curvearrowright T)$ for $\mathcal{G}$ with $G = \pi_1(\mathcal{G})$. In particular, $\mathcal{G}$ is trivial if and only if the action $G \curvearrowright T$ is bounded.

Definition 4.57. $\mathcal{G} \to (G \curvearrowright T) \to \mathcal{G}$ is the *Bass–Serre correspondence* between realizations of groups as fundamental groups of graphs of groups and group actions on trees without inversions.

We refer the reader to [**SW79**] and [**Ser80**] for further details on the Bass–Serre correspondence. Below is a simple, yet non-obvious, example of application of this correspondence:

Lemma 4.58. *Suppose that G is countable, but not finitely generated. Then G admits a nontrivial action on a simplicial tree.*

Proof. Using countability of G, enumerate the elements of the group G and define an exhaustion of G by finitely generated subgroups:

$$G_1 \leqslant G_2 \leqslant G_3 \leqslant \dots$$

where $G_{n+1} = \langle G_n, g_{n+1} \rangle$. The inclusion homomorphisms

$$\iota_n : G_n \hookrightarrow G_{n+1}$$

determine an infinite graph of groups, where vertices are labeled $v_n, n \in \mathbb{N}$, the vertex groups are $G_{v_n} = G_n$, the edge groups are

$$G_{e_n} = G_n, \quad e_n = [v_n, v_{n+1}]$$

and the map $G_{e_n} \to G_{v_n}$ is the identity, while the map $G_{e_n} \to G_{v_{n+1}}$ is ι_n. We claim that the fundamental group $\pi_1(\mathcal{G})$ of this graph of groups is G itself. Indeed, we have natural inclusion homomorphisms

$$f_n : G_n \to G.$$

If H is a group and $h_n : G_n \to H$ are homomorphisms, such that

$$h_{n+1}\big|_{G_n} = h_n,$$

then h_n's determine a homomorphism $h : G \to H$ by $h(g) = h_n(g)$ whenever $g \in G_n$. Uniqueness of h is also clear. Thus, G satisfies the universality property in the definition of $\pi_1(\mathcal{G})$ and, hence, $G \cong \pi_1(\mathcal{G})$.

Next, none of the vertex groups G_n maps onto G via the inclusion homomorphism ι_n. Therefore, the action $G \curvearrowright T$ of G on a simplicial tree, defined by the Bass–Serre correspondence, is nontrivial and without inversions. □

4.6. Ping-pong lemma. Examples of free groups

The ping-pong lemma is a simple, yet powerful, tool for constructing free groups acting on sets. We will see in Chapter 13 how ping-pong is used for the proof of the *Tits Alternative*.

We begin with the *ping-lemma*, a version of the ping-pong lemma for semigroups:

LEMMA 4.59 (Ping-pong for semigroups). *Let X be a set, and let $g : X \to X$ and $h : X \to X$ be two injective maps. Suppose that $A \subset X$ is a nonempty subset such that $g(A), h(A)$ are disjoint subsets of A. Then g, h generate a free subsemigroup of rank 2 in the semigroup of self-maps $X \to X$. Moreover, for two distinct words w, w' in the generators g, h,*

$$w(A) \cap w'(A) = \emptyset.$$

PROOF. Let w, w' be distinct nonempty words in the alphabet g, h. We claim that $w(A) \cap w'(A) = \emptyset$. We prove this by induction on the maximum of lengths $\ell(w), \ell(w')$ of w, w'. If both w, w' have unit length the claim is immediate. Suppose that the claim holds for all words w, w' such that $\max(\ell(w), \ell(w')) \leqslant n$. Let w, w' be distinct nonempty words in g, h such that $\ell(w) \leqslant \ell(w')) = n + 1$. The words w, w' either have the same first letter (the prefix), or distinct prefixes. Suppose first that w, w' have the same prefix $x \in \{g, h\}$; then

$$w = xu \quad w' = xu', \quad y \neq y', \quad \max(\ell(u), \ell(u')) \leqslant n.$$

Then, by the induction hypothesis,

$$u(A) \cap u'(A) = \emptyset.$$

Injectivity of x implies that the sets $w(A) = xu(A)$ and $w'(A) = xu'(A)$ are also disjoint, as claimed. Suppose, next, that w, w' have distinct prefixes:

$$w = xu \quad w' = x'u', \quad \{x, x'\} = \{g, h\}.$$

Then $w(A) \subset x(A)$, $w'(A) \subset x'(A)$ are disjoint and the claim follows. □

We next consider ping-pong for groups of bijections. The setup for the ping-pong lemma is a pair of bijections $g_1, g_2 \in Bij(X)$ ("ping-pong partners") and a quadruple of nonempty subsets

$$B_i^{\pm} \subset X, \quad i = 1, 2,$$

whose union is $B \subset X$. Define

$$C_i^{+} := B \setminus B_i^{-}, C_i^{-} := B \setminus B_i^{+}, \quad i = 1, 2.$$

It is clear that $B_i^{\pm} \neq C_j^{\pm}$ and $B_i^{\pm} \neq C_j^{\mp}$ for all choices of i, j and $+, -$.

EXERCISE 4.60. Suppose that $g \in Bij(X)$ is a bijection such that for some $A \subset X$,

$$g(A) \subsetneq A.$$

Then g has infinite order.

LEMMA 4.61 (Ping–pong, or table–tennis, lemma). *Let X, $B_i^{\pm}, C_i^{\pm}$ be as above, and suppose that*

$$g_i^{\pm 1}(C_i^{\pm}) \subset B_i^{\pm}, \quad i = 1, 2.$$

Then the bijections g_1, g_2 generate a rank 2 free subgroup of $Bij(X)$.

PROOF. Let w be a non-empty reduced word in $\{g, g^{-1}, h, h^{-1}\}$. In order to prove that w corresponds to a non-identity element of $Bij(X)$, it suffices to check that $w(C_j^{\pm}) \subset B_i^{\pm}$ for some i, j and for some choice of $+$ or $-$. We claim that whenever w has the form

$$w = g_i^{\pm 1} u g_j^{\pm 1},$$

we have

$$w(C_j^{\pm}) \subset B_i^{\pm}.$$

This would immediately imply that w does not represent the identity map $X \to X$. The claim is proven by induction on the length $\ell(w)$ of w as in the proof of Lemma 4.59. The statement is clear if $\ell(w) = 1$. Suppose it holds for all words w' of length n, we will prove it for words w or length $n + 1$. Such w has the form

$$w = g_i^{\pm 1} w', \quad \ell(w') = n.$$

Since the prefix of w' cannot equal $g_i^{\mp}$ (as w is a reduced word), it follows from the induction hypothesis that (for some j and a choice of $+, -$)

$$w'(C_j^{\pm}) \subset C_i^{\pm}.$$

Then $g_i^{\pm 1}(C_i^{\pm}) \subset B_i^{\pm}$ and the claim follows. □

Lemma 4.61 extends to the case of free products of subgroups. The setup for this extension is a collection $\{G_i : i \in I\}$ of subgroups of $Bij(X)$, and of nonempty subsets $A_i \subset X$ $(i \in I)$, whose union is denoted

$$A = \bigcup_{i \in I} A_i.$$

For each A_i define $A_i^c = A \setminus A_i$.

LEMMA 4.62 (The ping-pong lemma for free products). *Given the above data, suppose that for each $i \in I$ and all $g \in G_i \setminus \{1\}$, we have the inclusion*

$$g(A_i^c) \subset A_i.$$

Then the natural homomorphism

$$\phi : \star_{i \in I} G_i \to Bij(X), \quad \phi\big|_{G_i} = \mathrm{id}_{G_i}, i \in I,$$

is a monomorphism.

PROOF. Consider a non-trivial word w in the alphabet

$$\bigcup_{i \in I} G_i,$$

where no two consecutive letters belong to the same G_k. Suppose that w has the prefix $g_i \in G_i \setminus \{1\}$ and the suffix $g_j \in G_j \setminus \{1\}$. We claim that

$$w(A_j^c) \subset A_i.$$

The proof is the induction on the length $\ell(w)$ of w. The claim is clear for $\ell(w) = 1$. Suppose that the claim holds for all words w' of the length n and let w be a word of the length $n + 1$. Then w has the form

$$w = g_i w', \quad \ell(w') = n,$$

where the suffix of w' is $g_j \in G_j$. Since the prefix of w' cannot equal to an element of G_i, it follows from the induction hypothesis that

$$w'(A_j^c) \subset A_i^c.$$

Hence, $w(A_j^c) \subset g_i(A_i^c) \subset A_i$. It follows that the homomorphism ϕ is injective. □

In the following example we illustrate both form of ping-pong.

EXAMPLE 4.63. For any real number $r > 1$ the matrices

$$g_1 = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \text{ and } g_2 = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix}$$

generate a free subgroup of $SL(2,\mathbb{R})$.

1st proof. The group $SL(2,\mathbb{R})$ acts (with the kernel $\pm I$) on the upper half plane $\mathbb{H}^2 = \{z \in \mathbb{C} \mid \Im(z) > 0\}$ by linear fractional transformations

$$z \mapsto \frac{az+b}{cz+d}.$$

Define quater-planes

$$B_1^+ = \{z \in \mathbb{H}^2 : \Re(z) > r/2, \quad B_1^- = \{z \in \mathbb{H}^2 : \Re(z) < -r/2\}$$

and open disks

$$B_2^+ := \{z \in \mathbb{H}^2 : |z - \frac{1}{r}| < \frac{1}{r}\}, \quad B_2^- := \{z \in \mathbb{H}^2 : |z + \frac{1}{r}| < \frac{1}{r}\}.$$

The reader will verify that $g_k, B_k^{\pm}, k = 1, 2$ satisfy the assumptions of Lemma 4.61. It follows that the group $\langle g_1, g_2\rangle$ is free of rank 2.

FIGURE 4.3. Example of ping-pong. REDRAW

2nd proof. The group $SL(2,\mathbb{R})$ also acts linearly on $\mathbb{R}^2$. Consider the infinite cyclic subgroups $G_k = \langle g_k\rangle$, $i = 1, 2$ of $SL(2,\mathbb{R})$. Define the following subsets of $\mathbb{R}^2$

$$A_1 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : |x| > |y| \right\} \text{ and } A_2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : |x| < |y| \right\}.$$

Then for each $g \in G_1 \setminus \{1\}$, $g(A_2) \subset A_1$ and for each $g \in G_2 \setminus \{1\}$, $g(A_1) \subset A_2$. Therefore, the subgroup of $SL(2,\mathbb{R})$ generated by g_1, g_2 is free of rank 2 according to Lemma 4.62. □

Remark 4.64. The statement in the Example 4.63 no longer holds for $r = 1$. Indeed, in this case we have

$$g_1^{-1} g_2 g_1^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Thus, $(g_1^{-1} g_2 g_1^{-1})^2 = I$, and, hence, the group generated by g_1, g_2 is not free.

4.7. Free subgroups in $SU(2)$

As an application of ping-pong in $SL(2, \mathbb{R})$ and the formalism of algebraic groups, we will now give a "cheap" proof of the fact that the group $SU(2)$ contains a subgroup isomorphic to F_2, the free group on two generators:

Lemma 4.65. *The subset of monomorphisms $F_2 \to SU(2)$ is dense in the variety $Hom(F_2, SU(2)) = SU(2) \times SU(2)$.*

Proof. Consider the space $V = Hom(F_2, SL(2, \mathbb{C})) = SL(2, \mathbb{C}) \times SL(2, \mathbb{C})$; every element $w \in F_2$ defines a polynomial function

$$f_w : V \to SL(2, \mathbb{C}), \quad f_w(\rho) = \rho(w).$$

Since $SL(2, \mathbb{R}) \leqslant SL(2, \mathbb{C})$ contains a subgroup isomorphic to F_2 (see Example 4.63), it follows that for every $w \neq 1$, the function f_w takes values different from 1. In particular, the subset $E_w := f_w^{-1}(1)$ is a proper (complex) subvariety in V. Since $SL(2, \mathbb{C})$ is a connected complex manifold, the variety $SL(2, \mathbb{C})$ is irreducible; hence, V is irreducible as well. It follows that for every $w \neq 1$, E_w has empty interior (in the classical topology) in V. Suppose that for some $w \neq 1$, the intersection

$$E'_w := E_w \cap SU(2) \times SU(2)$$

contains a nonempty open subset U. In view of Exercise 3.49, $SU(2)$ is Zariski dense (over $\mathbb{C}$) in $SL(2, \mathbb{C})$; hence, U (and, thus, E_w) is Zariski dense in V. It then follows that $E_w = V$, which is false. Therefore, for every $w \neq 1$, the closed (in the classical topology) subset $E'_w \subset Hom(F_2, SU(2))$ has empty interior. Since F_2 is countable, by Baire category theorem, the union

$$E := \bigcup_{w \neq 1} E'_w$$

has empty interior in $Hom(F_2, SU(2))$. Since every $\rho \notin E$ is injective, lemma follows. ☐

Since the group $SU(2)/\{\pm I\}$ is isomorphic to $SO(3)$, we also obtain:

Corollary 4.66. *The subset of monomorphisms $F_2 \to SO(3)$ is dense in the variety $Hom(F_2, SO(3))$.*

4.8. Ping-pong on projective spaces

We will frequently use the Ping-pong lemma in the case when X is a projective space. This application of the ping-pong argument is the key for the proof of the Tits Alternative.

Let V be an n-dimensional space over a local field $\mathbb{K}$, the reader should think of $\mathbb{R}, \mathbb{C}$, or $\mathbb{Q}_p$. We endow the projective space $P(V)$ with the metric d as in §1.14. We refer the reader to §1.15 for the notion of *proximality*, attractive points $A_g \in P(V)$ and exceptional hyperplanes $E_g \subset P(V)$ for proximal projective transformations.

Definition 4.67. Two proximal elements $g, h \in GL(V)$ will be called *ping partners* if

$$A_g \notin E_h, \quad A_h \notin E_g.$$

Two very proximal elements $g, h \in GL(V)$ will be called *ping-pong partners* if all four pairs pairs $(g, h), (g, h^{-1}), (g^{-1}, h)$ and (g^{-1}, h^{-1}) are ping-partners. In particular, the four points $A_g, A_{g^{-1}}, A_h, A_{h^{-1}}$ are all distinct.

For instance, if $n = 2$, then g, h are ping-pong partners if and only if they are both proximal and their four fixed points in the projective line $P(V)$ are pairwise distinct.

Lemma 4.68. *Assume that $g, h \in GL(n, \mathbb{K})$ are ping partners. Then there exists a positive integer N such that for all $m \geqslant N$, the powers g^m and h^m generate a rank two free subsemigroup of $GL(n, \mathbb{K})$. Similarly, if g, h are ping-pong partners, then there exists N such that for all $m \geqslant N$, g^m and h^m generate a rank two free subgroup of $GL(n, \mathbb{K})$.*

Proof. We prove the statement about ping-pong partners, since its proof will contain the proof in the case of ping-partners. Define

$$\varepsilon = \frac{1}{2} \min \big(\text{dist}(A_g, H(g) \cup E_h \cup E_{h^{-1}}), \text{dist}(A_{g^{-1}}, E_{g^{-1}} \cup E_h \cup E_{h^{-1}}),$$
$$\text{dist}(A_h, E_h \cup H(g) \cup E_{g^{-1}}), \text{dist}(A_{h^{-1}}, E_{h^{-1}} \cup H(g) \cup E_{g^{-1}}) \big) .$$

Since g, h are ping-pong partners, $\varepsilon > 0$. Next, by Corollary 1.134, there exists N such that for all $m \geqslant N$ we have:

1. $g^{\pm m} : P(V) \to P(V)$ maps the complement of the ε-neighborhood of $E_{g^{\pm 1}}$ inside the ball of radius ε and center $A_{g^{\pm 1}}$.

2. $h^{\pm m}$ maps the complement of the ε-neighborhood of $E_{h^{\pm 1}}$ inside the ball of radius ε and center $A_{h^{\pm 1}}$.

Set

$$A := B(A_g, \varepsilon) \sqcup B(A_{g^{-1}}, \varepsilon)$$

and

$$B := B(A_h, \varepsilon) \sqcup B(A_{h^{-1}}, \varepsilon).$$

Clearly,

$$g^{km}(A) \subseteq B$$

and

$$h^{km}(B) \subseteq A$$

for every $k \in \mathbb{Z} \setminus \{0\}$. Hence, by Lemma 4.62, regarded as projective transformations, g^m and h^m generate a free subgroup of rank 2 in $PGL(n, \mathbb{K})$. Therefore, the same holds for $g^m, h^m \in GL(n, \mathbb{K})$, see Lemma 4.24. □

4.9. Cayley graphs

One of the central themes of Geometric Group Theory is treating groups as geometric objects. The oldest, and most common, way to 'geometrize' groups, by their *Cayley graphs*. Other 'geometrizations' of groups are given by simplicial complexes and Riemannian manifolds.

Every group may be turned into a geometric object (a graph) as follows. Given a group G and its generating set S, one defines the *Cayley graph* of G with respect to S. This is a directed graph $\text{Cayley}_{\text{dir}}(G, S)$ such that

- its set of vertices is G;
- its set of oriented edges is (g, gs), with $s \in S$.

Usually, the underlying non-oriented graph $\mathrm{Cayley}(G, S)$ of $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$, i.e., the graph such that:

- its set of vertices is G;
- its set of edges consists of all pairs of elements in G, $\{g, h\}$, such that $h = gs$, with $s \in S$,

is also called the *Cayley graph of G with respect to S*.

We will also denote the notation $\overline{gh}$ and $[g, h]$ for the edge $\{g, h\}$. In order to avoid the confusion with the notation for the commutator of the elements g and h we will always add the word *edge* in this situation.

EXERCISE 4.69. Show that the graph $\mathrm{Cayley}(G, S)$ is connected.

One can attach a *color* (*label*) from S to each oriented edge in $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$: the edge (g, gs) is labeled by s.

We endow the graph $\mathrm{Cayley}(G, S)$ with the standard length metric (where every edge has unit length). The restriction of this metric to G is called *the word metric associated to S* and it is denoted by dist_S or d_S.

NOTATION 4.70. For an element $g \in G$ and a generating set S we denote $\mathrm{dist}_S(1, g)$ by $|g|_S$, the *word norm* of g. With this notation, $\mathrm{dist}_S(g, h) = |g^{-1}h|_S = |h^{-1}g|_S$.

CONVENTION 4.71. In this book, unless stated otherwise, all Cayley graphs are defined for finite generating sets S.

Much of the discussion in this section, though, remains valid for arbitrary generating sets, including infinite ones.

REMARK 4.72. 1. Every group acts on itself, on the left, by the left multiplication:

$$G \times G \to G\,,\ (g, h) \mapsto gh\,.$$

This action extends to any Cayley graph: if $[x, xs]$ is an edge of $\mathrm{Cayley}(G, S)$ with the vertices x, xs, we extend g to the isometry

$$g : [x, xs] \to [gx, gxs]$$

between the unit intervals. Both actions $G \curvearrowright G$ and $G \curvearrowright \mathrm{Cayley}(G, S)$ are isometric. It is also clear that both actions are free, properly discontinuous and cocompact (provided that S is finite): The quotient $\mathrm{Cayley}(G, S)/G$ is homeomorphic to the bouquet of n circles, where n is the cardinality of S.

2. The action of the group on itself by the right multiplication defines maps

$$R_g : G \to G\,,\ R_g(h) = hg$$

that are, in general, not isometries with respect to a word metric, but are at finite distance from the identity map:

$$\mathrm{dist}(\mathrm{id}(h), R_g(h)) = |g|_S\,.$$

EXERCISE 4.73. Prove that the word metric on a group G associated to a generating set S may also be defined

(1) either as the unique maximal left-invariant metric on G such that

$$\mathrm{dist}(1,s) = \mathrm{dist}(1,s^{-1}) = 1\,,\ \forall s \in S\,;$$

(2) or by the following formula: $\mathrm{dist}(g,h)$ is the length of the shortest word w in the alphabet $S \cup S^{-1}$ such that $w = g^{-1}h$ in G.

Below are two simple examples of Cayley graphs.

EXAMPLE 4.74. Consider the group $\mathbb{Z}^2$ with the set of generators

$$S = \{a = (1,0), b = (0,1)\}.$$

The Cayley graph Cayley(G,S) is the square grid in the Euclidean plane: The vertices are points with integer coordinates, two vertices are connected by an edge if and only if either their first or their second coordinates differ by ± 1. See Figure 4.4.

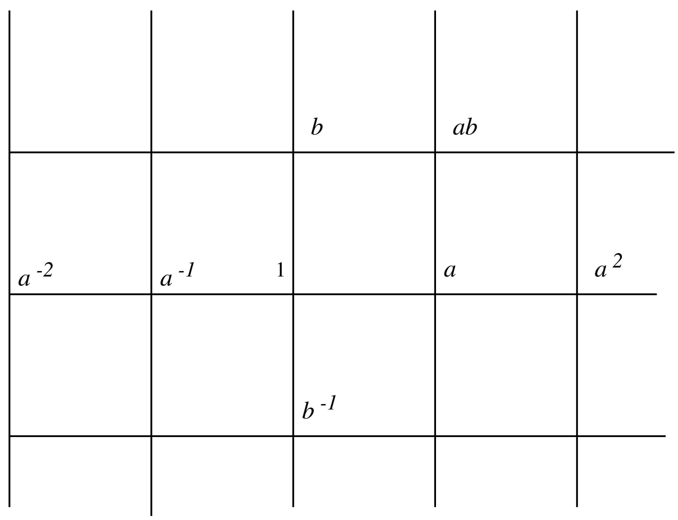


FIGURE 4.4. The Cayley graph of $\mathbb{Z}^2$.

The Cayley graph of $\mathbb{Z}^2$ with respect to the generating set$\{(1,0),(1,1)\}$ has the same set of vertices as the above, but the vertical lines are replaced by diagonal lines.

EXAMPLE 4.75. Let G be the free group on two generators a, b. Take $S = \{a, b\}$. The Cayley graph Cayley(G,S) is the 4-valent tree (there are four edges incident to each vertex). See Figure 4.5.

EXERCISE 4.76. 1. Show that every simplicial tree is contractible and, hence, simply-connected.

2. Show that, conversely, every simply-connected graph is a simplicial tree. (Hint: Verify that if a connected graph Γ is not a tree then $H_1(\Gamma) \neq 0$.)

THEOREM 4.77. *The fundamental group of every connected graph Γ is free.*

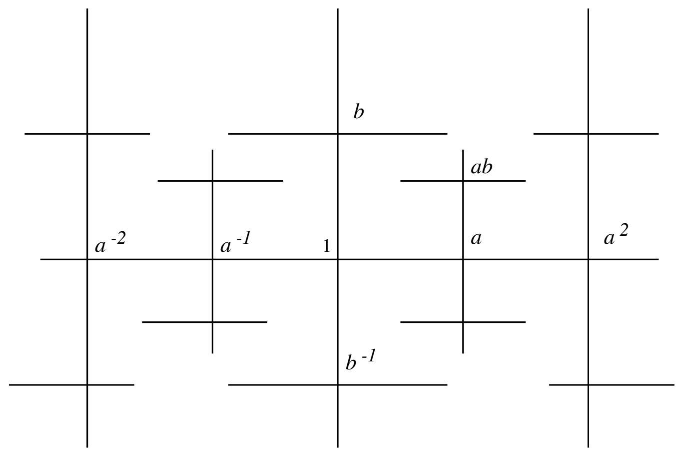


FIGURE 4.5. The Cayley graph of the free group F_2.

PROOF. By the Axiom of Choice (Zorn Lemma), Γ contains a maximal subtree $\Lambda \subset \Gamma$. Let Γ' denote the subdivision of Γ where very edge e in $\mathcal{E} = E(\Gamma) \setminus E(\Lambda)$ is subdivided in 3 sub-edges. For every such edge e let e' denote the middle 3rd. Now, add to Λ all the edges in $E(\Gamma')$ which are not of the form e' ($e \in \mathcal{E}$), and the vertices of such edges, of course, and let T' denote the resulting tree. Thus, we obtain a covering of Γ' by the simplicial tree T' and the subgraph $\Gamma_{\mathcal{E}}$ consisting of the pairwise disjoint edges e' ($e \in \mathcal{E}$), and the incident vertices. To this covering we can now apply Seifert— van Kampen Theorem and conclude (in view of the fact that T' is simply-connected) that $G = \pi_1(\Gamma)$ is free, with the free generators indexed by the set $\mathcal{E}$. □

COROLLARY 4.78. *1. Every free group $F(X)$ is the fundamental group of the bouquet (wedge) B of $|X|$ circles. 2. The universal cover of B is a tree T, which is isomorphic to the Cayley graph of $F(X)$ with respect to the generating set X.*

PROOF. 1. By Theorem 4.77, $G = \pi_1(B)$ is free; furthermore, the proof also shows that the generating set of G is identified with the set of edges of B. We now orient every edge of B using this identification. 2. The universal cover T of B is a simply-connected graph, hence, a tree. We lift the orientation of edges of B to orientation of edges of T. The group $F(X) = \pi_1(B)$ acts on T by covering transformations, hence, the action on the vertex $V(T)$ set of T is simply-transitive. Therefore, we obtain and identification of $V(T)$ with G. Let v be a vertex of T. By construction and the standard identification of $\pi_1(B)$ with covering transformations of T, every oriented edge e of B lifts to an oriented edge $\tilde{e}$ of T of the form $[v, w]$. Conversely, every oriented edge $[v, w]$ of T projects to an oriented edge of B. Thus, we labeled all the oriented edges of T with generators of $F(X)$. Again, by the covering theory, if an oriented edge $[u, w]$ of T is labeled with a generator $x \in F(X)$, then x sends u to w. Thus, T is isomorphic to the Cayley graph of $F(X)$. □

COROLLARY 4.79. *A group G is free if and only if it can act freely by automorphisms on a simplicial tree T.*

Proof. By the covering theory, $G \cong \pi_1(\Gamma)$ where $\Gamma = T/G$. Now, by Theorem 4.77, $G = \pi_1(\Gamma)$ is free. See **[Ser80]** for another proof and the more general discussion of group actions on trees. □

The concept of a simplicial tree generalizes to the one of a *real tree* (see Definition 2.59). There are non-free groups acting isometrically and freely on real trees, e.g., surface groups and free abelian groups. Rips proved that every finitely generated group acting freely and isometrically on a real tree is a free product of surface groups and free abelian groups, see e.g. [**BF95, Kap01, CR13**].

As an immediate application of Corollary 4.79 we obtain:

Corollary 4.80 (Nielsen–Schreier). *Every subgroup H of a free group F is itself free.*

Proof. Realize the free group F as the fundamental group of a bouquet B of circles; the universal cover T of B is a simplicial tree. The subgroup $H \leqslant F$ also acts on T freely. Thus, H is free. □

Proposition 4.81. *The free group of rank 2 contains an isomorphic copy of F_m for every finite m and for $m = \aleph_0$. Moreover, for finite m, we can find a subgroup $F_m < F_2$ of finite index.*

Proof. Let x, y denote the free generators of the group F_2.

1. Define the epimorphism $\rho_m : F_2 \to \mathbb{Z}_m$ by sending x to 1 and y to 0. Then the kernel K_m of ρ_m has index m in F_2. Then K_m is a finitely generated free group F. In order to compute the rank of F, it is convenient to argue topologically. Let R be a finite graph with the (free) fundamental group $\pi_1(R)$. Then $\chi(R) = 1 - b_1(R) = 1 - \text{rank}\,(\pi_1(R))$. Let R_2 be such a graph for F_2, then $\chi(R_2) = 1 - 2 = -1$. Let $R \to R_2$ be the m-fold covering corresponding to the inclusion $K_m \hookrightarrow F_2$. Then $\chi(R) = m\chi(R_2) = -m$. Hence, $\text{rank}\,(K_m) = 1 - \chi(R) = 1 + m$. Thus, for every $n = 1 + m \geqslant 2$, we have a finite-index inclusion $F_n \hookrightarrow F_2$.

2. Let x, y be the two generators of F_2. Let S be the subset consisting of all elements of F_2 of the form $x_k := y^k x y^{-k}$, for all $k \in \mathbb{N}$. We claim that the subgroup $\langle S \rangle$ generated by S is isomorphic to the free group of rank $\aleph_0$.

Indeed, consider the set A_k of all reduced words with prefix $y^k x$. With the notation of §4.2, the transformation $L_{x_k} : F_2 \to F_2$ has the property that $L_{x_k}(A_j) \subset A_k$ for every $j \neq k$. Obviously, the sets A_k, $k \in \mathbb{N}$, are pairwise disjoint. This and Lemma 4.62, imply that $\{L_{x_k} : k \in \mathbb{N}\}$ generate a free subgroup in $\text{Bij}(F_2)$, hence so do $\{x_k : k \in \mathbb{N}\}$ in F_2. □

Exercise 4.82. Let G and H be finitely generated groups, with S and X respective finite generating sets. Consider the wreath product $G \wr H$ (see Definition 3.31), endowed with the finite generating set canonically associated to S and X described in Exercise 4.13. For every function $f : H \to G$ denote by $\text{supp}\, f$ the set of elements $h \in H$ such that $f(h) \neq \mathbf{1}_G$.

Let f and g be arbitrary functions from H to G with finite support, and h, k arbitrary elements in H. Prove that the word distance in $G \wr H$ from (f, h) to (g, k) with respect to the generating set mentioned above is

$$\text{dist}\,((f,h),(g,k)) = \sum_{x \in H} \text{dist}_S(f(x), g(x)) + \text{length}(\text{supp}(g^{-1}f); h, k)\,, \tag{4.4}$$

where

$$\text{length}(\text{supp}(g^{-1}f); h, k)$$

is the length of the shortest path in Cayley(H, X) starting in h, ending in k and whose image contains the set supp$(g^{-1}f)$.

Thus, we succeeded in assigning to every finitely generated group G a metric space Cayley(G, S). The problem, however, is that this assignment

$$G \to \text{Cayley}(G, S)$$

is far from canonical: Different generating sets could yield completely different Cayley graphs. For instance, the trivial group has the presentations:

$$\langle \quad | \quad \rangle, \quad \langle a|a \rangle, \quad \left\langle a, b|ab, ab^2 \right\rangle, \ldots,$$

which give rise to the non-isometric Cayley graphs:

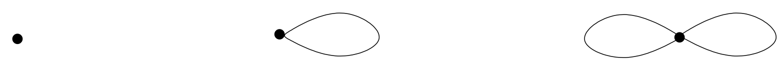


FIGURE 4.6. Cayley graphs of the trivial group.

The same applies to the infinite cyclic group:

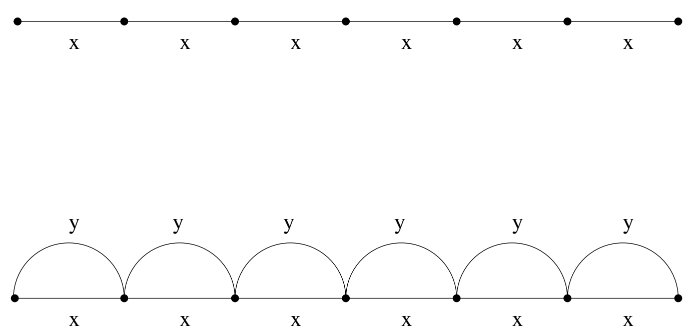


FIGURE 4.7. Cayley graphs of $\mathbb{Z} = \langle x| \rangle$ and $\mathbb{Z} = \left\langle x, y|xy^{-1} \right\rangle$.

In the above examples we did not follow the convention that $S = S^{-1}$.

Note, however, that all Cayley graphs of the trivial group have finite diameter; the same, of course, applies to all finite groups. The Cayley graphs of $\mathbb{Z}$ as above, although they are clearly non-isometric, are within finite distance from each other (when placed in the same Euclidean plane). Therefore, when seen from a (very) large distance (or by a person with a very poor vision), every Cayley graph of a finite group looks like a "fuzzy dot"; every Cayley graph of $\mathbb{Z}$ looks like a "fuzzy line," etc. Therefore, although non-isometric, they all "look alike".

EXERCISE 4.83. (1) Prove that if S and $\bar{S}$ are two finite generating sets of G, then the word metrics dist$_S$ and dist$_{\bar{S}}$ on G are bi-Lipschitz equivalent, i.e., there exists $L > 0$ such that

$$\frac{1}{L}\text{dist}_S(g, g') \leqslant \text{dist}_{\bar{S}}(g, g') \leqslant L\text{dist}_S(g, g')\,, \forall g, g' \in G\,. \tag{4.5}$$

Hint: Verify the inequality (4.5) first for $g' = 1_G$ and $g \in S$; then verify the inequality for arbitrary $g \in G$ and $g' = 1_G$. Lastly, verify the inequality for all g, g' using left-invariance of word-metrics.

(2) Prove that an isomorphism between two finitely generated groups is a bi-Lipschitz map when the two groups are endowed with word metrics.

Convention 4.84. From now on, unless otherwise stated, by a metric on a finitely generated group we mean a word metric coming from a finite generating set.

Exercise 4.85. Show that the Cayley graph of a finitely generated infinite group contains an isometric copy of $\mathbb{R}$, i.e., a bi-infinite geodesic. Hint: Apply Arzela-Ascoli theorem to a sequence of geodesic segments in the Cayley graph.

On the other hand, it is clear that no matter how poor one's vision is, the Cayley graphs of, say, $\{1\}$, $\mathbb{Z}$ and $\mathbb{Z}^2$ all look different: They appear to have different "dimension" (0, 1 and 2 respectively).

Telling apart the Cayley graph Cayley_1 of $\mathbb{Z}^2$ from the Cayley graph Cayley_2 of the Coxeter group

$$\Delta := \Delta(4,4,4) := \left\langle a, b, c | a^2, b^2, c^2, (ab)^4, (bc)^4, (ca)^4 \right\rangle$$

seems more difficult: They both "appear" 2-dimensional. However, by looking at the larger pieces of Cayley_1 and Cayley_2, the difference becomes more apparent: Within a given ball of radius R in Cayley_1, there seems to be less vertices than in Cayley_2. The former grows quadratically, while the latter grows exponentially fast as R goes to infinity.

Figure 4.8. Cayley graph of Δ.

The goal of the rest of the book is to make sense of this "fuzzy math".

In §5.1 we replace the notion of an *isometry* with the notion of a *quasi-isometry*, in order to capture what different Cayley graphs of the same group have in common.

Lemma 4.86. *A* finite index *subgroup of a finitely generated group is finitely generated.*

Proof. This lemma follows from Theorem 5.35 proven in the next chapter. We give here another proof, as the set of generators of the subgroup found here will be used in future applications.

Let G be a group and S a finite generating set of G, and let H be a finite index subgroup in G. Then

$$G = H \sqcup \bigsqcup_{i=1}^{k} H g_i$$

for some elements $g_i \in G$. Consider

$$R = \max_{1 \leqslant i \leqslant k} |g_i|_S \, .$$

Then $G = HB(1, R)$. We now prove that $X = H \cap B(1, 2R+1)$ is a generating set of H.

Let h be an arbitrary element in H and let $g_0 = 1, g_1, \ldots, g_n = h$ be the consecutive vertices on a geodesic in $\mathrm{Cayley}(G, S)$ joining 1 and h. In particular, this implies that $\mathrm{dist}_S(1, h) = n$.

For every $1 \leqslant i \leqslant n-1$ there exist $h_i \in H$ such that $\mathrm{dist}_S(g_i, h_i) \leqslant R$. Set $h_0 = 1$ and $h_n = h$. Then $\mathrm{dist}_S(h_i, h_{i+1}) \leqslant 2R + 1$, hence $h_{i+1} = h_i x_i$ for some $x_i \in X$, for every $0 \leqslant i \leqslant n-1$. It follows that $h = h_n = x_1 x_2 \cdots x_n$, whence X generates H and $|h|_X \leqslant |h|_S = n$. □

Other geometric models of groups. Let G be a finitely generated group. Then G is the quotient group of a free group F_n. Therefore, if Y is any connected space whose fundamental group surjects to F_n, we obtain the homomorphism

$$\phi : \pi_1(Y) \to G.$$

Therefore, if Y is, say, locally simply-connected, we obtain the regular covering map

$$p : X \to Y$$

associated to the kernel of ϕ, such that the group of covering transformations of p is isomorphic to G. The group G acts properly discontinuously on X. We will be primarily interested in two cases:

1. Y is a compact CW-complex.
2. Y is a compact Riemannian manifold.

The structure of a CW-complex/Riemannian manifold, lifts from Y to X and the action of G preserves this structure: The action of G is cellular in the former case and is isometric in the latter case. If X is a simplicial complex and the action $G \curvearrowright X$ is simplicial, the *standard* metric dist on X is G-invariant and, hence, (X, dist) is a *simplicial geometric model* for the group G. If X is a Riemannian manifold, taking dist to be the Riemannian distance function we obtain a *Riemannian geometric model* for the group G.

In order to construct a CW-complex Y we can take, for instance, Y equal to the bouquet of n circles; the space X is, then, a Cayley graph of G. In order to get a Riemannian manifold Y, we can take Y to be a compact Riemannian surface of genus n, the epimorphism $\pi_1(Y) \to F_n$ is then given by

$$\phi : \left\langle a_1, b_1, \ldots, a_n, b_n | \prod_{i=1}^{n} [a_i, b_i] \right\rangle \to \langle a_1, \ldots, a_n \rangle,$$

$$\phi(a_i) = a_i, \quad \phi(b_i) = 1, i = 1, \ldots, n.$$

In the case when G is finitely presented, one can do a bit better: Each finite presentation of G yields a finite *presentation complex* Y of G (see Definition 4.90), which is a finite CW-complex whose fundamental group is isomorphic to G. Hence, the universal cover X of Y is a simply-connected CW-complex and we obtain a cellular, free, properly discontinuous and cocompact action $G \curvearrowright X$. Since every compact CW-complex is homotopy-equivalent to a compact simplicial complex, we can also finite a simplicial complex X with the above properties.

Similarly, there exists a smooth closed m-manifold M ($m \geqslant 4$) whose fundamental group is isomorphic to G (see e.g. [**Hat02**]). Then we equip $Y = M$ with a Riemannian metric; lifting this metric to the universal cover $X \to Y$, we obtain a simply-connected complete Riemannian manifold X and a free, properly discontinuous, isometric and cocompact action $G \curvearrowright X$.

Working with geometric models (simplicial or Riemannian) of groups G is a major theme and a key technical tool of Geometric Group Theory. We will use this tool throughout this book. As one example, we will use both simplicial and Riemannian geometric models in Chapters 18 and 19 in order to prove group-theoretic theorems by Stallings and Dunwoody.

On the other hand, when replacing a group with its geometric model, we are faced with the inevitable:

QUESTION 4.87. What do all these geometric models have in common?

We will discuss this question in detail in the next chapter.

4.10. Volumes of maps of cell complexes and Van Kampen diagrams

The goal of this section is to describe several notions of volumes of maps and to relate them to each other and to the word reductions in finitely presented groups. It turns out that most of these notions are equivalent, but, in few cases, there subtle differences.

Recall that in §2.4 we defined volumes of maps between Riemannian manifolds. More generally, the same definition of volume of a map applies in the context of Lipschitz maps of simplicial complexes X equipped with the *standard metric*, see §2.8. Namely, in order to compute the n-volume of a map f, first compute volumes of restrictions $f\big|_{\Delta_i}$, of f to each n-dimensional simplex in X and then add up the results. Instead of this, intuitive and geometrically appealing, notion of volume, we will use the more combinatorial concepts.

4.10.1. Simplicial and combinatorial volumes of maps. Suppose that X, Y are simplicial/cell complexes, X is a finite complex and $f: X \to Y$ is a simplicial/cellular map. Then the *n-dimensional simplicial/cellular volume* $Vol_n(f)$ of f is just the number of n-dimensional simplices/cells in the domain X. The same definition of volume applies when X, Y are cell-complexes. Note that this, somewhat strange, concept, is independent of the map f but is, nevertheless, quite useful, see §5.10. The simplicial/cellular 1-volume is called *length* and the 2-volume is called *area*. They are denoted *length* and *Area* respectively.

The more intuitive concept is the one of the *combinatorial volume* of the map f. Assume that X, Y are simplicial complexes equipped with *standard metrics* and $f: X \to Y$ is a simplicial map. Define the *combinatorial volume* of f as

$$cVol_n(f) := \sum_{\Delta} \frac{1}{\varpi_n} Vol_n(f(\Delta)),$$

where the sum is taken over all n-simplices in X and ϖ_n is the volume of the standard Euclidean n-dimensional simplex. In other words, $cVol_n$ counts the number of n-simplices in X which are *not* mapped by f to simplices of lower dimension.

DEFINITION 4.88. Let X, Y be n-dimensional almost regular cell complexes. A cellular map $f : X \to Y$ is said to be *almost regular* if for every n-cell σ in X either:

(a) f *collapses* σ, i.e., $f(\sigma) \subset Y^{(n-1)}$, or

(b) f maps the interior of σ homeomorphically to the interior of an n-cell in Y.

An almost regular map is *regular* if only (b) occurs.

For instance, a simplicial map of simplicial complexes is almost regular, while a simplicial topological embedding of simplicial complexes is regular.

We define the *combinatorial* n-volume $cVol_n(f)$ of an almost regular map f to be the total number of n-cells in X which are not collapsed by f. Thus, this definition agrees with the notion of the combinatorial volume in the case of simplicial maps.

4.10.2. Topological interpretation of finite-presentability.

LEMMA 4.89. *A group G is isomorphic to the fundamental group of a finite cell complex Y if and only if G is finitely presented.*

PROOF. 1. Suppose that G has a finite presentation

$$\langle S|R\rangle = \langle x_1, \ldots, x_n | r_1, \ldots, r_m\rangle .$$

We construct a finite 2-dimensional cell-complex Y, as follows. The complex Y has unique vertex v. The 1-skeleton of Y is the n-rose, the bouquet of n circles $\gamma_1, \ldots, \gamma_n$ with the common point v, the circles are labeled $x_1, \ldots, x_n$. Observe that the free group $F(S)$ is isomorphic to $\pi_1(Y^{(1)}, v)$ where the isomorphism sends each x_i to the circle in $Y^{(1)}$ with the label x_i. Thus, every word w in X^* determines a based loop L_w in $Y^{(1)}$ with the base-point v. In particular, each relator r_i determines a loop $\alpha_i := L_{r_i}$. We then attach 2-cells $\sigma_1, \ldots, \sigma_m$ to $Y^{(1)}$ using the maps $\alpha_i : \mathbb{S}^1 \to Y^{(1)}$ as the attaching maps. Let Y be the resulting cell complex. It is clear from the construction that the complex Y is *almost regular*.

We obtain a homomorphism $\phi : F(S) \to \pi_1(Y^{(1)}) \to \pi_1(Y)$. Since each r_i lies in the kernel of this homomorphism, ϕ descends to a homomorphism $\psi : G \to \pi_1(Y)$. It follows from the Seifert–van Kampen theorem (see **[Hat02]** or **[Mas91]**) that ψ is an isomorphism.

2. Suppose that Y is a finite complex with $G \cong \pi_1(Y)$. Pick a maximal subtree $T \subset Y^{(1)}$ and let X be the complex obtained by contracting T to a point. Since T is contractible, the resulting map $Y \to X$ (contracting T to a point $v \in X^{(0)}$) is a homotopy-equivalence. The 1-skeleton of X is an n-rose with the edges $\gamma_1, \ldots, \gamma_n$ which we will label $x_1, \ldots, x_n$. It now again follows from the Seifert–van Kampen theorem that X defines a finite presentation of G: The generators x_i are the loops γ_i and the relators are the attaching maps $\mathbb{S}^1 \to X^{(1)}$ of the 2-cells of X. ☐

DEFINITION 4.90. The 2-dimensional complex Y constructed in the first part of the above proof is called the *presentation complex* of the presentation

$$\langle x_1, \ldots, x_n | r_1, \ldots, r_m\rangle .$$

4.10.3. Presentations of central coextensions. In this section we illustrate the concepts introduced earlier in the case of central coextensions. Let $f : F = F(S) \to G$ be an epimorphism. Consider a central coextension

$$0 \to A \to \tilde{G} = \tilde{\mathcal{G}}_\omega \xrightarrow{p} G \to 1$$

associated with a cohomology class $\omega \in H^2(G, A)$. Our goal is to describe a presentation of $\tilde{G}$ in terms of the presentation of G given by f. In §3.9.6, we discussed a *pull-back construction* for central coextensions. Applying this to the homomorphism f, we obtain a central coextension

$$0 \to A \to \tilde{\mathbb{F}} \xrightarrow{q} F \to 1$$

and a commutative diagram of homomorphisms

$$\begin{array}{ccc} \tilde{F} & \xrightarrow{\tilde{f}} & \tilde{G} \\ \Big\downarrow q & & \Big\downarrow p \\ F & \xrightarrow{\tilde{f}} & G \end{array}.$$

The homomorphism $\tilde{f}$ is surjective (Exercise 3.127, Part 1). Since F is free, its central coextension splits and there exists a homomorphism $s : F \to \tilde{F}$ right-inverse to q. Following our discussion in the end of §3.9.6, we pick a set-theoretic section s_1 of p, lift it to a set-theoretic section s_1 of q and observe that $\mathrm{Ker}(\tilde{f}) = s_1(K) \cong K$, where $K = \mathrm{Ker}(f)$, the normal closure of a set $R = \{R_i : i \in I\}$ of defining relators of G. Using the section s we define an isomorphism $\tilde{F} \cong F \times A$. With this identification, the restriction of the section s_1 to K is a homomorphism $\varphi : K \to A$ (invariant under conjugation by elements of F). By abusing the notation, we denote φ by φ_ω, even though, there are some choices involved in constructing φ from the central coextension. Then the group $\tilde{G}$ is isomorphic to the quotient of $F \times A$ by the normal closure of the subset

$$\{R_i \varphi(R_i)^{-1} : i \in I\}.$$

In order to describe the corresponding presentation of $\tilde{G}$, we fix a presentation

$$\langle T | Q \rangle$$

of the group A, $T = \{t_j : j \in J\}$, $Q = \{Q_\ell : \ell in L\}$. We then obtain the presentation

$$\langle S \sqcup T | Q, [x, t] = 1, x \in S, t \in T \rangle$$

of the group $F \times A$. Lastly, the presentation of the group $\tilde{G}$ is:

$$\langle S \sqcup T | Q, [x, t] = 1, x \in S, t \in T, R_i = \varphi(R_i), i \in I \rangle .$$

Example 4.91. Let G be the fundamental group of closed oriented surface Y of genus $n \geqslant 1$ with the standard presentation

$$\langle a_1, b_1, \ldots, a_p, b_p | \quad R = [a_1, b_1] \cdots [a_n, b_n] \rangle .$$

Since $Y = K(G, 1)$, $H^2(G) \cong H^2(Y) \cong \mathbb{Z}$. The space of F_{2n}-invariant homomorphisms

$$\langle \langle R \rangle \rangle \to A = \mathbb{Z}$$

is isomorphic to $\mathbb{Z}$ (since every such homomorphism is determined by its restriction to R). Thus, central coextensions $\tilde{G}_\omega$ of G are indexes by integers $e \in \mathbb{Z}$:

$$\varphi : R \mapsto e \in \mathbb{Z}.$$

The group $\tilde{G}_\omega$ has the presentation

$$\langle a_1, b_1, \ldots, a_n, b_n, t | \quad [a_1, b_1] \cdots [a_n, b_n] = t, [a_i, t] = 1, [b_i, t] = 1, i = 1, \ldots, n \rangle .$$

We next give (without a proof) two topological interpretations of the group $\tilde{G}$ and its presentation. Suppose that the presentation complex Y of $\langle S | R \rangle$ is aspherical, i.e. $\pi_2(Y) = 0$. Then $H^2(G, A) \cong H^2(Y, A)$. We will use cellular cohomology in order to compute $H^2(Y, A)$. Since Y is 2-dimensional, $Z^2(Y, A) = C^2(Y, A)$. The generators of $C_2(Y)$ are 2-cells e_i, which are labelled by the relators

$R_i \in R$. Let $\mathcal{R}$ denote the free abelian group with the basis R. Then we have the isomorphism

$$\Psi : Z^2(Y, A) \cong Hom(\mathcal{R}, A),$$

which sends a cocycle $c : e_i \mapsto c(e_i) \in A$ to the homomorphism $\psi_c : R_i \mapsto c(e_i)$. Altering c by a coboundary, results in a new element of $Hom(\mathcal{R}, A)$. In other words, only the coset $\psi_c\Psi(B^2(Y, A))$ is determined by the cohomology class $[c] \in H^2(Y, A)$.

The class $[c]$ maps to a cohomology class $\omega \in H^2(G, A)$ under the isomorphism $H^2(Y, A) \cong H^2(G, A)$. The class ω, as we say before, determines (subject to some ambiguity) a homomorphism φ_ω from the normal closure of R in F into A. This homomorphism is determined by its restriction to R (since φ_ω is F-invaraint). Thus, both $[c] \in H^2(Y, A)$ and ω determine (equivalence classes) of homomorphisms $\psi_c, \phi_\omega : \mathcal{R} \to A$. One can verify that *these equivalence classes are the same.* Therefore, we obtain a somewhat more concrete description of the presentation of the group $\tilde{G}$: In addition to the relators of $F \times A$, we have the relators

$$R_i = \psi_c(R_i), \quad i \in I.$$

This is the first topological interpretation of the presentation of $\tilde{G}$.

The second topological interpretation requires *complex line bundles* $\xi : L \to Y$. Such line bundles are parameterized by the elements of $H^2(Y) = H^2(Y, \mathbb{Z})$. The cohomology class defining L is called the *1st Chern class* $c_1(\xi)$ of the bundle ξ. We refer the reader to [**Che95**, pp. 33–34] for the details. Given a line bundle $L \to Y$, we define the space $L_o \subset L$ by removing the image of the zero section from L (then L_o is the total space of the $\mathbb{C}^*$-bundle associated with L). The fundamental group of $\mathbb{C}^*$ is infinite cyclic and $\pi_1(Y)$ acts trivially on this group. Therefore, taking into the account the long exact sequence of homotopy groups of the fibration

$$\mathbb{C}^* \to L_o \to Y,$$

we obtain that $\pi_1(L_o)$ is isomorphic to a certain central coextension $\tilde{G}$ of G:

$$0 \to \mathbb{Z} \to \tilde{G} \to G \to 1.$$

Then: *The cohomology class ω defining this central coextension maps to the 1st Chern class $c_1(\xi)$ under the isomorphism $H_1(G) \to H_1(Y)$.*

4.10.4. Dehn function and van Kampen diagrams. One of the oldest algorithmic problems in group theory is the *word problem.* This problem is *largely* controlled by the *Dehn function* of the group, which depends on the group presentation $\langle S|R\rangle$. In this section we define the Dehn function, van Kampen diagrams of finite presentations, and relate the latter to the word problem. We refer the reader to [**LS77**] for the more thorough treatment of this topic. The reader familiar with the treatment of van Kampen diagrams in [**LS77**] will notice that our definitions of diskoids and van Kampen diagrams are more general.

Suppose that w is a word in S, representing a trivial element of the group G with the presentation $\langle S|R\rangle$. How can we convince ourselves that, indeed, $w \equiv_G 1$? If R were empty, we could eliminate all the reductions in w, which will result in an empty word. In the case of nonempty R, we can try the same thing, namely, the reduction of w in $F = F(S)$. If the reduction results in an empty word, we are done; hence, we will work, in what follows, with nonempty reduced words. Thus, we will identify each w with a nontrivial element of the free group F. Any "proof" that such w is trivial in G would amount to finding a product decomposition of $w \in F$ of the form

$$w = \prod_{i=1}^{k} u_i r_i^{\pm 1} u_i^{-1}, \tag{4.6}$$

where r_i's are elements of R. Of course, it is to our advantage, to use as few defining relators r_i as we can, in order to get as short "proof" as possible. This leads us to

Definition 4.92. The *algebraic area* of the (reduced) word w, such that $w \equiv_G 1$, is defined as the least number k of relators r_i used to describe w as a product of conjugates of defining relators and their inverses. The algebraic area of w is denoted by $A(w)$.

The significance of this notion of area is that it captures the complexity of the word problem for the presentation $\langle S|R\rangle$ of the group G. In order to estimate "hardness" of the word problem, we then search for the words w of the largest area: The most reasonable way to do so, by analogy with the norms of linear operators, is to restrict to w's of bounded word-length. This leads us to

Definition 4.93 (Dehn function). The *Dehn function of the group G* (with respect to the finite presentation $\langle S|R\rangle$) is defined as

$$Dehn(\ell) := \max\{A(w) : |w| \leqslant \ell\}$$

where w's are elements in S^* representing trivial words in G.

Exercise 4.94. Let $\langle S, |R\rangle, \langle S'|R'\rangle$ be finite presentation of the same group G. Show that the resulting Dehn functions $Dehn, Dehn'$ are approximately equivalent in the sense of the Definition 1.3.

In view of this property, we will frequently use the notation $Dehn_G$ for the Dehn function of G (with respect to some unspecified finite presentation of G).

Our next task is to describe a *geometric interpretation* of areas of words and Dehn functions. The classical tool for this task is *van Kampen diagrams*, which give topological interpretation to the product decompositions (4.6).

Van Kampen diagrams. Suppose that Y is the presentation complex of the presentation $\langle S|R\rangle$. Each nonnempty reduced word w represents a certain closed edge-path c_w in $Y^{(1)}$ (here and in what follows, the base-point is the sole vertex y of Y): Each letter s in w corresponds to an oriented edge in $Y^{(1)}$ representing the corresponding generator of G or its inverse.

We will think of c_w as a regular map $\mathbb{S}^1 \to Y^{(1)}$, where $\mathbb{S}^1$ is the circle equipped with a certain fixed cell-complex structure as well as a base-vertex.

Example 4.95. If $w \in S \cup S^{-1}$, then the almost regular cell complex structure on $\mathbb{S}^1$ will consist of a single vertex and a single edge. The map c_w is a topological embedding.

As the map c_w is null-homotopic, one can extend the map c_w to a cellular map $\mathbb{D}^2 \to Y$. We will see below that one can find an extension of c_w which is almost regular; we then will define the *combinatorial area* of c_w as the least combinatorial area of the resulting extension. The extension will collapse some 2-cells in $\mathbb{D}^2$ into the 1-skeleton of Y: These cells contribute nothing to the combinatorial area and we would like to get rid of them. *Van Kampen diagrams* are a convenient (and traditional) way to eliminate these dimension reductions.

Definition 4.96. We say that a contractible finite planar almost regular cell complex $K \subset \mathbb{R}^2$ is a *diskoid* (a *tree of disks* or a *tree-graded disk*) if every edge of K is contained in the boundary of K in $\mathbb{R}^2$.

In other words, K is obtained from a finite simplicial tree by replacing some vertices with (cellulated) 2-disks, which is why we think of K as a "tree of disks". To simplify the picture, the reader can (at first) think of K as a single disk in $\mathbb{R}^2$ rather than a tree of disks. In what follows, we will assume that K is *nontrivial*: $K \neq \emptyset$ and K does not consist of a single vertex. (The case when K is a single vertex would correspond to the case of the empty word w.) Note that the boundary ∂K of K in $\mathbb{R}^2$ is also an almost regular cell complex (a planar graph). However, the graph ∂K may have some valence 1 vertices, the *leaves* of ∂K: These leaves will not exist in the case of van Kampen diagrams of reduced words.

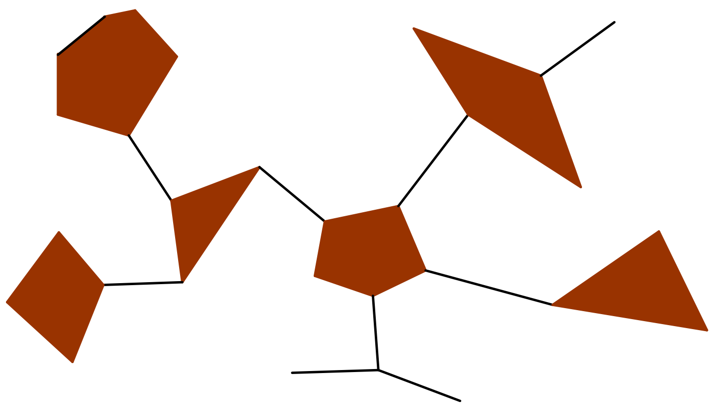


Figure 4.9. Example of a diskoid.

The complex K admits a *canonical enlargement* to a planar almost regular cell complex $\widehat{K}$ homeomorphic to the disk $\mathbb{D}^2$: The complement of K in $\widehat{K}$ is homeomorphic to the annulus

$$(0, 1] \times \mathbb{S}^1.$$

The 2-cells in $\widehat{K} \setminus \mathrm{int}(K)$, are rectangles. If e is an edge of ∂K which belongs to the closure of the interior of K in $\mathbb{R}^2$, then e is the boundary edge of exactly one such rectangle, otherwise, e is the boundary edge of exactly two rectangles in $\widehat{K} \setminus \mathrm{int}(K)$. Furthermore, every rectangular boundary face σ shares exactly one edge, called e_σ, with K. We refer to these 2-cells σ as the *boundary faces* of $\widehat{K}$. Thus, the number of boundary faces of $\widehat{K}$ is at most twice the number of edges in W. We have the canonical retraction

$$\kappa : \widehat{K} \to K$$

sending each boundary face σ to the edge e_σ. See Figure 4.10.

Figure 4.10. Collapsing map κ.

Restricting κ to the boundary circle of the disk $\widehat{K}$, we obtain a regular cellular map $b : \mathbb{S}^1 \to \partial K$ tracing the boundary of K according to the orientation induced on the boundary arcs of K from the Euclidean plane. Here $\mathbb{S}^1$ is given the structure of a regular cell complex C coming from $\widehat{K}$. We will refer to b as the *boundary map* of K. For each boundary edge e of K not contained in the closure of the interior of K in $\mathbb{R}^2$, the preimage $b^{-1}(e)$ contains exactly two edges in C.

We now describe a certain class of maps from diskoids to almost regular 2-dimensional cell complexes Y.

DEFINITION 4.97. A regular cellular map $h : K \to Y$ from a diskoid to an almost regular 2-dimensional cell complex Y is called a *van Kampen diagram in* Y. Suppose that $w \in S^*$ represents the identity in G and $c_w : \mathbb{S}^1 \to Y$ is the associated loop in the presentation complex Y of $\langle S|R\rangle$. If the composition $\partial h := h \circ b$ of h with the boundary map of K equals c_w, we will say that h is a *van Kampen diagram* of the word w.

It will be sometimes convenient to consider van Kampen diagrams not in presentation complexes but in their universal covers. This, of course, will make no difference as far as combinatorial areas and lengths are concerned.

It is customary to describe a van Kampen diagram by labeling *oriented edges* $\bar{e}$ of K by the elements of $S \cup S^{-1}$ which correspond to the edges $h(\bar{e})$ of Y. (Recall that some boundary edges of the diskoid K have two opposite orientations defined by the boundary map b: They will be labelled by a generator and its inverse respectively.) For instance, for the standard presentation

$$\left\langle a, b|aba^{-1}b^{-1}\right\rangle,$$

of the group $\mathbb{Z}^2$, a van Kampen diagram of the relator $[a, b]$ is described in the Figure 4.11.

A van Kampen diagram of the relator $[a, b]^2$ for the same presentation is described in the Figure 4.12.

EXERCISE 4.98. Note that each van Kampen diagram $h : K \to Y$ extends to the canonical enlargement of K:

$$\widehat{h} = h \circ \kappa : \widehat{K} \to Y.$$

Then

$$Area(h) = cArea(h) = cArea(\widehat{h}).$$

LEMMA 4.99 (Van Kampen lemma). *1. For every word w in the alphabet $S \cup S^{-1}$, representing the identity element 1_G, there exists a van Kampen diagram $h : K \to Y$ such that the maps ∂h and $c = c_w$ are homotopic as maps $\mathbb{S}^1 \to Y^{(1)}$, rel. the base-vertex in $\mathbb{S}^1$. Furthermore, $Area(h)$ equals $A(w)$.*

2. If the word w is reduced, then there exists a van Kampen diagram h in Part 1 such that $\partial h = c_w$.

PROOF. 1. According to the product decomposition (4.6) of $w \in F(S)$, the circle $\mathbb{S}^1$ is subdivided into cyclically ordered and oriented cellular subarcs

$$\alpha_1^+ \cup \beta_1 \cup \alpha_1^- \cup ... \cup \alpha_k^+ \cup \beta_k \cup \alpha_k^-,$$

so that:

(1) The path $c\big|_{\alpha_i^+}$ represents the word u_i.

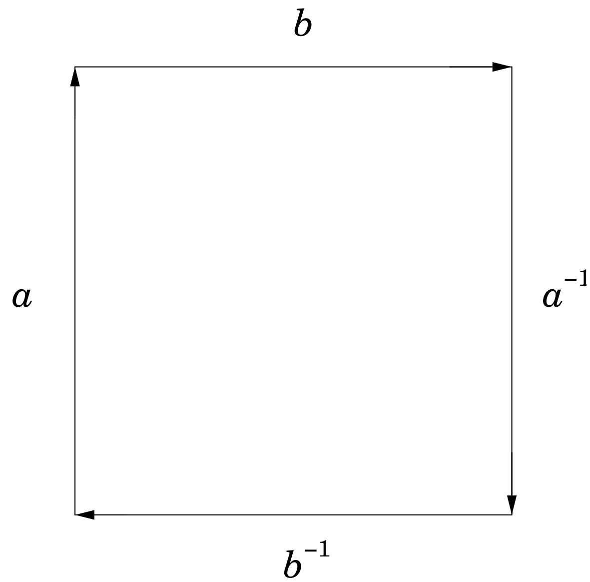


FIGURE 4.11. A van Kampen diagram of the commutator $[a, b]$.

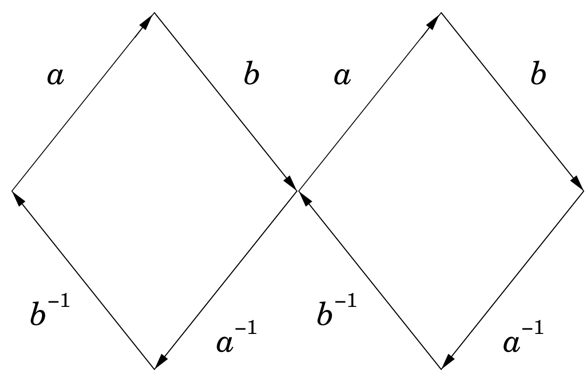


FIGURE 4.12. A van Kampen diagram of the relator $[a, b]^2$.

(2) The path $c|_{\alpha_i^-}$ represents the word u_i^{-1}.
(3) The path $c|_{\beta_i}$ represents the word $r_i^{\pm 1}$.

The orientation on the arcs $\alpha_i^{\pm}, \beta_i$ (induced from the standard orientation of unit circle), defines for each of these arcs the *head* and the *tail* vertex.

We then connect (some of) the vertices of C by chords in $\mathbb{D}^2$ as follows:

For each i, we connect the tail of α_i^+ to the head of α_i^- by the chord ϵ_i^+ and the head of α_i^+ to the tail of α_i^- by the chord ϵ_i^-.

The chords $\epsilon_i^{\pm}, \epsilon_j^{\pm}$ may cross only at the boundary circle $\mathbb{S}^1$. See Figure 4.13. The chords $\epsilon_i^{\pm}$, together with the original cell-complex structure C on $\mathbb{S}^1$, define a regular cell complex structure $\tilde{K}$ on $\mathbb{D}^2$, where every vertex is in $\mathbb{S}^1$. There are three types of 2-cells in $\tilde{K}$:

1 Cells A_i bounded by the "bigons" $\beta_i \cup \epsilon_i^-$.
2 Cells B_i bounded by "rectangles" $\alpha_i^+ \cup \epsilon_i^+ \cup \alpha_i^- \cup \epsilon_i^-$.
3 The rest, not having any edges in $\mathbb{S}^1$.

Note that in type (2) we allow the degenerate case when $\alpha_i^{\pm}$ is a single vertex: Then the corresponding "rectangle" degenerates to a triangle. It can even become a bigon in case when the word u_i is empty. Similarly, there will be one case when a "bigon" is actually a monogon: $w = r_1$.

We now collapse each type (3) cell to a vertex and collapse each type (2) cell to an edge e_i (so that each $\alpha_i^{\pm}$ maps homeomorphically onto this edge while the chords $\epsilon_i^{\pm}$ map to the end-points of e_i). Note that $\alpha_i^{\pm}$, with their orientation inherited from $\mathbb{S}^1$, define two opposite orientations on e_i. The quotient complex K will be our diskoid. We also obtain the quotient *collapsing* map $\kappa : \tilde{K} \to K$. Because "rectangles" can degenerate to triangles, or "bigons", the complex K is merely almost regular, not regular.

We define a map $h : K^{(1)} \to Y$ such that

$$h \circ \kappa\big|_{\alpha_i^{\pm}} = c_{u_i^{\pm 1}}$$

while

$$h \circ \kappa\big|_{\beta_i} = c_{r_i}.$$

Lastly, we extend h to the 2-cells $\kappa(A_i)$ in K: $h : \kappa(A_i) \to Y$ are the standard parameterizations of the 2-cells in Y corresponding to the defining relators r_i.

By the construction, h is a van Kampen diagram of w: The maps $h \circ \kappa$ and c_w are homotopic as based loops $\mathbb{S}^1 \to Y^{(1)}$. However, as maps they need not be the same, as the product decomposition (4.6) need not be a reduced word. The equality

$$Area(h) = k$$

is immediate from the construction.

2. Suppose now that the word w is reduced. The boundary map ∂h reads off a word w' in S^*: The word w' is obtained by reading off the boundary labels defined via h. The word w', by the construction, represents the same element of $F(S)$ as w. If w' were also reduced, we would be done. In general, however, w' is not reduced and, hence, we can find two adjacent boundary edges e_1, e_2 in ∂K, whose labels are inverses of each other. We then glue the edges e_1, e_2 together. The result is a new diskoid K_1, where the projection of e_1, e_2 is no longer a boundary edge. The map h descends to a van Kampen diagram $h_1 : K_1 \to Y$. By repeating this procedure inductively we eliminate all boundary reductions and obtain a new van Kampen diagram $h' : K' \to Y$ with the required properties. □

FIGURE 4.13

Lemma 4.99 shows that the algebraic area of w does not exceed the least combinatorial area of van Kampen diagrams of w.

EXERCISE 4.100. i. Show that $A(w)$ equals the least combinatorial area of all almost regular maps $f : \mathbb{D}^2 \to Y$ extending the cellular map c_w, where $\mathbb{D}^2$ is given the structure of an almost regular cell complex. Hint: Convert maps f into product decompositions of w as in (4.6).

ii. Combine Part (i) with the canonical extension $\widehat{h}$ of van Kampen diagrams $h : K \to Y$, to conclude that $A(w)$ equals

$$\min_{h:K\to Y} Area(h),$$

where the minimum is taken over all Van Kampen diagrams of w in Y. In other words, the algebraic area of the words w (trivial in G) equals the combinatorial area of the loops c_w in Y (the combinatorial area defined via van Kampen diagrams of w or, equivalently, almost regular extensions $\mathbb{D}^2 \to Y$ of c_w).

We will return to Dehn functions in §4.13 after discussing residual finiteness of groups.

4.11. Residual finiteness

Even though studying infinite groups is our primary focus, questions in group theory can be, sometimes, reduced to questions about finite groups. *Residual finiteness* is the concept that (sometimes) allows such reduction.

Definition 4.101. A group G is said to be *residually finite* if

$$\bigcap_{i\in I} G_i = \{1\},$$

where $\{G_i : i \in I\}$ is the set of all finite-index subgroups in G.

Clearly, subgroups of residually finite groups are also residually finite. In contrast, if G is an infinite simple group, then G cannot be residually-finite.

Lemma 4.102. *A finitely generated group G is residually finite if and only if for every $g \in G \setminus \{1\}$, there exists a finite group Φ and a homomorphism $\varphi : G \to \Phi$, such that $\varphi(g) \neq 1$.*

Proof. Suppose that G is residually finite. Then, for every $g \in G \setminus \{1\}$ there exists a finite-index subgroup $G_i \leqslant G$ so that $g \notin G_i$. It follows that G contains a normal subgroup of finite index $N_i \lhd G$, such that $N_i \leqslant G_i$. Clearly, $g \notin N_i$ and $|G : N_i| < \infty$. Now, setting $\Phi := G/N_i$, we obtain the required homomorphism $\varphi : G \to \Phi$.

Conversely, suppose that for every $g \neq 1$ we have a homomorphism $\varphi_g : G \to \Phi_g$, where Φ_g is a finite group, so that $\varphi_g(g) \neq 1$. Setting $N_g := \mathrm{Ker}(\varphi_g)$, we get

$$\bigcap_{g\in G} N_g = \{1\}.$$

The above intersection, of course, contains the intersection of all finite index subgroups in G. □

Exercise 4.103. Direct products of residually finite groups are again residually finite.

Lemma 4.104. *If a group G contains a residually finite subgroup of finite index, then G itself is residually finite.*

Proof. Let $H \leqslant G$ be a finite index residually finite subgroup. The intersection of all finite index subgroups

$$\bigcap_{i\in I} H_i \tag{4.7}$$

of H is $\{1\}$. Since H has finite index in G and each $H_i \leqslant H$ as above has finite index in G, the intersection of all finite index subgroups of G is contained in (4.7) and, hence, is trivial. $\square$

LEMMA 4.105. *Each group G virtually isomorphic to $\mathbb{Z}$ is residually finite and contains an infinite cyclic subgroup of finite index.*

PROOF. In view of Lemma 4.104, it suffices to show that if G is a finite coextension of the infinite cyclic group C,

$$1 \to F \to G \xrightarrow{p} C \to 1$$

(where F is finite), then G is residually finite. The group C contains a finite index subgroup C' such that the image of C in $Out(F)$ is trivial, i.e., the coextension

$$1 \to F \to G' \xrightarrow{p} C' \to 1$$

is central; here $G' = p^{-1}(C')$. Since C' is free, the central coextension splits and, hence, $G' \cong F \times C'$. It follows that G' is residually finite, hence, G is residually finite as well. $\square$

EXAMPLE 4.106. The group $\Gamma = GL(n, \mathbb{Z})$ is residually finite. Indeed, we take subgroups $\Gamma(p) \leqslant \Gamma$, $\Gamma(p) = \mathrm{Ker}(\varphi_p)$, where $\varphi_p : \Gamma \to GL(n, \mathbb{Z}_p)$ is the reduction modulo p. If $g \in \Gamma$ is a nontrivial element, we consider its nonzero off-diagonal entry $g_{ij} \neq 0$. Then $g_{ij} \neq 0 \mod p$, whenever $p > |g_{ij}|$. Thus, $\varphi_p(g) \neq 1$ and Γ is residually finite.

COROLLARY 4.107. *The free group F_2 of rank 2 is residually finite. Every free group of (at most) countable rank is residually finite.*

PROOF. As we saw in the Example 4.63 the group F_2 embeds in $SL(2, \mathbb{Z})$. Furthermore, every free group of (at most) countable rank embeds in F_2 (see Proposition 4.81). Now, the assertion follows from the Example 4.106. $\square$

We note that there are other proofs of residual finiteness of finitely generated free groups: Combinatorial (see [**Hal49**]), topological (see [**Sta83**]) and geometric (see [**Sco78**]).

EXERCISE 4.108. For arbitrary cardinality r, the free group F_r of rank r is residually finite.

Less trivially,

<u>THEOREM</u> 4.109 (K. W. Gruenberg [**Gru57**]). *Free products of residually finite groups are again residually finite.*

The simple argument for $GL(n, \mathbb{Z})$ is a model for a proof of a harder theorem:

THEOREM 4.110 (A. I. Mal'cev [**Mal40**]). *Let Γ be a finitely generated subgroup of $GL(n, R)$, where R is a commutative ring with unity. Then Γ is residually finite.*

Mal'cev's theorem is complemented by the following result proven by A. Selberg and known as *Selberg's Lemma* [**Sel60**]:

THEOREM 4.111 (Selberg's Lemma). *Let Γ be a finitely generated subgroup of $GL(n, F)$, where F is a field of characteristic zero. Then Γ contains a torsion-free subgroup of finite index.*

Proofs of Mal'cev's and Selberg's theorems, will be given in the Appendix, written by Bogdan Nica, in the end of this chapter.

A group Γ which is isomorphic to a subgroup of $GL(n, F)$, where F is a field, is called a *matrix group* or a *linear group*.

PROBLEM 4.112. It is known that all (finitely generated) Coxeter groups are linear, see e.g. [**Bou02**]. Is the same true for all Artin groups, Shephard groups, generalized von Dyck groups? (Note that even linearity of Artin Braid groups was unknown prior to [**Big01**].) Is it at least true that all these groups are residually finite?

Mal'cev's theorem implies that infinite finitely generated matrix groups cannot be simple. On the other hand, $PSL(2, \mathbb{Q})$ is a simple countable matrix group.

PROBLEM 4.113. Are there infinite simple discrete subgroups $\Gamma < SL(n, \mathbb{R}), n \gg 3$?

Here discreteness of Γ means that it is discrete with the subspace topology. One can prove that infinite discrete subgroups of $SO(n, 1)$, and, more generally, isometry groups of *rank* 1 *symmetric spaces*, cannot be simple: Given an infinite discrete subgroup $\Gamma < SO(n, 1)$, which does not preserve a line in $\mathbb{R}^{n+1}$, one shows (using a ping-pong argument) that there exists an infinite order element $g \in \Gamma$, such that normal closure Λ of $\{g\}$ in Γ is a free subgroup of Γ. If $\Lambda = \Gamma$ then Γ is not simple (since nontrivial free groups are never simple); otherwise, Λ is a proper normal subgroup of Γ.

4.12. Hopfian and cohopfian properties

A group G is called *hopfian* if every epimorphism $G \to G$ is injective. Mal'cev prove in [**Mal40**] that every residually finite group is hopfian. On the other hand, many Baumslag-Solitar groups are not hopfian. Collins and Levin [**CL83**] gave a criterion for $BS(m, n)$ to be hopfian (for $|m| > 1, |n| > 1$): The numbers m and n should have the same set of prime divisors.

An example of a hopfian group with a nonhopfian subgroup of finite index is the Baumslag–Solitar group

$$BS(2, 4) = \langle a, b | ab^2a^{-1} = b^4 \rangle .$$

According to the criterion of Collins and Levin, this group is hopfian. Meskin in [**Mes72**] proved that $BS(2, 4)$ contains a nonhopfian subgroup of finite index.

We now turn to *cohopfian property*, which is *dual* to the hopfian property: Every injective endomorphism $f : G \to G$ is surjective. Of course, every finite group is cohopfian. Sela proved in [**Sel97b**] that every torsion-free 1-ended *hyperbolic group* is cohopfian. On the other hand, every free abelian group $\mathbb{Z}^n$ is not cohopfian: The endomorphism $g \mapsto g^k$, $k > 1$, is injective but not surjective. However, there are finitely generated cohopfian nilpotent groups [**Bel03**]. Dekimpe and Deré [**DD14**] recently found a complete criterion for virtually nilpotent groups to be cohopfian, in particular, they proved that in this class of groups, cohopfian property is invariant under virtual isomorphisms.

EXERCISE 4.114. Each free group $F = F(X)$ is not cohopfian, provided that X is nonempty, of course.

We now give an example of a cohopfian virtually free group. Let

$$G_1 = \left\langle a, b, c | abc = 1, a^2 = 1, b^3 = 1, c^3 = 1 \right\rangle$$

be the alternating group A_4. We leave it to the reader to check that the subgroup $C = \langle c \rangle$ of G_1 is *malnormal*: For each $g \in G_1$,

$$gCg^{-1} \cap C \neq \{1\} \iff g \in C.$$

(One way to verify this is to let A_4 act as a group of symmetries of the regular 3-dimensional tetrahedron.) Define the amalgam

$$G = G_1 \star_C G_2,$$

where G_2 is another copy of G_1 and let T be the associated Bass-Serre tree. We let $v_i \in V(T)$ be the vertex fixed by G_i, $i = 1, 2$, and let $e = [v_1, v_2] \in E(T)$ be the edge fixed by C. Malnormality of C in G_1 translates to the fact that C does not fix any edges of T besides e.

We claim that the group G is cohopfian. Suppose that $f : G \to G$ is an injective endomorphism. (In fact, it suffices to assume that the restrictions of f to G_1 and G_2 are injective and that $f(G_1) \neq f(G_2)$.) Since the groups G_1, G_2 are finite, their images $f(G_i)$ fix vertices in the tree T. After composing f with an automorphism of G, we can assume that $f(G_1) = G_1$ (i.e., $f(G_1)$ fixes v_1) and $f(C) = C$ (i.e., $f(C)$ fixes e). Since C fixes only the edge e of T, the group $f(G_2)$ has to fix the vertex v_2 of e and, hence, $f(G_2) = G_2$. Surjectivity of f follows.

On the other hand, being an amalgam of finite groups, the group G is commensurable to the free group F_2, see Theorem 4.52.

4.13. Algorithmic problems in the combinatorial group theory

Presentations $G = \langle S|R \rangle$ provide a 'compact' form for defining the group G. They were introduced by Max Dehn in the early 20-th century. The main problem of the combinatorial group theory is to derive algebraic information about G from its presentation. Below is the list of such problems whose origin lies in the work of Max Dehn in the early 20th century.

Word Problem. Let $G = \langle S|R \rangle$ be a finitely presented group. Construct a Turing machine (or prove its non-existence) that, given a word w in the generating set X as its input, would determine if w represents the trivial element of G, i.e., if

$$w \in \langle\langle R \rangle\rangle .$$

Conjugacy Problem. Let $G = \langle S|R \rangle$ be a finitely presented group. Construct a Turing machine (or prove its non-existence) that, given a pair of word v, w in the generating set X, would determine if v and w represent conjugate elements of G, i.e., if there exists $g \in G$ so that

$$[w] = g^{-1}[v]g.$$

To simplify the language, we will state such problems below as: Given a finite presentation of G, determine if two elements of G are conjugate.

Simultaneous Conjugacy Problem. Given n-tuples pair of words

$$(v_1, \ldots, v_n), \quad (w_1, \ldots, w_n)$$

in the generating set X and a (finite) presentation $G = \langle S|R \rangle$, determine if there exists $g \in G$ so that

$$[w_i] = g^{-1}[v_i]g, i = 1, \ldots, n.$$

Triviality Problem. Given a (finite) presentation $G = \langle S|R\rangle$ as an input, determine if G is trivial, i.e., equals $\{1\}$.

Isomorphism Problem. Given two (finite) presentations $G_i = \langle X_i|R_i\rangle, i = 1, 2$ as an input, determine if G_1 is isomorphic to G_2.

Embedding Problem. Given two (finite) presentations $G_i = \langle X_i|R_i\rangle, i = 1, 2$ as an input, determine if G_1 is isomorphic to a subgroup of G_2.

Membership Problem. Let G be a finitely presented group, $h_1, \ldots, h_k \in G$ and H, the subgroup of G generated by the elements h_i. Given an element $g \in G$, determine if g belongs to H.

Note that a group with solvable conjugacy or membership problem, also has solvable word problem. It was discovered in the 1950-s in the work of P. S. Novikov, W. Boone and M. O. Rabin [**Nov58, Boo57, Rab58**] that all of the above problems are *algorithmically unsolvable*. For instance, in the case of the word problem, given a finite presentation $G = \langle S|R\rangle$, there is no algorithm whose input would be a (reduced) word w and the output YES is $w \equiv_G 1$ and NO if not. A. A. Fridman [**Fri60**] proved that certain groups have solvable word problem and unsolvable conjugacy problem. We will later see examples of groups with solvable word and conjugacy problems but unsolvable membership problem (Corollary 9.149). Furthermore, there are examples [**BH05**] of finitely presented groups with solvable conjugacy problem but unsolvable simultaneous conjugacy problem for every $n \geqslant 2$.

Nevertheless, the main message of the geometric group theory is that *under various geometric assumptions on groups (and their subgroups), all of the above algorithmic problems are solvable.* Incidentally, the idea that geometry can help solving algorithmic problems also goes back to Max Dehn. Here are two simple examples of solvability of word problem:

PROPOSITION 4.115. *Free group F of finite rank has solvable word problem.*

PROOF. Given a word w in free generators x_i (and their inverses) of F we cancel recursively all possible pairs $x_i x_i^{-1}$, $x_i^{-1} x_i$ in w. Eventually, this results in a reduced word w'. If w' is nonempty, then w represents a nontrivial element of F, if w' is empty, then $w \equiv 1$ in F. □

PROPOSITION 4.116. *Every finitely presented residually finite group has solvable word problem.*

PROOF. First, note that if Φ is a finite group, then it has solvable word problem (using the multiplication table in Φ we can "compute" every product of generators as an element of Φ and decide if this element is trivial or not). Given a residually finite group G with finite presentation $\langle S|R\rangle$ we will run two Turing machines T_1, T_2 simultaneously:

The machine T_1 will look for homomorphism $\varphi : G \to S_n$, where S_n is the symmetric group on n letters ($n \in \mathbb{N}$): The machine will try to send generators $x_1, \ldots, x_m$ of G to elements of S_m and then check if the images of the relators in G under this map are trivial or not. For every such homomorphism, T_1 will check if $\varphi(g) = 1$ or not. If T_1 finds φ so that $\varphi(g) \neq 1$, then $g \in G$ is nontrivial and the process stops.

The machine T_2 will list all the elements of the kernel N of the quotient homomorphism $F_m \to G$: It will multiply conjugates of the relators $r_j \in R$ by products of the generators $x_i \in X$ (and their inverses) and transforms the product to a reduced word. Every element of N is such a product, of course. We first write $g \in G$ as a reduced word w in generators x_i and their inverses. If T_2 finds that w equals one of the elements of N, then it stops and concludes that $g \equiv_G 1$.

The point of residual finiteness is that, eventually, one of the machines stops and we determine whether g is trivial or not. $\square$

The Dehn function $Dehn(n)$ of the group G (equipped with the finite presentation $\langle S|R\rangle$) quantifies (to some extent) the difficulty of solving the word problem in G:

THEOREM 4.117 (S. Gersten, [**Ger93a**].). *G has solvable word problem if and only if its Dehn function is recursive.*

We note that G has solvable word problem if and only if its Dehn function $Dehn(\ell)$ is merely bounded above by a recursive function $r(\ell)$. Indeed, given such a bound, one applies the machine T_2 from the proof of Proposition 4.116 to the word w of length ℓ. The number of van Kampen diagrams $h : W \to Y$ with reduced ∂h and of area $\leqslant r(\ell)$, is also bounded above by a recursive function of ℓ. Hence, the algorithm terminates in a finite amount of time either representing w as a product of conjugates of defining relators or verifying that such representation does not exist.

4.14. Appendix by Bogdan Nica: Three theorems on linear groups

Bogdan Nica
Mathematisches Institut, Georg-August Universität Göttingen, Germany.
Email: bogdan.nica@gmail.com

4.14.1. Introduction. Recall that a group is *linear* if it is (isomorphic to) a subgroup of $GL_n(\mathbb{K})$, where $\mathbb{K}$ is a field. If we want to specify the field, we say that the group is linear over $\mathbb{K}$. The following theorems are fundamental, at least from the perspective of the combinatorial group theory.

THEOREM 4.118 (A. I. Mal'cev, 1940). *A finitely generated linear group is residually finite.*

THEOREM 4.119 (A. Selberg, 1960). *A finitely generated linear group over a field of zero characteristic is virtually torsion-free.*

A group is *residually finite* if its elements are distinguished by the finite quotients of the group, i.e., if each non-trivial element of the group remains non-trivial in a finite quotient. A group is *virtually torsion-free* if some finite-index subgroup is torsion-free. As a matter of further terminology, Selberg's theorem is usually referred to as Selberg's lemma, and Mal'cev is alternatively transliterated as Malcev or Maltsev.

Residual finiteness and virtual torsion-freeness are related to a third property — roughly speaking, a "p-adic" refinement of residual finiteness. A theorem due to V. Platonov (1968) gives such refined residual properties for finitely generated linear groups. Both Mal'cev's theorem and Selberg's lemma are consequences of this more powerful, but lesser known, theorem of Platonov.

Once we have Platonov's theorem and its proof, we are not too far from our third theorem of interest. In order to formulate it, let us first observe that every non-trivial torsion element in a group G gives rise to a non-trivial idempotent in the complex group algebra $\mathbb{C}G$. Namely, if $g \in G$ has order $n > 1$, then

$$e = \tfrac{1}{n}(1 + g + \ldots + g^{n-1}) \in \mathbb{C}G$$

satisfies $e^2 = e$, and $e \neq 0, 1$. The *Idempotent Conjecture* is the bold statement that the converse also holds:

CONJECTURE 4.120 (Idempotent Conjecture). *If G is a torsion-free group, then the group algebra $\mathbb{C}G$ has no non-trivial idempotents.*

While not yet settled in general, this conjecture is known for many classes of groups. A particularly important partial result is proven by H. Bass in [**Bas76**]:

THEOREM 4.121. *Torsion-free linear groups satisfy the Idempotent Conjecture.*

4.14.2. Virtual and residual properties of groups. Virtual torsion-freeness and residual finiteness are instances of the following terminology. Let $\mathcal{P}$ be a group-theoretic property. A group is *virtually* $\mathcal{P}$ if it has a finite-index subgroup enjoying $\mathcal{P}$. A group is *residually* $\mathcal{P}$ if each non-trivial element of the group remains non-trivial in some quotient group enjoying $\mathcal{P}$. The virtually $\mathcal{P}$ groups and the residually $\mathcal{P}$ groups contain the $\mathcal{P}$ groups. It may certainly happen that a property is virtually stable (e.g., finiteness) or residually stable (e.g., torsion-freeness).

Besides virtual torsion-freeness and residual finiteness, we are interested in the hybrid notion of *virtual residual p-finiteness* where p is a prime. This is obtained by residualizing the property of being a finite p-group, followed by the virtual extension. The notion of virtual residual p-finiteness has, in fact, a leading role in this account for it relates both to residual finiteness and to virtual torsion-freeness.

Observe the following:

(Going down) If $\mathcal{P}$ is inherited by subgroups, then both virtually $\mathcal{P}$ and residually $\mathcal{P}$ are inherited by subgroups. In particular, virtual torsion-freeness, residual finiteness, and virtual residual p-finiteness are inherited by subgroups.

(Going up) Virtually $\mathcal{P}$ passes to finite-index supergroups. In particular, both virtual torsion-freeness and virtual residual p-finiteness pass to finite-index supergroups. Residual finiteness passes to finite-index supergroups.

Residual p-finiteness trivially implies residual finiteness. Going up, we obtain:

LEMMA 4.122. *Virtual residual p-finiteness for some prime p implies residual finiteness.*

On the other hand, residual p-finiteness imposes torsion restrictions. Namely, in a residually p-finite group, the order of a torsion element must be a p-th power. Hence, if a group is residually p-finite and residually q-finite for two different primes p and q, then it is torsion-free. Virtualizing this statement, we obtain:

LEMMA 4.123. *Virtual residual p-finiteness and virtual residual q-finiteness for two primes $p \neq q$ imply virtual torsion-freeness.*

4.14.3. Platonov's theorem. In light of Lemmas 4.122 and 4.123, we see that Mal'cev's theorem and Selberg's lemma are consequences of the following:

THEOREM 4.124 (Platonov, 1968). *Let G be a finitely generated linear group over a field $\mathbb{K}$. If $\operatorname{char} \mathbb{K} = 0$, then G is virtually residually p-finite for all but finitely many primes p. If $\operatorname{char} \mathbb{K} = p$, then G is virtually residually p-finite.*

Actually, the zero characteristic part of Platonov's theorem had been proved slightly earlier by Kargapolov (1967) and, independently, Merzlyakov (1967).

EXAMPLE 4.125. (Cf. Example 4.106) $SL_n(\mathbb{Z})$, where $n \geqslant 2$, is a finitely generated linear group over $\mathbb{Q}$. Reduction modulo a positive integer N defines a group homomorphism $SL_n(\mathbb{Z}) \to SL_n(\mathbb{Z}/N)$, whose kernel

$$\Gamma(N) := \operatorname{Ker}\big(SL_n(\mathbb{Z}) \to SL_n(\mathbb{Z}/N)\big) = \big\{X \in SL_n(\mathbb{Z}) : X \equiv 1_n \bmod N\big\}$$

is the *principal congruence subgroup of level N*. The principal congruence subgroups are finite-index, normal subgroups of $SL_n(\mathbb{Z})$. They are organized according to the divisibility of their levels: $\Gamma(M) \supseteq \Gamma(N)$ if and only if $M|N$, that is, "to contain is to divide". Hence the prime stratum $\{\Gamma(p) : p \text{ prime}\}$, and each descending chain $\{\Gamma(p^k) : k \geqslant 1\}$ corresponding to fixed prime p, stand out.

Elements of $SL_n(\mathbb{Z})$ can be distinguished both along the prime stratum,

$$\bigcap_p \Gamma(p) = \{1_n\},$$

as well as along each descending p-chain,

$$\bigcap_k \Gamma(p^k) = \{1_n\}.$$

We thus have two ways of seeing that $SL_n(\mathbb{Z})$ is residually finite.

There is no prime p for which $SL_n(\mathbb{Z})$ is residually p-finite, simply because the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$

has order 6. However, $SL_n(\mathbb{Z})$ is virtually residually p-finite for each prime p. The reason is that $\Gamma(p)$ is residually p-finite, and this is easily seen by noting that each successive quotient $\Gamma(p^k)/\Gamma(p^{k+1})$ in the descending chain $\{\Gamma(p^k) : k \geqslant 1\}$ is a p-group: for $X \in \Gamma(p^k)$ we have

$$X^p = 1_n + \sum_{i=1}^{p} \binom{p}{i} (X - 1_n)^i \in \Gamma(p^{k+1}).$$

EXAMPLE 4.126. $SL_n(\mathbb{F}_p[t])$, where $n \geqslant 2$, is linear over $\mathbb{F}_p(t)$ and finitely generated for $n \geqslant 3$ (though not for $n = 2$). A similar argument to the one of the previous example, this time involving the principal congruence subgroups corresponding to the descending chain of ideals (t^k) for $k \geqslant 1$, shows that $SL_n(\mathbb{F}_p[t])$ is virtually residually p-finite. On the other hand, $SL_n(\mathbb{F}_p[t])$ contains a copy of the infinite torsion group $(\mathbb{F}_p[t], +)$, and this prevents $SL_n(\mathbb{F}_p[t])$ from being virtually torsion-free. Consequently, $SL_n(\mathbb{F}_p[t])$ cannot be virtually residually q-finite for any prime $q \neq p$.

Platonov's theorem implies the following "p-adic" refinement of Mal'cev's theorem.

COROLLARY 4.127. *A finitely generated linear group is virtually residually p-finite for some prime p.*

This corollary, combined with Example 4.126, leads us to a simple example of a finitely generated group which is non-linear but residually finite:

$$SL_n(\mathbb{F}_p[t]) \times SL_n(\mathbb{F}_q[t]),$$

where p and q are different primes, and $n \geqslant 3$.

4.14.4. Proof of Platonov's theorem. Let G be a finitely generated linear group over a field $\mathbb{K}$, say $G \leqslant GL_n(\mathbb{K})$. In $\mathbb{K}$, consider the subring A generated by the multiplicative identity 1 and the matrix entries of a finite, symmetric set of generators for G. Thus A is a finitely generated domain, and G is a subgroup of $GL_n(A)$. Platonov's theorem is then a consequence of the following:

THEOREM 4.128. *Let A be a finitely generated domain. If* $\operatorname{char} A = 0$*, then $GL_n(A)$ is virtually residually p-finite for all but finitely many primes p. If* $\operatorname{char} A = p$*, then $GL_n(A)$ is virtually residually p-finite.*

Here, and for the remainder of the section, rings are commutative and unital. The proof of Theorem 4.128 is a straightforward variation on the example of $SL_n(\mathbb{Z})$, as soon as we know the following facts:

LEMMA 4.129. *Let A be a finitely generated domain. Then the following hold:*

i. *A is noetherian.*
ii. *$\cap_k I^k = 0$ for any ideal $I \neq A$.*
iii. *If A is a field, then A is finite.*
iv. *The intersection of all maximal ideals of A is 0.*
v. *If* $\operatorname{char} A = 0$*, then only finitely many primes $p = p \cdot 1$ are invertible in A.*

Let us postpone the proof of Lemma 4.129 for the moment, and focus instead on deriving Theorem 4.128. The principal congruence subgroup of $GL_n(A)$ corresponding to an ideal I of A is defined by

$$\Gamma(I) = \mathrm{Ker}\big(GL_n(A) \to GL_n(A/I)\big).$$

If π is a maximal ideal then A/π is a finite field, by Lemma 4.129 iii, so $\Gamma(\pi)$ has finite index in $GL_n(A)$. Also,

$$\bigcap_{\pi} \Gamma(\pi) = \{1_n\}$$

as π runs over the maximal ideals of A, by Lemma 4.129 iv. This shows that $GL_n(A)$ is residually finite, thereby proving Mal'cev's theorem.

For each $k \geqslant 1$, the quotient π^k/π^{k+1} is naturally an A/π-module. It inherits finite generation from the finite generation of the A-module π^k, the latter due to A being noetherian. As A/π is finite, π^k/π^{k+1} is finite as well. It follows that the ring A/π^k is finite, and so $\Gamma(\pi^k)$ has finite index in $GL_n(A)$. Furthermore,

$$\bigcap_{k} \Gamma(\pi^k) = \{1_n\}$$

by Lemma 4.129 ii, which shows once again that $GL_n(A)$ is residually finite. Now let p denote the characteristic of A/π, so $p = p \cdot 1 \in \pi$. Then $\Gamma(\pi^k)/\Gamma(\pi^{k+1})$ is a p-group: for $X \in \Gamma(\pi^k)$ we have

$$X^p = 1_n + \sum_{i=1}^{p} \binom{p}{i} (X - 1_n)^i \in \Gamma(\pi^{k+1}).$$

To conclude, $GL_n(A)$ is virtually residually p-finite for each prime p not invertible in A. By Lemma 4.129 v, this happens for all but finitely many primes p in the zero characteristic case. In characteristic p, there is only such prime, namely p itself. Theorem 4.128 is proved. □

We now return to the proof of the lemma.

Proof of Lemma 4.129. The first two points are standard: i) follows from the Hilbert Basis Theorem, and ii) is the Krull Intersection Theorem for domains.

iii) We claim the following: If $F \subseteq F(u)$ is a field extension with $F(u)$ finitely generated as a ring, then $F \subseteq F(u)$ is a finite extension and F is finitely generated as a ring.

We use the claim as follows. Let F be the prime field of A and let $a_1, \dots, a_k$ be generators of A as a ring. Thus $A = F(a_1, \dots, a_k)$. Going down the chain

$$A = F(a_1, \dots, a_k) \supseteq F(a_1, \dots, a_{k-1}) \supseteq \dots \supseteq F,$$

we obtain that $F \subseteq A$ is a finite extension, and that F is finitely generated as a ring. Then F is a finite field, as $\mathbb{Q}$ is not finitely generated as a ring, and so A is finite.

Now let us prove the claim. Assume that u is transcendental over F, i.e., $F(u)$ is the field of rational functions in u. Let $P_1/Q_1, \dots, P_k/Q_k$ generate $F(u)$ as a ring, where $P_i, Q_i \in F[u]$. The multiplicative inverse of $1 + u \cdot \prod Q_i$ is a polynomial expression in the P_i/Q_i's, which can be written as $R/\prod Q_i^{s_i}$. Therefore,

$$\prod Q_i^{s_i} = (1 + u \cdot \prod Q_i)R$$

in $F[u]$. But this is impossible, since $\prod Q_i^{s_i}$ is relatively prime to $1 + u \cdot \prod Q_i$.

Thus u is algebraic over F. Let

$$X^d + \alpha_1 X^{d-1} + \cdots + \alpha_d$$

be the minimal polynomial of u over F. Let also $a_1, \ldots, a_k$ be ring generators of $F(u) = F[u]$. We may write each a_i as

$$\sum_{0 \leqslant m \leqslant d-1} \beta_{i,m}\, u^m$$

with $\beta_{i,m} \in F$. We claim that the α_j's and the $\beta_{i,m}$'s are ring generators of F. Let $c \in F$. Then c is a polynomial in $a_1, \ldots, a_k$ over F, hence a polynomial in u over the subring of F generated by the $\beta_{i,m}$'s, hence a polynomial in u of degree less than d over the subring of F generated by the α_j's and the $\beta_{i,m}$'s. By the linear independence of $\{1, u, \ldots, u^{d-1}\}$, the latter polynomial is actually of degree 0. Hence c ends up in the subring of F generated by the α_j's and the $\beta_{i,m}$'s.

iv) Let $a \neq 0$ in A. To find a maximal ideal of A not containing a, we rely on the basic avoidance: maximal ideals do not contain invertible elements. Consider the localization $A' = A[1/a]$. Let π' be a maximal ideal in A', so $a \notin \pi'$. The restriction $\pi = \pi' \cap A$ is an ideal in A, and $a \notin \pi$. We show that π is maximal. The embedding $A \hookrightarrow A'$ induces an embedding $A/\pi \hookrightarrow A'/\pi'$. As A'/π' is a field which is finitely generated as a ring, in follows from iii) that A'/π' is finite field. Therefore the subring A/π is a finite domain, hence a field as well.

v) We shall use Noether's Normalization Theorem: If R is a finitely generated algebra over a field $F \subseteq R$, then there are elements $x_1, \ldots, x_k \in R$ algebraically independent over F such that R is integral over $F[x_1, \ldots, x_k]$.

In our case, $\mathbb{Z}$ is a subring of A, and A is an integral domain which is finitely generated as a $\mathbb{Z}$-algebra. Extending to rational scalars, we have that $A_{\mathbb{Q}} = \mathbb{Q} \otimes_{\mathbb{Z}} A$ is a finitely generated $\mathbb{Q}$-algebra. By the Normalization Theorem, there exist elements $x_1, \ldots, x_k$ in $A_{\mathbb{Q}}$ which are algebraically independent over $\mathbb{Q}$, and such that $A_{\mathbb{Q}}$ is integral over $\mathbb{Q}[x_1, \ldots, x_k]$. Up to replacing each x_i by an integral multiple of itself, we may assume that $x_1, \ldots, x_k$ are in A. There is some positive $m \in \mathbb{Z}$ such that each ring generator of A is integral over $\mathbb{Z}[1/m][x_1, \ldots, x_k]$. Thus $A[1/m]$ is integral over the subring $\mathbb{Z}[1/m][x_1, \ldots, x_k]$. If a prime p is invertible in A, then it is also invertible in $A[1/m]$ while at the same time $p \in \mathbb{Z}[1/m][x_1, \ldots, x_k]$.

Now we use the following general fact. Let R be a ring which is integral over a subring S. If $s \in S$ is invertible in R, then s is already invertible in S. The proof is easy. Let $r \in R$ with $rs = 1$. We have

$$r^d + s_1 r^{d-1} + \cdots + s_{d-1} r + s_d = 0$$

for some $s_i \in S$, since r is integral over S. Multiplying through by s^{d-1} yields $r \in S$.

Returning to our proof, we infer that p is invertible in $\mathbb{Z}[1/m][x_1, \ldots, x_k]$. By the algebraic independence of $x_1, \ldots, x_k$, it follows that p is actually invertible in $\mathbb{Z}[1/m]$. But only finitely many primes have this property, namely the prime factors of m. □

4.14.5. The Idempotent Conjecture for linear groups. Our approach to Bass's theorem relies on the following criterion of E. Formanek [**For73**], whose proof is postponed till the next section.

Theorem 4.130 (E. Formanek, 1973). *Let G be a torsion-free group with the property that, for infinitely many primes p, G has no p-self-similar elements. Then the Idempotent Conjecture holds for G.*

Given a group G, we say that a non-trivial element $g \in G$ is *self-similar* if g is conjugate in G to a proper power g^N, where $N \geqslant 2$. Clearly, torsion elements are self-similar. It turns out that the converse holds for linear groups in positive characteristic.

Lemma 4.131. *In a linear group over a field of positive characteristic, every self-similar element is torsion.*

Proof. Let char $\mathbb{K} = p$, and consider the relation $g^N = x^{-1}gx$ in $GL_n(\mathbb{K})$, where $N \geqslant 2$. Without loss of generality, $\mathbb{K}$ is algebraically closed and g is in Jordan normal form. Each Jordan block is of the form $\lambda \cdot 1_k + \Delta_k$, where Δ_k is the $k \times k$-matrix with 1's on the super-diagonal and 0's everywhere else. Since

$$(\lambda \cdot 1_k + \Delta_k)^{p^s} = \lambda^{p^s} \cdot 1_k + \Delta_k^{p^s}$$

and $\Delta_k^{p^s} = 0$ for large enough s, it follows that g^{p^s} is diagonal for large enough s. Thus, up to replacing g by g^{p^s}, we may assume that g is diagonal. So let g have $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ along the diagonal, and write out the relation $gx = xg^N$ in matrix form: $(x_{ij}\ \lambda_i) = (x_{ij}\ \lambda_j^N)$. Compare the i-th row on the two sides. At least one of $x_{i1}, x_{i2}, \ldots, x_{in}$ is non-zero, hence $\lambda_i = \lambda_{\sigma(i)}^N$ for some $\sigma(i) \in \{1, \ldots, n\}$. Since $\sigma^s = \sigma^{s+t}$ for some positive integers s and t, it follows that

$$\lambda_i = \lambda_{\sigma^{s+t}(i)}^{N^{s+t}} = \left(\lambda_{\sigma^s(i)}^{N^s}\right)^{N^t} = \lambda_i^{N^t}$$

for each i. We conclude that $g^{N^t-1} = 1$ in $GL_n(\mathbb{K})$. □

In characteristic zero, a linear group may contain self-similar elements of infinite order. A simple example in, say, $GL_2(\mathbb{R})$ is provided by the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

which is conjugate into its N-th power by

$$\begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix}.$$

Exercise 4.132. Show that the entire subgroup generated by these two matrices is torsion-free.

The analogue of Lemma 4.131 in characteristic zero involves the following refined notion of self-similarity. Given a group G and a prime p, let us say that a non-trivial element $g \in G$ is *p-self-similar* if g is conjugate in G to a proper p-th power g^{p^k}, where $k \geqslant 1$.

Lemma 4.133. *In a finitely generated linear group over a field of characteristic zero, the following holds for all but finitely many primes p: Every p-self-similar element is torsion.*

Proof. The characteristic zero case of Platonov's theorem reduces the claim to showing that, in a virtually residually p-finite group, every p-self-similar element is torsion. This easily follows from the observation that a residually p-finite group has no p-self-similar elements. □

The upshot of Lemmas 4.131 and 4.133 is that a finitely generated, torsion-free linear group comfortably meets the requirement of Formanek's criterion, and so it satisfies the Idempotent Conjecture. The theorem of Bass follows.

4.14.6. Proof of Formanek's criterion. The proof of Theorem 4.130 uses tracial methods. Let us first recall that a *trace* on a $\mathbb{K}$-algebra $\mathcal{A}$ is a $\mathbb{K}$-linear map $\tau : \mathcal{A} \to \mathbb{K}$ with the property that $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}$. In short, traces are linear functionals which vanish on commutators. The ersatz commutativity afforded by a trace is extremely valuable in a noncommutative world.

On a group algebra $\mathbb{K}G$, the *standard trace* $\mathrm{tr} : \mathbb{K}G \to \mathbb{K}$ is the linear functional which records the coefficient of the identity element:

$$\mathrm{tr}\big(\sum a_g\, g\big) = a_1.$$

In general, traces on $\mathbb{K}G$ are in bijective correspondence with maps $G \to \mathbb{K}$ which are constant on conjugacy classes. The characteristic map $1_C : G \to \mathbb{K}$ of a conjugacy class $C \subseteq G$ defines the trace

$$\tau_C\big(\sum a_g\, g\big) = \sum_{g \in C} a_g,$$

thus, $\mathrm{tr} = \tau_{\{1\}}$ with this notation. The traces τ_C, where C runs over the conjugacy classes of G, provide a natural basis for the $\mathbb{K}$-linear space formed by the traces of $\mathbb{K}G$. Another distinguished trace is the *augmentation map* $\epsilon : \mathbb{K}G \to \mathbb{K}$, given by

$$\epsilon\big(\sum a_g\, g\big) = \sum a_g.$$

This is the trace on $\mathbb{K}G$ defined by the constant map $1 : G \to \mathbb{K}$. The augmentation map is in fact a unital $\mathbb{K}$-algebra homomorphism, hence ϵ is a trace which is $\{0,1\}$-valued on idempotents.

Understanding the range of the standard trace on idempotents is much more difficult. The following theorem addresses this problem in the case of complex group algebras.

Theorem 4.134 (I. Kaplansky, 1969). *Let e be an idempotent in $\mathbb{C}G$. Then* $\mathrm{tr}(e) \in [0,1]$. *Furthermore,* $\mathrm{tr}(e) = 0$ *if and only if* $e = 0$, *and* $\mathrm{tr}(e) = 1$ *if and only if* $e = 1$.

Now let us return to the proof of Formanek's criterion. It consists of two steps.

(Positive characteristic claim) Fix a prime p. If G has no p-self-similar elements and $\mathbb{K}$ is a field of characteristic p, then the standard trace is $\{0,1\}$-valued on the idempotents of $\mathbb{K}G$.

It is a familiar fact that the identity $(a+b)^p = a^p + b^p$ holds in any *commutative* $\mathbb{K}$-algebra. Its noncommutative generalization, somewhat lesser known, says that, in a $\mathbb{K}$-algebra, $(a+b)^p - a^p - b^p$ is a sum of commutators. Indeed, we may assume that we are in the free $\mathbb{K}$-algebra on a and b. We expand $(a+b)^p$ into monomials of degree p in a and b, and we let the cyclic group of order p act on these monomials by cyclic permutations. We see orbits of size p, except for a^p and b^p, which are fixed by the action. Now we observe that the sum of monomials corresponding to each orbit of size p is a sum of commutators. This follows from the identity

$$\begin{aligned} &x_1x_2\dots x_{p-1}x_p + x_2x_3\dots x_px_1 + \dots + x_px_1\dots x_{p-2}x_{p-1} \\ &\quad = p\cdot x_1x_2\dots x_{p-1}x_p - [x_1, x_2\dots x_p] - [x_1x_2, x_3\dots x_p] - \dots - [x_1\dots x_{p-1}, x_p]. \end{aligned}$$

Next, let us iterate: We show by induction that $(a+b)^{p^k} - a^{p^k} - b^{p^k}$ is a sum of commutators for every positive integer k. For the induction step we write

$$(a+b)^{p^{k+1}} = \Big(a^{p^k} + b^{p^k} + \sum [u_i, v_i]\Big)^p = a^{p^{k+1}} + b^{p^{k+1}} + \sum [u_i, v_i]^p + \sum [u'_j, v'_j]$$

and

$$[u, v]^p = (uv)^p - (vu)^p + \sum [y_l, z_l] = \big[(uv)^{p-1}u, v\big] + \sum [y_l, z_l].$$

In particular, a trace τ on a $\mathbb{K}$-algebra has the property that

$$\tau\big((a+b)^{p^k}\big) = \tau\big(a^{p^k}\big) + \tau\big(b^{p^k}\big)$$

for every positive integer k. For a basic trace τ_C, where $C \neq \{1\}$, and an idempotent $e \in \mathbb{K}G$, we obtain

$$\tau_C(e) = \tau_C(e^{p^k}) = \tau_C\Big(\big(\sum e_g\, g\big)^{p^k}\Big) = \sum \tau_C\big((e_g\, g)^{p^k}\big) = \sum e_g^{p^k} 1_C\big(g^{p^k}\big)$$

for each positive integer k. The hypothesis that G has no p-self-similar elements implies that, for each g in the support of e, there is at most one k so that $g^{p^k} \in C$. Thus, taking k large enough, we see that $\tau_C(e) = 0$. Using the relation

$$\epsilon = \mathrm{tr} + \sum_{C \neq \{1\}} \tau_C,$$

we conclude that tr is $\{0,1\}$-valued on the idempotents of $\mathbb{K}G$.

(Zero characteristic claim) Assume that, for infinitely many primes p, the following holds: The standard trace is $\{0,1\}$-valued on the idempotents of $\mathbb{K}G$, whenever $\mathbb{K}$ is a field of characteristic p. Then the standard trace is $\{0,1\}$-valued on the idempotents of $\mathbb{C}G$.

Arguing by contradiction, we assume that e is an idempotent in $\mathbb{C}G$ with $e_1 = \mathrm{tr}(e) \notin \{0,1\}$. Let $A \subseteq \mathbb{C}$ be the subring generated by the support of e together with $1/e_1$ and $1/(1-e_1)$, and view e as an idempotent in the group ring AG. By Lemma 4.129 v, for all but finitely many primes p there is a quotient map $A \to \mathbb{K}$, $a \mapsto \overline{a}$, onto a field of characteristic p. Note that $\overline{e}_1 \neq 0, 1$ in $\mathbb{K}$, since e_1 and $1-e_1$ are invertible in A. The induced ring homomorphism $AG \to \mathbb{K}G$ sends e to an idempotent $\overline{e}$ in $\mathbb{K}G$ with $\mathrm{tr}(\overline{e}) \neq 0, 1$, thereby contradicting our hypothesis.

The proof of Theorem 4.130 is concluded by invoking Kaplansky's theorem. □

4.14.7. Notes. Platonov's theorem. Besides the Russian original [**Pla68**], the only other source in the literature for Platonov's theorem appears to be the presentation by B. A. F. Wehrfritz in [**Weh73**]. The proof presented herein seems considerably simpler. It is mainly influenced by the discussion of Mal'cev's theorem in lecture notes by Stallings [**Sta00**], and it has a certain degree of similarity with Platonov's own arguments in [**Pla68**].

Selberg's lemma. It is important to note that Selberg's lemma is just a minor step in Selberg's paper [**Sel60**], whose true importance is that it started the rich stream of rigidity results for lattices in higher rank semisimple Lie groups. An alternative road to Selberg's lemma is to use valuations. This is the approach taken by J. W. Cassels in [**Cas86**] and by J. Ratcliffe in [**Rat06**].

The Idempotent Conjecture. The Idempotent Conjecture is usually attributed to Kaplansky, but a reference seems elusive. What Kaplansky did state on more than one occasion (Problem 1, p. 122 in [**Kap69**], and Problem 6, p. 448 in [**Kap70**]) is a problem nowadays referred to as the

Conjecture 4.135 (Zero-Divisor Conjecture). *If G is a torsion-free group and $\mathbb{K}$ is a field, then the group algebra $\mathbb{K}G$ has no zero-divisors, i.e., $ab \neq 0$ whenever $a, b \neq 0$ in $\mathbb{K}G$.*

The Zero-Divisor Conjecture over the complex field, which clearly implies the Idempotent Conjecture, is still not settled for the class of (torsion-free) linear groups.

Kaplansky's theorem. We refer to M. Burger and A. Valette [**BV98**] for a proof, as well as for a nice complementary reading. The main insight of Kaplansky's analytic proof is to pass from the group algebra $\mathbb{C}G$ to a completion afforded by the regular representation on $\ell^2 G$. One can use the weak completion, that is the von Neumann algebra $\mathrm{L}G$, or the norm completion, the so-called reduced C^*-algebra $\mathrm{C}^*_r G$. Kaplansky's proof, while remarkable in itself, is perhaps more important for suggesting what came to be known as the **Kadison Conjecture**:

Conjecture 4.136 (Kadison Conjecture). *For every torsion-free group G, the reduced C^*-algebra $\mathrm{C}^*_r G$ has no non-trivial idempotents.*

At the time of writing, the Kadison Conjecture for the class of (torsion-free) linear groups is still open.

Bass's theorem. As we have seen, the step from Formanek's criterion to the theorem of Bass is rather short, and it uses results on linear groups which were known — certainly on the eastern side of the Iron Curtain, but probably also on its western side — at the time of [**For73**]. Ascribing the theorem to Bass *and* Formanek is therefore not entirely unwarranted. The hard facts, however, are that Bass [**Bas76**] actually proves much more whereas Formanek [**For73**] states less.

CHAPTER 5

Coarse geometry

In this chapter we will *coarsify* familiar geometric concepts: In the context of the coarse geometry the exact geometric computations will not matter, what matters are the *asymptotics* of various geometric quantities. For instance, the exact computations of distances become irrelevant, as long as we have uniform linear bounds on the distances; accordingly, isometries will be coarsified to *quasiisometries*. In the process of coarsification, metric spaces will be frequently replaced with nets which *approximate* of *discretize* these metric spaces:

$$\text{coarsification} : (X, \text{dist}_X) \to \quad \text{a net } N \subset X,$$

while maps between metric spaces will be replaced with maps between the respective nets. We will coarsify the notions of area, volume and isoperimetric inequalities: Various geometric quantities will be replaced with cardinalities of certain nets.

The drawback of the coarse geometry is that we will be missing the beauty of the precise formulae and sharp inequalities of the classical geometry: We will be unable to tell apart the Euclidean $n \times n$ square from the Euclidean disk of the radius n. Accordingly, we will think of n^2 as their (coarse) areas. What we gain, however, is equivalence of different geometric models of groups and equivalence of the associated geometric invariants.

5.1. Quasi-isometry

In this section define an important equivalence relation between metric spaces: The *quasiisometry*. It is this concept that will relate different geometric models of finitely generated groups which were introduced in the previous chapter. The quasiisometry of spaces has two equivalent definitions (both useful): One which is easy to visualize and the other which makes it easier to understand why it is an equivalence relation. We begin with the first definition, continue with the second and then prove their equivalence.

The notion of *quasiisometry*

$$f : (X, \text{dist}_X) \to (Y, \text{dist}_Y)$$

between two metric spaces appeared first in work of Mostow on *strong rigidity* of lattices in semisimple Lie groups, see e.g. **[Mos73]**. Mostow's notion (the one of a *pseudo-isometry*) was slightly more restrictive than the one we will be using:

$$L^{-1}\text{dist}_X(x, x') - A \leqslant \text{dist}_Y(f(x), f(x')) \leqslant L\text{dist}_X(x, x').$$

In particular, pseudo-isometries used by Mostow were continuous maps. Later on, it became clear that it makes sense to add an additive constant in this equation on the right hand side as well, and work with (typically) discontinuous maps.

Definition 5.1. Two metric spaces (X, dist_X) and (Y, dist_Y) are called *quasi-isometric* if and only if there exist separated nets $A \subset X$ and $B \subset Y$, such that (A, dist_X) and (B, dist_Y) are bi-Lipschitz equivalent.

Thus, if we think of a separated net as a *discretization* of a metric space, then quasiisometric spaces are the ones which admit bi-Lipschitz discretizations.

Examples 5.2. (1) A nonempty metric space of finite diameter is quasi-isometric to a point.

(2) The space $\mathbb{R}^n$ endowed with a norm is quasiisometric to $\mathbb{Z}^n$ with the metric induced by that norm.

Historically, quasiisometry was introduced in order to formalize the relationship between some discrete metric spaces (most of the time, groups) and some "non-discrete" (or continuous) metric spaces like for instance Riemannian manifolds, etc. Examples of this is the relationship between finitely generated (or, finitely presented) groups and their *geometric models* introduced in §4.9.

When trying to prove that the quasiisometry relation is an equivalence relation, reflexivity and symmetry are straightforward, but, when attempting to prove transitivity, the following question naturally arises:

Question 5.3 (M. Gromov, [**Gro93**], p. 23). Can a space contain two separated nets that are not bi-Lipschitz equivalent?

Gromov's question was answered by

Theorem 5.4 (D. Burago, B. Kleiner, [**BK98**]). *There exists a separated net N in $\mathbb{R}^2$ which is not bi-Lipschitz equivalent to $\mathbb{Z}^2$.*

Along the same lines, Gromov asked whether two infinite finitely generated groups G and H that are quasiisometric are also bi-Lipschitz equivalent. The negative answer to this question was given by T. Dymarz [**Dym10**]. We discuss Gromov's questions in more detail in Chapter 23.

Fortunately, there is a second equivalent way of defining quasiisometry of two metric spaces, based on loosening (coarsifying) the Lipschitz concept. The reader can think of the coarse Lipschitz notion defined below as a generalization of the traditional notion of continuity. Unlike the notion continuity, we will not care about behavior of maps on the small scale, as long as they behave "well" on the large scale.

Definition 5.5. Let X, Y be metric spaces. A map $f : X \to Y$ is called (L, C)*–coarse Lipschitz* if

$$\mathrm{dist}_Y(f(x), f(x')) \leqslant L\mathrm{dist}_X(x, x') + C, \tag{5.1}$$

for all $x, x' \in X$. A map $f : X \to Y$ is called an (L, C)*–quasiisometric embedding* if

$$L^{-1}\mathrm{dist}_X(x, x') - C \leqslant \mathrm{dist}_Y(f(x), f(x')) \leqslant L\mathrm{dist}_X(x, x') + C, \tag{5.2}$$

for all $x, x' \in X$. Note that a quasiisometric embedding does not have to be an embedding in the usual sense, however distant points have distinct images.

Example 5.6. 1. The floor function $f : \mathbb{R} \to \mathbb{Z} \subset \mathbb{R}$, $f(x) = \lfloor x \rfloor$, is $(0, 1)$-coarse Lipschitz. This function is a quasiisometric embedding $\mathbb{R} \to \mathbb{Z}$.

2. The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, is not coarse Lipschitz.

LEMMA 5.7. *Suppose that G_1, G_2 are finitely generated groups equipped with word-metrics. Then every coarse Lipschitz map $f : G_1 \to G_2$ is K-Lipschitz for some K.*

PROOF. Let S_1, S_2 be finite generating sets of the groups G_1, G_2. Suppose that $f : (G_1, \mathrm{dist}_{S_1}) \to (G_2, \mathrm{dist}_{S_2})$ is (L, C)-coarse Lipschitz. For every $s \in S_1$ and $g \in G_1$, we have

$$\mathrm{dist}_{S_2}(f(g), f(sg)) \leqslant L + C.$$

Therefore, by the triangle inequalities, for all $g, h \in G_1$,

$$\mathrm{dist}_{S_2}(f(g), f(h)) \leqslant (L + C)\mathrm{dist}_{S_1}(g, h).$$

Hence, f is K-Lipschitz with $K = L + C$. □

Nice thing about Lipschitz and coarse Lipschitz maps is that these classes of maps can be recognized *locally*.

LEMMA 5.8. *Consider a map $f : (X, \mathrm{dist}_X) \to (Y, \mathrm{dist}_Y)$ between metric spaces, where X is a geodesic metric space (but Y is not required to be geodesic). Suppose that r is a positive number such that for all $x, x' \in X$*

$$\mathrm{dist}_X(x, x') \leqslant r \Rightarrow \mathrm{dist}_Y(f(x), f(x')) \leqslant A.$$

Then f is $(\frac{A}{r}, A)$-coarse Lipschitz.

PROOF. For points $x, x' \in X$ consider a geodesic $\gamma \subset X$ connecting x to x'. There exists a finite sequence

$$x_1 = x, x_2, \ldots, x_n, x_{n+1} = x'$$

along the geodesic γ, such that

$$r(n-1) \leqslant D = \mathrm{dist}_X(x, x') < rn, \quad \mathrm{dist}_X(x_i, x_{i+1}) \leqslant r, i = 1, \ldots, n.$$

Then, applying the triangle inequality and the fact that $\mathrm{dist}_Y(f(x_i), f(x_{i+1})) \leqslant A$, we obtain:

$$\mathrm{dist}_Y(f(x), f(x')) \leqslant nA \leqslant \frac{DA}{r} + A = \frac{A}{r}\mathrm{dist}_X(x, x') + A. \quad \square$$

If X is a finite interval $[a, b]$ then an (L, C)–quasiisometric embedding $\mathfrak{q} : X \to Y$ is called an (L, C)-*quasigeodesic (segment)*. If $a = -\infty$ or $b = +\infty$ then $\mathfrak{q}$ is called an (L, C)-*quasigeodesic ray*. If both $a = -\infty$ and $b = +\infty$, then $\mathfrak{q}$ is called an (L, C)-*quasigeodesic line*. By abuse of terminology, the same names are used for the image of $\mathfrak{q}$.

In line with loosening the Lipschitz concept, we will also loosen the concept of the inverse map:

DEFINITION 5.9. Maps of metric spaces $f : X \to Y, \bar{f} : Y \to X$ are said to be *C-coarse inverse to each other* if

$$\mathrm{dist}_X(\bar{f} \circ f, id_X) \leqslant C, \quad \mathrm{dist}_Y(f \circ \bar{f}, \mathrm{dist}_Y) \leqslant C. \tag{5.3}$$

In particular, a 0-coarse inverse map is the inverse map in the usual sense. Lastly, we can define quasiisometries:

DEFINITION 5.10. A map $f : X \to Y$ between metric spaces is called a *quasi-isometry* if it is coarse Lipschitz and admits a coarse Lipschitz coarse inverse map. More precisely, f is an (L, C)–*quasiisometry* if f is (L, C)–coarse Lipschitz and there exists a (L, C)–coarse Lipschitz map $\bar{f} : Y \to X$ such that the maps $f, \bar{f}$ are C-coarse inverse to each other.

Two metric spaces X, Y are *quasiisometric* if there exists a quasiisometry $X \to Y$.

A metric space X is called *quasigeodesic* if there exist constants (L, C) so that every pair of points in X can be connected by an (L, C)-quasigeodesic.

Most of the time, the *quasiisometry constants* L, C do not matter, hence, we shall use the words *quasiisometries*, *quasigeodesic* and *quasiisometric embeddings* without specifying the constants. We will frequently abbreviate *quasiisometry*, *quasiisometric* and *quasiisometrically* to QI.

EXERCISE 5.11.
(1) Prove that every quasiisometry $f : X \to Y$ is a quasi-isometric embedding.
(2) Prove that the coarse inverse of a quasiisometry is also a quasiisometry.
(3) Prove that the composition of two quasiisometric embeddings is a quasi-isometric embedding, and that the composition of two quasiisometries is a quasiisometry.
(4) If $f, g : X \to Y$ are within finite distance from each other, i.e.,
$$\mathrm{dist}(f, g) < \infty,$$
and f is a quasiisometry, then g is also a quasiisometry.
(5) Let $f_i : X \to X, i = 1, 2, 3$ be maps such that f_3 is (L_3, A_3) coarse Lipschitz and $\mathrm{dist}(f_2, id_X) \leqslant A_2$. Then
$$\mathrm{dist}(f_3 \circ f_1, f_3 \circ f_2, \circ f_1) \leqslant L_3 A_2 + A_3.$$
(6) Prove that quasiisometry of metric spaces (defined as in Definition 5.5) is an equivalence relation.

EXERCISE 5.12. 1. Suppose that Y and Z are subsets of a metric space (X, dist) such that Z is contained in the r-neighborhood $\mathcal{N}_r(Y)$. Define the "nearest point projection" $\pi_Z : Y \to Z$, sending each $y \in Y$ to a point $z \in Z$ such that
$$\mathrm{dist}_X(y, z) \leqslant r.$$
Show that π is a quasiisometric embedding.

2. Suppose that $f : X \to Y$ is a quasiisometric embedding such that $f(X)$ is r-dense in Y for some $r < \infty$. Show that f is a quasiisometry. *Hint:* Construct a coarse inverse $\bar{f}$ to the map f by mapping a point $y \in Y$ to $x \in X$ such that
$$\mathrm{dist}_Y(f(x), y) \leqslant r.$$

Maps $f : X \to Y$ such that $f(X)$ is r-dense in Y for some $r < \infty$, are *coarsely surjective.* Thus, we obtain:

COROLLARY 5.13. *A map $f : X \to Y$ is a quasiisometry if and only if f is a coarsely surjective quasiisometric embedding.*

EXAMPLE 5.14. The cylinder $X = \mathbb{S}^n \times \mathbb{R}$ with a product metric is quasiisometric to $Y = \mathbb{R}$; the quasiisometry is the projection to the second factor.

EXAMPLE 5.15. Let $h : \mathbb{R} \to \mathbb{R}$ be an L–Lipschitz function. Then the map

$$f : \mathbb{R} \to \mathbb{R}^2, \quad f(x) = (x, h(x)),$$

is a QI embedding.

Indeed, f is $\sqrt{1+L^2}$–Lipschitz. On the other hand, clearly,

$$\mathrm{dist}(x, y) \leqslant \mathrm{dist}(f(x), f(y)),$$

for all $x, y \in \mathbb{R}$.

EXAMPLE 5.16. Let $\varphi : [1, \infty) \to \mathbb{R}_+$ be a differentiable function such that

$$\lim_{r\to\infty} \varphi(r) = \infty,$$

and there exists $C \in \mathbb{R}$ for which $|r\varphi'(r)| \leqslant C$ for all r. For instance, take $\varphi(r) = \log(r)$. Define the function $F : \mathbb{R}^2 \setminus B(0,1) \to \mathbb{R}^2 \setminus B(0,1)$ which, in the polar coordinates, takes the form

$$(r, \theta) \mapsto (r, \theta + \varphi(r)).$$

Hence F maps radial straight lines to spirals. Let us check that F is L–bi-Lipschitz for $L = \sqrt{1+C^2}$. Indeed, the Euclidean metric in the polar coordinates takes the form

$$ds^2 = dr^2 + r^2 d\theta^2.$$

Then

$$F^*(ds^2) = ((r\varphi'(r))^2 + 1)dr^2 + r^2 d\theta^2$$

and the assertion follows. Extend F to the unit disk by the zero map. Therefore, $F : \mathbb{R}^2 \to \mathbb{R}^2$, is a QI embedding. Since F is onto, it is a quasiisometry $\mathbb{R}^2 \to \mathbb{R}^2$.

PROPOSITION 5.17. *Two metric spaces* (X, dist_X) *and* (Y, dist_Y) *are quasiisometric in the sense of Definition 5.1 if and only if there exists a quasiisometry* $f : X \to Y$.

PROOF. Assume there exists an (L, C)–quasiisometry $f : X \to Y$. Let $\delta = L(C+1)$ and let A be a δ–separated ε–net in X. Then $B = f(A)$ is a 1–separated $(L\varepsilon + 2C)$–net in Y. Moreover, for any $a, a' \in A$,

$$\mathrm{dist}_Y(f(a), f(a')) \leqslant L\mathrm{dist}_X(a, a') + C \leqslant \left(L + \frac{C}{\delta}\right) \mathrm{dist}_X(a, a')$$

and

$$\mathrm{dist}_Y(f(a), f(a')) \geqslant \frac{1}{L}\mathrm{dist}_X(a, a') - C \geqslant \left(\frac{1}{L} - \frac{C}{\delta}\right) \mathrm{dist}_X(a, a') =$$

$$\frac{1}{L(C+1)} \mathrm{dist}_X(a, a').$$

It follows that f, restricted to A and with target B, is a bi-Lipschitz map.

Conversely, assume that $A \subset X$ and $B \subset Y$ are two ε–separated δ–nets, and that there exists a surjective bi-Lipschitz map $g : A \to B$. We define a map $f : X \to Y$ as follows: For every $x \in X$ we choose a point $a_x \in A$ at distance at most δ from x and define $f(x) = g(a_x)$.

REMARK 5.18. The Axiom of Choice makes here yet another important appearance. We will discuss Axiom of Choice in more detail in Chapter 7. Nevertheless, when X is proper (for instance X is a finitely generated group with a word metric), there are finitely many possibilities for the point a_x. Hence, the Axiom of Choice is not required in this situation, as in the finite case it follows from the Zermelo–Fraenkel axioms.

Since $f(X) = g(A) = B$ it follows that Y is contained in the ε–tubular neighborhood of $f(X)$. For every $x, y \in X$,

$$\mathrm{dist}_Y(f(x), f(y)) = \mathrm{dist}_Y(g(a_x), g(a_y)) \leqslant L\mathrm{dist}_X(a_x, a_y) \leqslant L(\mathrm{dist}_X(x, y) + 2\varepsilon)\,.$$

Also

$$\mathrm{dist}_Y(f(x), f(y)) = \mathrm{dist}_Y(g(a_x), g(a_y)) \geqslant \frac{1}{L}\mathrm{dist}_X(a_x, a_y) \geqslant \frac{1}{L}(\mathrm{dist}_X(x, y) - 2\varepsilon)\,.$$

Now the proposition follows from Exercise 5.12. □

Below is yet another variation on the definition of quasiisometry, based on relations.

First, some terminology: Given a relation $R \subset X \times Y$, for $x \in X$ let $R(x)$ denote $\{(x, y) \in X \times Y : (x, y) \in R\}$. Similarly, define $R(y)$ for $y \in Y$. Let π_X, π_Y denote the projections of $X \times Y$ to X and Y respectively.

DEFINITION 5.19. Let X and Y be metric spaces. A subset $R \subset X \times Y$ is called an (L, A)–*quasiisometric relation* if the following conditions hold:

1. Each $x \in X$ and each $y \in Y$ are within distance $\leqslant A$ from the projection of R to X and Y, respectively.
2. For all $x, x' \in \pi_X(R)$,

$$\mathrm{dist}_{Haus}(\pi_Y(R(x)), \pi_Y(R(x'))) \leqslant L\mathrm{dist}(x, x') + A.$$

3. Similarly, for all $y, y' \in \pi_Y(R)$,

$$\mathrm{dist}_{Haus}(\pi_X(R(y)), \pi_X(R(y'))) \leqslant L\mathrm{dist}(y, y') + A.$$

Observe that for any (L, A)–quasiisometric relation R, for all pair of points $x, x' \in X$, and $y \in R(x), y' \in R(x')$ we have

$$\frac{1}{L}\mathrm{dist}(x, x') - \frac{A}{L} \leqslant \mathrm{dist}(y, y') \leqslant L\mathrm{dist}(x, x') + A.$$

The same inequality holds for all pairs of points $y, y' \in Y$, and $x \in R(y), x' \in R(y')$.

In particular, by using the axiom of choice as in the proof of Proposition 5.17, if R is an (L, A)–quasiisometric relation between nonempty metric spaces, then it induces an (L_1, A_1)–quasiisometry $X \to Y$. Conversely, every (L, A)–quasiisometry is an (L_2, A_2)–quasiisometric relation.

Quasi-isometry group of a space. Some quasiisometries $X \to X$ are more interesting than others. The *boring* quasiisometries are the ones which are within finite distance from the identity:

DEFINITION 5.20. Given a metric space (X, dist) we denote by $\mathcal{B}(X)$ the set of maps $f : X \to X$ (not necessarily bijections) which are *bounded perturbations of the identity*, i.e., maps such that

$$\mathrm{dist}(f, id_X) = \sup_{x \in X} \mathrm{dist}(f(x), x) < \infty.$$

In order to mod out the semigroup of quasiisometries $X \to X$ by $\mathcal{B}(X)$, one introduces a group $QI(X)$ defined below. Given a metric space (X, dist), consider the set $QI(X)$ of equivalence classes $[f]$ of quasiisometries $f : X \to X$, where two quasiisometries f, g are equivalent if and only if

$$\text{dist}(f, g) < \infty.$$

In particular, the set of quasiisometries equivalent to id_X is $\mathcal{B}(X)$. Clearly, the composition is an associative binary operation on $QI(X)$.

EXERCISE 5.21. Show that the coarse inverse defines an inverse in $QI(X)$, and, hence, $QI(X)$ is a group.

DEFINITION 5.22. The group $(QI(X), \circ)$ is called the *group of quasiisometries of the metric space* X. When G is a finitely generated group, then $QI(G)$ will denote the group of quasiisometries of G equipped with the word metric.

Note that if S, S' are two finite generating sets of a group G then the identity map $(G, \text{dist}_S) \to (G, \text{dist}_{S'})$ is a quasiisometry, see Exercise 4.83.

EXERCISE 5.23. If $h : X \to X'$ is a quasiisometry of metric spaces, then the groups $QI(X), QI(X')$ are isomorphic; the isomorphism is given by the map

$$[f] \mapsto [h \circ f \circ \bar{h}],$$

where $\bar{h}$ is a coarse inverse to h. Conclude that the group $QI(G)$ is independent of the generating set of G.

More importantly, we will see (Corollary 5.62) that every group quasiisometric to G admits a natural homomorphism to $QI(G)$.

Isometries and virtual isomorphisms. For every metric space X there is a natural homomorphism $q_X : \text{Isom}(X) \to QI(X)$, given by $f \mapsto [f]$. In general, this homomorphism is not injective. For instance, if $X = \mathbb{R}^n$, then the kernel of q_X is the full group of translations $\mathbb{R}^n$. Similarly, the entire group $G = \mathbb{Z}^n \times F$, where F is a finite group, maps trivially to $QI(G)$.

Suppose now that G is an arbitrary finitely generated group. Since G acts isometrically on (G, dist_S) (where S a finite generating set of G), we obtain a homomorphism $q_G : G \to QI(G)$. We will prove in Lemma 14.20 that the kernel K of this homomorphism is a subgroup such that for every $k \in K$ the G-centralizer of k has finite index in G. In particular, if $G = K$ then G is virtually abelian.

The group $VI(G)$ of virtual automorphisms of G defined in §3.2 also maps naturally to $QI(G)$. Indeed, suppose that an isomorphism

$$\phi : G_1/K_1 \to G_2/K_2$$

is a virtual automorphism of G; here G_1, G_2 are finite index subgroups of G and $K_i \lhd G_i$ are finite normal subgroups, $i = 1, 2$. Then ϕ is a quasiisometry; it lifts to a map

$$\psi : G_1 \to G_2, \quad \phi(gK_1) = \psi(g)K_2, g \in G_1.$$

We leave it to the reader to verify that ψ is also a quasiisometry. Since G_i's are finite-index subgroups in G, they are nets in G. Therefore, as in the proof of Proposition 5.17, ψ extends to a quasiisometry

$$f = f_\phi : G \to G.$$

EXERCISE 5.24. Show that the map defined by $\phi \mapsto [f_\phi]$ is a homomorphism $VI(G) \to QI(G)$.

Uniformly proper maps. Sometimes, in order to show that a map $f : X \to Y$ is a quasiisometry, it suffices to check a weaker version of (5.2). We discuss this weaker version below.

Let X, Y be topological spaces. Recall that a (continuous) map $f : X \to Y$ is called *proper* if the inverse image $f^{-1}(K)$ of each compact in Y is a compact in X. The next definition is a "coarsification" of the notion of a proper map:

DEFINITION 5.25. A map $f : X \to Y$ between proper metric spaces is called *uniformly proper* if f is coarse Lipschitz and there exists a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\operatorname{diam}(f^{-1}(B(y, R))) \leqslant \zeta(R)$ for each $y \in Y, R \in \mathbb{R}_+$. Equivalently, there exists a proper continuous function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\operatorname{dist}(f(x), f(x')) \geqslant \eta(\operatorname{dist}(x, x')).$$

The functions ζ and η are called *upper* and *lower distortion functions* of f respectively.

Clearly, every QI embedding is uniformly proper. Conversely, if f is uniformly proper with linear lower distortion function η, then f is a QI embedding. In Lemma 5.28 we will see how uniformly proper maps appear naturally in group theory.

EXERCISE 5.26. 1. Consider the arc-length parameterization $f : \mathbb{R} \to \mathbb{R}^2$ of the parabola $y = x^2$. Then f is uniformly proper but is not a QI embedding.

2. The following function is L-Lipschitz, proper, but not uniformly proper:

$$f(x) = (|x|, \arctan(x)), f : \mathbb{R} \to \mathbb{R}^2.$$

3. The function $\log : (0, \infty) \to \mathbb{R}$ is not uniformly proper.

4. Composition of uniformly proper maps is again uniformly proper.

5. If $f_1, f_2 : X \to Y$ are such that $\operatorname{dist}(f_1, f_2) < \infty$ and f_1 is uniformly proper, then so is f_2.

Even though, uniform properness is weaker than the requirement of a QI embedding, sometimes, the two notions coincide:

LEMMA 5.27. *Suppose that Y is a geodesic metric space, $f : X \to Y$ is a uniformly proper map whose image is r-dense in Y for some $r < \infty$. Then f is a quasiisometry.*

PROOF. We have to construct a coarse inverse to the map f. Given a point $y \in Y$ pick a point $\bar{f}(y) := x \in X$ such that $\operatorname{dist}(f(x), y) \leqslant r$. Let us check that $\bar{f}$ is coarse Lipschitz. Since Y is a geodesic metric space it suffices to verify that there is a constant A such that for all $y, y' \in Y$ with $\operatorname{dist}(y, y') \leqslant 1$, one has:

$$\operatorname{dist}(\bar{f}(y), \bar{f}(y')) \leqslant A.$$

Pick $t > 2r + 1$ which is in the image of the lower distortion function η. Then take $A \in \eta^{-1}(t)$. Hence, $\bar{f}$ is also coarse Lipschitz. It is also clear that the maps $f, \bar{f}$ are coarse inverse to each other. Hence, f is a quasiisometry. $\square$

LEMMA 5.28. *Suppose that G is a finitely generated group equipped with word metric and $G \curvearrowright X$ is a properly discontinuous isometric action on a metric space X. Then for every $o \in X$ the* orbit map *$f : G \to X$, $f(g) = g \cdot o$, is uniformly proper.*

PROOF. 1. Let S denote the finite generating set of G; set

$$L = \max_{s \in S}(d_X(s(o), o).$$

Then for every $g \in G, sinS$, $d_S(gs, g) = 1$, while

$$d_X(gs(o), g(o)) = d_X(s(o), o) \leqslant L.$$

Therefore, by applying triangle inequalities, we conclude that f is L-Lipschitz.

2. Define the function

$$\eta(n) = \min\{d_X(go, o) : |g| = n\}.$$

Since the action $G \curvearrowright X$ is properly discontinuous,

$$\lim_{n \to \infty} \eta(n) = \infty.$$

We extend η linearly to unit intervals $[n, n+1] \subset \mathbb{R}$ and retain the notation η for the extension. The extension $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and proper. By definition of the function η, for every $g \in G$,

$$d_X(f(g), f(1)) = d_X(go, o) \geqslant \eta(d_S(g, 1)).$$

Since G acts on itself and on X isometrically, it follows that

$$d_X(f(g), f(h)) \geqslant \eta(d_S(g, h)), \quad \forall g, h \in G.$$

Thus, the map f is uniformly proper. $\square$

COROLLARY 5.29. *Let $H \leqslant G$ is a finitely generated subgroup of a finitely generated group G. Then the inclusion map $H \to G$ is uniformly proper, where we are using word metrics on G and H associated with their respective finite generating sets.*

We will discuss distortion of subgroups of finitely generated groups in more detail in §5.9.

Coarse convergence. So far, we coarsified geometric concepts. Below is a useful coarsification of an analytical concept.

DEFINITION 5.30. Suppose that (Y, d_Y) is a metric space and X is a set. A sequence of maps $f_i : X \to Y$ is said to *coarsely uniformly converge* to a map $f : X \to Y$ if there exists $R \in \mathbb{R}_+$ and $i_0 \in \mathbb{N}$ such that for all $i > i_0$ and all $x \in X$,

$$d_Y(f(x), f_i(x)) \leqslant R.$$

In other words, there exists i_0 such that for all $i > i_0$, $\mathrm{dist}(f, f_i) < \infty$.

Note that the difference with the usual notion of uniform convergence is just one quantifier: $\forall R$ is replaced with $\exists R$.

Similarly, one defines coarse uniform convergence on compact subsets:

DEFINITION 5.31. Suppose that X is a topological. A sequence (f_i) of maps $X \to Y$ is said to *coarsely uniformly converge to a map $f : X \to Y$ on compact subsets*, if:

There exists a number $R < \infty$ so that for every compact $K \subset X$, there exits i_K so that for all $i > i_K$,

$$\forall x \in K, \quad d(f_i(x), f(x)) \leqslant R.$$

We will use the notation

$$\lim_{i\to\infty}^{c} f_i = f.$$

to denote the fact that the sequence (f_i) coarsely converges to f.

PROPOSITION 5.32 (Coarse Arzela–Ascoli theorem.). *Fix real numbers L, A and D and let X, Y be proper metric spaces so that X admits a separated R-net. Let $f_i : X \to Y$ be a sequence of (L_1, A_1)-Lipschitz maps, such that for some points $x_0 \in X, y_0 \in Y$ we have $d(f(x_0), y_0) \leqslant D$. Then there exists a subsequence (f_{i_k}), and an (L_2, A_2)–Lipschitz map $f : X \to Y$, such that*

$$\lim_{k\to\infty}^{c} f_{i_k} = f.$$

Furthermore, if the maps f_i are (L_1, A_1)–quasiisometries, then f is also an (L_3, A_3)–quasiisometry.

PROOF. Let $N \subset X$ be a separated net. We can assume that $x_0 \in N$. Then the restrictions $f_i\big|_N$ are L'-Lipschitz maps and, by the usual Arzela–Ascoli theorem, the sequence $(f_i\big|_N)$ subconverges (uniformly on compact subsets) to an L' -Lipschitz map $f : N \to Y$. We extend f to X by the rule:

For $x \in X$ pick $x' \in N$ so that $d(x, x') \leqslant R$ and set $f(x) := f(x')$.

Then the map $f : X \to Y$ is (L_2, A_2)–Lipschitz. For a metric ball $B(x_0, r) \subset X, r \geqslant R$, there exists i_r so that for all $i \geqslant i_r$ and all $x \in N \cap B(x_0, r)$, we have $d(f_i(x), f(x)) \leqslant 1$. For arbitrary $x \in K$, we find $x' \in N \cap B(x_0, r + R)$ such that $d(x', x) \leqslant R$. Then

$$d(f_i(x), f(x)) \leqslant d(f_i(x'), f(x')) \leqslant L_1(R + 1) + A.$$

This proves coarse convergence. The statement about quasiisometries follows from the Exercise 5.11, part (4). ☐

5.2. Group-theoretic examples of quasiisometries

We begin by noting that given a finitely generated group G endowed with a word metric, the set $\mathcal{B}(G)$ is particularly easy to describe. To begin with, it contains all the right translations $R_g : G \to G$, $R_g(x) = xg$ (see Remark 4.72).

LEMMA 5.33. *For a finitely generated group (G, dist_S) endowed with a word metric, the set of maps $\mathcal{B}(G)$ consists of* piecewise right translations. *That is, given a map $f \in \mathcal{B}(G)$ there exist finitely many elements $h_1, \ldots, h_n$ in G and a decomposition $G = T_1 \sqcup T_1 \sqcup \ldots \sqcup T_n$ such that f restricted to T_i coincides with R_{h_i}.*

PROOF. Since $f \in \mathcal{B}(G)$, there exists a constant $R > 0$ such that for every $x \in G$, $\mathrm{dist}(x, f(x)) \leqslant R$. This implies that $x^{-1}f(x) \in B(1, R)$. The ball $B(1, R)$ is a finite set. We enumerate its distinct elements $\{h_1, \ldots, h_n\}$. Thus, for every $x \in G$ there exists h_i such that $f(x) = xh_i = R_{h_i}(x)$ for some $i \in \{1, 2, \ldots, n\}$. We define $T_i = \{x \in X \; ; \; f(x) = R_{h_i}(x)\}$. If there exists $x \in T_i \cap T_j$ then $f(x) = xh_i = xh_j$, which implies $h_i = h_j$, a contradiction. ☐

The main example of a quasiisometry, which partly justifies the interest in such maps, is given by Theorem 5.35, proved in the context of Riemannian manifolds first by A. Schwarz [**Šva55**] and, 13 years later, by J. Milnor [**Mil68b**]. At the time, both were motivated by relating volume growth in universal covers of compact Riemannian manifolds and growth of their fundamental groups. Note that sometimes, in the literature it is this theorem (stating the equivalence between the growth function of the fundamental group of a compact manifold and that of the universal cover of the manifold) that is referred to as the Milnor–Schwarz Theorem, and not Theorem 5.35 below.

In fact, it had been observed already by V.A. Efremovich in [**Efr53**] that two growth functions as above (i.e., of the volume of metric balls in the universal cover of a compact Riemannian manifold, and of the cardinality of balls in the fundamental group with a word metric) increase at the same rate.

REMARK 5.34 (What is in the name?). Schwarz is a German-Jewish name which was translated to Russian (presumably, at some point in the 19-th century) as Шварц. In the 1950-s, the AMS, in its infinite wisdom, decided to translate this name to English as Švarc. A. Schwarz himself eventually moved to the United States and is currently a colleague of the second author at University of California, Davis. See http://www.math.ucdavis.edu/∼schwarz/bion.pdf for his mathematical autobiography. The transformation

$$\text{Schwarz} \to \text{Шварц} \to \text{Švarc}$$

is a good example of a composition of a quasiisometry and its coarse inverse.

THEOREM 5.35 (Milnor–Schwarz). *Let* (X, dist) *be a proper geodesic metric space (which is equivalent, by Theorem 1.60, to* X *being a length metric space which is complete and locally compact) and let* G *be a group acting geometrically on* X*. Then:*

(1) *The group* G *is finitely generated.*
(2) *For any word metric* dist_W *on* G *and any point* $x \in X$*, the orbit map* $G \to X$ *given by* $g \mapsto gx$ *is a quasiisometry.*

PROOF. We denote the orbit of a point $y \in X$ by Gy. Given a subset A in X we denote by GA the union of all orbits Ga with $a \in A$.

Step 1: The generating set.

As every geometric action, the action $G \curvearrowright X$ is cobounded: There exists a closed ball $\overline{B}$ of radius D such that $G\overline{B} = X$. Since X is proper, $\overline{B}$ is compact. Define

$$S = \{s \in G \ : \ s\overline{B} \cap \overline{B} \neq \emptyset\}\,.$$

Note that S is finite because the action of G is proper, and that $1 \in S^{-1} = S$ by the definition of S. If $S = G$, then there is nothing to prove; we assume, therefore, that $G \neq S$.

Step 2: Outside of the generating set.

Now consider

$$2d := \inf\{\mathrm{dist}(\overline{B}, g\overline{B}) \ ; \ g \in G \setminus S\}.$$

Pick $g \in G \setminus S$; the distance $\mathrm{dist}(\overline{B}, g\overline{B})$ is a positive constant R, by the definition of S. The subset $H \subset G$ consisting of elements $h \in G$ such that $\mathrm{dist}(\overline{B}, h\overline{B}) \leqslant R$, is contained in the set

$$\{g \in G \ : \ g\overline{B}(x, D+R) \cap \overline{B}(x, D+R) \neq \emptyset\}$$

and, hence, the subset H is finite. Now,

$$\inf\{\text{dist}(\overline{B}, g\overline{B}) \;:\; g \in G \setminus S\} = \inf\{\text{dist}(\overline{B}, g\overline{B}) \;:\; g \in H \setminus S\}$$

and the latter infimum is over finitely many positive numbers. Therefore, there exists $h_0 \in H \setminus S$ such that $\text{dist}(\overline{B}, h_0\overline{B})$ realizes that infimum, which is, therefore, positive. By the definition, $\text{dist}(\overline{B}, g\overline{B}) < 2d$ implies that $g \in S$.

Step 3: G is finitely generated.

Consider a geodesic $[x, gx] \subset X$ and define

$$k = \left\lfloor \frac{\text{dist}(x, gx)}{d} \right\rfloor.$$

Then there exists a finite sequence of points on the geodesic $[x, gx]$,

$$y_0 = x, y_1, \ldots, y_k, y_{k+1} = gx,$$

such that $\text{dist}(y_i, y_{i+1}) \leqslant d$ for every $i \in \{0, \ldots, k\}$. For every $i \in \{1, \ldots, k\}$ let $h_i \in G$ be such that $y_i \in h_i\overline{B}$. We take $h_0 = 1$ and $h_{k+1} = g$. As

$$\text{dist}(\overline{B}, h_i^{-1}h_{i+1}\overline{B}) = \text{dist}(h_i\overline{B}, h_{i+1}\overline{B}) \leqslant \text{dist}(y_i, y_{i+1}) \leqslant d,$$

it follows that $h_i^{-1}h_{i+1} = s_i \in S$, that is, $h_{i+1} = h_i s_i$. Then

$$g = h_{k+1} = s_0 s_1 \cdots s_k. \tag{5.4}$$

We have thus proved that G is generated by S, consequently, G is finitely generated.

Step 4: The quasiisometry.

Since all word metrics on G are bi-Lipschitz equivalent, it suffices to prove Part (2) for the word metric dist_S on G, where S is the finite generating set found as above for the chosen point x. The space X is contained in the $2D$–tubular neighborhood of the image Gx of the orbit map$G \to X$. It, therefore, remains to prove that the orbit map is a quasiisometric embedding. By the equation (5.4),

$$|g|_S \leqslant k + 1 \leqslant \frac{1}{d}\text{dist}(x, gx) + 1.$$

On the other hand, by Lemma 5.28, the orbit map of an isometric properly discontinuous action of a finitely generated group, is L-Lipschitz for some L. Therefore,

$$d|g|_S - d \leqslant \text{dist}(x, gx) \leq L|g|_S,$$

equivalently,

$$d \cdot dist_S(1_G, g) - d \leqslant \text{dist}(x, gx) \leq L dist_S(1_G, g)$$

Since both the word metric dist_S and the metric dist on X are left-invariant with respect to the action of G, in the above inequality, 1_G can be replaced by any element $h \in G$. □

Exercise 5.36. Verify that the orbit map in this proof is $2D$-Lipschitz.

Corollary 5.37. *Given M a compact connected Riemannian manifold, let $\widetilde{M}$ be its universal cover endowed with the pull-back Riemannian metric, so that the fundamental group $\pi_1(M)$ acts isometrically on $\widetilde{M}$.*

Then the group $\pi_1(M)$ is finitely generated, and the metric space $\widetilde{M}$ is quasi-isometric to $\pi_1(M)$ with some word metric.

Thus, the Milnor–Schwarz Theorem provides an answer to the question about the relation between different geometric models of a finitely generated group G: Different models are *quasiisometric* to each other and to the group G equipped with the word metric.

EXERCISE 5.38. Prove the Milnor–Schwarz Theorem replacing the assumption that X is a geodesic metric space by the hypothesis that X is a *quasigeodesic* metric space.

Our next goal is to prove several corollaries and generalizations of Theorem 5.35.

LEMMA 5.39. *Let (X, dist_i), $i = 1, 2$, be proper geodesic metric spaces. Suppose that the action $G \curvearrowright X$ is geometric with respect to both metrics $\mathrm{dist}_1, \mathrm{dist}_2$. Then the identity map*

$$\mathrm{id} : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

is a quasiisometry.

PROOF. The group G is finitely generated by Theorem 5.35; choose a word metric dist_G on G corresponding to any finite generating set. Pick a point $x_0 \in X$; then the orbit maps

$$f_i : (G, \mathrm{dist}_G) \to (X, \mathrm{dist}_i), \quad f_i(g) = g(x_0)$$

are quasiisometries, let $\bar{f}_i$ denote their coarse inverses. Then the map

$$\mathrm{id} : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

is within finite distance from the quasiisometry $f_2 \circ \bar{f}_1$. □

COROLLARY 5.40. *Let $\mathrm{dist}_1, \mathrm{dist}_2$ be as in Lemma 5.39. Then any geodesic γ with respect to the metric dist_1 is a quasigeodesic with respect to the metric dist_2.*

LEMMA 5.41. *Let $G \curvearrowright X$ be a geometric action on a proper geodesic metric space X. Suppose, in addition, that we have an isometric properly discontinuous action $G \curvearrowright X'$ on another metric space X' and a G-equivariant coarsely Lipschitz map $f : X \to X'$. Then f is uniformly proper.*

PROOF. Pick a point $p \in X$ and set $o := f(p)$. We equip G with a word metric corresponding to a finite generating set S of G; then the orbit map $\phi : g \mapsto g(p), \phi : G \to X$ is a quasiisometry by Milnor–Schwarz theorem. We have the second orbit map $\psi : G \to X'$, $\psi(g) = g(p)$. The map ψ is uniformly proper according to Lemma 5.28. We leave it to the reader to verify that

$$\mathrm{dist}(f \circ \phi, \psi) < \infty.$$

Thus, the map $f \circ \phi$ is uniformly proper as well (see Exercise 5.26). Taking $\bar{\phi} : X \to G$, a coarse inverse to ϕ, we see that the composition

$$f \circ \phi \circ \bar{\phi}$$

is uniformly proper too. Since

$$\mathrm{dist}(f \circ \phi \circ \bar{\phi}, f) < \infty,$$

we conclude that f is also uniformly proper. □

Let $G \curvearrowright X, G \curvearrowright X'$ be isometric actions and let $f : X \to X'$ be a quasiisometric embedding. We say that f is *G-quasiequivariant* if for every $g \in G$

$$\mathrm{dist}(g \circ f, f \circ g) \leqslant C,$$

where $C < \infty$ is independent of g.

EXERCISE 5.42. (1) Composition of quasiequivariant maps is again quasiequivariant.

(2) If $f : X \to X'$ is a quasiequivariant quasiisometry, then every coarse inverse $\bar{f} : X' \to X$ to the map f is also quasiequivariant.

LEMMA 5.43. *Suppose that X, X' are proper geodesic metric spaces, G is a group acting geometrically on X and X' respectively. Then there exists a G–quasiequivariant quasiisometry $f : X \to X'$.*

PROOF. Pick points $x \in X$, $x' \in X'$. According to Theorem 5.35, the orbit maps

$$G \to G \cdot x \hookrightarrow X, \quad G \to G \cdot x' \hookrightarrow X'$$

are quasiisometries. The statement now follows from the Exercise 5.42. □

EXERCISE 5.44. Construct an example when in the setting of the lemma there is no quasiisometry $X \to X'$ which is G-equivariant in the traditional sense, i.e.,

$$f \circ g = g \circ g$$

for all $g \in G$.

Below we discuss the relation between quasiisometry and virtual isomorphism, in view of Milnor–Schwarz Theorem.

COROLLARY 5.45. *Let G be a finitely generated group.*

(1) *If G_1 is a finite index subgroup in G, then G_1 is also finitely generated; moreover the groups G and G_1 are quasiisometric.*
(2) *Given a finite normal subgroup N in G, the groups G and G/N are quasiisometric.*
(3) *Thus, two virtually isomorphic (VI) finitely generated groups are quasiisometric (QI).*

PROOF. (1) is a special case of Theorem 5.35, with $G_2 = G$ and X a Cayley graph of G.

(2) follows from Theorem 5.35 applied to the action of the group G on a Cayley graph of the group G/N.

(3) The last part is an immediate consequence of parts (1) and (2). □

The next example shows that VI is not equivalent to QI.

EXAMPLE 5.46. Let A be a matrix diagonalizable over $\mathbb{R}$ in $SL(2, \mathbb{Z})$ so that $A^2 \neq I$. Thus, the eigenvalues λ, λ^{-1} of A have the absolute value $\neq 1$. We will use the notation $Hyp(2, \mathbb{Z})$ for the set of such matrices. Define the action of $\mathbb{Z}$ on $\mathbb{Z}^2$ so that the generator $1 \in \mathbb{Z}$ acts by the automorphism given by A. Let G_A denote the associated semidirect product $G_A := \mathbb{Z}^2 \rtimes_A \mathbb{Z}$. We leave it to the reader to verify that $\mathbb{Z}^2$ is a unique maximal normal abelian subgroup in G_A. By diagonalizing the matrix A, we see that the group G_A embeds as a discrete cocompact subgroup in the Lie group

$$Sol_3 = \mathbb{R}^2 \rtimes_D \mathbb{R}$$

where

$$D(t) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, t \in \mathbb{R}.$$

In particular, G_A is torsion-free. The group Sol_3 has its left-invariant Riemannian metric; hence, G_A acts isometrically on Sol_3, regarded as a metric space. Therefore, every group G_A as above is QI to Sol_3. We now construct two groups G_A, G_B of the above type which are not VI to each other. Pick two matrices $A, B \in Hyp(2, \mathbb{Z})$ such that for every $n, m \in \mathbb{Z} \setminus \{0\}$, A^n is not conjugate to B^m. For instance, take

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}.$$

(The above property of the powers of A and B follows by considering the eigenvalues of A and B and observing that the fields they generate are different quadratic extensions of $\mathbb{Q}$.) The group G_A is QI to G_B since they are both QI to Sol_3. Let us check that G_A is not VI to G_B. First, since both G_A, G_B are torsion-free, it suffices to show that they are not commensurable, i.e., do not contain isomorphic finite index subgroups. Let $H = H_A$ be a finite index subgroup in G_A. Then H intersects the normal rank 2 free abelian subgroup of G_A along a rank 2 free abelian subgroup L_A. The image of H under the quotient homomorphism $G_A \to G_A/\mathbb{Z}^2 = \mathbb{Z}$ has to be an infinite cyclic subgroup, generated by some $n \in \mathbb{N}$. Therefore, H_A is isomorphic to $\mathbb{Z}^2 \rtimes_{A^n} \mathbb{Z}$. For the same reason, $H_B \cong \mathbb{Z}^2 \rtimes_{B^m} \mathbb{Z}$. Any isomorphism $H_A \to H_B$ has to carry L_A isomorphically to L_B. However, this would imply that A^n is conjugate to B^m. Contradiction.

Example 5.47. Another example where QI does not imply VI is as follows. Let S be a closed oriented surface of genus $n \geqslant 2$. Let $G_1 = \pi_1(S) \times \mathbb{Z}$. Let M be the total space of the unit tangent bundle $UT(S)$ of S. Then the fundamental group $G_2 = \pi_1(M)$ is a nontrivial central extension of $\pi_1(S)$:

$$1 \to \mathbb{Z} \to G_2 \to \pi_1(S) \to 1,$$

$$G_2 = \left\langle a_1, b_1, \ldots, a_n, b_n, t | [a_1, b_1] \cdots [a_n, b_n] t^{2n-2}, [a_i, t], [b_i, t], i = 1, \ldots, n \right\rangle.$$

We leave it to the reader to check that passing to any finite index subgroup in G_2 does not make it a trivial central extension of the fundamental group of a hyperbolic surface. On the other hand, since the group $\pi_1(S)$ is hyperbolic, the groups G_1 and G_2 are quasiisometric, see §9.19.

One more example of quasiisometry (which comes from a virtual isomorphism) is the following:

Example 5.48. All non-abelian free groups of finite rank are quasiisometric to each other.

Proof. We present two proofs: One is algebraic and the other is geometric.

1. **Algebraic proof.** We claim that all free groups $F_n, 2 \leqslant n < \infty$, are virtually isomorphic. By Proposition 4.81, for every $1 < m < \infty$, the group F_2 contains a finite index subgroup isomorphic to F_m. Since virtual isomorphism is a transitive relation, which implies quasiisometry, the claim follows.

2. **Geometric proof.** The Cayley graph of F_n with respect to a set of n generators and their inverses is the regular simplicial tree of valence $2n$.

We claim that all regular simplicial trees of valence at least 3 (equipped with the standard metrics) are quasiisometric. Let $\mathcal{T}_k$ denote the regular simplicial tree of valence k; we will show that $\mathcal{T}_3$ is quasiisometric to $\mathcal{T}_k$ for every $k \geqslant 4$.

We construct a countable collection $\mathcal{C}$ of pairwise-disjoint embedded edge-paths c of length $k-3$ in $\mathcal{T}_k$, such that every vertex in $\mathcal{T}_k$ belongs to exactly one such path. See in Figure 5.1, where the paths $c \in \mathcal{C}$ are drawn in tick lines. Let $\mathcal{T}$ denote the tree obtained from $\mathcal{T}_k$ by collapsing each path $c \in \mathcal{C}$ to a single vertex. We leave it to the reader to verify that the quotient tree $\mathcal{T}$ is isomorphic to the valence k tree $\mathcal{T}_k$. The quotient map

$$\mathfrak{q} = \mathfrak{q}_k : \mathcal{T}_3 \to \mathcal{T}_k$$

is a morphism of trees. We also leave it to the reader to verify that the map $\mathfrak{q}$ satisfies the inequality

$$\frac{1}{k-2}\operatorname{dist}(x,y) - 1 \leqslant \operatorname{dist}(\mathfrak{q}(x), \mathfrak{q}(y)) \leqslant \operatorname{dist}(x,y)$$

for all vertices $x, y \in V(\mathcal{T}_3)$. Therefore, $\mathfrak{q}$ is a surjective quasiisometric embedding and, hence, a quasiisometry. □

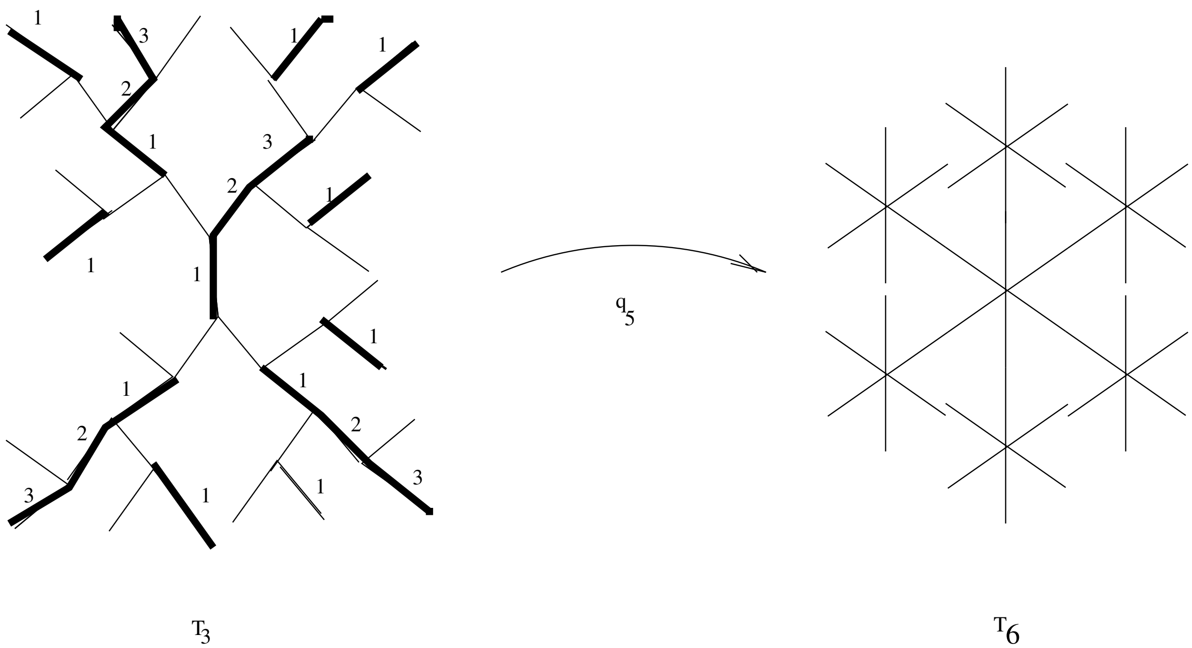


FIGURE 5.1. All regular simplicial trees are quasiisometric.

5.3. A metric version of the Milnor–Schwarz Theorem

In the case of a Riemannian manifold, or more generally a metric space, without a geometric action of a group, one can still use a purely metric argument and create a discretization of the space, that is a simplicial graph quasiisometric to the space. We begin with a few simple observations.

LEMMA 5.49. *Let X and Y be two discrete metric spaces that are bi-Lipschitz equivalent. If X is uniformly discrete, then so is Y.*

PROOF. Assume $f : X \to Y$ is an L–bi-Lipschitz bijection, where $L \geqslant 1$, and assume that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a function such that for every $r > 0$ every closed ball $\overline{B}(x, r)$ in X contains at most $\phi(r)$ points. Every closed ball $\overline{B}(y, R)$ in Y is in 1-to-1 correspondence with a subset of $B(f^{-1}(y), LR)$, whence it contains at most $\phi(LR)$ points. □

Notation: Let A be a subset in a metric space. We denote by $\mathcal{G}_\kappa(A)$ the simplicial graph with set of vertices A and set of edges

$$\{(a_1, a_2) \mid a_1, a_2 \in A,\ 0 < \mathrm{dist}(a_1, a_2) \leqslant \kappa\}\,.$$

In other words, $\mathcal{G}_\kappa(A)$ is the 1-skeleton of the Rips complex $\mathrm{Rips}_\kappa(A)$.

As usual, we will equip each component of the graph $\mathcal{G}_\kappa(A)$ with the standard metric.

THEOREM 5.50. (1) *Let (X, dist) be a proper geodesic metric space (equivalently, a complete, locally compact length metric space, see Theorem 1.60). Let $N \subset X$ be an ε–separated δ–net, where $0 < \varepsilon < 2\delta < 1$ and let $\mathcal{G}$ be the metric graph $\mathcal{G}_{8\delta}(N)$. Then the graph $\mathcal{G}$ is connected, and the metric space (X, dist) is quasiisometric to the graph $\mathcal{G}$. More precisely, for all $x, y \in N$ we have*

$$\frac{1}{8\delta}\mathrm{dist}_X(x, y) \leqslant \mathrm{dist}_{\mathcal{G}}(x, y) \leqslant \frac{3}{\varepsilon}\mathrm{dist}_X(x, y)\,. \tag{5.5}$$

(2) *If, moreover, (X, dist) is either a complete Riemannian manifold of bounded geometry or a metric simplicial complex of bounded geometry, then $\mathcal{G}$ is a graph of bounded geometry (see Definition 2.33).*

PROOF. (1) Our proof is modeled on the one of Milnor–Schwarz Theorem. Let x, y be two points in N. If $\mathrm{dist}_X(x, y) \leqslant 8\delta$ then, by construction, $\mathrm{dist}_{\mathcal{G}}(x, y) = 1$ and both inequalities in (5.5) hold. Let us suppose that $\mathrm{dist}_X(x, y) > 8\delta$.

The distance $\mathrm{dist}_{\mathcal{G}}(x, y)$ is the length s of an edge-path $e_1 e_2 \ldots e_s$, where x is the tail vertex of e_1 and y is the head vertex of e_s. It follows that

$$\mathrm{dist}_{\mathcal{G}}(x, y) = s \geqslant \frac{1}{8\delta}\mathrm{dist}_X(x, y)\,.$$

The distance $\mathrm{dist}_X(x, y)$ is the length of a geodesic $\mathfrak{c} \colon [0, \mathrm{dist}_X(x, y)] \to X$. Let

$$t_0 = 0, t_1, t_2, \ldots, t_m = \mathrm{dist}_X(x, y)$$

be a sequence of numbers in $[0, \mathrm{dist}_X(x, y)]$ such that $5\delta \leqslant t_{i+1} - t_i \leqslant 6\delta$, for every $i \in \{0, 1, \ldots, m-1\}$.

Let $x_i = \mathfrak{c}(t_i)$, $i \in \{0, 1, 2, \ldots, m\}$. For every $i \in \{0, 1, 2, \ldots, m\}$ there exists $w_i \in N$ such that $\text{dist}_X(x_i, w_i) \leqslant \delta$. We note that $w_0 = x, w_m = y$. The choice of t_i implies that

$$3\delta \leqslant \text{dist}_X(w_i, w_{i+1}) \leqslant 8\delta\,, \quad \text{for every } i \in \{0, \ldots, m-1\}$$

In particular:

- w_i and w_{i+1} are the endpoints of an edge in $\mathcal{G}$, for every $i \in \{0, \ldots, m-1\}$;
- $\text{dist}_X(x_i, x_{i+1}) \geqslant \text{dist}(w_i, w_{i+1}) - 2\delta \geqslant \text{dist}(w_i, w_{i+1}) - \frac{2}{3}\text{dist}(w_i, w_{i+1}) = \frac{1}{3}\text{dist}(w_i, w_{i+1})$.

We can then write

(5.6)

$$\text{dist}_X(x, y) = \sum_{i=0}^{m-1} \text{dist}_X(x_i, x_{i+1}) \geqslant \frac{1}{3}\sum_{i=0}^{m-1} \text{dist}(w_i, w_{i+1}) \geqslant \frac{\varepsilon}{3} m \geqslant \frac{\varepsilon}{3}\text{dist}_{\mathcal{G}}(x, y)\,.$$

This inequality implies both connectivity of the graph $\mathcal{G}$ and the required quasi-isometry estimates.

(2) According to the Example 2.34, the graph $\mathcal{G}$ has bounded geometry if and only if its set of vertices with the induced simplicial distance is uniformly discrete. Lemma 5.49 implies that it suffices to show that the set of vertices of $\mathcal{G}$ (i.e., the net N) with the metric induced from X is uniformly discrete.

When X is a Riemannian manifold, this follows from Lemma 2.31. When X is a simplicial complex this follows from the fact that the set of vertices of X is uniformly discrete. □

Note that one can also discretize a Riemannian manifold M (i.e., of replace M by a quasiisometric simplicial complex) using Theorem 2.36, which implies:

THEOREM 5.51. *Every Riemannian manifold M of bounded geometry is quasi-isometric to a bounded geometry simplicial complex homeomorphic to M.*

5.4. Topological coupling

In this section we describe an alternative, and sometimes useful, *dynamical* criterion for quasiisometry between groups. This alternative definition then motivates the notion of *measure-equivalence* between groups. Neither notion will be used elsewhere in the book.

We begin with Gromov's interpretation of quasiisometry between groups using the language of topological actions.

A *topological coupling* of topological groups G_1, G_2, is a metrizable locally compact topological space X, together with two commuting cocompact properly discontinuous topological actions $\rho_i : G_i \to Homeo(X), i = 1, 2$. (The actions commute if and only if $\rho_1(g_1)\rho_2(g_2) = \rho_2(g_2)\rho_1(g_1)$ for all $g_i \in G_i$, $i = 1, 2$.) Note that the actions ρ_i are not required to be isometric. Below we will see some natural examples of topological couplings.

The following theorem was first proven by Gromov in [**Gro93**]; see also [**dlH00**, page 98].

THEOREM 5.52. *If G_1, G_2 are finitely generated groups, then G_1 is QI to G_2 if and only if there exists a topological coupling between these discrete groups.*

Proof. 1. Suppose that G_1 is QI to G_2. Then there exists an (L, A)-quasiisometry $\mathfrak{q} : G_1 \to G_2$. The map $\mathfrak{q}$ is $(L+A)$–Lipschitz (see Lemma 5.7). Consider the space X of (L, A)-quasiisometric maps $G_1 \to G_2$. We equip X with the topology of pointwise convergence. By the Arzela–Ascoli Theorem, X is locally compact.

The groups G_1, G_2 act on X as follows:

$$\rho_1(g_1)(f) := f \circ g_1^{-1}, \quad \rho_2(g_2)(f) := g_2 \circ f, \quad f \in X.$$

It is clear that these actions commute and are topological. For each (L, A)-quasiisometry $f \in X$, there exist $g_1 \in G_1, g_2 \in G_2$ such that

$$g_2 \circ f(1_{G_1}) = 1_{G_1}, f \circ g_1^{-1}(1_{G_2}) \in B(1, A) \subset G_2.$$

Therefore, by the Arzela–Ascoli Theorem, both actions (of G_1 and of G_2) are cocompact. We will check that ρ_2 is properly discontinuous as the case of ρ_1 is analogous. Let $K \subset X$ be a compact subset. Then there exists $R < \infty$ so that for every $f \in K$,

$$f(1_{G_1}) \in B(1, R).$$

If $g_2 \in G_2$ is such that $g_2 \circ f \in K$ for some $f \in K$, then

$$g_2(B(1_{G_2}, R)) \cap B(1_{G_2}, R) \neq \emptyset. \tag{5.7}$$

Since the action of G_2 on itself is free, it follows that the collection of $g_2 \in G_2$ satisfying (5.7) is finite. Hence, ρ_2 is properly discontinuous.

Lastly, the space X is metrizable, since it is locally compact, 2nd countable and Hausdorff; more explicitly, one can define distance between functions as the Gromov–Hausdorff distance between their graphs. (Note that this metric is G_1–invariant.)

2. Suppose that X is a topological coupling of G_1 and G_2. If X were a geodesic metric space and the actions of G_1, G_2 were isometric, we would not need commutation of these action (as Milnor-Schwarz Theorem would apply). However, there are examples of QI groups which do not act geometrically on the same geodesic metric space, see Theorem 5.35. Nevertheless, the construction of a quasiisometry below is pretty much the same as in the proof of the Milnor-Schwarz Theorem.

Since $G_i \curvearrowright X$ is cocompact, there exists a compact $K \subset X$ so that $G_i K = X$; pick a point $p \in K$. Then for each $g_i \in G_i$ there exists $\phi_i(g_i) \in G_{i+1}$ such that $g_i(p) \in \phi_i(g_i)(K)$; here and below i is taken modulo 2. We, thus, have the maps $\phi_i : G_i \to G_{i+1}$, $i = 1, 2$.

a. Let us check that these maps are Lipschitz. Let $s \in S_i$, a finite generating set of G_i, we will use the word metric on G_i with respect to S_i, $i = 1, 2$. Define C to be the union

$$\bigcup_{s \in S_1} s(K) \cup \bigcup_{s \in S_2} s(K).$$

Since ρ_i's are properly discontinuous actions, the sets

$$G_i^C := \{h \in G_i : h(C) \cap C \neq \emptyset\}, i = 1, 2,$$

are finite. Therefore, the word-lengths of the elements of these sets are bounded by some $L < \infty$. Suppose now that $g_{i+1} = \phi_i(g_i)$, $s \in S_i$. Then

$$g_i(p) \in g_{i+1}(K), s g_i(p) \in g'_{i+1}(K),$$

for some $g'_{i+1} \in G_{i+1}$. Therefore,

$$s g_{i+1}(K) \cap g'_{i+1}(K) \neq \emptyset$$

and, hence,

$$g_{i+1}^{-1} g'_{i+1}(K) \cap s(K) \neq \emptyset.$$

(This is where we are using the fact that the actions of G_1 and G_2 on X commute.) Therefore, $g_{i+1}^{-1} g'_{i+1} \in G_{i+1}^C$, which implies that $d(g_{i+1}, g'_{i+1}) \leqslant L$. Consequently, ϕ_i is L–Lipschitz.

b. Set

$$\phi_i(g_i) = g_{i+1}, \phi_{i+1}(g_{i+1}) = g'_i.$$

Then $g_i(K) \cap g'_i(K) \neq \emptyset$ and, hence, $g_i^{-1} g'_i \in G_i^C$. Therefore, we conclude that

$$\text{dist}(\phi_{i+1} \circ \phi_i, \text{id}_{G_i}) \leqslant L,$$

and, thus, the maps ϕ_1, ϕ_2 are coarse inverse to each other. Thus, $\phi_1 : G_1 \to G_2$ is a quasiisometry. ☐

The more useful direction of this theorem is, of course, from QI to a topological coupling, see e.g. [**Sha04, Sau06**].

Definition 5.53. Two groups G_1, G_2 are said to have a *common geometric model* if there exists a proper quasigeodesic metric space X such that G_1, G_2 both act geometrically on X.

In view of Theorem 5.35, if two groups have a common geometric model then they are quasiisometric. The following theorem shows that the converse is false:

Theorem 5.54 (L. Mosher, M. Sageev, K. Whyte, [**MSW03**]). *Consider the groups*

$$G_1 := \mathbb{Z}_p * \mathbb{Z}_p, \quad G_2 := \mathbb{Z}_q * \mathbb{Z}_q,$$

where p, q are distinct odd primes. Then the groups G_1, G_2 are quasiisometric (since they are virtually isomorphic to the free group on two generators) but do not have a common geometric model.

This theorem, in particular, implies that in Theorem 5.52 one cannot assume that both group actions are isometric (for the same metric).

Measure–equivalence. The interpretation of quasiisometry of groups in terms of topological couplings was generalized by M. Gromov [**Gro93**] in the measure-theoretic context:

Definition 5.55. A *measurable coupling* for two groups G_1, G_2 is a measure space (Ω, μ) such that G_1, G_2 admit commuting measure-preserving free actions on (Ω, μ), which both admit a finite measure fundamental set in (Ω, μ). Groups G_1, G_2 are called *measure-equivalent* if they admit a measurable coupling.

We refer the reader to [**Gab05, Gab10**] for further discussion of this fruitful concept.

5.5. Quasi-actions

The notion of an *action* of a group on a space is frequently replaced, in the context of quasiisometries, by the one of a *quasiaction*. Recall that an *action* of a group G on a set X is a homomorphism $\phi : G \to Aut(X)$, where $Aut(X)$ is the group of bijections $X \to X$. Since quasiisometries are defined only up to "bounded error", the concept of a homomorphism has to be modified when we use quasiisometries.

Definition 5.56. Let G be a group and X be a metric space. An (L, A)-quasiaction of G on X is a map $\phi : G \to Map(X, X)$, such that:

- $\phi(g)$ is an (L, A)-quasiisometry of X for all $g \in G$.
- $d(\phi(1_G), \mathrm{id}_X) \leqslant A$.
- $d(\phi(g_1 g_2), \phi(g_1)\phi(g_2)) \leqslant A$ for all $g_1, g_2 \in G$.

By abusing the notation, we will denote quasiactions by $\phi : G \curvearrowright X$, even though, what we have is not an action.

The last two conditions can be informally summarized as: ϕ is "almost" a homomorphism with the error A.

Similarly, a *quasihomomorphism* from a group to another group equipped with a left-invariant metric is a map

$$\phi : G_1 \to (G_2, \mathrm{dist})$$

which satisfies properties (2) and (3) of a quasiaction with respect to the metric dist on G_2 (the property (1) is automatic since G_1 will quasiact via isometries on G_2).

Example 5.57. Suppose that G is a group and $\phi : G \to \mathbb{R} \subset \mathrm{Isom}(\mathbb{R})$ is a function. Then ϕ, of course, satisfies (1), while properties (2) and (3) are equivalent to the single condition:

$$|\phi(g_1 g_2) - \phi(g_1) - \phi(g_2)| \leqslant A.$$

In other words, such maps ϕ are *quasimorphisms*, see Definition 3.106.

We refer the reader to [**FK13**] for the discussion of quasihomomorphisms with noncommutative targets.

We can also define proper discontinuity and cocompactness for quasiactions by analogy with isometric actions:

Definition 5.58. Let $\phi : G \curvearrowright X$ be a quasiaction.

1. We say that ϕ is *properly discontinuous* if for every $x \in X, R \in \mathbb{R}_+$, the set

$$\{g \in G | d(x, \phi(g)(x)) \leqslant R\}$$

is finite. Note that if X proper and ϕ is an isometric action, this definition is equivalent to proper discontinuity of the action $\phi : G \curvearrowright X$.

2. We say that ϕ is *cobounded* if there exists $x \in X, R \in \mathbb{R}_+$ such that for every $x' \in X$ there exists $g \in G$, for which $d(x', \phi(g)(x)) \leqslant R$.

3. Lastly, we say that a quasiaction ϕ is *geometric* if it is both properly discontinuous and cobounded.

Exercise 5.59. Let $QI(X)$ denote the group of (equivalence classes of) quasi-isometries $X \to X$. Show that every quasiaction $G \curvearrowright X$ determines a homomorphism $\widehat{\phi} : G \to QI(X)$ given by composing ϕ with the projection to $QI(X)$.

The *kernel* of the quasiaction $\phi : G \curvearrowright X$ is the kernel of the homomorphism $\widehat{\phi}$.

Exercise 5.60. Construct an example of a geometric quasiaction $G \curvearrowright \mathbb{R}$ whose kernel is the entire group G.

Below we explain how quasiactions appear in the context of QI rigidity problems. Suppose that G_1, G_2 are groups, $\psi_i : G_i \curvearrowright X_i$ are isometric actions; for instance, X_i could be G_i or its Cayley graph. Suppose that $f : X_1 \to X_2$ is a quasiisometry with coarse inverse $\bar{f}$. We then define a *conjugate* quasiaction $\phi = f^*(\psi_2)$ of G_2 on X_1 by

$$\phi(g) = \bar{f} \circ g \circ f. \tag{5.8}$$

More generally, we say that two quasiactions $\psi_i : G \curvearrowright X_i$ are (quasi) *conjugate* if there exists a quasiisometry $f : X_1 \to X_2$, such that ψ_1 and $f^*(\psi_2)$ project to the same homomorphism

$$G \to QI(X_1).$$

LEMMA 5.61. *Suppose that $\psi : G \curvearrowright X_2$ is a quasiaction, $f : X_1 \to X_2$ is a quasiisometry and $\phi = f^*(\psi)$ is defined by the formula* (5.8). *Then:*

1. $\phi = f^*(\psi)$ is a quasiaction of G on X_1.

2. If ψ is properly discontinuous (respectively, cobounded, or geometric), then so is ϕ.

PROOF. 1. Suppose that f is an (L, A)-quasiisometry with coarse inverse $\bar{f}$. In view of Exercise 5.11, it is clear that ϕ satisfies Parts 1 and 2 of the definition of a quasiaction; we only have to verify Part (3):

$$\operatorname{dist}(\phi(g_1 g_2), \phi(g_1)\phi(g_2)) = \operatorname{dist}(\bar{f} g_1 g_2 f, \bar{f} g_1 f \bar{f} g_2 f) \leqslant LA + A.$$

2. We will verify the statement about properly discontinuous quasiactions, since the proof for cobounded quasiactions is similar. Pick $x \in X, R \in \mathbb{R}_+$, and consider the subset

$$G_{x,R} = \{g \in G | d(x, \phi(g)(x)) \leqslant R\} \subset G.$$

By the definition, $\phi(g)(x) = \bar{f} g f(x)$. Thus, $d(x, g(x)) \leqslant LR + 2A$. Hence, by proper discontinuity of the action $\psi : G \curvearrowright X_2$, the set $G_{x,R}$ is finite. ☐

COROLLARY 5.62. *Let G_1 and G_2 be finitely generated quasiisometric groups and let $f : G_1 \to G_2$ be a quasiisometry. Then:*

1. The quasiisometry f induces (by conjugating actions and quasiactions on G_2) an isomorphism $QI(f) : QI(G_2) \to QI(G_1)$ and a homomorphism $f_* : G_2 \to QI(G_1)$

2. The homomorphism f_* is quasiinjective: For every $K \geqslant 0$, the set of $g \in G_2$ such that $\operatorname{dist}(f_*(g), \mathrm{id}_{G_1}) \leqslant K$, is finite.

PROOF. The isomorphism $QI(f) : QI(G_2) \to QI(G_1)$ is defined *via* the formula (5.8). The inverse to this homomorphism is defined by switching the roles of f and $\bar{f}$. We leave it to the reader to verify that $QI(f)$ is an isomorphism. To define f_* we compose the homomorphism $G_2 \to QI(G_2)$ with $QI(f)$. Quas-injectivity of f_* follows from the proper discontinuity of the action $G_2 \curvearrowright G_2$ *via* the left multiplication. ☐

REMARK 5.63. For many groups $G = G_1$, if $h : G \to G$ is an (L, A)-quasiisometry which belongs to $\mathcal{B}(G)$, we also have $\operatorname{dist}(f, Id_G) \leqslant D(L, A)$, where $D(L, A)$ depends only on L, A and (G, d_S) but not on f. For instance, this holds when G is a non-elementary hyperbolic group, see Lemma 9.112. This is also true for isometry groups of irreducible symmetric spaces and Euclidean buildings and many other spaces, see e.g. [**KKL98**]. In this situation, the kernel of f_* above is actually finite.

The following theorem is a weak converse to the construction of a conjugate quasiaction:

THEOREM 5.64 (B. Kleiner, B. Leeb, [**KL09**]). *Suppose that $\phi : G \curvearrowright X_1$ is a quasiaction. Then there exists a metric space X_2, a quasiisometry $f : X_1 \to X_2$ and an isometric action $\psi : G \curvearrowright X_2$, such that f conjugates ψ to ϕ.*

Thus, every quasiaction is conjugate to an isometric action, but, *a priori*, on a different metric space. The main issue of the QI (quasiisometric) rigidity, discussed in the next section is:

Can one, under some conditions, take $X_2 = X_1$? More precisely: Given a quasiaction $G \curvearrowright X$ of a group G on a space $X = X_1$, can one find a conjugate isometric action $G \curvearrowright X$?

5.6. Quasi-isometric rigidity problems

So far, we succeeded in converting finitely generated groups into metric spaces, i.e., *treating groups as geometric objects.* All these spaces are quasiisometric to each other, but we would like to *reconstruct* (to the extent possible) the group G, as an *algebraic object*, from its geometric models (defined only up to a quasiisometry). In other words, we would like to know, to which extent the "geometrization map"

$$geo : \text{Finitely generated groups} \to \text{metric spaces/quasiisometry}$$

is injective?

Corollary 5.45 establishes a limitation on injectivity of *geo*: Virtually isomorphic groups are quasiisometric to each other. Therefore, the best we can hope for, is to recover a group from its (coarse) geometry up to virtual isomorphisms.

DEFINITION 5.65. 1. A (finitely generated) group G is called *QI rigid* if every group G' which is quasiisometric to G is, in fact, virtually isomorphic to G.

2. A group G is called *strongly QI rigid* if the natural map $Comm(G) \to QI(G)$ (from the group of virtual automorphisms of G to the group of self-quasiisometries of G) is surjective.

3. A subclass $\mathcal{G}$ of the class of all (finitely generated) groups is called *QI rigid* if each group G which is quasiisometric to a member of $\mathcal{G}$, is virtually isomorphic to a member of $\mathcal{G}$.

4. A group G in a subclass $\mathcal{G}$ (of all groups) is *QI rigid within* $\mathcal{G}$ if any $G' \in \mathcal{G}$ which is quasiisometric to G, is virtually isomorphic to G.

In the purely geometric context, one can ask if a quasiisometry between metric spaces is within finite distance from an isometry of these spaces:

DEFINITION 5.66. 1. A metric space X is called *strongly QI rigid* if the natural map $\mathrm{Isom}(X) \to QI(X)$ is surjective.

2. A more quantitative version of this property is the *uniform QI rigidity*: A space X is *uniformly QI rigid* if there exists $D(X, L, A) \in \mathbb{R}_+$ such that every (L, A)-quasiisometry $X \to X$ is within distance $\leqslant D(X, L, A)$ from an isometry $X \to X$.

3. More restrictively, one talks about QI rigidity *within a subclass* $\mathcal{M}$ of the class of all metric spaces, by requiring that any two quasiisometric spaces in $\mathcal{M}$ are, in fact, isometric.

A *QI rigidity theorem* is a theorem which establishes QI rigidity in the sense of any of the above definitions.

Most proofs of QI rigidity theorems proceed along the following route:

1. Suppose that the groups G_1, G_2 are quasiisometric. Find a "nice space" X_1 on which G_1 acts geometrically. Take a quasiisometry $f : X_1 \to X_2 = G_2$, where $\psi : G_2 \curvearrowright G_2$ is the action *via* the left multiplication.

2. Define the conjugate quasiaction $\phi = f^*(\psi)$ of G_2 on X_1.

3. Show that the quasiaction ϕ has finite kernel (or, at least, identify the kernel, prove that it is, say, abelian).

4. Extend, if necessary, the quasiaction $G_2 \curvearrowright X_1$ to a quasiaction $\widehat{\phi}$ on a larger space $\widehat{X}_1$.

5. Show that $\widehat{\phi}$ has the same projection to $QI(\widehat{X}_1)$ as a isometric action $\phi' : G_2 \curvearrowright \widehat{X}_1$ by verifying, for instance, that $\widehat{X}_1$ has very few quasiisometries, namely, every quasiisometry of X is within finite distance from an isometry. (Well, maybe no all quasiisometries of $\hat{X}_1$, but the ones which extend from X_1.) Then conclude either that $G_2 \curvearrowright \widehat{X}_1$ is geometric, or, that the isometric actions of G_1, G_2 are commensurable, i.e., the images of G_1, G_2 in $\mathrm{Isom}(\widehat{X}_2)$ have a common subgroup of finite index.

We will see how R. Schwartz's proof of QI rigidity for nonuniform lattices follows this line of arguments: X_1 will be a truncated hyperbolic space and $\widehat{X}_1$ will be the hyperbolic space itself. The same is true for QI rigidity of higher rank non-uniform lattices (A. Eskin's theorem [**Esk98**]). This is also true for uniform lattices in the isometry groups of nonpositively curved symmetric spaces other than $\mathbb{H}^n$ and $\mathbb{CH}^n$ (P. Pansu, [**Pan89**], B. Kleiner and B. Leeb [**KL98b**]; A. Eskin and B. Farb [**EF97b**]), except one does not have to enlarge X_1. Another example of such argument is the proof by M. Bourdon and H. Pajot [**BP00**] and X. Xie [**Xie06**] of QI rigidity of groups acting geometrically on 2-dimensional hyperbolic buildings.

5'. Part 5 may fail if X has too many quasiisometries, e.g. if $X_1 = \mathbb{H}^n$ or $X_1 = \mathbb{CH}^n$. Then, instead, one shows that every geometric quasiaction $G_2 \curvearrowright X_1$ is quasiconjugate to a geometric (isometric!) action. We will see such a proof in the case of the Sullivan–Tukia rigidity theorem for uniform lattices in $\mathrm{Isom}(\mathbb{H}^n), n \geqslant 3$. Similar arguments apply in the case of groups quasiisometric to the hyperbolic plane.

Not all quasiisometric rigidity theorems are proven in this fashion. An alternative route is to show QI rigidity of a certain algebraic property (P) is to show that it is equivalent to some geometric property (P'), which is QI invariant. Examples of such proofs are QI rigidity of the class of virtually nilpotent groups and of virtually free groups. The first property is equivalent, by Gromov's theorem, to the polynomial growth. The argument in the second case is less direct (see Theorem 18.44), but the key fact is that the geometric condition of having infinitely many ends is equivalent to the algebraic condition that a group splits (as a graph of groups) over a finite subgroup.

5.7. The growth function

Suppose that X is a discrete metric space (see Definition 1.50) and $x \in X$ is a base-point. We define the *growth function*

$$\mathfrak{G}_{X,x}(R) := \text{card } \bar{B}(x, R),$$

the cardinality of the closed R-ball centered at x. Similarly, given a connected simplicial complex X or a graph (equipped with the standard metric) and a vertex v as a base-point, the *growth function* of X is the growth function of its set of vertices with the base-point v.

We refer the reader to Notation 1.4 for the equivalence relation $\asymp$ between functions used below.

Lemma 5.67 (Equivalence class of growth is QI invariant.). *If* (X, x_0) *and* (Y, y_0) *are quasiisometric uniformly discrete pointed spaces, then* $\mathfrak{G}_{X,x_0} \asymp \mathfrak{G}_{Y,y_0}$.

Proof. Let $f : X \to Y, \bar{f} : Y \to X$ be L-Lipschitz maps which are coarse inverse to each other (see Definition 5.5). We assume that $f, \bar{f}$ satisfy

$$L^{-1}d(x, x') - A \leqslant d(f(x), f(x')), \quad L^{-1}d(y, y') - A \leqslant d(\bar{f}(y), \bar{f}(y')).$$

Let $D = \max(d(f(x_0), y_0), d(x_0, \bar{f}(y_0))$. Then for each $R > 0$,

$$f(\bar{B}(x_0, R)) \subset \bar{B}(y_0, LR + D), \quad \bar{f}(\bar{B}(y_0, R)) \subset \bar{B}(x_0, LR + D),$$

while $f(x) = f(x')$ implies $d(x, x') \leqslant AL$. The same applies to the map $\bar{f}$. Since the spaces X and Y are uniformly discrete, both maps f, $\bar{f}$ have multiplicity $\leqslant m$, where m is an upper bound for the cardinalities of closed LA-balls in X and Y. It follows that

$$\text{card } \bar{B}(x_0, R) \leqslant m \, \text{card } \bar{B}(y_0, LR + D)$$

and

$$\text{card } \bar{B}(y_0, R) \leqslant m \, \text{card } \bar{B}(x_0, LR + D). \quad \square$$

Corollary 5.68. $\mathfrak{G}_{X,x} \asymp \mathfrak{G}_{X,x'}$ *for all* $x, x' \in X$.

Exercise 5.69. Prove that the lemma and the corollary also hold for simplicial complexes and graphs of bounded geometry.

Henceforth we will suppress the choice of the base-point in the notation for the growth function.

Exercise 5.70. Show that for each (uniformly discrete) space X, $\mathfrak{G}_X(R) \preceq e^R$.

For a group G endowed with the word metric dist_S corresponding to a finite generating set S we sometimes will use the notation $\mathfrak{G}_S(R)$ for $\mathfrak{G}_G(R)$. Since G acts transitively on itself, this function does not depend on the choice of a base-point.

Examples 5.71.
(1) If $G = \mathbb{Z}^k$ then $\mathfrak{G}_S \asymp x^k$ for every finite generating set S.
(2) If $G = F_k$ is the free group of finite rank $k \geqslant 2$ and S is the set of k generators then

$$\mathfrak{G}_S(n) = 1 + (q^n - 1)\frac{q + 1}{q - 1}, \quad q = 2k - 1.$$

Exercise 5.72.
(1) Prove the two statements above.

(2) Conclude that $\mathbb{Z}^m$ is quasiisometric to $\mathbb{Z}^n$ if and only if $n = m$. (Cf. Lemma 5.67.)

(3) Compute the growth function for the group $\mathbb{Z}^2$ equipped with the generating set x, y, where $\{x, y\}$ is a basis of $\mathbb{Z}^2$.

(4) Prove that for every $n \geqslant 2$ the group $SL(n, \mathbb{Z})$ has exponential growth.

Proposition 5.73. (1) *If S, S' are two finite generating sets of G then $\mathfrak{G}_S \asymp \mathfrak{G}_{S'}$. Thus one can speak about the growth function $\mathfrak{G}_G$ of a group G, well defined up to the equivalence relation $\asymp$.*

(2) *If G is infinite, $\mathfrak{G}_S|_{\mathbb{N}}$ is strictly increasing.*

(3) *The growth function is* sub-multiplicative*:*

$$\mathfrak{G}_S(r+t) \leqslant \mathfrak{G}_S(r)\mathfrak{G}_S(t)\,.$$

(4) *For each finitely generated group G, $\mathfrak{G}_G(r) \preceq 2^r$.*

Proof. (1) follows immediately from Lemma 5.67 and Milnor–Schwarz theorem.

(2) Consider two integers $n < m$. As G is infinite there exists $g \in G$ at distance $d \geqslant m$ from 1. The shortest path joining 1 and g in Cayley(G, S) can be parameterized as an isometric embedding $p : [0, d] \to \text{Cayley}(G, S)$. The vertex $p(n+1)$ is an element of $\bar{B}(1, m) \setminus \bar{B}(1, n)$.

(3) follows immediately from the fact that

$$\bar{B}(1, n+m) \subseteq \bigcup_{y \in \bar{B}(1,n)} \bar{B}(y, m)\,.$$

(4) follows from the existence of an epimorphism $\pi_S : F(S) \to G$, where S is a finite generating set of G. □

The property (3) implies that the function $\ln \mathfrak{G}_S(n)$ is sub-additive, hence by the Fekete's Lemma, see e.g. [**HP74**, Theorem 7.6.1], there exists a (finite) limit

$$\lim_{n\to\infty} \frac{\ln \mathfrak{G}_S(n)}{n}.$$

Hence, we also get a finite limit

$$\gamma_S = \lim_{n\to\infty} \mathfrak{G}_S(n)^{\frac{1}{n}}\,,$$

called *growth constant*. The property (2) implies that $\mathfrak{G}_S(n) \geqslant n$; whence, $\gamma_S \geqslant 1$.

Definition 5.74. If $\gamma_S > 1$ then G is said to be of *exponential growth*. If $\gamma_S = 1$ then G is said to be of *sub-exponential growth*.

Note that by Proposition 5.73, (1), if there exists a finite generating set S such that $\gamma_S > 1$ then $\gamma_{S'} > 1$ for every other finite generating set S'. Likewise for equality to 1.

The notion of subexponential growth makes sense for (some classes of) general metric spaces.

DEFINITION 5.75. Let (X, dist) be a metric space for which the growth function is defined (e.g. a Riemannian manifold equipped with the Riemannian distance function, a discrete proper metric space, a locally finite simplicial complex). The space X is said to be of *sub-exponential growth* if for some basepoint $x_0 \in X$

$$\limsup_{n\to\infty} \frac{\ln \mathfrak{G}_{x_0,X}(n)}{n} = 0\,.$$

Since for every other basepoint y_0, $\mathfrak{G}_{y_0,X}(n) \leqslant \mathfrak{G}_{x_0,X}\left(n + \mathrm{dist}(x_0, y_0)\right)$, it follows that the definition is independent of the choice of basepoint.

PROPOSITION 5.76. (a) *If H is a finitely generated subgroup in a finitely generated group G then $\mathfrak{G}_H \preceq \mathfrak{G}_G$.*

(b) *If H is a subgroup of finite index in G then $\mathfrak{G}_H \asymp \mathfrak{G}_G$.*

(c) *If N is a normal subgroup in G then $\mathfrak{G}_{G/N} \preceq \mathfrak{G}_G$*

(d) *If N is a finite normal subgroup in G then $\mathfrak{G}_{G/N} \asymp \mathfrak{G}_G$.*

PROOF. (a) If X is a finite generating set of H and S is a finite generating set of G containing X then $\mathrm{Cayley}(H, X)$ is a subgraph of $\mathrm{Cayley}(G, S)$ and $\mathrm{dist}_X(1, h) \geqslant \mathrm{dist}_S(1, h)$ for every $h \in H$. In particular the closed ball of radius r and center 1 in $\mathrm{Cayley}(H, X)$ is contained in the closed ball of radius r and center 1 in $\mathrm{Cayley}(G, S)$.

(b) and (d) are immediate corollaries of Lemma 5.67 and the Milnor–Schwarz theorem.

(c) Let S be a finite generating set in G, and let $\bar{S} = \{sN \mid s \in S, s \notin N\}$ be the corresponding finite generating set in G/N. The epimorphism $\pi : G \to G/N$ maps the ball of center 1 and radius r onto the ball of center 1 and radius r.

□

Let G and H be two groups with finite generating sets S and X, respectively. A homomorphism $\varphi : G \to H$ is called *expanding* if there exist constants $\lambda > 1$ and $C \geqslant 0$ such that for every $g \in G$ with $|g|_S \geqslant C$

$$|\varphi(g)|_X \geqslant \lambda |g|_S\,.$$

Such homomorphisms generalize the notion of Euclidean similarities, which expand lengths of all vectors by a fixed constant.

EXERCISE 5.77. Let G be a group with a finite generating set S and $H \leqslant G$ a finite index subgroup. We equip G with the word metric d_S and equip H with the metric which is the restriction of d_S. Assume that there exists an expanding homomorphism $\varphi : H \to G$ such that $\varphi(H)$ has finite index in G. Prove *Franks' Lemma*, that such group G has polynomial growth.

More importantly, one has the following generalization of Efremovich's theorem **[Efr53]**:

PROPOSITION 5.78 (Efremovich–Schwarz–Milnor). *Let M be a connected complete Riemannian manifold with bounded geometry. If M is quasiisometric to a graph $\mathcal{G}$ with bounded geometry, then the growth function $\mathfrak{G}_{M,x_0}$ and the growth function of $\mathcal{G}$ with respect to an arbitrary vertex v, are equivalent in the sense of the equivalence relation $\asymp$.*

Proof. The manifold M has bounded geometry, therefore its sectional curvature is at least a and at most b for some constants $a \leqslant b$; moreover, there exists a uniform lower bound $2\rho > 0$ on the injectivity radius of M at every point. Let n denote the dimension of M. We let $V(x,r)$ denote volume of r-ball centered at the point $x \in M$ and let $V_a(r)$ denote the volume of the r-ball in the complete simply-connected n-dimensional manifold of constant curvature a.

The fact that the sectional curvature is at least a implies, by Theorem 2.23, Part (1), that for every $r > 0$, $V(x,r) \leqslant V_a(r)$. Similarly, Theorem 2.23, Part (2), implies that the volume $V(x,\rho) \geqslant V_b(\rho)$.

Since M and $\mathcal{G}$ are quasiisometric, by Definition 5.1 it follows that there exist $L \geqslant 1$, $C \geqslant 0$, two $2C$–separated nets A in M and B in $\mathcal{G}$, respectively, and a L–bi-Lipschitz bijection $\mathfrak{q} : A \to B$. Without loss of generality we may assume that $C \geqslant \rho$; otherwise we choose a maximal 2ρ–separated subset A' of A and then restrict $\mathfrak{q}$ to A'.

According to Remark 2.16, (2), we may assume without loss of generality that the base-point x_0 in M is contained in the net A, and that $\mathfrak{q}(x_0) = v$, the base vertex in $\mathcal{G}$.

For every $r > 0$ we have that

$$\mathfrak{G}_{M,x_0}(r) \geqslant \text{card}\,[A \cap B_M\,(x_0, r-C)]\,V_b(\rho) \geqslant \text{card}\left[B \cap B_{\mathcal{G}}\left(1, \frac{r-C}{L}\right)\right] V_b(\rho)$$
$$\geqslant \mathfrak{G}_{\mathcal{G}}\left(\frac{r-C}{L}\right)\frac{V_b(\rho)}{\mathfrak{G}_{\mathcal{G}}(2C)}\,.$$

Conversely,

$$\mathfrak{G}_{M,x_0}(r) \leqslant \text{card}\,[A \cap B_M\,(x_0, r+2C)]\,V_a(2C) \leqslant$$
$$\text{card}\,[B \cap B_{\mathcal{G}}\,(1, L(r+2C))]\,V_a(2C) \leqslant \mathfrak{G}_G\,(L(r+2C))\,V_a(2C)\,.$$

□

Thus, it follows from Theorem 4.35 that considering $\asymp$–equivalence classes of growth functions of universal covers of compact Riemannian manifolds is not different from considering equivalence classes of growth functions of finitely presented groups.

Remark 5.79. Note that in view of Theorem 5.50, every connected Riemannian manifold of bounded geometry is quasiisometric to a graph of bounded geometry.

Question 5.80. What is the set $Growth(groups)$ of the equivalence classes of growth functions of finitely generated groups?

Question 5.81. What are the equivalence classes of growth functions for finitely presented groups?

This question is equivalent to

Question 5.82. What is the set $Growth(manifolds)$ of equivalence classes of growth functions for universal covers of compact connected Riemannian manifolds?

Clearly, $Growth(manifolds) \subset Growth(groups)$. This inclusion is proper since R. Grigorchuk **[Gri84a]** proved that there exist uncountably many nonequivalent growth functions of finitely generated groups, while there are only countably many nonisomorphic finitely presented groups.

We will see later on that:

$$\{\exp(t), t^n, n \in \mathbb{N}\} \subset Growth(manifolds) \subset Growth(groups)$$

One can refine Question 5.82 by defining $Growth_n(manifolds)$ as the set of equivalence classes of growth functions of universal covers of n-dimensional compact connected Riemannian manifolds. Since every finitely presented group is the fundamental group of a closed smooth 4-dimensional manifold and growth function depends only on the fundamental group, we obtain:

$$Growth_4(manifolds) = Growth_n(manifolds), \quad \forall n \geqslant 4.$$

On the other hand:

THEOREM 5.83. *$Growth_2(manifolds) = \{1, t^2, e^t\}$, $Growth_3(manifolds) = \{1, t, t^3, t^4, e^t\}$.*

Below is an outline of the proof. Firstly, in view of classification of surfaces, for every closed connected oriented surface S we have:

(1) If $\chi(S) = 2$ then $\pi_1(S) = \{1\}$ and growth function is trivial.
(2) If $\chi(S) = 0$ then $\pi_1(S) = \mathbb{Z}^2$ and growth function is equivalent to t^2.
(3) If $\chi(S) < 0$ then $\pi_1(S)$ contains a free nonabelian subgroup, so growth function is exponential.

In the case of 3-dimensional manifolds, one has to appeal to Perelman's Geometrization Theorem. We refer to [**Kap01**] for the precise statement and definitions which appear below:

For every closed connected 3-dimensional manifold M one of the following holds:

(1) M admits a Riemannian metric of constant positive curvature, in which case $\pi_1(M)$ is finite and has trivial growth.
(2) M admits a Riemannian metric locally isometric to the product metric $\mathbb{S}^2 \times \mathbb{R}$. In this case growth function is linear.
(3) M admits a flat Riemannian metric, so universal cover of M is isometric to $\mathbb{R}^3$ and growth function is t^3.
(4) M is homeomorphic to the quotient H_3/Γ, where H_3 is the 3-dimensional Heisenberg group and Γ is a uniform lattice in H_3. In this case, in view of Exercise 12.4, growth function is t^4.
(5) The fundamental group of M is solvable but not virtually nilpotent, thus, by Wolf's Theorem (theorem 12.30), the growth function is exponential.
(6) In all other cases, $\pi_1(M)$ contains free nonabelian subgroup; hence, its growth is exponential.

QUESTION 5.84 (J. Milnor [**Mil68b**]). Is it true that the growth of a finitely generated group is either polynomial (i.e., $\mathfrak{G}_S(t) \preceq t^d$ for some integer d) or exponential (i.e., $\gamma_S > 1$)?

R. Grigorchuk in [**Gri83**] (see also [**Gri84a, Gri84b**]) proved that Milnor's question has negative answer, by constructing finitely generated groups of *intermediate growth*, i.e., their growth is superpolynomial but subexponential. More precisely, Grigorchuk proved that for every sub-exponential function f there exists a group G_f of intermediate growth equipped with a finite generating set S_f whose growth function $\mathfrak{G}_{S_f}(n)$ is larger than $f(n)$ for infinitely many n. A. Erschler in [**Ers04**] adapted Grigorchuk's arguments to show that for every such function f, a direct product $G_f \times G_f$, equipped with the generating set $S = S_f \sqcup S_f$, has the growth function $\mathfrak{G}_S(n)$ satisfying $\mathfrak{G}_S(n) \geqslant f(n)$ for all but finitely many n.

The first explicit of computations of growth functions (up to the equivalence relation $\asymp$) some groups of intermediate growth were done by L. Bartholdi and A. Erschler in [**BE12**]. For every $k \in \mathbb{N}$, they construct examples of torsion groups G_k and of torsion-free groups H_k such that their growth functions satisfy

$$\mathfrak{G}_{G_k}(x) \asymp \exp\left(x^{1-(1-\alpha)^k}\right) ,$$

and

$$\mathfrak{G}_{H_k}(x) \asymp \exp\left(\log x \, (x^{1-(1-\alpha)^k}\right) ,$$

Here, α is the number satisfying $2^{3-\frac{3}{\alpha}} + 2^{2-\frac{2}{\alpha}} + 2^{1-\frac{1}{\alpha}} = 2$.

We note that all currently known groups of intermediate growth have growth larger than $2^{\sqrt{n}}$. Existence of finitely presented groups of intermediate growth is unknown. In particular the the currently known examples of groups of intermediate growth do not answer Question 5.82.

5.8. Codimension one isoperimetric inequalities

One can define, in the setting of graphs, the following concepts, inspired by, and closely connected to, notions introduced in Riemannian geometry (see Definitions 2.19 and 2.21). Recall that for a subset $F \subset V, F^c$ denotes its complement in V.

DEFINITION 5.85. An *isoperimetric inequality* in a graph $\mathcal{G}$ of bounded geometry is an inequality satisfied by all finite subsets F of vertices, of the form

$$\text{card}\,(F) \leq f(F) g\,(\text{card}\, E(F, F^c)) ,$$

where f and g are real-valued functions, g defined on $\mathbb{R}_+$.

DEFINITION 5.86. Let Γ be a graph of bounded geometry, with the vertex set V and edge set E. The *Cheeger constant* or the *Expansion Ratio* of the graph Γ is defined as

$$h(\Gamma) = \inf\left\{ \frac{|E(F, F^c)|}{|F|} \ : \ F \text{ is a finite nonempty subset of } V,\ |F| \leqslant \frac{|V|}{2} \right\} .$$

Here $E(F, F^c)$ is edge boundary for both F and F^c, i.e., the set of edges connecting F to F^c (see Definition 1.41). Thus, the condition $|F| \leqslant \frac{|V|}{2}$ insures that, in case V is finite, one picks the smallest of the two sets F and F^c in the definition of the Cheeger constant. Intuitively, finite graphs with small Cheeger constant can be separated by vertex sets which are relatively small comparing to the size of (the smallest component of) their complements. In contrast, graphs with large Cheeger constant are "hard to separate."

EXERCISE 5.87. a. Let Γ be a single circuit with n vertices. Then $h(\Gamma) = \frac{2}{n}$.

b. Let $\Gamma = K_n$ be the complete graph on n vertices, i.e., Γ is the 1-dimensional skeleton of the $n-1$-dimensional simplex. Then

$$h(\Gamma) = \left\lfloor \frac{n}{2} \right\rfloor .$$

The inequalities in (1.4) imply that in every isoperimetric inequality, the edge-boundary can be replaced by the vertex boundary, if one replaces the function

g by an asymptotically equal function (respectively the Cheeger constants by bi-Lipschitz equivalent values). Therefore, in what follows we choose freely whether to work with the edge-boundary or with the vertex-boundary, depending on which one is more convenient.

There exists an isoperimetric inequality satisfied in every Cayley graph of an infinite group.

PROPOSITION 5.88. *Let $\mathcal{G}$ be the Cayley graph of a finitely generated infinite group. For every finite set F of vertices*

$$\operatorname{card}(F) \leqslant [\operatorname{diam}(F) + 1] \operatorname{card}(\partial_V F) . \tag{5.9}$$

PROOF. Assume that $\mathcal{G}$ is the Cayley graph of an infinite group G with respect to a finite generating set S.

Let d be the diameter of F with respect to the word metric dist_S, and let g be an element in G such that $|g|_S = d + 1$. Let $g_0 = 1, g_1, g_2, \ldots, g_d, g_{d+1} = g$ be the set of vertices on a geodesic joining 1 to g.

Given an arbitrary vertex $x \in F$, the element xg is at distance $d + 1$ from x; therefore, by the definition of d it follows that $xg \in F^c$. In the finite sequence of vertices $x, xg_1, xg_2, \ldots, xg_d, xg_{d+1} = xg$ consider the largest i such that $xg_i \in F$. Then $i < d+1$ and $xg_{i+1} \in F^c$, whence $xg_{i+1} \in \partial_V F$, equivalently, $x \in [\partial_V F]\, g_{i+1}^{-1}$.

We have thus proved that $F \subseteq \bigcup_{i=1}^{d+1} [\partial_V F]\, g_i^{-1}$, which implies the inequality (5.9). □

An argument similar in spirit, but more elaborate, allows to relate isoperimetric inequalities and growth functions:

PROPOSITION 5.89 (Coulhon-Saloff-Coste inequality). *Let $\mathcal{G}$ be the Cayley graph of an infinite group G with respect to a finite generating set S, and let d be the cardinality of S. For every finite set F of vertices*

$$|F| \leqslant 2d\, k \operatorname{card}(\partial_V F) , \tag{5.10}$$

where k is the unique integer such that $\mathfrak{G}_S(k-1) \leqslant 2|F| < \mathfrak{G}_S(k)$.

PROOF. Our goal is to show that with the given choice of k, there exists an element $g \in B_S(1, k)$ such that for a certain fraction of the vertices x in F, the right-translates xg are in F^c. In what follows we omit the subscript S in our notation.

We consider the sum

$$\mathcal{S} = \frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \operatorname{card}\{x \in F \mid xg \in F^c\} = \frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \sum_{x \in F} \mathbf{1}_{F^c}(xg) =$$

$$\frac{1}{\mathfrak{G}(k)} \sum_{x \in F} \sum_{g \in B(1,k)} \mathbf{1}_{F^c}(xg) = \frac{1}{\mathfrak{G}(k)} \sum_{x \in F} \operatorname{card}\,[B(x,k) \setminus F] .$$

By the choice of k, the cardinality of each ball $B(x, k)$ is larger than $2|F|$, whence

$$\operatorname{card}\,[B(x,k) \setminus F] \geqslant |F| .$$

The denominator $\mathfrak{G}(k) \leqslant d\mathfrak{G}(k-1) \leqslant 2d|F|$. We, therefore, find as a lower bound for the sum $\mathcal{S}$, the value

$$\frac{1}{2d|F|} \sum_{x \in F} |F| = \frac{|F|}{2d} .$$

It follows that

$$\frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \operatorname{card}\{x \in F \mid xg \in F^c\} \geqslant \frac{|F|}{2d}\,.$$

The latter inequality implies that there exists $g \in B(1,k)$ such that

$$\operatorname{card}\{x \in F \mid xg \in F^c\} \geqslant \frac{|F|}{2d}\,.$$

We now argue as in the proof of Proposition 5.88, and for the element $g \in B(1,k)$ thus found, we consider $g_0 = 1, g_1, g_2, \dots, g_{m-1}, g_m = g$ to be the set of vertices on a geodesic joining 1 to g, where $m \leqslant k$. The set $\{x \in F \mid xg \in F^c\}$ is contained in the union $\bigcup_{i=1}^m [\partial_V F]\, g_i^{-1}$; therefore, we obtain

$$\frac{|F|}{2d} \leqslant k\, |\partial_V F|\ .$$

□

Remarks 5.90. Proposition 5.89 was initially proved in [**VSCC92**] for nilpotent groups using random walks. The proof reproduced above follows [**CSC93**].

Corollary 5.91. *Let G be an infinite finitely generated group and let F be an arbitrary set of elements in G.*

(1) *If $\mathfrak{G}_G \asymp x^n$ then*

$$|F| \leqslant K\, [\operatorname{card}\,(\partial_V F)]^{\frac{n}{n-1}}\ .$$

(2) *If $\mathfrak{G}_G \asymp \exp(x)$ then*

$$\frac{|F|}{\ln(\operatorname{card} F)} \leqslant K \operatorname{card}\,(\partial_V F)\ .$$

In both inequalities above, the boundary $\partial_V F$ is considered in the Cayley graph of G with respect to a finite generating set S, and K depends on S.

5.9. Distortion of a subgroup in a group

So far, we were primarily interested in quasiisometries and quasiisometric embeddings. In this section we will consider coarse Lipschitz maps which fail to be quasiisometric embeddings and our goal is to quantify failure of the quasiisometric embedding property. While our primary interest comes from finitely generated subgroups of finitely generated groups, we start with general definitions.

Definition 5.92. Let $f : Y \to X$ be a coarse Lipschitz map. The *distortion* Δ_f of the map f is defined as

$$\Delta_f(t) = \sup\{\operatorname{dist}_Y(y, y') : \operatorname{dist}_X(f(y), f(y')) \leqslant t\}.$$

Note that the function Δ_f, in general, takes infinite values. It is clear from the definition, that f is uniformly proper if and only if Δ_f takes values in $\mathbb{R}$. It is also clear that f is a quasiisometric embedding if and only if Δ_f is bounded above by a linear function.

Example 5.93. The function $f : \mathbb{R} \to \mathbb{R}$, $f(y) = \sqrt{y}$ has quadratic distortion.

An interesting special case of the distortion function is when X is a path-metric space, Y is a rectifiable connected subspace of X, equipped with the induced path-metric and f is the identity embedding:

$$\mathrm{dist}(y, y') = \inf_{\gamma} \mathrm{length}_X(\gamma)$$

where the infimum is taken over all paths in Y connecting y to y', while the length of these paths is computed using the metric of X. In this setting, the distortion function of f is denoted Δ_X^Y. (Note that f itself is 1-Lipschitz.)

EXERCISE 5.94. 1. Compute the distortion of the parabola $y = x^2$ in the Euclidean plane.

2. Compute the distortion of the cubic curve $y = x^3$ in the Euclidean plane.

3. Composing $f : X \to Y$ with quasiisometries $X \to X'$ and $Y' \to Y$, preserves the $\asymp$ equivalence class of the distortion function.

We now specialize to the group-theoretic setting. Suppose that G is a finitely generated group and $H \leqslant G$ is a finitely generated subgroup; we let S be a finite generating set of G and T be a finite generating set of H. We then have word-metrics dist_S on G and dist_T on H, and the identity embedding $f : H \to G$. In order to analyze the distortion of H in G (up to the $\asymp$ equivalence relation), we are free to choose the generating sets S and T (see Part 3 of Exercise 5.94); in particular, we can assume that $T \subset S$ and, hence, Cayley(H, T) is a subgraph of Cayley(G, S). Since H acts transitively on itself via left multiplication, we obtain that

$$\Delta_G^H(n) = \max \{\mathrm{dist}_T(1, h) \mid h \in H\,,\ \mathrm{dist}_S(1, h) \leqslant n\}\ . \tag{5.11}$$

The subgroup H is called *undistorted* (in G) if $\Delta_G^H(n) \asymp n$, equivalently, the inclusion map $H \to G$ is a quasiisometric embedding.

In general, distortion functions for subgroups can be as bad as one can imagine, for instance, nonrecursive.

EXAMPLE 5.95. [Mikhailova's construction] Let Q be a finitely presented group with Dehn function $\delta(n)$. Let $a_1, \ldots, a_m$ be generators of Q and $\phi : F_m \to Q$ be the epimorphism from the free group of rank m sending free generators of F_m to the elements a_i, $i = 1, \ldots, m$. Consider the group $G = F_m \times F_m$ and its subgroup

$$H = \langle (g_1, g_2) \in G | \phi(g_1) = \phi(g_2) \rangle\,.$$

This construction of H is called *Mikhailova's construction*, it is a source of many pathological examples in group theory. The subgroup H is finitely generated and its distortion in G is $\asymp \delta(n)$. In particular, if Q has unsolvable word problem then its distortion in G is nonrecursive. We refer the reader to [**OS01**, Theorem 2] for further details.

Below are the basic properties of the distortion function:

PROPOSITION 5.96. (1) *If $\widetilde{X}$ and $\widetilde{S}$ are finite generating sets of H and G, respectively, and $\widetilde{\Delta}_G^H$ is the distortion function with respect to these generating sets, then $\widetilde{\Delta}_G^H \asymp \Delta_G^H$. Thus up to the equivalence relation $\asymp$, the distortion function of the subgroup H in the group G is uniquely defined by H and G.*

(2) *For every finitely generated subgroup H in a finitely generated group G,* $\Delta_G^H(n) \succeq n$.

(3) *If H has finite index in G then* $\Delta_G^H(n) \asymp n$.

(4) *Let $K \lhd G$ is a finite normal subgroup and let $H \leqslant G$ be a finitely generated subgroup; set $\bar{G} := G/K, \bar{H} := H/K$. Then*

$$\Delta_G^H \asymp \Delta_{\bar{G}}^{\bar{H}}.$$

(5) *If $K \leqslant H \leqslant G$ then*

$$\Delta_G^K \preceq \Delta_H^K \circ \Delta_G^H .$$

(6) *Subgroups of finitely generated abelian groups are undistorted.*

Proof. (1) follows from Part 3 of Exercise 5.94.

(2) If we take finite generating sets S and T of G and H, respectively such that $T \subset S$, then the embedding $H \to G$ is 1-Lipschitz with respect to the resulting word metrics. Whence $\Delta_G^H(n) \geqslant n$.

(3) The statement follows immediately from the fact that the inclusion map $H \to G$ is a quasiisometry.

(4) This equivalence follows from the fact that the projections $G \to \bar{G}$ and $H \to \bar{H}$ are quasiisometries.

(5) Consider dist_K, dist_H and dist_G three word metrics, and an arbitrary element $k \in K$ such that $\mathrm{dist}_G(1,k) \leqslant n$. Then $\mathrm{dist}_H(1,k) \leqslant \Delta_G^H(n)$ whence

$$\mathrm{dist}_K(1,k) \leqslant \Delta_H^K\left(\Delta_G^H(n)\right) .$$

(6) By the classification theorem of finitely generated abelian groups (Theorem 11.7), every subgroup $H \leqslant G$ of an abelian group G is isomorphic to the direct product of a finite group and free abelian group. In particular, every finitely generated abelian group is virtually torsion-free. Therefore, by combining (3) and (5), it suffices to consider the case where G is torsion-free of rank n. Then G acts by translations geometrically on $\mathbb{R}^n$; its rank m subgroup H also acts geometrically on a subspace $\mathbb{R}^m \subset \mathbb{R}^n$. Since $\mathbb{R}^m$ is isometrically embedded in $\mathbb{R}^n$, it follows that the embedding $H \to G$ is quasiisometric. Hence, H is undistorted in G and $\Delta_G^H(n) \asymp n$. □

5.10. Metric filling functions

This is a technical section. Here we define *coarse* notions of loops, filling disks, isoperimetric functions and minimal filling area in the setting of geodesic metric spaces, following **[Bow91b]** and **[Gro93]**. We then relate them to the notions of volume, area and Dehn functions defined earlier, in sections 2.4, 4.10.1, 4.10.4. We further show that growth rates of the functions thus defined are stable under quasiisometry.

Throughout this section, (X, dist) will be a quasigeodesic metric space. Thus, there exists a constant ρ, which we fix once and for all, such that the Rips complex $\mathrm{Rips}_\rho(X)$ is connected. Given any pair of points $x, y \in X$, consider the shortest edge-path

$$p_{xy} = [x = q_{xy}(0), q_{xy}(1)] \cup ... \cup [q_{xy}(n-1), q_{xy}(n) = y],$$

in $\text{Rips}_\rho(X)$ connecting x to y. The map $q_{xy} : \{0, \dots, n\} \to X$ is the corresponding *vertex-path* connecting x to y. We let $\ell(p_{x,y}) = n$ denote the combinatorial lengthp_{xy}.

Since X is a quasigeodesic metric space, the map

$$X \to \text{Rips}_\rho(X)$$

is a quasiisometry. In particular,

$$\frac{1}{\rho}\text{dist}_X(x, y) \leqslant \text{length}(p_{x,y}) \leqslant k\text{dist}_X(x, y) + a$$

for some uniform constants k and a.

5.10.1. Coarse isoperimetric functions and coarse filling radius. Our first goal is to discretize/coarsify the usual notions of Lipschitz maps of the unit circle and the unit disk to X. We fix a number $\delta > 0$, the *scale of coarsification*.

For a triangulation $\mathcal{T}$ of the circle $\mathbb{S}^1$, we let $V(\mathcal{T})$ and $E(\mathcal{T})$ denote the vertex and edge sets of $\mathcal{T}$. Similarly, if $\mathcal{D}$ is a triangulation of the disk $\mathbb{D}^2$, extending $\mathcal{T}$, then $V(\mathcal{D}), E(\mathcal{D}), F(\mathcal{D})$ will denote the sets of vertices, edges and 2-dimensional faces of $\mathcal{D}$.

A *(coarse) δ-loop* in X is a pair consisting of a triangulated circle $(\mathbb{S}^1, \mathcal{T})$ and a map

$$\mathfrak{c} : V(\mathcal{T}) \to X$$

such that for every edge $e = [u, w] \in E(\mathcal{T})$,

$$\text{dist}_X(\mathfrak{c}(u), \mathfrak{c}(v)) \leqslant \delta. \tag{5.12}$$

If X is geodesic, one can define a *geodesic extension* of $\mathfrak{c}$ to a Lipschitz map of the entire circle, sending each edge of the triangulation to a geodesic segment connecting images of its end-points. In view of non-uniqueness of geodesics in X, the geodesic extension is not unique, nevertheless, by abusing the notation, we will denote it by $\tilde{\mathfrak{c}}$.

We let $\Omega_\delta(X)$ denote the space of δ-loops in X. We then define the δ-length function

$$\ell = \ell_\delta : \Omega_\delta(X) \to \mathbb{N}.$$

The *δ-length* $\ell_\delta(\mathfrak{c})$ of the δ-loop $\mathfrak{c}$ is the number of vertices in the triangulation $\mathcal{T}$.

Similarly, a *(coarse) δ-disk* is a map

$$\mathfrak{d} : V(\mathcal{D}) \to X$$

satisfying the inequality (5.12) for every edge of $\mathcal{D}$. The δ-disk $\mathfrak{d}$ is a *(coarse) filling disk* of the coarse loop $\mathfrak{c}$, if the triangulations $\mathcal{T}$ and $\mathcal{D}$ agree on the boundary circle and

$$\mathfrak{c} = \mathfrak{d}\big|_{V(\mathcal{T})}.$$

Thus, a filling disk is a discretization of the notion of a Lipschitz-continuous extension $\mathbb{D}^2 \to X$ of a Lipschitz-continuous map $\mathbb{S}^1 \to X$.

Let τ be a 2-face of $\mathcal{D}$, with the vertex set $V(\tau)$. The restrictions

$$\mathfrak{d}\big|_{V(\tau)}, \tau \in F(\mathcal{D}),$$

are called *bricks* of the coarse filling disk $\mathfrak{d}$.

The simplicial area of a coarse filling disk (see §4.10.1) is the number of 2-simplices in the triangulation $\mathcal{D}$, i.e., the number of bricks.

Definition 5.97. The δ-*filling area* of the coarse loop $\mathfrak{c}$ is defined to be the minimum of simplicial areas $Area(\mathfrak{d})$ of δ-filling disks $\mathfrak{d}$ of $\mathfrak{c}$. We will use both notation $\mathrm{Ar}_\delta(\mathfrak{c})$ and $\mathrm{P}(\mathfrak{c}, \delta)$ for the δ-filling area.

To motivate the definition, suppose for a moment that X is a Riemannian manifold of bounded geometry. Then every brick in a δ-filling disk can be filled in with a smooth triangle of the area $\leqslant C\delta^2$, where C is a uniform constant. Therefore, the filling area, in this case, approximates (as δ tends to zero) δ^2 times the least area of a singular disk in X bounded by the loop $\tilde{\mathfrak{c}}$.

We, likewise, define the δ-*filling radius function* as

$$\mathrm{r}_\delta : \Omega_\delta(X) \to \mathbb{R}_+,$$

$$\mathrm{r}_\delta(\mathfrak{c}) = \inf \left\{ \max_{x \in V(\mathcal{D})} \mathrm{dist}_X \left(\mathfrak{d}(x), \mathfrak{c}(V(\mathcal{T}))\right) \ : \ \mathfrak{d} \text{ is a } \delta\text{-filling disk of the loop } \mathfrak{c} \right\}.$$

Again, in the case when X is a Riemannian manifold of bounded geometry, the function r_δ approximates (as δ tends to zero) the radius of the least radius singular 2-disk bounding the geodesic extension of the loop $\mathfrak{c}$.

Both functions Ar_δ and r_δ depend on the parameter δ, and may take infinite values. In order to obtain real-valued functions, we add the hypothesis that X is *coarsely simply connected*, i.e., there exists $\mu \geqslant \rho > 0$, such that for all $\delta \geqslant \mu$, every δ-loop in X admits a δ-filling disk. Such spaces will be called μ-*simply connected*. The notion of coarse simple connectivity will be generalized in Chapter 6. In particular, in Corollary 6.35 we will prove quasiisometries preserve coarse simple connectivity.

Suppose that X is μ-simply connected and $\delta \geqslant \mu$. We define the δ-*isoperimetric function*

$$Ar_\delta = Ar_{\delta,X} : \mathbb{Z}_+ \to \mathbb{Z}_+$$

by

$$Ar_\delta(\ell) := \sup\{\mathrm{Ar}_\delta(\mathfrak{c}) : \mathfrak{c} \in \Omega_\delta(X), \ell_X(\mathfrak{c}) \leqslant \ell\},$$

i.e., the maximal δ-area needed to fill in a coarse loop of δ-length at most ℓ. When δ is fixed, we will also refer to the function Ar_δ as the *coarse isoperimetric function* or the *coarse filling area function*.

The function Ar_δ is a coarsification of the classical isoperimetric functions from the Riemannian geometry $IP_M = IP_{M,1}$ defined via maps of 2-disks into a Riemannian manifold M, see §2.5.

The following theorem relates the coarse isoperimetric functions to the Riemannian ones:

Theorem 5.98. *(Cf.* **[BT02]**.*) If M is a simply-connected Riemannian manifold of bounded geometry, and $X = V(\mathcal{G})$, where $\mathcal{G}$ is a graph approximating M as in §5.3, then for all $\delta > 0$,*

$$IP_M(\ell) \approx Ar_{\delta,X}(\ell).$$

Likewise, using the radius function we define the δ-*filling radius function* as

$$r = r_{\delta,X} : \mathbb{R}_+ \to \mathbb{R}_+, \ r_\delta(\ell) = \sup\{\mathrm{r}_\delta(\mathfrak{c}) \ : \ \mathfrak{c} \in \Omega_\delta(X), \ell(\mathfrak{c}) \leqslant \ell\}.$$

Again, we regard r_δ as a *coarse filling radius function*.

EXAMPLE 5.99. In order to get a better feel for the δ-filling area function, let us estimate Ar_δ (from below) in the case $X = \mathbb{R}^2$.

Suppose that $\mathfrak{c}$ is a δ-loop in $\mathbb{R}^2$ with a δ-filling disk $\mathfrak{d} : V(\mathcal{D}) \to \mathbb{R}^2$. We let

$$g : \mathbb{D}^2 \to \mathbb{R}^2$$

denote the unique extension of $\mathfrak{d}$, which is affine on every simplex in $\mathcal{D}$. Then, in view of Heron's formula for the triangle area, for every 2-simplex τ in $\mathcal{D}$ we obtain the inequality:

$$\mathrm{Area}(g(\tau)) \leqslant \frac{\sqrt{3}}{4}\delta^2$$

Therefore, summing up over all 2-simplices τ in $\mathcal{D}$, we obtain an upper bound on the area of the polygon in $\mathbb{R}^2$, which is the image of g:

$$\mathrm{Area}(g(\mathbb{D}^2)) \leqslant \sum_\tau \delta^2 \frac{\sqrt{3}}{4} \leqslant \delta^2 \frac{\sqrt{3}}{4} \mathrm{Ar}_\delta(\mathfrak{c}). \tag{5.13}$$

5.10.2. Quasi-isometric invariance of filling functions. We first consider the behavior of the isoperimetric functions and the filling radius under the change of the parameter δ:

LEMMA 5.100. *Suppose that X is a μ-simply connected quasigeodesic metric space. Then there exists $K = K(\delta_1, \delta_2, k, a)$ such that for all $\delta_2 \geqslant \delta_1 \geqslant \mu$,*

$$Ar_{\delta_1}(\ell) \leqslant Ar_{\delta_2}(\ell) \leqslant Ar_{\delta_1}(K)\, Ar_{\delta_2}(\ell)$$

and

$$r_{\delta_1}(\ell) \leqslant r_{\delta_2}(\ell) \leqslant r_{\delta_2}(\delta_1)\, r_{\delta_1}(\ell)\,.$$

In particular,

$$Ar_{\delta_1} \asymp Ar_{\delta_2} \quad \text{and} \quad r_{\delta_1} \asymp r_{\delta_2}.$$

PROOF. We will consider only the isoperimetric function as the proof for the filling radius function is nearly the same. The inequality

$$Ar_{\delta_1} \leqslant Ar_{\delta_2}$$

is immediate from the definition, as each δ_1-filling disk for a δ_1-loop $\mathfrak{c}$ is also a δ_2-filling disk for the same loop. Consider a coarse loop $\mathfrak{c} \in \Omega_{\delta_1}(X)$. Since $\delta_1 \leqslant \delta_2$, we can treat $\mathfrak{c}$ as a δ_2-loop in X.

Let $\mathfrak{d}$ be a δ_2-filling disk of $\mathfrak{c}$. Our goal is to replace the triangulation $\mathcal{D}$ (this is a triangulation of the 2-disk associated with the map $\mathfrak{d}$) with its subdivision $\widetilde{\mathcal{D}}$ and extend the map $\mathfrak{d}$ to a δ_1-filling disk

$$\tilde{\mathfrak{d}} : V(\widetilde{\mathcal{D}}) \to X.$$

Let τ be one of the 2-simplices of $\mathcal{D}$ and

$$\sigma : V(\tau) = \{u_1, u_2, u_3\} \to X, \sigma(u_j) = x_j, \quad j = 1, 2, 3,$$

be the corresponding brick of $\mathfrak{d}$. Since $\delta_1 \geqslant \mu \geqslant \rho$, we obtain vertex-paths

$$q_j = q_{x_j x_{j+1}} : \{0, \ldots, N_j + 1\} \to X,$$

connecting x_j to x_{j+1} ($j \in \{1, 2, 3\}$), defined in the beginning of this section. (Here and in what follows, the we compute $j + 1$ modulo 3.) Thus,

$$N_j \leqslant \lambda = \lceil k\delta_2 + a \rceil,$$

and for all $i = 0, \dots, N_j$,

$$\text{dist}_X(q_j(i), q_j(i+1)) \leqslant \delta_1,$$

$j = 1, 2, 3$. We then subdivide each edge of the simplex τ into at most λ new edges and using the maps q_1, q_2, q_3 we define a map

$$\mathfrak{c}_\tau : V(\mathcal{T}_\tau) \to X,$$

where $\mathcal{T}_\tau$ is the resulting triangulation of the boundary of τ.

Figure 5.2

The new map $\mathfrak{c}_\tau$ agrees with $\mathfrak{c}$ on the vertices u_1, u_2, u_3 and sends vertices on the edge $[u_j, u_{j+1}]$ to the points $q_j(i)$, $i = 0, \dots, N_j$, $j = 1, 2, 3$. By the construction, the map $\mathfrak{c}_\tau$ is a δ_1-loop in X. Furthermore,

$$\ell_{\delta_1}(\mathfrak{c}_\tau) \leqslant \delta_1(N_1 + N_2 + N_3) \leqslant K := 3\delta_1(k\delta_2 + a + 1)$$

Therefore, we can fill in this coarse loop with a δ_1-disk of δ_1-area at most

$$\text{A}r_{\delta_1}(\mathfrak{c}_\tau) \leqslant Ar_{\delta_1}(K).$$

By repeating this filling for each brick in $\mathcal{D}$, we construct a δ_1-filling disk $\widetilde{\mathfrak{d}}$ for the δ_1-loop $\mathfrak{c}$, such that

$$\text{A}rea(\widetilde{\mathfrak{d}}) \leqslant Ar_{\delta_1}(K) Ar_{\delta_2}(\ell_{\delta_1}(\mathfrak{c})).$$

Therefore,

$$Ar_{\delta_1}(\ell) \leqslant Ar_{\delta_1}(K) Ar_{\delta_2}(\ell),$$

for every ℓ. $\square$

We can now prove quasiisometric invariance of the filling area and filling radius functions:

Theorem 5.101. *Suppose that X_1, X_2 are quasiisometric μ-simply connected quasigeodesic metric spaces. Then their coarse isoperimetric functions and, respectively their filling radii, functions, are approximately equivalent in the sense of Definition 1.3.*

Proof. We again consider only the coarse isoperimetric function and leave the case of the filling radius function as an exercise to the reader. Our proof is parallel to the one of Corollary 6.35. Let $f : X_1 \to X_2$ be an (L, A)-quasiisometry with coarse inverse $\bar{f} : X_1 \to X_2$. Consider a δ_1-loop $\mathfrak{c}_1 \in \Omega_{\delta_1}(X)$ of the length ℓ.

The composition $\mathfrak{c}_2 = f \circ \mathfrak{c}_1$ is a δ_2-loop in X_2, where

$$\delta_2 = L\delta_1 + A.$$

Since $\delta_2 \geq \delta_1 \geqslant \mu$, the coarse loop $\mathfrak{c}_2$ admits a δ_2-filling disk

$$\mathfrak{d}_2 : V(\mathcal{D}_2) \to X_2,$$

where $\mathcal{D}_2$ is a triangulation of $\mathbb{D}^2$.

Next, apply the coarse inverse map $\bar{f}$ to the coarse disk $\mathfrak{d}_2$: The composition

$$\mathfrak{d}_3 := \bar{f} \circ \mathfrak{d}_2$$

is a δ_3-filling disk for the coarse loop

$$\mathfrak{c}_3 = \bar{f} \circ \mathfrak{c}_2,$$

where

$$\delta_3 = L\delta_2 + A.$$

The δ_3-length of $\mathfrak{c}_3$ is the same as the one of $\mathfrak{c}_1$ and

$$\mathrm{Ar}_{\delta_3}(\mathfrak{c}_3) = \mathrm{Ar}_{\delta_1}(\mathfrak{c}_1),$$

since we did not change the triangulation of the unit circle and the unit disk. Of course, the coarse loop $\mathfrak{c}_3$ is not the same as $\mathfrak{c}_1$, but they are with distance $\leqslant A$ from each other:

$$\mathrm{dist}_{X_1}(\mathfrak{c}_1(v), \mathfrak{c}_3(v)),$$

for every vertex v of the triangulation of the circle $\mathbb{S}^1$. Observe now that A does not exceed δ_3. Therefore, we can add to the coarse disk $\mathfrak{d}_3$ a "coarse annulus" $\mathfrak{a} : \mathcal{A} \to X$, as in the Figure 5.3. This requires adding ℓ vertices, 3ℓ edges and 2ℓ faces to the original triangulation $\mathcal{D}$ of the disk $\mathbb{D}^2$. We let $\widetilde{\mathcal{D}}$ denote the new triangulation of the 2-disk. We let $\mathfrak{d}_4$ denote the extension of the map $\mathfrak{d}_3$ via the map $\mathfrak{c}_1$ of the boundary vertices.

FIGURE 5.3. A coarse annulus

The result is a δ_3-coarse disk

$$\mathfrak{d}_4 : V(\widetilde{\mathcal{D}}) \to X$$

extending the map $\mathfrak{c}_1$; the simplicial area of this disk is

$$2\ell + \mathrm{Area}(\mathfrak{d}_2) \leqslant 2\ell + Ar_{\delta_2, X_2}(\ell)$$

This proves that

$$Ar_{\delta_3, X_1}(\ell) \leqslant 2\ell + Ar_{\delta_2, X_2}(\ell).$$

Taking into account Lemma 5.100, we conclude that

$$Ar_{\delta, X_1} \precsim Ar_{\delta, X_2}$$

for any $\delta \geqslant \mu$. Therefore, the spaces X_1, X_2 have approximately equivalent coarse isoperimetric functions:

$$Ar_{X_1} \approx Ar_{X_2}. \quad \square$$

An immediate corollary of this theorem is that the approximate growth rates of the filling area and filling radius, are quasiisometry invariants of the metric space X. The order of the filling function of a metric space X is also called *the filling order of* X. If the coarse isoperimetric function $Ar(\ell)$ of a metric space X satisfies $Ar(\ell) \prec \ell$ or ℓ^2 or e^ℓ, it is said that the space X *satisfies a linear, quadratic or exponential isoperimetric inequality*.

Filling area/radius in the Rips complex. Suppose that X is a μ-simply connected metric space and $\delta \geqslant \mu$. Instead of filling coarse loops in X by δ-disks, one can fill in polygonal loops in $P = Rips_\delta(X)$ by simplicial disks. Let $\mathfrak{c}$ be a δ-loop in X. Then we have a triangulation of the circle $\mathbb{S}^1$ such that $\mathrm{diam}(\mathfrak{c}(\partial e)) \leqslant \delta$ for every edge e of the triangulation. Thus, we define an edge-loop $\mathfrak{c}_\delta = \tilde{\mathfrak{c}}$ in P by connecting points $\mathfrak{c}(\partial e)$ by the edges in P (provided that these points are distinct, of course). We will think of $\mathfrak{c}_\delta$ as a simplicial map $\mathbb{S}^1 \to P$ (this map may send some edges to vertices). Then

$$\mathrm{length}(\mathfrak{c}_\delta) = \ell(\mathfrak{c}).$$

It is clear that for $\delta > 0$ the map

$$\{\delta\text{-loops in } X \text{ of length} \leqslant \ell\} \to \{\text{edge-loops in } P \text{ of length} \leqslant \ell\}$$
$$\mathfrak{c} \mapsto \mathfrak{c}_\delta$$

is surjective. Furthermore, every δ-disk $\mathcal{D}$ which fills in $\mathfrak{c}$, yields a simplicial map $\mathcal{D}_\delta : \mathbb{D}^2 \to P$ which is an extension of $\mathfrak{c}_\delta$. The area is preserved under this construction:

$$Area(\mathcal{D}_\delta) = \mathrm{Area}(\mathcal{D}).$$

We leave it to the reader to verify that the above procedure yields *all* simplicial maps $\mathbb{D}^2 \to P$ extending $\mathfrak{c}_\delta$ and we obtain

$$Area(\mathfrak{c}_\delta) = \mathrm{Ar}_\delta(\mathfrak{c}).$$

Summarizing all this, we obtain

$$A_{Rips_\delta(X)}(\ell) = Ar_\delta(\ell),$$

where the left hand side is defined analogously to the function Ar, only using simplicial maps to the Rips complex instead of δ-maps to X itself. The same argument applies to the filling radius and we obtain:

Observation 5.102. Studying filling area and filling radius functions in X (up to the equivalence relation $\approx$) is equivalent to studying the simplicial filling area and filling radius functions in $Rips_\delta(X)$.

Lastly, we relate the filling area function to the Dehn function:

Theorem 5.103. *For every finitely presented group G, the coarse isoperimetric function and the Dehn function are also approximately equivalent.*

Proof. Let G be a finitely-presented group with the finite presentation $\langle S|R\rangle$, and equipped with the word metric dist_S. We let μ denote the length of the longest relator in R and let $Dehn_G$ denote the Dehn function associated with the presentation $\langle S|R\rangle$.

Exercise 5.104. The metric space (G, dist_S) is μ-simply connected.

We will prove the approximate inequality

$$Ar_{\mu,G} \precsim Dehn_G$$

and leave the opposite inequality as an exercise to the reader.

We let Y denote the presentation complex of $\langle S|R\rangle$ and let $\tilde{Y}$ denote its universal cover: The vertex set of the complex $\tilde{Y}$ is the group G, the 1-skeleton of $\tilde{Y}$ is the Cayley graph $\mathrm{Cayley}(G, S)$ of G (with respect to the generating set S).

Let $\mathfrak{c}$ be a coarse loop in G, an element of $\Omega_1(G)$; this coarse loop defines an (almost) regular cellular map $c = \tilde{\mathfrak{c}} : \mathbb{S}^1 \to \mathrm{Cayley}(G, S)$, form the triangulated unit circle. Our goal is to estimate above the filling area of $\mathfrak{c}$ via van Kampen diagrams extending the map c.

The loop c projects to a map $c_w : \mathbb{S}^1 \to Y^{(1)}$ corresponding to some, possibly nonreduced, word w in S. (See §4.10.4.) Replacing w with its free reduction w' will change very little:

$$A(w') = A(w)$$

and the lift c' of the loop $c_{w'}$ will satisfy

$$\mathrm{Ar}_\mu(\mathfrak{c}) \leqslant \mathrm{Ar}_\mu(\mathfrak{c}') + \ell(\mathfrak{c}).$$

Here $\mathfrak{c}'$ is the restriction of c' to the vertex set of the triangulation of $\mathbb{S}^1$. Therefore, in what follows, we will assume that w is reduced.

Every van Kampen diagram $h : K \to Y$ of the word w lifts to a map

$$f : K \to \tilde{Y},$$

whose boundary value $\partial f : \mathbb{S}^1 \to \tilde{Y}$ is a lift of c_w. The van Kampen diagram f extends to an almost regular map $g = \widehat{f} : \widehat{K} \to \tilde{Y}$ as in §4.10.4 and

$$cArea(g) = Area(h).$$

By the construction, $\widehat{f}$ sends cells of $\tilde{K}$ to cells of Cayley(G, S); the only problem is that $\tilde{K}$ is not a simplicial complex. However, the 2nd barycentric subdivision of $\tilde{K}$ is a triangulation $\mathcal{D}$ of the disk $\mathbb{D}^2$. The total number of faces of $\mathcal{D}$ is at most $12\mu Area(h)$. In order to extend g to the vertices of $\mathcal{D}$, for each vertex $v \in V(\mathcal{D})$, we let σ_v be the smallest cell of $\widehat{K}$ containing v. Lastly, let $g(v)$ be an arbitrarily chosen vertex in $g(\sigma_v)$. Thus, we define the new map $\mathfrak{d} : V(\mathcal{D}) \to G = V(\tilde{Y})$, equal to the restriction of g to $V(\mathcal{D})$. The map $\mathfrak{d}$ is a μ-disk in G extending the coarse loop $\mathfrak{c}$. We obtain

$$\mathrm{Ar}_\mu(\mathfrak{c}) \leqslant 12\mu\, \mathrm{Area}(h) \leqslant 12\mu Dehn_G(\ell) \leqslant 12\mu Dehn_G(\ell(\mathfrak{c})).$$

where ℓ is the word-length of w. The approximate inequality

$$Ar_\mu \precsim Dehn_G$$

follows. □

The filling radius function is not as commonly used in Geometric Group Theory as the coarse isoperimetric function and the Dehn function. As we noted earlier, Gersten proved that solvability of the word problem for G is equivalent to the recursivity of its Dehn function. In the same paper [**Ger93a**] Gersten also proved:

PROPOSITION 5.105. *For a finitely presented group G the following are equivalent.*

1. *G has solvable word problem.*
2. *The filling radius function of G is recursive.*

5.10.3. Higher Dehn functions. The Dehn functions $Dehn(\ell)$ generalize to "higher Dehn functions" $Dehn_n$ for groups G of type $\mathbf{F}_n$, with $Dehn = Dehn_1$. All the definitions amount to a coarsification of the Riemannian isoperimetric function $IP_{M,n}$ responsible for the least volume extension of maps of n-spheres to maps of $n+1$-balls, provided that $\pi_n(M) = 0$, see §2.5. As in Theorem 5.98, the resulting higher Dehn function $Dehn_n$ are approximately equivalent to their Riemannian counterparts, under the assumption that G acts isometrically and cocompactly on M.

Below is one of the many equivalent definitions of $Dehn_n$. For sufficiently large δ's consider maps $c : \mathbb{S}^{n-1} \to \mathrm{Rips}_\delta(G)$ of the triangulated $n-1$-sphere into $\mathrm{Rips}_\delta(G)$. For each map c find the extension $f : \mathbb{D}^n \to \mathrm{Rips}_\delta(X)$ which has the least simplicial n-volume. Define

$$FVol_n(c) := cVol_n(f),$$

the filling volume of c. Then take the supremum over all c's:

$$Dehn_{n-1}(\ell) = \sup_{c: cVol_{n-1}(c) \leqslant \ell} FVol_n(c).$$

We refer the reader to [**Gro93**, Chapter 5], [**ECH**$^+$**92**, Chapter 10] and [**Pap00**] for further detail. The following result was proven by P. Papasoglou:

THEOREM 5.106 (P. Papasoglou, [**Pap00**]). *The second Dehn function* $Dehn_2$ *of a group of type* $\mathbf{F}_3$ *is bounded above by a recursive function.*

This theorem represents a striking contrast with the fact that there are finitely-presented groups with unsolvable word problem and, hence, Dehn function which is not bounded above by any recursive function.

Here is the idea of the proof of Theorem 5.106. Let Y be a finite cell complex with $\pi_1(Y) \cong G$ and $\pi_2(Y) = 0$. Consider cellular maps $s : \mathbb{S}^2 \to Y$. Every such map s is null-homotopic and for every ℓ, there are only finitely many such maps with $cVol_2(s) \leqslant \ell$. The key then is to design an algorithm which, for each s, finds *some* extension $f : \mathbb{D}^3 \to Y$: This algorithm uses the algorithmic recognition of 3-dimensional balls (and, hence, fails for $Dehn_n$, $n \geqslant 3$). One then computes the combinatorial volume $cVol_3(f)$ and takes the maximum $\Delta(\ell)$, over all maps s. The resulting function $\Delta(\ell)$ gives the required recursive upper bound:

$$Dehn_2 \precsim \Delta(\ell).$$

We observe that one gets only an upper bound on $Dehn_2$ since the filling maps f above might not be optimal. Observe also that this proof also fails for the ordinary Dehn function since the presentation complex Y is (usually) not simply connected and the recognizing which loops in Y are null-homotopic is algorithmically impossible.

5.10.4. Coarse Besikovitch inequality. In this section we will prove coarse analogues of the classical *Besikovitch inequality* (see e.g. [**BZ88**]).

Let $Q \subset \mathbb{R}^2$ denote the unit square; then the topological circle $C = \partial Q$ has the natural structure of a simplicial complex with the consecutive edges $e_1, \ldots, e_4$. Subdividing the edges of Q further, we obtain a triangulated topological circle $(\mathbb{S}^1, \mathcal{T})$, which will be used in the proposition below. We will regard the sides e_i $(i = 1, 2, 3, 4)$ of Q as subcomplexes of $\mathcal{T}$. A *topological quadrilateral* in a topological space X is a continuous map $f : C \to X$. Similarly, one defines a *topological triangle* in X as a continuous map form $f : T \to X$, where $T \subset \mathbb{R}^2$ is a triangle with the edges e_1, e_2, e_3 (recall that triangles are always treated as 1-dimensional objects). Again, we will regard the edges of T as subcomplexes of a fixed triangulation $\mathcal{T}$ of T, refining the original simplicial structure.

Given a topological quadrilateral $f : \partial Q \to X$ in a metric space X, we define its *separation* $\mathrm{sep}(f)$ as the pair (d_1, d_2), where

$$d_i = \mathrm{dist}(f(e_i), f(e_{i+2})), \quad i = 1, 2,$$

where dist is the minimal distance between subsets of X (see §1.8.1). For instance, suppose that $f : Q \to \mathbb{R}^2$ is an affine map, whose image is the parallelogram P, with the side-lengths s_1 (the length of $f(e_1)$) and s_2 (the length of $f(e_2)$) and the angle α between $f(e_1), f(e_2)$. Then the separation of $f\big|_{\partial Q}$ equals

$$(d_1, d_2) = (s_2 \sin(\alpha), s_1 \sin(\alpha)).$$

It is then immediate that

$$\mathrm{Area}(P) \leqslant d_1 d_2.$$

Besikovitch proved that the same inequality holds for topological quadrilaterals in arbitrary metric spaces, where Area of a topological quadrilateral is understood as the least area of 2-disks that its bounds in X.

The notion of *minsize* for topological triangles defined below is an analogue of separation for topological quadrilaterals.

DEFINITION 5.107. The *minimal size (minsize)* of a topological triangle $f : T \to X$ is defined as

$$\operatorname{minsize}(f) = \inf\{\operatorname{diam}\{f(y_1),\, f(y_2),\, f(y_3)\}\ ;\ y_i \in e_i,\, i = 1, 2, 3\}\ .$$

Next, we coarsify the notions of topological triangles and quadrilaterals, their minsize and separation. In what follows, we fix X, a μ-coarsely simply connected metric space, $\delta \geqslant \mu$ and $(C, \mathcal{T})$, a triangulated topological circle (a subdivided quadrilateral or a triangle). A δ-loop $\mathfrak{c} : V(\mathcal{T}) \to X$ will be regarded as a *coarse quadrilateral*, resp. a *coarse triangle* in X.

DEFINITION 5.108. The *separation* of a coarse quadrilateral $\mathfrak{c}$ is defined as the pair (d_1, d_2), where

$$d_i = \operatorname{dist}\left(\mathfrak{c}(V(e_i)), \mathfrak{c}(V(e_{i+2}))\right), \quad i = 1, 2.$$

The *minsize* of a coarse triangle $\mathfrak{c}$ is defined as

$$\operatorname{minsize}(\mathfrak{c}) = \min\{\operatorname{diam}\{\mathfrak{c}(y_1),\, \mathfrak{c}(y_2),\, \mathfrak{c}(y_3)\}\ ;\ y_i \in V(e_i),\, i = 1, 2, 3\}\ .$$

PROPOSITION 5.109 (The coarse Besikovitch inequality). *With the notation as above, for each coarse quadrilateral* $\mathfrak{c} \in \Omega_\delta(X)$ *we have*

$$\operatorname{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2}{\sqrt{3}\delta^2} d_1 d_2\ .$$

PROOF. Our proof follows closely the proof of the classical Besikovitch inequality. Consider the plane $\mathbb{R}^2$, whose points will be denoted (s, t). Define the map $\beta : X \to \mathbb{R}^2$,

$$\beta(x) = \left(\operatorname{dist}(x, \mathfrak{c}(V(e_1))\,,\, \operatorname{dist}(x, \mathfrak{c}(V(e_2))\right).$$

Since each component of β is a 1–Lipschitz map, the map β itself is $\sqrt{2}$–Lipschitz. Define the composition

$$\beta \circ \mathfrak{c} : V(\mathcal{T}) \to \mathbb{R}^2$$

and its geodesic extension f. Then the image $f(e_1) \subset \mathbb{R}^2$ is a vertical segment connecting the origin to a point $(0, t_1)$, with $t_1 \geqslant d_2$, while $f(e_2)$ is a horizontal segment connecting the origin to a point $(s_2, 0)$, with $s_2 \geqslant d_1$.

Similarly, the image $f(e_3)$ is a path to the right of the vertical line $\{s = d_1\}$ and $f(e_4)$ is another path above the horizontal line $t = d_2$. Thus, the rectangle $R \subset \mathbb{R}^2$ with the vertices

$$(0, 0), (d_1, 0), (d_1, d_2), (0, d_2),$$

is separated from the infinity by the curve $\beta \circ \tilde{\mathfrak{c}}(\mathbb{S}^1)$ (see Figure 5.4).

In particular, the image of any continuous extension g of the map f to the entire square Q, contains the rectangle R. Thus,

$$\operatorname{Area}(g(Q)) \geqslant \operatorname{Area}(R) = d_1 d_2.$$

By taking into the account the fact that the map β is $\sqrt{2}$-Lipschitz, and the inequality (5.13), we obtain

$$d_1 d_2 \leqslant \epsilon^2 \frac{\sqrt{3}}{4} \mathrm{Ar}_\epsilon(\beta \circ \mathfrak{c}),$$

where $\epsilon = \sqrt{2}\delta$.

Consider a δ-filling disk $\mathfrak{d}$ of the δ-loop $\mathfrak{c}$ and let g be the extension of the map $\beta \circ \mathfrak{d}$, defined as in Example 5.99. The simplicial area of $\beta \circ \mathfrak{d}$ is, of course, exactly the same as the one of the map $\mathfrak{d}$. Putting this all together:

$$\mathrm{sArea}(\mathfrak{d}) \geqslant \mathrm{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2}{\sqrt{3}\delta^2} d_1 d_2,$$

as required. □

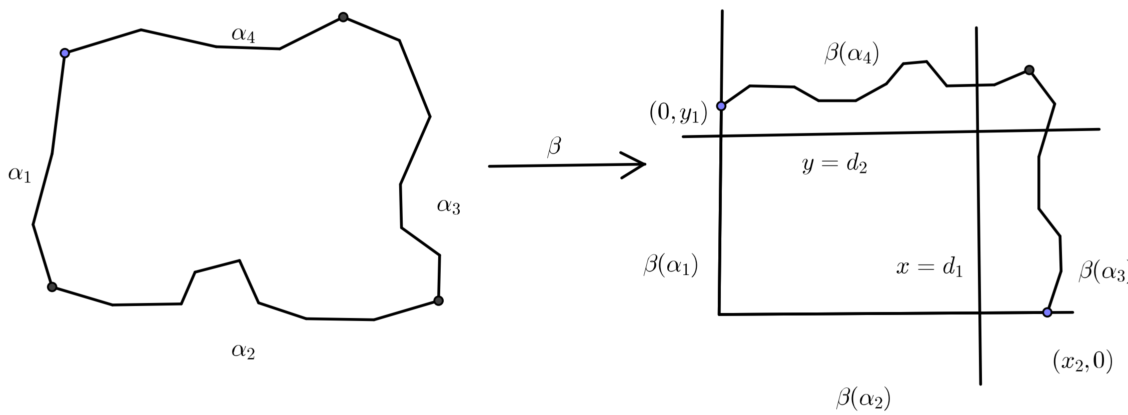


FIGURE 5.4. The map β.

Besikovitch's inequality generalizes from maps of squares to maps of triangles: This generalization has interesting applications to δ-hyperbolic spaces which will be discussed in §9.22.1.

PROPOSITION 5.110 (Minsize inequality). *Let X be a μ–simply connected metric space and let $\delta \geqslant \mu$. Then each coarse topological triangle $\mathfrak{c} : V(\mathcal{T}) \to X, \mathfrak{c} \in \Omega_\delta(X)$, satisfies the* minsize inequality

$$\mathrm{Ar}_\delta(\mathfrak{c}) \geqslant \frac{1}{2\sqrt{3}\delta^2} [\mathrm{minsize}(\mathfrak{c})]^2 .$$

PROOF. The proof is analogous to the one for coarse quadrilaterals. Define the $\sqrt{2}$–Lipschitz map $\beta : X \to \mathbb{R}^2$,

$$\beta(x) = (\beta_1, \beta_2) = (\mathrm{dist}(x, \mathfrak{c}(V(e_1)), \mathrm{dist}(x, \mathfrak{c}(V(e_2)))) ,$$

the composition $\beta \circ \mathfrak{c} : V(\mathcal{T}) \to \mathbb{R}$ and the geodesic extension $f = (f_1, f_2) : T \to \mathbb{R}^2$ of the latter.

As in the proof of the coarse Besikovitch inequality for quadrilaterals, f sends the edges e_1, e_2 to coordinate segments, while the restriction of f to e_3 satisfies:

$$\max(f_1(x), f_2(x)) \geqslant \frac{m}{2}, \forall x \in e_3,$$

where $m = \text{minsize}(\mathfrak{c})$. Therefore, the image of any continuous extension g of of f contains the square with the vertices

$$(0,0), (\frac{m}{2}, 0), (\frac{m}{2}, \frac{m}{2}), (0, \frac{m}{2}).$$

Arguing as in the case of coarse quadrilaterals, we obtain the estimate

$$\frac{m^2}{4} \leqslant \text{Area}(g)$$

and, hence,

$$\text{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2}{\sqrt{3}\delta^2} \frac{m^2}{4} = \frac{1}{2\sqrt{3}\delta^2} m^2 .$$

□

CHAPTER 6

Coarse topology

So far, we succeeded in coarsifying Riemannian manifolds and groups, while treating metric spaces *up to quasiisometry*. The trouble is that, in a way, we succeeded all too well, and, seemingly, lost all the topological tools in the process. Indeed, quasiisometries lack continuity and uniformly discrete spaces have very boring (discrete) topology. The goal of this chapter is to describe tools of algebraic topology for studying quasiisometries and other concepts of the geometric group theory. We will see how to define *coarse topological invariants* of metric spaces, which are robust enough to be stable under quasiisometries. The price we have to pay for this stability is that we will be forced to work not with simplicial/cell complexes and their (co)homology groups as it is done in algebraic topology, but with direct/inverse systems of such complexes and groups.

In this chapter we also introduce *metric cell complexes with bounded geometry*, which will provide a class of spaces for which application of algebraic topology (in the coarse setting) is possible.

Note that the coarse algebraic topology invariants defined and used in this chapter and in this book are quite basic (homology, coarse separation, Poincaré duality).

QUESTION 6.1. Are there any interesting coarse topology applications of other invariants of algebraic topology?

6.1. Ends

In this section we review the oldest coarse topological notion, the one of *ends* of a topological space. Even though we are primarily interested in coarse topology of metric spaces, we will also define ends in the more general, topological, setting. We refer the reader to [**BH99**] and [**Geo08**] for a more detailed treatment of ends of spaces.

6.1.1. The number of ends. We begin with the motivation. One of the simplest topological invariants of a space X is the number of its connected components or, more precisely, the cardinality of the set of its connected components. Alternatively, one can use the set $\pi_0(X)$ of path-connected components of X. Suppose, however, we are dealing with a connected (or path-connected) topological space. The next topological invariant one can try, is the number of connected components of complements to points or, more generally, finite subsets, of X. For instance, if one space can be disconnected by a point and the other cannot, then the two spaces are not homeomorphic. In the coarse setting (of metric spaces) a point is undistinguishable from a bounded subset. Therefore, one naturally looks for complementary components of bounded subsets, say, metric balls.

Remark 6.2. In the topological setting, metric balls will be replaced with compact subsets. In order to maintain consistency between the two notions (metric and topological), we will later restrict to proper geodesic metric spaces on the metric side and locally compact, locally path-connected Hausdorff topological spaces on the topological side.

The trouble is that, say, a point might fail to disconnect a metric space, while a larger bounded (or compact) subset, might disconnect X. Moreover, some complementary components C of a bounded subset might be bounded themselves and, hence, such C "disappears" if we remove a larger bounded subset from X. Such bounded complementary components should be discarded, of course. This leads to the first, *numerical*, definition below, which suffices for many purposes. In what follows, for a subset B in X, B^c will denote the complement of B in X. For each closed subset $B \subset X$ we define the set $\pi_0^u(B^c) := \pi_0(U_B)$, where U_B is the union of unbounded path-connected components of B^c. (The letter u stays for *unbounded*). In the topological setting, being unbounded, of course, makes no sense. Thus, for a Hausdorff topological space X, we let U_B denote the union of path-components of $B^c = X \setminus B$ which are not relatively compact in X. We retain the notation $\pi_0^u(B^c)$ for the set $\pi_0(U_B)$.

From now on, let X be nonempty, locally compact, connected, locally path-connected, 2nd countable Hausdorff topological space, e.g., a proper geodesic metric space.

Definition 6.3. The *number of ends* of X is the supremum, taken over all compact subsets $K \subset X$, of cardinalities of $\pi_0^u(K^c)$. We will denote the number of ends of X by $\eta(X)$.

The reader has to be warned at this points that we, eventually, will define a certain set $\epsilon(X)$, called the *set of ends*, of the space X. The cardinality of this set equals $\eta(X)$ if either one of them is finite; in the infinite case, $\text{card}\,(\epsilon(X)) \geqslant \eta(X)$. In the group-theoretic setting, $\epsilon(X)$ will have the cardinality of continuum, once it is infinite. Nevertheless, what we will really care about, as far as groups are concerned, is finiteness or infiniteness of the number of ends. Thus, the distinction between $\eta(X)$ and $\text{card}\,(\epsilon(X))$ will not be that important.

According to our definition, X has zero number of ends iff X is compact; X has one end (is *one-ended*) iff X is noncompact and for each compact $K \subset X$, the complement K^c has exactly one unbounded component. The space X is *disconnected at infinity* iff X has at least two ends. The space X has infinitely many ends iff for every $n \in \mathbb{N}$, there exists a compact $K \subset X$ such that K^c has at least n unbounded complementary components.

It is clear that the number of ends is a topological invariant of X. Note also that for any compact subsets $K_1 \subset K_2 \subset X$ we have

$$\text{card}\,(\pi_0^u(K_2^c)) \geqslant \text{card}\,(\pi_0^u(K_1^c)).$$

In particular, in the definition of the number of ends of a proper geodesic metric space, we can take the supremum of cardinalities $\pi_0^u(B^c)$ over all metric balls in X; equivalently, over all bounded subsets of X.

Exercise 6.4. (1) The real line $\mathbb{R}$ is 2-ended.
(2) $\mathbb{R}^n$ is one-ended for $n \geqslant 2$.

(3) Suppose that X is a regular tree of finite valence $k \geqslant 3$. Then X has infinitely many ends.

The proof of the following lemma is a model for many arguments appearing in this chapter.

LEMMA 6.5. *The number of ends $\eta(X)$ is a quasiisometry invariant of X.*

PROOF. Let $f : X \to Y$ be an (L, A)-quasiisometry of (proper, geodesic) metric spaces. Suppose that $\eta(X) \geqslant n$, $n \in \mathbb{N}$. This means that there exists a metric ball $B = B(x, R) \subset X$ such that B^c consists of at least n unbounded components. The image of a bounded subset under quasiisometry is again bounded, while the image of an unbounded complementary component C is still unbounded. The trouble is that $f(C)$, of course, may fail to be connected and be contained in the complement of $f(B)$; moreover, images of distinct complementary components under f might be contained in the same complementary component of $f(B)$.

We will deal with these three problems one at a time. Consider another metric ball $B' = B(x, R')$, $R' \geqslant R$.

1. If $R' - R \geqslant t$, where $L^{-1}t - A > 0$, then for each component C' of $(B')^c$, its image $f(C')$ is disjoint from $f(B)$. (Thus, it suffices to take $R' > R + AL$.)

2. If $x_1, x_2 \in X$ are within distance $\leqslant 1$ from each other, then

$$\mathrm{dist}_Y(f(x_1), f(x_2)) \leqslant r := L + A.$$

Therefore, the r-neighborhood $\mathcal{N}_r(f(C'))$ of $f(C')$ in Y will be path-connected. In order for this neighborhood to be disjoint from $f(B)$, we need to increase R' a little bit: It suffices to take t such that $L^{-1}t - A > r$, i.e., $R' > R + L(A + r)$.

3. The last issue we have to address is slightly more difficult: So far, we only used the fact that f is a QI embedding. Considering the example of an isometric embedding of the line into the plane, we see what can go wrong without the assumption of coarse surjectivity of f. Suppose that C_1, C_2 are distinct unbounded components of B and $x_i \in C_i, i = 1, 2$, are points which are mapped to points $y_i = f(x_i)$ which are in the same complementary (path-connected) component of $\mathrm{cl}(f(B))^c$. Pick a path p connecting y_1, y_2 and avoiding $\mathrm{cl}(f(B))$. The composition of p with the coarse inverse $\bar{f}$ to f, is not a path in X, so we have to coarsify p. We find a finite sequence $z_1 = y_1, z_2, \ldots, z_n = y_2$ in the image of p, such that

$$\mathrm{dist}_Y(z_i, z_{i+1}) \leqslant 1, \quad i = 1, \ldots, n-1.$$

Then points $w_i = \bar{f}(z_i) \in X$ satisfy

$$\mathrm{dist}_X(w_i, w_{i+1}) \leqslant L + A, \quad i = 1, \ldots, n-1.$$

The points $x_1' := w_1, x_2' := w_n$ are within distance $\leqslant A$ from the points $w_0 := x_1, w_{n+1} := x_2$, respectively. Connecting the consecutive points $w_i, w_{i+1}, i = 0, \ldots, n$, by geodesic segments in X results in a path q, connecting x_1 to x_2. This is our replacement for the (likely discontinuous) path $\bar{f} \circ p$. We would like to ensure that the image of q is disjoint from B: This would result in a contradiction, as we assumed that $C_1 \neq C_2$ are distinct components of B^c. If the image $\mathrm{Im}(p)$ of the path p lies outside of the ball $B(y, r')$, $y = f(x)$, then

$$\mathrm{dist}_X(x, \mathrm{Im}(q)) \geqslant R'' := L^{-1}r' - 3A - L.$$

We choose r' such that $R'' \geqslant R$. Therefore, if x_1, x_2 are sufficiently far away form x (and this is certainly possible to achieve since we assume that the sets C_1, C_2

are unbounded), then y_1, y_2 lie in distinct complementary components of $B(y, r')$. Thus, there exists a bounded subset $B' = B(y, r') \subset Y$ whose complement contains at least n unbounded components. We proved that $\eta(Y) \geqslant \eta(X)$.

Reversing the roles of X and Y, we conclude that $\eta(X) = \eta(Y)$. □

In particular, we now can define the number of ends of finitely generated groups:

DEFINITION 6.6. Let G be a finitely generated group. Then the number of ends $\eta(G)$ is the number of ends of its Cayley graph.

In view of Lemma 6.5, the quantity $\eta(G)$ is well-defined, as the number of ends is independent of the generating set of G. Moreover, $\eta(G)$ is a quasiisometry invariant of G.

6.1.2. The space of ends. Our next goal is to define a set $\epsilon(X)$ of ends of a topological space X, such that $\operatorname{card}(\epsilon(X)) = \eta(X)$ if either one is finite. We will also equip $\epsilon(X)$ with a topology, which we then use in order to compactify X by adding to it the set of ends. The idea is that the ends of X are encoded by *decreasing* families of complementary components of compact subsets of X. We refer the reader to §1.5 for the required background on inverse limits.

We again let X be a nonempty, locally compact, connected, locally path-connected, 2nd countable Hausdorff topological space. In particular, X admits an exhaustion by a countable family $(B_n)_{n\in\mathbb{N}}$ of compact subsets as in Proposition 1.20. For instance, if X is a proper metric space (the case we are mostly interested in), we can take $B_n = \bar{B}(x, n)$, where $x \in X$ is a fixed point, $n \in \mathbb{N}$.

Define $\mathcal{K} = \mathcal{K}_X$, the poset of compact subsets of X with the partial order $\leqslant$ given by the inclusion. It is clear that the poset $\mathcal{K}$ is directed, as the union of two compact sets is again compact. For each $K \in \mathcal{K}$ we have the set $\pi_0(K^c)$ whose elements are connected (equivalently, path-connected) components of K^c. Whenever $K_1 \leqslant K_2$ are compact subsets of X, we have the associated map

$$f_{K_1,K_2} : \pi_0(K_2^c) \to \pi_0(K_1^c),$$

sending each component C_2 of K_2^c to the unique component C_1 of K_1^c such that C_2 is contained in C_1.

EXERCISE 6.7. Verify that the resulting collection of maps f_{K_2,K_1} is an inverse system, i.e.,

$$f_{K_1,K_2} \circ f_{K_2,K_3} = f_{K_1,K_3}, \quad f_{K,K} = \mathrm{id}\,.$$

We will use the notation $\pi_0(\mathcal{K}^c)$ for this inverse system.

DEFINITION 6.8. The *set of ends* of X, denoted $\epsilon(X)$, is the inverse limit of the inverse system $\pi_0(\mathcal{K}^c)$. We will equip $\epsilon(X)$ with the initial topology, where each $\pi_0(K^c)$ is equipped with the discrete topology.

EXERCISE 6.9. Show that the space $\epsilon(X)$ is totally disconnected and Hausdorff.

PROPOSITION 6.10. *For every compact $K \subset X$, the set $\pi_0^u(K^c)$ is finite.*

PROOF. We will assume that K is nonempty since the proof is clear otherwise. Since X admits an exhaustion by compact subsets, there exists a compact $K' \subset X$ whose interior contains K. We claim that only finitely many components of U_K have nonempty intersection with $X \setminus K'$. It suffices to exclude the case when U_K has countably infinitely many components $U_i, i \in \mathbb{N}$. Since X is path-connected, for each component U_i, there exists a path connecting some $x \in K$ to $x_i \in U_i \setminus K'$. Let y_i be a point in this path which belongs to $\partial K'$. Since $\partial K'$ is compact, after passing to a subsequence, we can assume that

$$\lim_{i\to\infty} y_i = y \in \partial K'.$$

Then $V := X \setminus K$ is a neighborhood of y. Since X is locally path-connected, there exists a neighborhood W of y contained in V, such that for all $i \geqslant i_0$ points y, y_i are connected by a path contained in W. It follows that $U_i = U_{i_0}$ for all $i \geqslant i_0$. □

In addition to the inverse system $(\pi_0(K^c))_{K\in\mathcal{K}}$, we also have, similarly defined, inverse systems

$$(\pi_0^u(K^c))_{K\in\mathcal{K}}$$

and

$$(\pi_0^u(B_n^c))_{n\in\mathbb{N}}$$

where we use the standard order on $\mathbb{N}$. Inclusion maps

$$\{B_n : n \in \mathbb{N}\} \hookrightarrow \mathcal{K}$$

and

$$\pi_0^u(K^c) \to \pi_0(K^c)$$

induce maps of inverse limits

$$\phi : \varprojlim \pi_0^u(B_n^c) \to \varprojlim \pi_0^u(K^c)$$

and

$$\psi : \varprojlim \pi_0^u(K^c) \to \varprojlim \pi_0(K^c) = \epsilon(X).$$

We again equip the inverse limits $\varprojlim \pi_0^u(B_n^c)$ and $\varprojlim \pi_0(K^c)$ with the initial topology.

Since each $\pi_0^u(K^c)$ is finite, the inverse limit

$$\varprojlim \pi_0^u(K^c)$$

is compact by Tychonoff's theorem, cf. Exercise 1.22.

In view of Proposition 6.10, each $K_1 \in \mathcal{K}$ is contained in the interior of $K_2 \in \mathcal{K}$, such that the image of the map $\pi_0(K_2^c) \to \pi_0(K_1^c)$ is contained in $\pi_0^u(K_1^c)$. Combined with the fact that (B_n) is cofinal in $\mathcal{K}$, Exercises 1.23 and 1.23 now imply that the maps ϕ and ψ are continuous bijections. Since the domain of each map is compact and the range is Hausdorff, it follows that the maps ϕ, ψ are homeomorphisms.

Therefore, we can identify elements of $\epsilon(X)$ with decreasing sequences, called *chains*,

$$C_1 \supset C_2 \supset \dots$$

of components of the sets B_i^c, $i \in \mathbb{N}$, defined with respect to a fixed exhaustion of X as above.

One way to think about ends of X according to the definition, is that an end of X is a map $e : \mathcal{K} \to 2^X$, which sends each compact $K \subset X$ to a component C of K^c, such that

$$K_1 \subset K_2 \Rightarrow e(K_2) \subset e(K_1).$$

The topology on $\epsilon(X)$ extends to a topology on $\bar{X} = X \cup \epsilon(X)$: The basis of topology at $e \in \epsilon(X)$ is the collection of subsets $B_{K,e} \subset \bar{X}, K \in \mathcal{K}$, where $B_{K,e} \cap X = e(K)$

and $B_{K,e} \cap \epsilon(X)$ consists of all maps $e' : \mathcal{K} \to 2^X$, such that $e'(K) = e(K)$. We will also refer to each set $e(K)$ as a *neighborhood of* e *in* X.

The topology on X is, of course, the original one. It is then immediate that X is open and dense in $\bar{X}$.

We will say that a compact subset $K \subset X$ *separates* ends e, e' of X if e, e' belong to distinct components of $\bar{X} \setminus K$. Equivalently, there are unbounded components C, C' of K^c such that (C, e) is a neighborhood of e and (C', e') is a neighborhood of e' in $\epsilon(X)$.

Exercise 6.11. Every topological action $G \curvearrowright X$ extends to a topological action of G on $\bar{X}$.

Remark 6.12. There is a terminological confusion here coming from the literature in differential geometry and geometric analysis, where X is a complete Riemannian manifold: An analyst would call each unbounded set C_i above, an end of X.

Here is yet another alternative description of the space $\epsilon(X)$. From each C_i we pick a point x_i. Then, for each chain (C_i) defining the end $e \in \epsilon(X)$, the sequence (x_i), denoted $x_\bullet$, *represents* the end e. Given a sequence $x_\bullet$ representing e, we connect each x_i to x_{i+1} by a path contained in C_i. The concatenation of these paths is a *ray* in X, i.e., a proper continuous map

$$r : \mathbb{R}_+ \to X, \quad r(i) = x_i.$$

Conversely, given a ray r in X, every sequence $t_i \in \mathbb{R}_+$ monotonically diverging to infinity, defines the sequence $x_\bullet$ (with $x_i = r(t_i)$) which represents an end e of X. This end is independent of the choice of a sequence t_i.

Two rays r_1, r_2 represent the same end of X if and only if for every compact $K \subset X$ there exists T such that for all $t \geq T$ the points $r_1(t), r_2(t)$ lie in the same component of K^c. We refer the reader to [**BH99**] and [**Geo08**] for more detailed description of $\epsilon(X)$ and topology on $\bar{X}$ using this interpretation of ends.

Exercise 6.13. 1. The space $\bar{X}$ is Hausdorff.

2. If X is 2nd countable, so is $\bar{X}$.

3. A sequence $x_\bullet$ in X represents the end e if and only if it converges to e in the topology of $\bar{X}$.

4. If X is a metric space and $(x_i), (x'_i)$ are sequences within bounded distance from each other:

$$\sup_i \operatorname{dist}(x_i, x'_i) < \infty$$

and (x_i) represents $e \in \epsilon(X)$ then (x'_i) also represents e.

An example of the space of ends is given by the Figure 6.1. The space X in this picture has five visibly different ends: $\epsilon_1, ..., \epsilon_5$. We have $K_1 \subset K_2 \subset K_3$. The compact K_1 separates the ends ϵ_1, ϵ_2. The next compact K_2 separates ϵ_3 from ϵ_4. Finally, the compact K_3 separates ϵ_4 from ϵ_5.

Lemma 6.14. *The space* $\bar{X} = X \cup \epsilon(X)$ *is compact.*

Proof. Let $\mathcal{V} = \{V_j\}_{j \in J}$ be an open cover of $\bar{X}$. Since $\epsilon(X)$ is compact, there is a finite subset

$$\{(K_i, e_i) : i = 1, \ldots, n\} = \mathcal{V}_1 \subset \mathcal{V},$$

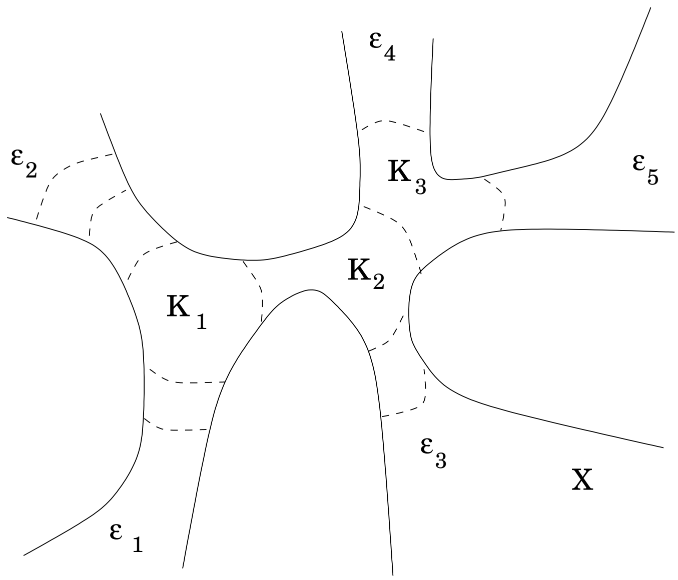


Figure 6.1. *E*nds of X.

which still covers $\epsilon(X)$. Here $K_i \in \mathcal{K}$ and $e_i \in \epsilon(X)$. Consider the open sets $C_i = e_i(K_i)$ (here we think of ends of X as maps $\mathcal{K} \to 2^X$). We claim that the closed set

$$A = X \setminus \bigcup_{i=1}^{n} C_i$$

is compact in X. If not, then there exists a sequence $x_k \in A$ which is not contained in any of the compact $K_i, i \in \mathbb{N}$. After passing to a subsequence in the sequence (x_k), we get a decreasing sequence of complementary sets

$$C_{k_l} \subset X \setminus K_{k_l},$$

such that $x_{k_l} \in C_{k_l}$. This sequence of complementary sets defines an end $e \in \epsilon(X)$ not covered by any of the sets $(K_i, e_i), i = 1, \ldots, n$, which is a contradiction.

Thus, $A \subset X$ is compact. Then there exists another finite subset $\mathcal{V}_2 \subset \mathcal{V}$, which covers A. Therefore,

$$\mathcal{V}_1 \cup \mathcal{V}_2$$

is a finite subcover of $\bar{X}$. □

Corollary 6.15. *1. $\bar{X}$ is a compactification of X.*
2. The space $\bar{X}$ is normal.

Exercise 6.16. 1. $\eta(X)$ is finite if and only if $\epsilon(X)$ is finite.
2. If $\eta(X)$ is finite, then $\eta(X)$ equals the cardinality of $\epsilon(X)$.

Proofs of 1 and 2 amount to simply following the definitions of $\eta(X)$ and $\epsilon(X)$.

Remark 6.17. One can think of the space of ends of X as its "π_0 at infinity." One can also define higher homotopy and (co)homology groups of X "at infinity", by replacing $\pi_0(K^c)$ with suitable homotopy or (co)homology groups and then taking inverse (or direct) limit. See **[Geo08]**.

PROPOSITION 6.18. *Every quasiisometry of proper geodesic metric spaces $X \to X'$ induces a homeomorphism $\epsilon(X) \to \epsilon(X')$.*

PROOF. The proof of this proposition follows the proof of Lemma 6.5 and below we only sketch the proof, leaving details to the reader.

Let $f : X \to X'$ be an (L, A)-quasiisometry. As in the proof of Lemma 6.5, we observe that there exists $r = r(L, A)$ such that for every connected subset $C \subset X$, the subset $\mathcal{N}_r(f(C)) \subset X'$ is also connected.

We define a map $\epsilon(f) : \epsilon(X) \to \epsilon(X')$ as follows. Suppose that $e \in \epsilon(X)$ is represented by a nested sequence (C_i), where C_i is a component of K_i^c, $K_i = \bar{B}(x, i)$. Each connected subset $\mathcal{N}_r(f(C_{i'}))$ will be disjoint from $\bar{B}(x', i)$ $(x' = f(x))$, where $i \mapsto i'$ is a nonconstant linear function, depending only on L and A. Let D_i denote the (unbounded) component of $\bar{B}(x', i)$ containing $\mathcal{N}_r(f(C_{i'}))$. The sets D_i are nested: $D_{i+1} \subset D_i$, $i \in \mathbb{N}$. Therefore, the sequence (D_i) defines an end e' of X', and we set $\epsilon(f)(e) := e'$. Proof of injectivity of the map $\epsilon(f)$ is the same as the 3rd part of the argument in Lemma 6.5.

In order to verify continuity of $\epsilon(f)$, let $D_i \subset Y$ be a neighborhood of $e' = \epsilon(f)(e)$, as above. Then, as we noted, $f(C_{i'}) \subset D_i$, where $i \to i'$ is a nonconstant linear function. Thus, if (C_j) is a chain representing e, then for all $j \geq i'$, $f(C_j) \subset D_i$. Therefore, the entire neighborhood of $e \in \epsilon(X)$ defined by the pair $(K_{i'}, e)$, is mapped by $\epsilon(f)$ into the neighborhood of e', defined by the pair $(\bar{B}(x', i), e')$. Continuity of $\epsilon(f)$ follows.

In order to prove surjectivity of $\epsilon(f)$, take r such that $\mathcal{N}_r(f(X)) = X'$. Then, given a sequence (x'_i) in X' representing an end $e' \in \epsilon(X')$, find a sequence (x_i) in X such that

$$\mathrm{dist}_Y(f(x_i), x'_i) \leq r.$$

Then the sequence (x_i) will converge to an end e of X and $\epsilon(f)(e) = e'$. □

EXERCISE 6.19. 1. Show that every bounded perturbation of the identity $f : X \to X$ extends to the identity map $\epsilon(f) : \epsilon(X) \to \epsilon(X)$.

2. Suppose that $f : X \to Y, g : Y \to Z$ are quasiisometries. Show that

$$\epsilon(g \circ f) = \epsilon(g) \circ \epsilon(f).$$

3. Conclude that if $g : Y \to X$ is a coarse inverse to $f : X \to Y$, then $\epsilon(g)$ is the inverse of $\epsilon(f)$. This gives another proof, in Proposition 6.18, of the claim that $\epsilon(f)$ is invertible.

Proposition 6.18 immediately implies:

COROLLARY 6.20. *Quasi-isometric spaces have homeomorphic spaces of ends.*

EXERCISE 6.21. Suppose that X is a simplicial tree of finite valence, where all but finitely many vertices have the same valence $k \geq 3$. Then $\epsilon(X)$ is homeomorphic to the Cantor set.

Hint: In order to prove this, consider first the case when X is a binary rooted tree, i.e., a tree with one distinguished vertex v_0 (the root) of valence 2 and the rest of the vertices of the valence 3. Then consider the ternary Cantor set E. This set is obtained by intersecting closed subsets $A_i \subset [0, 1]$, $i \in \mathbb{N}$; each A_i is the disjoint union of 2^i closed intervals $J_{i,k}$. Similarly, for $i \in \mathbb{N}$, the complement to the closed ball $K_i = \bar{B}(v_0, i) \subset X$ consists of 2^i components $C_{i,k}$. Points in E are encoded by decreasing sequences of intervals $J_{i,k}$, while points in $\epsilon(X)$ are encoded by chains $(C_{i,k})$. Now use bijections between the sets $\{J_{i,k} : k = 1, \dots, 2^i\}$ and $\{C_{i,k} : k = 1, \dots, 2^i\}$. For more general simplicial trees follow the geometric proof of the Example 5.48.

6.1.3. Ends of groups. Suppose that G is a finitely generated group. Then we define the space of ends $\epsilon(G)$ as the space of ends of its Cayley graph X. Corollary 6.20 shows that $\epsilon(G)$ is independent of the generating set. It follows from the Exercise 6.11 that the group G acts topologically on $\bar{X} = X \cup \epsilon(G)$. The same applies if instead of the Cayley graph we use as X a Riemannian manifold M, on which G acts isometrically, properly discontinuously and cocompactly.

A proof of the following theorem can be found for instance [**BH99**, Theorem 8.32]:

THEOREM 6.22 (Properties of $\epsilon(X)$). *1. Suppose that G is a finitely generated group. Then $\epsilon(G)$ consists of 0, 1, or 2 points, or is infinite. In the latter case, the topological space $\epsilon(G)$ is perfect. In particular, $\epsilon(G)$ is homeomorphic to the Cantor set.*

2. $\epsilon(G)$ is empty iff G is finite. $\epsilon(G)$ consists of 2-points if and only if G is virtually (infinite) cyclic. In particular, G is quasiisometric to $\mathbb{Z}$ if and only if G is virtually isomorphic to $\mathbb{Z}$.

3. If G splits nontrivially over a finite subgroup then $|\epsilon(G)| > 1$.

Below we prove Part 2 of this theorem. Our proof (which we learned from Mladen Bestvina) is differential-geometric, in line with the arguments in Chapters 18 and 19. A combinatorial argument can be found in [**BH99**].

PROPOSITION 6.23. *Every 2-ended group G is virtually isomorphic to $\mathbb{Z}$ and, hence, contains a finite index subgroup isomorphic to $\mathbb{Z}$.*

PROOF. Let M be an n-dimensional oriented Riemannian manifold on which G acts isometrically, properly discontinuously and cocompactly, preserving orientation. We let $\omega \in \Lambda^n(M)$ denote the volume form of M. Since M is QI to G, the manifold M is also 2-ended. After passing to an index 2 subgroup of G, we can assume that G fixes the ends e_1, e_2 of M. Every compact connected hypersurface $S \subset M$ separating the ends of M has a canonical coorientation, such that the end e_1 lies to the left of S. Since M is oriented, we, therefore, obtain a canonical orientation on S. This orientation is preserved under the action of G. The oriented hypersurface S will be regarded below as a smooth singular cycle in M, an element of $Z_{n-1}(M)$. (This cycle is the image of the fundamental cycle of S under the map $Z_{n-1}(S) \to Z_{n-1}(M)$.) Accordingly, $-S$ is the hypersurface S with reversed orientation. We claim that for every $g \in G$, the oriented hypersurfaces $S, g(S)$, represent the same homology class in $H_{n-1}(M)$. Indeed, the hypersurface $g(S)$ still separates the ends of M. If $g(S) \cap S = \emptyset$, then, since M is 2-ended, there exists a compact submanifold $B \subset M$ whose (oriented) boundary equals $-S \cup g(S)$. Hence, $[S] = [g(S)] \in H_{n-1}(M)$. For arbitrary $g \in G$ we take $h \in G$ such that $h(S) \cap S = \emptyset$ and $h(S) \cap g(S)$ and obtain

$$[S] = [h(S)] = [g(S)].$$

We, thus, obtain a homomorphism

$$\phi : G \to \mathbb{R},$$

defined by

$$\phi(g) = \int_B \omega$$

where $B \in C_n(M)$ is a smooth singular chain such that

$$g(S) - S = B.$$

As we observed above, if $g(S) \cap S \neq \emptyset$, then B is realized by a submanifold in M, which implies that $\phi(g) \neq 0$ in this case. Since the action of G on M is properly discontinuous, the map $\phi : G \to \mathbb{R}$ is proper. In particular, its image is an infinite cyclic group and its kernel is finite. Therefore, the group G is virtually isomorphic to $\mathbb{Z}$. The existence of a finite index subgroup of G isomorphic to $\mathbb{Z}$ was proven in Lemma 4.105. ☐

Part 3 of Theorem 6.22 has a deep and important converse:

THEOREM 6.24. *If $|\epsilon(G)| > 1$ then G splits nontrivially over a finite subgroup.*

This theorem is due to Stallings [**Sta68**] (in the torsion-free case) and Bergman [**Ber68**] for groups with torsion. To this day, there is no simple proof of this result. A geometric proof could be found in Niblo's paper [**Nib04**]. For finitely presented groups, there is an easier combinatorial proof due to Dunwoody using minimal tracks, [**Dun85**]; a combinatorial version of this argument could be found in [**DD89**]. In Chapters 18 and 19 we prove Theorem 6.24 first for finitely presented, and then for all finitely generated groups. We will also prove QI rigidity of the class of virtually free groups.

An immediate corollary of Theorem 6.24 (and QI invariance of the number of ends) is

COROLLARY 6.25. *Suppose that a finitely generated group G splits nontrivially as $G_1 \star G_2$ and G' is a group quasiisometric to G. Then G' splits nontrivially as $G'_1 \star_F G'_2$ (amalgamated product) or as $G'_1 \star_F$ (HNN splitting), where F is a finite group.*

We conclude this section with a technical result which will be used in §19.3 for the proof of Stallings theorem via harmonic functions.

LEMMA 6.26. *Suppose that M is a complete connected n-dimensional Riemannian manifold and X is the corresponding metric space. Let $\chi : \epsilon(X) \to \{0, 1\}$ be the characteristic function of a clopen subset $A \subset \epsilon(X)$. Then χ admits a continuous extension $\varphi : \bar{X} \to [0, 1]$ which is smooth on M and $d\varphi\big|_M$ is compactly supported in M.*

PROOF. Since the space $\bar{X}$ is normal, the disjoint closed sets A and $B = \epsilon(X) \setminus A$ admit disjoint open neihborhoods $U \subset \bar{X}$ and $\bar{V} \subset \bar{X}$, respectively. We first extend χ to a function $\psi : U \cup V \to \{0, 1\}$, which is constant on U and on V. Next, by Tietze–Urysohn extension theorem (Theorem 1.14) the function $\psi : U \cup V \to \{0, 1\}$ admits a continuous extension $\zeta : \bar{X} \to \mathbb{R}$. After replacing ζ with $\zeta_0 = \max(\zeta, 0)$ and, afterwards, with $\zeta_1 = \min(\zeta, 1)$, we may assume that $\zeta : \bar{X} \to [0, 1]$. To get a smooth extension, consider a smooth partition of unity $\{\eta_i\}_{i \in I}$ corresponding to a locally finite open covering $\{U_i\}_{i \in I}$ of M via subsets diffeomorphic to the unit open ball $\mathbb{D} \subset \mathbb{R}^n$. We choose the functions η_i to have unit integrals over U_i (with respect to the Lebesgue measure coming from $\mathbb{D}$). Using the diffeomorphisms $f_i : U_i \to \mathbf{B}$, we define convolutions

$$\zeta \star \eta_i(x) = \int_{\mathbb{D}} \zeta(f_i(y))\eta_i(f_i(x))dx.$$

The sum

$$\varphi = \sum_{i\in I} \zeta \star \eta_i$$

is the required extension. □

6.2. Rips complexes and coarse homotopy theory

Connecting the dots. In the proof of Lemma 6.5, we saw an important principle of coarse topology: In order to recover a useful topological object from the image $f(C)$ of a set under a quasiisometry f, we first discretize C (replace C with a net $C' \subset C$) and then "connect the dots" in $f(C')$: Connect certain points (which are not too far from each other) in $f(C')$ by geodesic segments in the ambients space. How far the "connected dots" should be from each other is determined by geometry of the metric spaces involved and quasiisometric constants of f. The same principle will reappear in this section: "Connecting dots" will be replaced by taking a subcomplex of a suitable Rips complex Rips_R. The ambiguity in choosing the scale R (how far apart the "dots" can be) forces us to work with direct systems of Rips complexes and direct/inverse systems of the associated homotopy, homology and cohomology groups.

6.2.1. Rips complexes. Recall (Definition 1.71) that the R-Rips complex of a metric space X is the simplicial complex whose vertices are the points of X; vertices $x_1, ..., x_n$ span a simplex if and only if

$$\mathrm{dist}(x_i, x_j) \leqslant R, \forall i, j.$$

For each pair $0 \leqslant R_1 \leqslant R_2 < \infty$ we have a natural simplicial embedding

$$\iota_{R_1,R_2} : \mathrm{Rips}_{R_1}(X) \to \mathrm{Rips}_{R_2}(X),$$

such that

$$\iota_{R_1,R_3} = \iota_{R_2,R_3} \circ \iota_{R_1,R_2},$$

provided that $R_1 \leqslant R_2 \leqslant R_3$. Thus, the collection of Rips complexes of X forms a direct system $\mathrm{Rips}_\bullet(X)$ of simplicial complexes indexed by positive real numbers.

Following the construction in §2.8, we metrize (connected) Rips complexes $\mathrm{Rips}_R(X)$ using the *standard metric* on simplicial complexes. Then, each embedding ι_{R_1,R_2} is isometric on every simplex and is 1-Lipschitz overall. Note that if X is uniformly discrete (see Definition 1.50), then for every R, the complex $\mathrm{Rips}_R(X)$ is a simplicial complex of bounded geometry (Definition 2.33).

EXERCISE 6.27. 1. Suppose that $X = G$, a finitely generated group with a word metric. Show that for every R, the action of G on itself extends to a simplicial action of G on $\mathrm{Rips}_R(G)$. Show that this action is geometric.

2. Show that a metric space X is quasigeodesic (see §5.1) if and only if for all sufficiently large R the Rips complex $\mathrm{Rips}_R(X)$ is connected and the inclusion $X \to \mathrm{Rips}_R(X)$ is a quasiisometry.

The following simple observation explains why Rips complexes are useful for analyzing quasiisometries:

Lemma 6.28. *Let $f : X \to Y$ be an (L, A)–coarse Lipschitz map. Then f induces a simplicial map $\mathrm{Rips}_R(X) \to \mathrm{Rips}_{LR+A}(Y)$ for each $R \geqslant 0$. We retain the notation f for this simplicial map.*

Proof. Consider an m-simplex σ in $\mathrm{Rips}_R(X)$; the vertices of σ are distinct points $x_0, x_1, ..., x_m \in X$ within distance $\leqslant R$ from each other. Since f is (L, A)–coarse Lipschitz, the points $f(x_0), ..., f(x_m) \in Y$ are within distance $\leqslant LR + A$ from each other, hence, they span a simplex σ' of dimension $\leqslant m$ in $\mathrm{Rips}_{LR+A}(Y)$. The map f sends vertices of σ to vertices of σ'. Thus, we have a simplicial map of simplicial complexes $\mathrm{Rips}_R(X) \to \mathrm{Rips}_{LR+A}(Y)$. □

The idea behind the next definition is that the "coarse homotopy groups" of a metric space X are the homotopy groups of the Rips complexes $\mathrm{Rips}_R(X)$ of X. Literally speaking, this does not make much sense since the above homotopy groups depend on R. To eliminate this dependence, we have to take into account the maps $\iota_{r,R}$.

Definition 6.29. 1. A metric space X is *coarsely connected* if $\mathrm{Rips}_r(X)$ is connected for some r. (Equivalently, $\mathrm{Rips}_R(X)$ is connected for all sufficiently large R.)

2. A metric space X is *coarsely k-connected* if it is coarsely connected and for each r there exists $R \geqslant r$ such that the mapping $\mathrm{Rips}_r(X) \to \mathrm{Rips}_R(X)$ induces trivial maps of the homotopy groups

$$\pi_i(\mathrm{Rips}_r(X), x) \to \pi_i(\mathrm{Rips}_R(X), x),$$

for all $1 \leqslant i \leqslant k$ and $x \in X$.

In particular, X is *coarsely simply-connected* if it is coarsely 1-connected.

For instance, X is coarsely connected if there exists a number R such that each pair of points $x, y \in X$ can be connected by an R-chain of points $x_i \in X$, i.e., a finite sequence of points x_i, where $\mathrm{dist}(x_i, x_{i+1}) \leqslant R$ for each i.

The definition of coarse k-connectedness is not quite satisfactory since it only deals with "vanishing" of coarse homotopy groups without actually defining these groups for a general metric space X. One way to deal with this issue is to consider *pro-groups*, which are direct systems

$$\pi_i(\mathrm{Rips}_r(X)), r \in \mathbb{N},$$

of groups. Given such algebraic objects, one can define their *pro-homomorphisms*, *pro-monomorphisms*, etc., see [**KK05**] where this is done in the category of abelian groups (the homology groups). Alternatively, one can work with the direct limit of the homotopy groups.

6.2.2. Direct system of Rips complexes and coarse homotopy.

Lemma 6.30. *Let X be a metric space. Then for $r, c < \infty$, each simplicial spherical cycle σ of diameter $\leqslant c$ in $\mathrm{Rips}_r(X)$ bounds a singular disk of diameter $\leqslant r+c$ within $\mathrm{Rips}_{r+c}(X)$. More precisely, every simplicial map of a triangulated $n-1$ sphere, $\sigma : \mathbb{S}^{n-1} \to \mathrm{Rips}_r(X)$, extends to a simplicial map $\tau : \mathbb{D}^n \to \mathrm{Rips}_{r+c}(X)$, where $\mathbb{D}^n$ is a triangulated n-disk whose triangulation agrees with that one of $\mathbb{S}^{n-1}$.*

Proof. Pick a vertex $x \in \mathrm{Im}(\sigma)$. Then $\mathrm{Rips}_{r+c}(X)$ contains the simplicial cone $C = \tau(\mathbb{D}^n)$ over $\mathrm{Im}(\sigma)$ with vertex at x. Clearly, $\mathrm{diam}(C) \leqslant r + c$. Coning off the map σ from the vertex x, defines an extension τ of σ to the n-disk, which we identify with the cone over $\mathbb{S}^{n-1}$. □

Recall that the product of simplicial complexes $C \times [0,1]$ admits a certain standard triangulation (determined by an ordering of vertices of X and the set $\{0,1\}$). We will always equip this product simplicial complex with the standard metric.

Proposition 6.31. *Let $f, g : X \to Y$ be maps within distance $\leqslant c$ from each other, which extend to simplicial maps*

$$f, g : \mathrm{Rips}_{r_1}(X) \to \mathrm{Rips}_{r_2}(Y).$$

Then for $r_3 = r_2 + c$, the maps

$$f, g : \mathrm{Rips}_{r_1} \to \mathrm{Rips}_{r_3}(Y)$$

are homotopic via a 1-Lipschitz homotopy $F : \mathrm{Rips}_{r_1}(X) \times I \to \mathrm{Rips}_{r_3}(Y)$. Furthermore, tracks of this homotopy have length $\leqslant (n+1)$, where $n = \dim(\mathrm{Rips}_{r_1}(X))$.

Proof. The map F of the zero-skeleton of $\mathrm{Rips}_{r_1}(X) \times I$ is, of course, just $F(x,0) = f(x), F(x,1) = g(x)$. Let $\sigma \subset \mathrm{Rips}_{r_1}(X) \times I$ be an i–simplex. Then

$$\mathrm{diam}\,(F(V(\sigma))) \leqslant r_3 = r_2 + c,$$

where $V(\sigma)$ is the vertex set of σ. Therefore, F extends (linearly) from σ^0 to a (1-Lipschitz) map $F : \sigma \to \mathrm{Rips}_{r_3}(Y)$ whose image is the simplex spanned by $F(\sigma^0)$.

To estimate the lengths of the tracks of the homotopy F, we note that for each $x \in \mathrm{Rips}_{r_1}(X)$, the path $F(x,t)$ has length $\leqslant 1$ since the interval $x \times I$ is covered by $\leqslant (n+1)$ simplices, each of which has unit diameter. □

In view of the above lemma, we make the following definition:

Definition 6.32. Maps $f, g : X \to Y$ are *coarsely homotopic* if for all r_1, r_2, such that f and g extend to

$$f, g : \mathrm{Rips}_{r_1}(X) \to \mathrm{Rips}_{r_2}(Y),$$

there exist r_3 and r_4 so that the maps

$$f, g : \mathrm{Rips}_{r_1}(X) \to \mathrm{Rips}_{r_3}(Y)$$

are homotopic via a homotopy whose tracks have lengths $\leqslant r_4$.

We then say that a map $f : X \to Y$ determines a *coarse homotopy equivalence* (between the direct systems of Rips complexes of X, Y), if there exists a map $g : Y \to X$ such that the compositions $g \circ f, f \circ g$ are coarsely homotopic to the identity maps.

The next two corollaries, then, are immediate consequences of Proposition 6.31.

Corollary 6.33. *Let $f, g : X \to Y$ be L–Lipschitz maps within finite distance from each other. Then they are coarsely homotopic.*

Corollary 6.34. *If $f : X \to Y$ is a quasiisometry, then f induces a coarse homotopy-equivalence of the Rips complexes:* $\mathrm{Rips}_\bullet(X) \to \mathrm{Rips}_\bullet(Y)$.

The following corollary is a coarse analogue of the familiar fact that homotopy equivalence preserves connectivity properties of a space:

COROLLARY 6.35. *Coarse k-connectedness is a QI invariant.*

PROOF. Suppose that X' is a coarsely k-connected metric space and $f : X \to X'$ is an L–Lipschitz quasiisometry with L–Lipschitz quasiinverse $\bar{f} : X' \to X$. Let γ be a spherical i-cycle in $\mathrm{Rips}_r(X)$, $0 \leqslant i \leqslant k$. Then we have the spherical i-cycle $f(\gamma) \subset \mathrm{Rips}_{Lr}(X')$. Since X' is coarsely k-connected, there exists $r' \geqslant Lr$ such that $f(\gamma)$ bounds a singular $(i+1)$–disk β within $\mathrm{Rips}_{r'}(X')$. Consider now $\bar{f}(\beta) \subset \mathrm{Rips}_{L^2r}(X)$. The boundary of this singular disk is a singular i-sphere $\bar{f}(\gamma)$. Since $\bar{f} \circ f$ is homotopic to id within $\mathrm{Rips}_{r''}(X)$, $r'' \geqslant L^2 r$, there exists a singular cylinder σ in $\mathrm{Rips}_{r''}(X)$ which cobounds γ and $\bar{f}(\gamma)$. Note that r'' does not depend on γ. By combining σ and $\bar{f}(\beta)$ we get a singular $(i+1)$–disk in $\mathrm{Rips}_{r''}(X)$ whose boundary is γ. Hence, X is coarsely k-connected. □

6.3. Metric cell complexes

We now introduce a generalization of metric simplicial complexes, where the notion of bounded geometry does not imply finite-dimensionality. The objects that we will consider are called *metric cell complexes*, they are hybrids of metric spaces and CW complexes. The advantage of metric cell complexes over metric simplicial complexes is the same as of CW complexes over simplicial complexes in the traditional algebraic topology: CW complexes are more flexible.

A *metric cell complex* is a cell complex X together with a metric d defined on its 0-skeleton $X^{(0)}$. Note that if X is connected, its 1-skeleton $X^{(1)}$ is a graph, and, hence, can be equipped with the standard metric dist. The map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$, in general, need not be a quasiisometry. However, in the most interesting cases, coming from finitely generated groups, this map is actually an isometry. Therefore, we impose, from now on, the condition:

Axiom 1. The map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$ is a quasiisometry. Equivalenty, X is a quasigeodesic metric space.

Even though this assumption could be avoided in what follows, restricting to complexes satisfying this axiom makes our discussion more intuitive.

Our first goal is to define, using the metric d, certain metric concepts on the entire complex X. We define inductively a map c, which sends cells in X to finite subsets of $X^{(0)}$ as follows. For a vertex $v \in X^{(0)}$ we set $c(v) = \{v\}$. Suppose that c is defined on $X^{(i)}$. For each closed $i+1$-cell e, the *support* of e is the smallest subcomplex $\mathrm{Supp}(e)$ of $X^{(i)}$, containing the image of the attaching map of e to $X^{(i)}$. We then set

$$c(\sigma) = c(\mathrm{Supp}(e)).$$

For instance, for every 1-cell σ, $c(\sigma)$ consists of one or two vertices of X to which σ is attached.

REMARK 6.36. The reader familiar with the concepts of *controlled topology*, see e.g. [**Ped95**], will realize that the coarsely defined map $c : X \to X^{(0)}$ is a *control map* for X and $(X^{(0)}, d)$ is the *control space*. Metric cell complexes form a subclass of *metric chain complexes* defined in [**KK05**].

The *diameter* $\mathrm{diam}(\sigma)$ of a cell σ in X is defined to be the diameter of $c(\sigma)$.

Example 6.37. Take a connected simplicial complex X and restrict its standard metric to $X^{(0)}$. Then, the diameter of a cell in X (as a simplicial complex) is the same as its diameter in the sense of metric cell complexes.

Definition 6.38. A metric cell complex X is said to have *bounded geometry* if there exists a collections of increasing functions $\phi_k(r)$ and numbers $D_k < \infty$ such that the following axioms hold:

Axiom 2. For each ball $B(x,r) \subset X^{(0)}$, the set of k-cells σ such that $c(\sigma) \subset B(x,r)$, contains at most $\phi_k(r)$ cells.

Axiom 3. The diameter of each k-cell is at most $D_k = D_{k,X}$, $k \in \mathbb{N}$.

Axiom 4. $D_0 := \inf\{d(x,x') : x \neq x' \in X^{(0)}\} > 0$.

Note that we allow X to be infinite-dimensional. We will refer to the function $\phi_k(r)$ and the numbers D_k as *geometric bounds* on X, and set

$$D_X = \sup_{k>0} D_{k,X}. \tag{6.1}$$

The basic examples of metric cell complexes of bounded geometry are:

1. Simplicial complexes of bounded geometry.

2. Let M be a connected Riemannian manifold of bounded geometry and X is a simplicial complex defined in Theorem 2.36. Now, equip $X^{(0)}$ with the distance function d which is restriction of the Riemannian distance function on M to the vertex set of X.

3. $X^{(0)} := G$ is a finitely generated group with its word metric and X is the Cayley graph of G with the standard metric.

4. A covering space X of a connected finite cell complex Y. Equip $X^{(0)}$ with the restriction of the distance function dist on $X^{(1)}$.

5. Consider the spheres $\mathbb{S}^n$ with the standard CW complex structure (single 0-cell and single n-cell). Then, the cellular embeddings $\mathbb{S}^n \hookrightarrow \mathbb{S}^{n+1}$ give rise to an infinite-dimensional cell complex S^∞. This complex has bounded geometry (since it has only one cell in every dimension). In view of this trivial example, the concept of metric cell complexes is more flexible than the one of simplicial complexes.

Exercise 6.39. 1. Suppose that X is a simplicial complex. Then the two notions of bounded geometry coincide for X. We will use this special class of metric cell complexes in §6.6.

2. If X is a metric cell complex of bounded geometry and $S \subset X$ is a connected subcomplex, then for every two vertices $u, v \in S$ there exists a chain $x_0 = u, x_1, ..., x_m = v$, such that $d(x_i, x_{i+1}) \leqslant D_1$ for every i. In particular, if X is connected, the identity map $(X^{(0)}, d) \to (X^{(1)}, \text{dist})$ is D_1-Lipschitz.

Exercise 6.40. Let X, Y be metric cell complexes. Then the product cell–complex $X \times Y$ is also a metric cell complex, where we equip the zero-skeleton $X^{(0)} \times Y^{(0)}$ of $X \times Y$ with the product–metric. Furthermore, if X, Y have bounded geometry, then so does $X \times Y$.

We now continue defining metric concepts for metric cell complexes. The (coarse) R-ball $\mathbf{B}(x,R)$ centered at a vertex $x \in X^{(0)}$ is defined as the union of the cells σ in X such that $c(\sigma) \subset B(x,R)$.

We will say that the *diameter* $\mathrm{diam}(S)$ of a subcomplex $S \subset X$ is the diameter of $c(S)$. Given a subcomplex $W \subset X$, we define the *closed R-neighborhood* $\bar{\mathcal{N}}_R(W)$ of W in X to be the largest subcomplex $S \subset X$ such that for every $\sigma \in S$, there exists a vertex $\tau \in W$ such that $\mathrm{dist}_{Haus}(c(v), c(w)) \leqslant R$. A cellular map $f : X \to Y$ between metric cell complexes is called *L–Lipschitz* if for every cell σ in X, we have $\mathrm{diam}(f(\sigma)) \leqslant L$. In particular, the map

$$f : (X^{(0)}, d) \to (Y^{(0)}, d)$$

is $\frac{L}{D_0}$-Lipschitz as a map of metric spaces.

EXERCISE 6.41. Suppose that $f_i : X_i \to X_{i+1}$ are L_i-Lipschitz for $i = 1, 2$. Show that $f_2 \circ f_1$ is L_3-Lipschitz with

$$L_3 = L_2 \max_k \left(\phi_{X_2,k}(L_1)\right).$$

EXERCISE 6.42. Construct examples of a cellular map $f : X \to Y$ between metric graphs of bounded geometry, such that the restriction $f\big|_{X^{(0)}}$ is L–Lipschitz, but f is not L'–Lipschitz, for any $L' < \infty$.

The following definition is a version of the notion of uniformly proper maps of metric spaces in Definition 5.25. A map $f : X \to Y$ of metric cell complexes is called a *uniformly proper cellular map*, if f is cellular, L-Lipschitz for some $L < \infty$ and $f\big|_{X^{(0)}}$ is uniformly proper: There exists a proper function $\eta(R)$ such that

$$d(f(x), f(x')) \geqslant \eta(d(x, x')),$$

for all $x, x' \in X^{(0)}$. The function $\eta(R)$ is called the *(lower) distortion function* of f. We will frequently omit the adjective *cellular* when talking about uniformly proper maps of metric cell complexes.

For instance, suppose that H is a finitely generated group and $G \leqslant H$ is a finitely generated subgroup, whose generating set is contained in the one of H. Let X and Y denote the Cayley graphs of G and H, respectively. Then the inclusion map $X \to Y$ is a uniformly proper cellular map. As another example, suppose that G is the fundamental group of a finite cell complex X_1, H is the fundamental group of a finite cell complex Y_1 and $f_1 : X_1 \to Y_1$ is a cellular map inducing the inclusion of fundamental groups $G \hookrightarrow H$. Let $f : X \to Y$ be a lift of f_1 to the universal covers X, Y of X_1, Y_1, respectively. Then f is a uniformly proper cellular map.

We now relate metric cell complexes of bounded geometry to simplicial complexes of bounded geometry:

EXERCISE 6.43. Let X be a *finite-dimensional* metric cell complexes of bounded geometry. Then there exists a simplicial complex Y of bounded geometry and a cellular homotopy-equivalence $X \to Y$ which is a quasiisometry in the following sense: f and has homotopy-inverse $\bar{f}$ so that:

1. Both $f, \bar{f}$ are L-Lipschitz for some $L < \infty$.
2. $f \circ \bar{f}, \bar{f} \circ f$ are homotopic to the identity.
3. The maps $f : X^{(0)} \to Y^{(0)}, \bar{f} : Y^{(0)} \to X^{(0)}$ are quasiinverse to each other:

$$d(f \circ \bar{f}, \mathrm{id}) \leqslant A, \quad d(\bar{f} \circ f, \mathrm{id}) \leqslant A.$$

Hint: Apply the standard construction which converts a finite-dimensional CW-complex into a simplicial complex, see e.g. **[Hat02]**.

Recall that quasiisometries are not necessarily continuous. In order to use algebraic topology, we, thus, have to approximate quasiisometries by cellular maps in the context of metric cell complexes, as it was done for Rips complexes (Lemma 6.28). In general, such approximation is, of course, impossible, since one complex in question can be, say, 0-dimensional and the other 1-dimensional. The *uniform contractibility* hypothesis allows one to resolve this issue.

Suppose that X and Y are call complexes and $f : X \to Y$ is a cellular map. We will say that the map f is *k-null* if it induces zero map $\tilde{H}_0(X) \to \tilde{H}_0(Y)$ and trivial maps of all homotopy groups

$$\pi_i(X) \to \pi_i(Y), \quad 1 \leqslant i \leqslant k.$$

Definition 6.44. A metric cell complex X is said to be *uniformly contractible* if there exists a continuous function $\psi(R)$ such that for every $x \in X^{(0)}$ the map

$$\mathbf{B}(x, R) \to \mathbf{B}(x, \psi(R))$$

is null-homotopic. Similarly, X is *uniformly k-connected* if there exists a function $\psi_k(R)$ such that for every $x \in X^{(0)}$ the map

$$\mathbf{B}(x, R) \hookrightarrow \mathbf{B}(x, \psi_k(R)) \tag{6.2}$$

is k-null. We will refer to ψ, ψ_k as the *contractibility functions* of X. By the abuse of terminology, we will say that the inclusion of the balls (6.2) induces *trivial maps of homotopy groups* π_i, $i \leqslant k$. (The abuse comes from the fact that for $k = 0$ we use the reduced homology.)

The above definition implies, for instance, that the entire ball $\mathbf{B}(x, R)$ is contained in a single connected component of $\mathbf{B}(x, \psi_0(R))$, every loop in $\mathbf{B}(x, R)$ bounds a singular disk in $\mathbf{B}(x, \psi_1(R))$.

Example 6.45. Suppose that X is a connected metric graph with the standard metric. Then X is uniformly 0-connected.

In general, even for simplicial complexes of bounded geometry, contractibility does not imply uniform contractibility. For instance, start with a triangulated 2-torus T^2, let X be an infinite cyclic cover of T^2. Of course, X is not contractible, but we attach a triangulated disk $\mathbb{D}^2$ to X along a simple homotopically nontrivial loop in $X^{(1)}$. The result is a contractible 2-dimensional simplicial complex Y which clearly has bounded geometry.

Exercise 6.46. Show that Y is not uniformly contractible.

We will see, nevertheless, in Lemma 6.50, that under certain assumptions (presence of a cocompact group action) contractibility implies uniform contractibility.

The reader uncomfortable with metric cell complexes in the proofs below, can think instead of Riemannian manifolds equipped with structures of CW-complexes, which appear as Riemannian cellular coverings of compact Riemannian manifolds, which are given structures of finite CW complexes. Instead of the notions of diameter in metric cell complexes used in the book, the reader can think of the ordinary Riemannian diameters.

The following proposition is a metric analogue of the cellular approximation theorem:

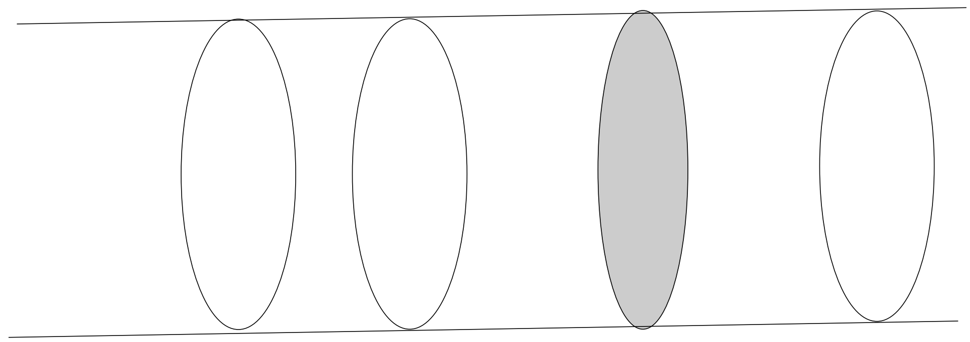


FIGURE 6.2. Contractible but not uniformly contractible space.

PROPOSITION 6.47 (Lipschitz cellular approximation). *Suppose that X, Y are metric cell complexes, where X is finite-dimensional and has bounded geometry, Y is uniformly contractible, and $f : X^{(0)} \to Y^{(0)}$ is an L–Lipschitz map. Then f admits a (continuous) cellular extension $f : X \to Y$, which is an L'–Lipschitz map, where L' depends on L and geometric bounds on the complex X and the uniform contractibility function of Y. Furthermore, $f(X) \subset \bar{N}_{L'}(f(X^{(0)}))$.*

PROOF. The proof of this proposition is a prototype of most of the proofs which appear in this and the following sections: It is a higher-dimensional version of the "collect the dots" process. The proof essentially amounts to a quantitive version of the proof of Whitehead's theorem (see e.g. Theorem 4.5 in [**Hat02**]).

We extend f by induction on skeleta of X. We claim that (for certain constants C_i, C'_{i+1}, $i \geqslant 0$) we can construct a sequence of extensions $f_k : X^{(k)} \to Y^{(k)}$ such that:

1. $\text{diam}(f(\sigma)) \leqslant C_k$ for every k-cell $\sigma = \hat{e}(\mathbb{D}^k)$ in $X^{(k)}$.
2. $\text{diam}(f(\partial\tau)) \leqslant C'_{k+1}$, for every $(k+1)$–cell τ in X.

Base of the induction. We already have $f = f_0 : X^{(0)} \to Y^{(0)}$ satisfying (1) with $C_0 = 0$. If x, x' belong to the boundary of a 1-cell τ in X then

$$\text{dist}(f(x), f(x')) \leqslant LD_1,$$

where $D_1 = D_{1,X}$ is the upper bound on the diameters of 1-cells in X. This establishes (2) in the base case.

Inductively, assume that $f = f_k$ was defined on $X^{(k)}$, so that (1) and (2) hold. Let $\sigma = \hat{e}(\mathbb{D}^{k+1})$ be a $(k+1)$–cell in X. Note that

$$\text{diam}(f(\partial\sigma)) \leqslant C'_{k+1},$$

by the induction hypothesis. Then, using uniform contractibility of Y, we extend f to σ so that the diameter of the image of σ in Y is bounded above by C_{k+1} where $C_{k+1} = \psi(C'_k)$. Namely, the composition $f \circ e : \partial\mathbb{D}^{n+1} \to Y$ is null-homotopic and, hence, extends to a map $\mathbb{D}^{n+1} \to Y$ of controlled diameter. Without loss of generality (cf. Whitehead's cellular approximation theorem, [**Hat02**, Theorem 4.8]), we can assume that this extension h is cellular, i.e., its image is contained in

$Y^{(n+1)}$. The extension of $f \circ e$ to $\mathbb{D}^{n+1}$ determines the required extension of f to $\hat{e}(\mathbb{D}^{k+1})$:

$$f(x) := \tilde{f}(\hat{e}^{-1}(x)), x \in \sigma.$$

We thus obtain a cellular map $f : X^{(k+1)} \to Y^{(k+1)}$.

Let us verify that the new map $f : X^{(k+1)} \to Y^{(k+1)}$ satisfies (2).

Suppose that τ is a $(k+2)$–cell in X. Then, since X has bounded geometry, $\operatorname{diam}(\tau) \leqslant D_{k+2} = D_{k+2,X}$. In particular, $\partial\tau$ is connected and is contained in the union of at most $\phi(D_{k+2}, k+1)$ cells of dimension $k+1$. Therefore,

$$\operatorname{diam}(f(\partial\tau)) \leqslant C_{k+1} \cdot \phi(D_{k+2}, k+1) =: C'_{k+2}.$$

This proves (2).

Since X is, say, n-dimensional, the induction terminates after n steps. The resulting map $f : X \to Y$ satisfies

$$L' := \operatorname{diam}(f(\sigma)) \leqslant \max_{i=1,\ldots,n} C_i,$$

for every cell σ in X. Therefore, $f : X \to Y$ is L'–Lipschitz. The second assertion of the proposition follows from the definition of C_i's. ☐

We note that Proposition 6.47 can be *relativized*:

LEMMA 6.48. *Suppose that X, Y are metric cell complexes, X is finite-dimensional and has bounded geometry, Y is uniformly contractible, and $Z \subset X$ is a subcomplex. Suppose that $f : Z \to Y$ is a continuous cellular map which extends to an L–Lipschitz map $f : X^{(0)} \to Y^{(0)}$. Then $f : Z \cup X^{(0)} \to Y$ admits a (continuous) cellular extension $g : X \to Y$, which is an L'–Lipschitz map, where L' depends on L, geometric bounds on X and contractibility function of Y.*

PROOF. The proof is the same induction on skeleta argument as in Proposition 6.47. ☐

COROLLARY 6.49. *Suppose that X, Y are as above and $f_0, f_1 : X \to Y$ are L–Lipschitz cellular maps such that* $\operatorname{dist}(f_0, f_1) \leqslant C$ *in the sense that*

$$d(f_0(x), f_1(x)) \leqslant C, \quad \forall x \in X^{(0)}.$$

Then there exists an L'–Lipschitz homotopy $f : X \times I \to Y$ between the maps f_0, f_1.

PROOF. Consider the map $f_0 \cup f_1 : X \times \{0, 1\} \to Y$, where $X \times \{0, 1\}$ is a subcomplex in the metric cell complex $W := X \times I$ (see Exercise 6.40). Then the required extension $f : W \to Y$ of this map exists by Lemma 6.48. ☐

6.4. Connectivity and coarse connectivity

Our next goal is to find a large supply of examples of metric spaces which are coarsely m-connected.

LEMMA 6.50. *If X is a finite-dimensional m-connected complex which admits a properly discontinuous, cocompact, cellular, isometric (on $X^{(0)}$) group action $G \curvearrowright X$, then X is uniformly m-connected.*

PROOF. Existence of the action $G \curvearrowright X$ implies that X is locally finite. Pick a base-vertex $x \in X$ and let $r < \infty$ be such that the G-orbit of $B(x, r) \subset X^{(0)}$ is the entire $X^{(0)}$. Therefore, if a subcomplex $C \subset X$ has diameter $\leqslant R/2$, there exists $g \in G$ such that $C' = g(C) \subset \mathbf{B}(x, r + R)$.

Since C is finite, its fundamental group $\pi_1(C')$ is finitely generated. Thus, simple connectivity of X implies that there exists a finite subcomplex $C'' \subset X$ such that each generator of $\pi_1(C')$ vanishes in $\pi_1(C'')$. Consider now $\pi_i(C'), 2 \leqslant i \leqslant m$. Then, by Hurewicz theorem, the image of $\pi_i(C')$ in $\pi_i(X) \cong H_i(X)$, is contained in the image of $H_i(C')$ in $H_i(X)$. Since C' is a finite complex, we can choose C'' above such that the map $H_i(C') \to H_i(C'')$ is zero. To summarize, there exists a finite connected subcomplex C'' in X containing C', such that all maps $\pi_i(C') \to \pi_i(C'')$ are trivial, $1 \leqslant i \leqslant m$.

Since C'' is a finite complex, there exists $R' < \infty$ such

$$C'' \subset \mathbf{B}(x, r + R + R').$$

Hence, the inclusion map

$$C' \to \mathbf{B}(x, r + R + R')$$

is m-null.

Set $\psi(k, r) = \rho = r + R'$. Therefore, taking into account action of G on X, we conclude that for each subcomplex $C \subset X$ of diameter $\leqslant R/2$, the inclusion map

$$C \to \mathcal{N}_\rho(C)$$

is m-null. □

Our next goal is to relate the notion of coarse n-connectivity from §6.2.2 to uniform n-connectivity for metric cell complexes.

THEOREM 6.51. *Suppose that X is a uniformly n-connected metric cell complex of bounded geometry. Then $Z := X^{(0)}$ is coarsely n-connected.*

PROOF. Let $\gamma : \mathbb{S}^k \to \mathrm{Rips}_R(Z)$ be a spherical m-cycle in $\mathrm{Rips}_R(Z)$, $0 \leqslant k \leqslant n$. Without loss of generality (using simplicial approximation) we can assume that γ is a simplicial cycle, i.e., the sphere $\mathbb{S}^k$ is given a triangulation τ such that γ is a simplicial map.

LEMMA 6.52. *There exists a cellular map $\gamma' : (\mathbb{S}^k, \tau) \to X$ which agrees with γ on the vertex set of τ and such that $\mathrm{diam}(\gamma'(\sigma)) \leqslant R'$, for each simplex $\sigma \in \tau$, where $R' \geqslant R$ depends only on R and contractibility functions $\psi_i(k, \cdot)$ of X, $i = 0, \ldots, k$.*

PROOF. We construct γ' by induction on skeleta of $(\mathbb{S}^k, \tau)$. The map is already defined on the 0-skeleton, namely, it is the map γ and images of all vertices of τ are within distance $\leqslant R$ from each other. This map extends to the 1-skeleton of τ: Given an edge $\sigma = [v, w]$ of τ, we extend γ to σ using a path of length $\leqslant \psi(1, R)$ in $X^{(1)}$ connecting the vertices $\gamma(v), \gamma(w) \in X^{(0)}$.

Suppose we constructed the required extension

$$\gamma' : \tau^{(i)} \to X$$

such that

$$\mathrm{diam}(\gamma'(\sigma)) \leqslant R_i = R_i(R, \psi(k, \mathrm{diam}(\sigma)))$$

for each i-simplex σ.

Let σ be an $i+1$-simplex in τ. We already have a map γ' defined on the boundary of σ and $\mathrm{diam}(\gamma'(\partial\sigma)) \leqslant R_i$. Then, as in the proof of Proposition 6.47, using uniform contractibility of X, we extend γ' to the simplex σ, so that the new cellular map γ' map satisfies

$$\mathrm{diam}(\gamma'(\sigma)) \leqslant \psi(i + 1, R_i).$$

This implies that the image $\gamma'(\sigma)$ is contained in $\mathbf{B}(\gamma(v), 2\psi(i+1, R_i))$, where v is a vertex of σ. Thus,

$$\operatorname{diam}(\gamma'(\tau^{i+1})) \leqslant R_{i+1} := R + \psi(i+1, R_i).$$

Now, the lemma follows by induction. $\square$

Since X is k-connected, the map γ' extends to a cellular map $\gamma' : \mathbb{D}^{k+1} \to X^{(k+1)}$, where $\mathbb{D}^{k+1}$ is a triangulated disk whose triangulation $\mathcal{T}$ extends the triangulation τ of $\mathbb{S}^k$. Our next goal is to "push" γ' into a map $\gamma'' : \mathbb{D}^{k+1} \to \operatorname{Rips}_{R'}(Z)$, relative to the restriction of γ to the vertex set of τ.

Let σ be a simplex in $\mathcal{T}$. A simplicial map is determined by images of vertices. By definition of the number R', images of the vertices of σ under γ' are within distance $\leqslant R'$ from each other. Therefore, we have a canonical extension γ'' of $\gamma'\big|_{\sigma^{(0)}}$ to a map $\sigma \to \operatorname{Rips}_{R'}(Z)$. If σ_1 is a face of σ_2, then $\gamma'' : \sigma_1 \to \operatorname{Rips}_{R'}(Z)$ agrees with the restriction of $\gamma'' : \sigma_2 \to \operatorname{Rips}_{R'}(Z)$, since maps are determined by their vertex values. We thus obtain a simplicial map

$$\gamma'' : \mathbb{D}^{k+1} \to \operatorname{Rips}_{R'}(Z)$$

which agrees with γ on the boundary sphere. We conclude that the inclusion map $\operatorname{Rips}_R(Z) \to \operatorname{Rips}_{R'}(Z)$ is n-null, i.e., Z is coarsely n-connected. $\square$

So far we have seen, how to go from uniform k-connectivity of a metric cell complex X to coarse k-connectivity of its 0-skeleton. Our goal now is to go in the opposite direction: Convert a coarsely k-connected space to a uniformly k-connected metric cell complex.

LEMMA 6.53. *Let G be a finitely generated group with word metric. Then G is coarsely simply-connected if and only if* $\operatorname{Rips}_R(G)$ *is simply-connected for all sufficiently large R.*

PROOF. One direction is clear, we only need to show that coarse simple connectivity of G implies that $\operatorname{Rips}_R(G)$ is simply-connected for all sufficiently large R. Our argument is similar to the proof of Theorem 6.51. Note that the 1-skeleton of $\operatorname{Rips}_1(G)$ is just a Cayley graph of G. Using coarse simple connectivity of G, we find $D \geqslant 1$ such that the map

$$\operatorname{Rips}_1(G)) \to \operatorname{Rips}_D(G)$$

is 1-null (i.e., induces trivial map of fundamental groups). We claim that for all $R \geqslant D$ the Rips complex $\operatorname{Rips}_R(G)$ is simply-connected. Let $\gamma \subset \operatorname{Rips}_R(G)$ be a simplicial loop. For every edge $\gamma_i := [x_i, x_{i+1}]$ of γ we let $\gamma_i' \subset \operatorname{Rips}_1(X)$ denote a geodesic path from x_i to x_{i+1}. The path γ_i' necessarily has length $\leqslant R$. Therefore, all the vertices of γ_i' are contained in the ball $B(x_i, R) \subset G$ and, hence, span a simplex in $\operatorname{Rips}_R(G)$. Thus, the paths γ_i, γ_i' are homotopic in $\operatorname{Rips}_R(G)$, relative their end-points. Let γ' denote the loop in $\operatorname{Rips}_1(G)$ which is the concatenation of the paths γ_i'. Then, by the above observation, γ' is freely homotopic to γ in $\operatorname{Rips}_R(G)$. On the other hand, γ' is null-homotopic in $\operatorname{Rips}_R(G)$ since the map

$$\pi_1(\operatorname{Rips}_1(G)) \to \pi_1(\operatorname{Rips}_R(G))$$

is trivial. We conclude that γ is null-homotopic in $\operatorname{Rips}_R(G)$ as well. $\square$

Corollary 6.54. *Suppose that G is a finitely generated group with the word metric. Then G is finitely presented if and only if G is coarsely simply-connected. In particular, finite-presentability is a QI invariant.*

Proof. Suppose that G is finitely presented and let Y be its finite presentation complex (see Definition 4.90). Then the universal cover X of Y is simply-connected. Hence, by Lemma 6.50, X is uniformly simply-connected and, by Theorem 6.51, the group G is coarsely simply-connected.

Conversely, suppose that G is coarsely simply-connected. By Lemma 6.53, the simplicial complex $\mathrm{Rips}_R(G)$ is simply-connected for some R. The group G acts on $X := \mathrm{Rips}_R(G)$ simplicially, properly discontinuously and cocompactly. Therefore, by Corollary 3.95, G admits a properly discontinuous, free cocompact action on another simply-connected cell complex Z. It follows that G is finitely presented. $\square$

We now proceed to $k \geqslant 2$. Recall (see Definition 3.93) that a group G has type $\mathbf{F}_n$ ($n \leqslant \infty$) if its admits a free cellular action on a cell complex X such that for each $k \leqslant n$:

1. $X^{(k+1)}/G$ is compact.
2. $X^{(k+1)}$ is k-connected.

Theorem 6.55 (See 1.C2 in [**Gro93**]). *Type $\mathbf{F}_n$ is a QI invariant for each $n \le \infty$.*

Proof. Our argument is similar to the proof of Corollary 6.54, except we cannot rely on $n-1$-connectivity of Rips complexes $\mathrm{Rips}_R(G)$ for large R. If G has type $\mathbf{F}_n$, then it admits a free cellular action $G \curvearrowright X$ on some $n-1$-connected cell complex X, so that the quotient of each skeleton (of dimension $\le n$) is a finite complex. (In the case $n = \infty$, we require, of course, the entire X to be contractible and the quotient of each skeleton to be finite.) By combining Lemma 6.50 and Theorem 6.51, we see that the group G is coarsely $n-1$-connected. It remains to prove

Proposition 6.56. *If G is a coarsely $n-1$-connected group, then G has type $\mathbf{F}_n$.*

Proof. Note that we already proved this statement for $n = 2$: Coarsely simply-connected groups are finitely presented (Corollary 6.54). The proof below follows [**KK05**]. We break the argument in three parts: We first consider the case when G is torsion-free and $n < \infty$, then the case when G is still torsion-free but $n = \infty$ and, lastly, the general case.

Our goal is to build a complex X on which G acts as required by the definition of type $\mathbf{F}_n$. We construct this complex and the action by induction on skeleta $X^{(0)} \subset ... \subset X^{(n-1)} \subset X^{(n)}$. Furthermore, we will inductively construct cellular G-equivariant maps

$$f_i : X_i = X^{(i)} \to Y_{R_i} = \mathrm{Rips}_{R_i}(G)$$

$$\bar{f}_i : Y_{R_i}^{(i)} \to X_i, i = 0, ..., n,$$

and (cellular G-equivariant) homotopies

$$H_i : X_{i-1} \times [0,1] \to X_i$$

of

$$h_{i-1} := \bar{f}_{i-1} \circ f_{i-1} : X_{i-1} \to X_{i-1} \subset X_i$$

to the inclusion maps $X_{i-1} \hookrightarrow X_i$.

1. Torsion-free case, $n < \infty$. In this case the G-action on every Rips complex is free and cocompact. Our construction is by induction on i.

$i = 0$. We let $X_0 = G, R_0 = 0$ and let $f_0 = \bar{f}_0 : G \to G$ be the identity map.

$i = 1$. We let $R_1 = 1$ and let $X_1 = Y_{R_1}^{(1)}$ be the Cayley graph of G. Again, $f_1 = \bar{f}_1 = \mathrm{id}$, and, of course, $H_0(x,t) = x$.

$i = 2$. According to Lemma 6.53, there exists R_2 so that Y_R is simply-connected for all $R \geqslant R_2$. We then take $X_2 := Y_{R_2}^{(2)}$. Again, we let $f_2 = \bar{f}_2 = \mathrm{id}$, $H_1(x,t) = x$.

$i \Rightarrow i+1$. Suppose now that $3 \leqslant i \leqslant n-1$, $X_i, f_i, \bar{f}_i, H_i$ are constructed and R_i chosen; we will construct $X_{i+1}, f_{i+1}, \bar{f}_{i+1}$ and H_{i+1}.

In the arguments below we will be using unbased spherical cycles when dealing with homotopy groups of X_i: This is harmless since X_i is $i-1$-connected and we can identify homotopy and homology groups (in degree i) via Hurewicz theorem. Our first task is to extend the homotopy H_i from $X_{i-1} \times [0,1]$ to $X_i \times [0,1]$. This is impossible without increasing the dimension of X_i and this will be the first step of our construction.

Lemma 6.57. *There exists a bounded geometry cell complex Z_{i+1} of dimension $i+1$ whose i-skeleton is X_i, such that:*

1. The G-action extends from X_i to a free cellular properly discontinuous cocompact action on Z_{i+1}.

2. The homotopy $H_i : X_{i-1} \times [0,1] \to X_i$ extends to a G-equivariant homotopy $H_{i+1} : X_i \times [0,1] \to Z_{i+1}$ between the map h_i and the inclusion map.

Proof. There are only finitely many i-cells in X_i modulo the G-action. It suffices to extend H_i to the finitely many cells $\hat{e}_\gamma : \mathbb{D}^i \to X_i$ in each G-orbit. Consider the $i+1$-ball $\mathbb{D}^i \times [0,1]$. The homotopy H_i lifts to a homotopy $\hat{H}_i : \partial\mathbb{D}^i \times [0,1] \to X_i$ between the map $h_{i-1} \circ e_\gamma$ and the attaching map e_γ; furthermore, we are also given maps e_γ and $h_i \circ e_\gamma$ on $\mathbb{D}^i \times \{1\}$ and $\mathbb{D}^i \times \{1\}$ respectively. If we knew that the resulting map of the boundary sphere of $\mathbb{D}^i \times [0,1]$

$$\epsilon_\gamma : \partial(\mathbb{D}^i \times [0,1]) \to X_i$$

is null-homotopic, we would be able to construct the required extension $\hat{H}_{i+1}$. There is no reason, of course, for this null-homotopy (since X_i is only required to be $i-1$-connected and not i-connected). Therefore, we attach an $i+1$-cell to X_i along the map ϵ_γ.

Since the homotopy H_i was G-equivariant, we can attach these cells in G-equivariant fashion. The result is the G-complex Z_{i+1}. Proper discontinuity of the action of G on X_i ensures that Z_{i+1} has bounded geometry and the action $G \curvearrowright Z_{i+1}$ is properly discontinuous and cocompact. Freeness of the action of G follows from the fact that G is torsion-free. □

The next step is to construct X_{i+1} by enlarging Z_{i+1}. Let $R' > R = R_i$ be such that the inclusion map

$$Y_R = \mathrm{Rips}_R(G) \to Y_{R'} = \mathrm{Rips}_{R'}(G)$$

is i-null. Since X_i is $i-1$-connected, the map

$$\bar{f}_i : Y_R^{(i)} \to X_i$$

extends to a cellular G-equivariant map

$$\tilde{f}_i : Y_{R'}^{(i)} \to X_i,$$

as in the proof of Proposition 6.47.

LEMMA 6.58. *There exists a finite set of spherical classes $[\sigma_\alpha], \alpha \in A'$, in $H_i(Z_{i+1})$, which generates $H_i(Z_{i+1})$ as a G-module.*

PROOF. We let $\{\Delta_\alpha : \alpha \in A\}$ denote the set of $i+1$-simplices in $Y_{R'}$. For each simplex Δ_α we let

$$\tau_\alpha : \partial\Delta_\alpha \to Y_{R'}$$

denote the inclusion map. We will identify the boundary of Δ_α with a triangulated sphere $\mathbb{S}^i$ and think of the maps τ_α as spherical cycles in $Y_{R'}$. Since the map $H_i(Y_R) \to H_i(Y_{R'})$ is trivial, each $[\eta] \in H_i(Y_R)$ has the form

$$[\eta] = \sum_{\alpha \in A} z_\alpha [\partial\Delta_\alpha], \quad z_\alpha \in \mathbb{Z}.$$

In other words, in the group $H_i(Y_{R'}^{(i)})$ we have the equality

$$[\eta] = \sum_{\alpha \in A} z_\alpha [\tau_\alpha].$$

Since the action of G on Y_R is cocompact, there exists a finite subset $A' \subset A$, such that each cycle $\tau_\beta, \beta \in A$ belongs to the G-orbit of some $\tau_\alpha, \alpha \in A'$. In other words, the image M of $H_i(Y_R^{(i)})$ in $H_i(Y_{R'}^{(i)})$ is a G-submodule of the finitely generated G-module M' with the generators

$$\{[\tau_\alpha] : \alpha \in A'\}.$$

Each $[\sigma] \in H_i(Z_{i+1})$ is represented by a (spherical) cycle σ in X_i and

$$(\bar{f}_i \circ f_i)_*([\sigma]) = [\sigma]$$

in $H_i(Z_{i+1})$ because of the homotopy H_{i+1}. Therefore, $[\sigma]$ belongs to the finitely generated G-module $(\tilde{f}_i)_*(M')$ whose generators are represented by spherical cycles

$$\sigma_\alpha := \tilde{f}_i(\tau_\alpha), \alpha \in A'.$$

We conclude that the G-module $H_i(Z_{i+1})$ is generated by the finite set $[\sigma_\alpha], \alpha \in A'$. □

We now use the maps $e_\alpha = g \circ \sigma_\alpha, \alpha \in A'$, as attaching maps for $i+1$-cells, and let X_{i+1} denote the cell complex obtained by (equivariantly) attaching cells to Z_{i+1} along these maps. Recall, for a future reference, that $\hat{e}_\alpha : \mathbb{D}^{i+1} \to X_{i+1}$ denotes the $i+1$-cell defined via the attaching map e_α. The G-action extends from Z_{i+1} to a free cocompact properly discontinuous action on X_{i+1}. By the construction X_{i+1} is i-connected, since we killed $\pi_i(Z_{i+1})$ by attaching $i+1$-cells along its generators.

We next construct the maps f_{i+1} and $\bar{f}_{i+1}$. To construct the map $f_{i+1} : X_{i+1} \to Y_{R'}$, for each $\alpha \in A', g \in G$, we extend the map $f_i \circ g\sigma_\alpha : \mathbb{S}^i \to X_i$ to $\mathbb{D}^{i+1}$ G-equivariantly using vanishing of the map

$$\pi_i(Y_R) \to \pi_i(Y_{R'}).$$

The construction of $\bar{f}_{i+1}$ is similar: We already have an equivariant map $\tilde{f}_i : Y^{(i)}_{R'} \to X_i$. We extend this map to each $i+1$-simplex $g\Delta_\alpha \cong \mathbb{D}^{i+1}, \alpha \in A'$, using the map $g \circ \hat{e}_\alpha : \mathbb{D}^{i+1} \to X^{(i+1)}$. This concludes the proof in the case when G is torsion-free and n is finite. Note that at the last step of the construction, we only get a homotopy H_n between h_{n-1} and id: As we noted above, there is no reason for the map h_n to be homotopic to the identity.

2. $n = \infty$**.** The inductive construction described in the proof, runs indefinitely. We obtain an increasing sequence of $i-1$-connected i-dimensional G-complexes X_i. Let X be the union

$$\bigcup_{i\geq 0} X_i$$

equipped with the weak topology. Since each X_i is is $i-1$-connected, the complex X is contractible. The group G acts cellularly and freely on X, since it acts this way on each i-skeleton. The quotients X_i/G are finite for every $i \in \mathbb{N}$ and the action of G on each X_i is free and properly discontinuous. This concludes the proof in the case of torsion-free groups G.

3. General Case. We now explain what to do in the case when G is not torsion-free. The main problem is that a group G with torsion will not act freely on its Rips complexes. Thus, while equivariant maps f_i still exist, we would be unable to construct equivariant maps $\bar{f}_i : \mathrm{Rips}_R(G) \to X_i$. Furthermore, it could happen that, for large R, the complex Y_R is contractible: This is clearly true if G is finite, it also holds for all Gromov-hyperbolic groups. If we were to have f_i and $\bar{f}_i$ as before, we would be able to conclude that X_i is contractible for large i, while a group with torsion cannot act freely on a contractible cell complex.

We, therefore, have to modify the construction. For each R we let W_R denote the barycentric subdivision of $Y^{(i)}_R = \mathrm{Rips}_R(G)^{(i)}$. Then G acts on W_R without inversions (see Definition 3.89). Let $\widehat{W_R}$ denote the regular cell complex obtained by applying the *Borel construction* to W_R, see §3.8. The complex $\widehat{W_R}$ is infinite-dimensional if G has torsion, but this does not cause trouble since at each step of induction we work only with finite skeleta. The action $G \curvearrowright W_R$ lifts to a free (properly discontinuous) action $G \curvearrowright \widehat{W_R}$ which is cocompact on each skeleton. We then can apply the arguments from the torsion-free case to the complexes $\widehat{W_R}$ instead of $\mathrm{Rips}_R(G)$. The key is that, since the action of G on $\widehat{W_R}$ is free, the construction of the equivariant maps $\bar{f}_i : Y^{(i)}_{R_i} \to X^{(i)}$ goes through. Note also that in the first steps of the induction we used the fact that Y_R is simply-connected for sufficiently large R in order to construct $X^{(2)}$. Since the projection $\widehat{W_R} \to W_R$ is a homotopy-equivalence, the 2-skeleton of $\widehat{W_R}$ is simply-connected for the same values of R. □

This finishes the proof of Theorem 6.55 as well. □

COROLLARY 6.59. *The condition of having the type* $\mathbf{F}_n$*,* $1 \leqslant n \leqslant \infty$*, is a VI invariant.*

The condition $\mathbf{F}_n$ has *cohomological analogues*, for instance, the condition $\mathbf{FP}_n$, see [**Bro82b**]. The arguments used in this section apply in the context of $\mathbf{FP}_n$-groups as well, see Proposition 11.4 in [**KK05**]. The main difference is that instead of metric cell complexes, one works with metric chain complexes and instead of

k-connectedness of the system of Rips complexes, one uses acyclicity over commutative rings R:

THEOREM 6.60 (Kapovich, Kleiner, [**KK05**]). *Let R be a commutative ring with a unit. Then the property of being $\mathbf{FP}_n$ over R is QI invariant.*

QUESTION 6.61. 1. Is the homological dimension (over $\mathbb{Q}$) of a group a quasi-isometric invariant?

2. Suppose that G has geometric dimension $n < \infty$. Is there a bounded geometry uniformly contractible n-dimensional metric cell complex with free G-action $G \curvearrowright X$?

3. Is the geometric dimension a quasiisometric invariant for torsion-free groups?

4. Is the property of having the type $\mathbf{F}$ invariant under quasiisometries?

Note that cohomological dimension is (mostly) known to equal geometric dimension, except there could be groups satisfying

$$2 = cd(G) \leqslant gd(G) \leqslant 3,$$

see [**Bro82b**]. On the other hand,

$$cd(G) \leqslant hd(G) \leqslant cd(G) + 1,$$

see [**Bie76a**]. Here cd stands for cohomological dimension, gd is the geometric dimension and hd is the homological dimension. QI invariance of cohomological dimension (over $\mathbb{Q}$) was proven by R. Sauer:

THEOREM 6.62 (R. Sauer [**Sau06**]). *The cohomological dimension $cd_{\mathbb{Q}}$ of a group (over $\mathbb{Q}$) is a QI invariant. Moreover, if G_1, G_2 are groups and $f : G_1 \to G_2$ is a quasiisometric embedding, then $cd_{\mathbb{Q}}(G_1) \leqslant cd_{\mathbb{Q}}(G_2)$.*

Note that partial results on QI invariance of cohomological dimension were proven earlier by P. Pansu [**Pan83**] (for virtually nilpotent groups), S. Gersten, [**Ger93b**] (for groups of type FP_n) and Y. Shalom, [**Sha04**] (for amenable groups).

6.5. Retractions

The goal of this section is to give a non-equivariant version of the construction of the retractions ρ_i from the proof of Proposition 6.56 in the previous section.

Suppose that X, Y are uniformly contractible finite-dimensional metric cell complexes of bounded geometry. Consider a uniformly proper map $f : X \to Y$. Our goal is to define a *coarse left-inverse* to f, a *retraction* ρ which maps an r-neighborhood of $V := f(X)$ back to X.

LEMMA 6.63. *Under the above assumptions, there exist numbers L, L', A, a function $R = R(r)$ which depend only on the distortion function of f and on the geometry of X and Y, such that:*

1. For every $r \in \mathbb{N}$ there exists a cellular L–Lipschitz map $\rho = \rho_r : \mathcal{N}_r(V) \to X$ so that $\operatorname{dist}(\rho \circ f, \operatorname{id}_X) \leqslant A$. *Here and below we equip $W^{(0)}$ with the restriction of the path-metric on the metric graph $W^{(1)}$ in order to satisfy Axiom 1 of metric cell complexes.*

2. $\rho \circ f$ is homotopic to the identity by an L'–Lipschitz cellular homotopy.

3. The composition $h = f \circ \rho : \mathcal{N}_r(V) \to V \subset \mathcal{N}_R(V)$ is homotopic to the identity embedding $\operatorname{id} : V \to \mathcal{N}_R(V)$.

4. If $r_1 \leqslant r_2$, then $\rho_{r_2}|_{\mathcal{N}_{r_1}(V)} = \rho_{r_1}$.

Proof. Let $D_0 = 0, D_1, D_2, ...$ denote the geometric bounds on Y and

$$\max_{k>0} D_k = D < \infty.$$

Since f is uniformly proper, there exists a proper monotonic function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\eta(d(x,x')) \leqslant d(f(x), f(x')), \forall x, x' \in X^{(0)}.$$

Let A_0, A_1 denote real numbers for which

$$\eta(t) > 0, \quad \forall t > A_0,$$

$$\eta(t) > 2r + D_1, \quad \forall t > A_1.$$

Recall that the neighborhood $W := \bar{\mathcal{N}}_r(V)$ is a subcomplex of Y. For each vertex $y \in W^{(0)}$ we pick a vertex $\rho(y) := x \in X^{(0)}$ such that the distance $\mathrm{dist}(y, f(x))$ is the smallest possible. If there are several such points x, we pick one of them arbitrarily. The fact that f is uniformly proper, ensures that

$$\mathrm{dist}(\rho \circ f, \mathrm{id}_{X^{(0)}}) \leqslant A := A_0.$$

Indeed, if $\rho(f(x)) = x'$, then $f(x) = f(x')$; if $d(x,x') > A_0$, then

$$0 < \eta(d(x,x')) \leqslant d(f(x), f(x')),$$

contradicting that $f(x) = f(x')$. Thus, by our choice of the metric on $W^{(0)}$ coming from $W^{(1)}$, we conclude that ρ is A_1-Lipschitz.

Next, observe also that for each 1-cell σ in W, $\mathrm{diam}(\rho(\partial\sigma)) \leqslant A_1$. Indeed, if $\partial\sigma = \{y_1, y_2\}$, then $d(y_1, y_2) \leqslant D_1$, by the definition of a metric cell complex. For $y_i' := f(x_i)$, $d(y_i, y_i') \leqslant r$. Thus, $d(y_1', y_2') \leqslant 2r + D_1$ and $d(x_1, x_2) \leqslant A_1$, by the definition of A_1. Now, existence of L-Lipschitz extension $\rho : W \to X$ follows from Proposition 6.47. This proves (1).

Part (2) follows from Corollary 6.49. To prove Part (3), observe that $h = f \circ \rho : \bar{\mathcal{N}}_r(V) \to V$ is L''-Lipschitz (see Exercise 6.41), $\mathrm{dist}(h, \mathrm{id}) \leqslant r$. Now, (3) follows from Corollary 6.49 since Y is also uniformly contractible.

Lastly, in order to guarantee (4), we can construct the retractions ρ_r by induction on the values of r and using the Extension Lemma 6.48. ☐

Corollary 6.64. *There exists a function $\alpha(r) \geqslant r$, such that for every r the map $h = f \circ \rho : \mathcal{N}_r(V) \to \mathcal{N}_{\alpha(r)}(V)$ is properly homotopic to the identity, where $V = f(X)$.*

We will think of this lemma and its corollary as a proper homotopy-equivalence between X and the direct system of metric cell complexes $\mathcal{N}_R(V)$, $R \geqslant 1$. Recall that the usual proper homotopy-equivalence induces isomorphisms of compactly supported cohomology groups. In our case we get an "approximate isomorphism" of $H_c^*(X)$ to the inverse system of compactly supported cohomology groups $H_c^*(\mathcal{N}_R(V))$:

Corollary 6.65. *1. The induced maps $\rho_R^* : H_c^*(X) \to H_c^*(\mathcal{N}_R(W))$ are injective.*

2. The induced maps ρ_R^* are approximately surjective *in the sense that the subgroup coker$(\rho_{\alpha(R)}^*)$ maps to zero under the map induced by restriction map*

$$rest_R : H_c^*(\mathcal{N}_{\alpha(R)}(V)) \to H_c^*(\mathcal{N}_R(V)).$$

Proof. 1. Follows from the fact that $\rho \circ f$ is properly homotopic to the identity and, hence, induces the identity map of $H_c^*(X)$, which means that f^* is the right-inverse to ρ_R^*.

2. By Corollary 6.64 the restriction map $rest_R$ equals the map $\rho_R^* \circ f^*$. Therefore, the cohomology group $H_c^*(\mathcal{N}_{\alpha(R)}(W))$ maps *via* $rest_R$ to the image of ρ_R^*. The second claim follows. □

6.6. Poincaré duality and coarse separation

In this section we discuss coarse implications of Poincaré duality in the context of triangulated manifolds. For a more general version of Poincaré duality, we refer the reader to [**Roe03**]; this concept was coarsified in [**KK05**], where coarse Poincaré duality was introduced and used in the context of metric cell complexes. We will be working work with metric cell complexes which are simplicial complexes, the main reason being that Poincaré duality has cleaner statement in this case.

Let X be a connected simplicial complex of bounded geometry, which is a triangulation of a (possibly noncompact) n-dimensional manifold without boundary. Suppose that $W \subset X$ is a subcomplex, which is a triangulated manifold (possibly with boundary). We will use the notation W' to denote its 1st barycentric subdivision. We then have the Poincaré duality isomorphisms

$$P_k : H_c^k(W) \to H_{n-k}(W, \partial W) = H_{n-k}(X, X \setminus W).$$

Here, H_c^* are the cohomology groups with compact support. The Poincaré duality isomorphisms are *natural* in the sense that they commute with proper embeddings of manifolds and manifold pairs. Furthermore, the isomorphisms P_k move cocycles by *uniformly bounded amount*: Suppose that $\zeta \in Z_c^k(W)$ is a simplicial cocycle supported on a compact subcomplex $K \subset W$. Then the corresponding relative cycle $P_k(\zeta) \in Z_{n-k}(W, \partial W)$ is represented by a simplicial chain in W', where each simplex has nonempty intersection with K.

Exercise 6.66. If $W \subsetneq X$ is a proper subcomplex, then $H_c^n(W) = 0$.

We will also have to use the Poincaré duality in the context of subcomplexes $V \subset X$ which are not submanifolds with boundary. Such V, nevertheless, admits a (closed) regular neighborhood $W = \mathcal{N}(V)$, which is a submanifold with boundary. The neighborhood W is homotopy-equivalent to V.

In this section we will present two applications of Poincaré duality to the coarse topology of X.

Coarse surjectivity

Theorem 6.67. *Let X, Y be uniformly contractible simplicial complexes of bounded geometry homeomorphic to $\mathbb{R}^n$. Then every uniformly cellular proper map $f : X \to Y$ is surjective.*

Proof. Assume to the contrary, i.e., $V = f(X) \neq Y$ is a proper subcomplex. Thus, $H_c^n(V) = 0$ by Exercise 6.66. Let $\rho : V \to X$ be a retraction constructed in Lemma 6.63. By Lemma 6.63, the composition $h = \rho \circ f : X \to X$ is properly homotopic to the identity. Thus, this map has to induce an isomorphism $H_c^*(X) \to H_c^*(X)$. However, $H_c^n(X) \cong \mathbb{Z}$ since X is homeomorphic to $\mathbb{R}^n$, while $H_c^n(V) = 0$. Contradiction. □

Corollary 6.68. *Let X, Y be as above and let $f : X^{(0)} \to Y^{(0)}$ be a quasiisometric embedding. Then f is a quasiisometry.*

Proof. Combine Proposition 6.47 with Theorem 6.67. □

Coarse separation.

Suppose that X is a simplicial complex and $W \subset X$ is a subcomplex. Consider $\mathcal{N}_R(W)$, the open metric R-neighborhoods of W in X, and their complements C_R in X.

For a component $C \subset C_R$ define the *inradius*, $\mathrm{Inrad}(C)$, of C to be the supremum of radii of balls $\mathbf{B}(x, R)$ in X contained in C. A component C is called *shallow* if $\mathrm{Inrad}(C)$ is finite and *deep* if $\mathrm{Inrad}(C) = \infty$.

Example 6.69. Suppose that W is compact. Then deep complementary components of C_R are components of infinite diameter. These are the components which appear as neighborhoods of ends of X.

A subcomplex W is said to *coarsely separate* X if there is R such that $\mathcal{N}_R(W)$ has at least two distinct deep complementary components.

Example 6.70. A simple properly embedded curve Γ in $\mathbb{R}^2$ need not coarsely separate $\mathbb{R}^2$ (see Figure 6.3). A straight line in $\mathbb{R}^2$ coarsely separates $\mathbb{R}^2$.

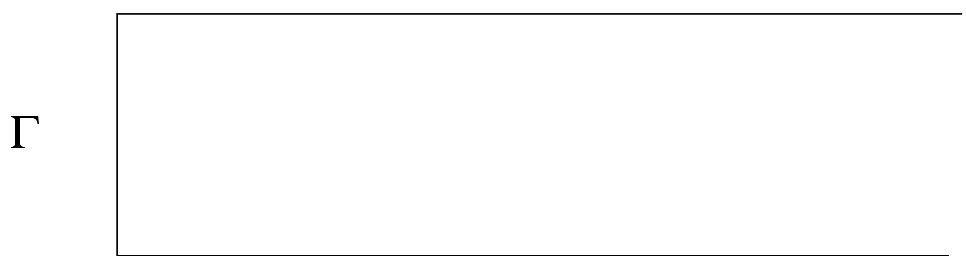


Figure 6.3. A separating curve which does not coarsely separate the plane.

The following theorem is a coarse analogue of the Jordan separation theorem which states that for the image of an arbitrary proper embedding $f : \mathbb{R}^{n-1} \to \mathbb{R}^n$ separates $\mathbb{R}^n$ into exactly two components. This topological theorem follows immediately from the Jordan separation theorem for spheres, since we can take the one-point compactifications of $A = f(\mathbb{R}^{n-1})$ and $\mathbb{R}^n$. Properness of f ensures that the compactification of A is homeomorphic to $\mathbb{S}^{n-1}$. The proof of the coarse Jordan separation theorem follows the same arguments and the proof of the topological separation theorem (via Poincaré duality).

Theorem 6.71 (Coarse Jordan separation). *Suppose that X and Y are uniformly contractible simplicial complexes of bounded geometry, homeomorphic to $\mathbb{R}^{n-1}$ and $\mathbb{R}^n$, respectively. Then for each uniformly proper simplicial map $f : X \to Y$, the image $V = f(X)$ coarsely separates Y. Moreover, for all sufficiently large R, $Y \setminus \mathcal{N}_R(V)$ has exactly two deep components.*

Proof. Actually, our proof will use the assumption on the topology of X only weakly: To get coarse separation it suffices to assume that $H^{n-1}_c(X) \neq 0$.

Recall that in §6.5 we constructed a system of retractions

$$\rho_R : \mathcal{N}_R(V) \to X, \quad R \in \mathbb{N}$$

and proper homotopy-equivalences $f \circ \rho \equiv \mathrm{id}$, for which

$$\rho_R \circ f\big|_{\mathcal{N}_R(V)} \equiv \mathrm{id} : \mathcal{N}_R(V) \to \mathcal{N}_{\alpha(R)}(V).$$

Furthermore, we have the *restriction maps*

$$rest_{R_1,R_2} : H_c^*(\bar{\mathcal{N}}_{R_2}(V)) \to H_c^*(\bar{\mathcal{N}}_{R_1}(V)), \quad R_1 \leqslant R_2.$$

These maps satisfy

$$rest_{R_1,R_2} \circ \rho_{R_2}^* = \rho_{R_1}^*$$

by Part 4 of Lemma 6.63. We also have the projection maps

$$proj_{R_1,R_2} : H_*(Y, Y - \bar{\mathcal{N}}_{R_2}(V)) \to H_*(Y, Y - \bar{\mathcal{N}}_{R_1}(V)), \quad R_1 \leqslant R_2,$$

induced by inclusion maps of pairs $(Y, Y - \bar{\mathcal{N}}_{R_2}(V)) \hookrightarrow (Y, Y - \bar{\mathcal{N}}_{R_1}(V))$. Poincaré duality in $\mathbb{R}^n$ also gives us a system of isomorphisms

$$P : H_c^{n-1}(\bar{\mathcal{N}}_R(V)) \cong H_1(X, X \setminus \mathcal{N}_R(V)).$$

By naturality of Poincaré duality, we have a commutative diagram:

$$\begin{array}{ccc} H_c^*(\bar{\mathcal{N}}_{R_2}(V)) & \xrightarrow{P} & H_{n-*}(Y, C_{R_2}) \\ \downarrow{\scriptstyle rest_{R_1,R_2}} & & \downarrow{\scriptstyle proj_{R_1,R_2}} \\ H_c^*(\bar{\mathcal{N}}_{R_1}(V)) & \xrightarrow{P} & H_{n-*}(Y, C_{R_1}) \end{array}$$

where P's are the Poincaré duality isomorphisms.

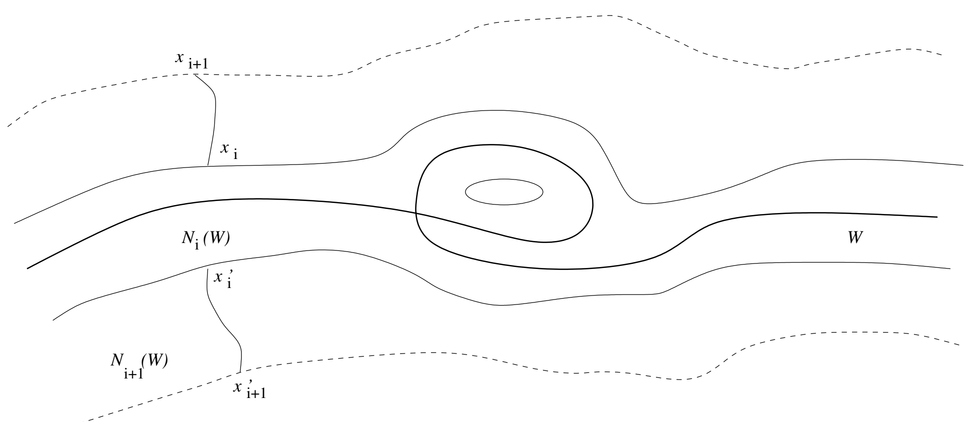


FIGURE 6.4. Coarse separation.

Let ω be a generator of $H_c^{n-1}(X) \cong \mathbb{R}$. Given $R > 0$ consider the pull-back $\omega_R := \rho_R^*(\omega)$ and the relative cycle $\sigma_R = P(\omega_R)$. Then $\omega_r = rest_{r,R}(\omega_R)$ and

$$\sigma_r = proj_{r,R}(\sigma_R) \in H_1(Y, C_r),$$

for all $r < R$, see Figure 6.4. Observe that for every r, ω_r is non-zero, since $f^* \circ \rho^* = \text{id}$ on the compactly supported cohomology of X. Hence, every σ_r is nonzero as well.

Contractibility of Y and the long exact sequence of the homology groups of the pair (Y, C_r) implies that

$$H_1(Y, C_r) \cong \tilde{H}_0(C_r).$$

We let τ_r denote the image of σ_r under this isomorphism. The class τ_r is represented by a 0-cycle, the boundary of the chain representing σ_r. Running the Poincaré duality in the reverse and using the fact that ω is a generator of $H_c^{n-1}(X)$, we see that τ_r is represented by the difference $y'_r - y''_r$, where $y'_r, y''_r \in C_r$. Nontriviality of τ_r means that y'_r, y''_r belong to distinct components C'_r, C''_r of C_r. Furthermore, since for $r < R$,

$$proj_{r,R}(\sigma_R) = \sigma_r,$$

it follows that

$$C'_R \subset C'_r, \quad C''_R \subset C''_r.$$

Because this can be done for arbitrarily large r, R, we conclude that both components C'_r, C''_r are deep. The same argument run in the reverse implies that there are exactly two deep complementary components. □

We refer to **[FS96]** and **[KK05]** for further discussion and generalization of coarse separation and coarse Poincaré/Alexander duality.

CHAPTER 7

Ultralimits of Metric Spaces

Let $(X_i)_{i\in I}$ be an indexed family of metric spaces. The goal of this chapter is to describe the asymptotic behavior of the family (X_i) by studying limits of indexed families of finite subsets $Y_i \subset X_i$. Ultrafilters are an efficient technical device for simultaneously taking limits of all such families of subspaces and putting them together to form one object, namely an *ultralimit* of (X_i). The price to pay for this efficiency is that our discussion will have to rely upon a version of the Axiom of Choice.

7.1. The axiom of choice and its weaker versions

We first recall that the *Zermelo–Fraenkel axioms (ZF)* form a list of axioms which are the basis of axiomatic set theory in its standard form, see for instance [**Kun80**], [**HJ99**], [**Jec03**].

The Axiom of Choice (AC) can be seen as a rule of building sets from other sets. It was first formulated by Ernesto Zermelo in [**Zer04**]. According to work of Kurt Gödel and Paul Cohen, the Axiom of Choice is logically independent of the Zermelo–Fraenkel axioms (i.e., neither it nor its negation can be proven in ZF).

Given a non-empty collection $\mathcal{S}$ of non-empty sets, a *choice function* defined on $\mathcal{S}$ is a function $f : \mathcal{S} \to \cup_{A\in\mathcal{S}} A$, such that for every set A in $\mathcal{S}$, $f(A)$ is an element of A. In other words, a choice function on $\mathcal{S}$ is an element of the Cartesian product $\prod_{A\in\mathcal{S}} A$.

Axiom of choice: *On any non-empty collection of non-empty sets there exists a choice function. Equivalently, the Cartesian product of a nonempty family of non-empty sets is non-empty:*

$$\mathcal{S} \neq \emptyset \quad \& \quad \forall A \in \mathcal{S}, A \neq \emptyset \Rightarrow \prod_{A\in\mathcal{S}} A \neq \emptyset.$$

We will use the abbreviation ZFC for the Zermelo–Fraenkel axioms plus the Axiom of Choice.

REMARK 7.1. If $\mathcal{S} = \{A\}$ then the existence of f follows from the fact that A is non-empty. If $\mathcal{S}$ is finite or countable, the existence of a choice function can be proved by induction. Thus, if the collection $\mathcal{S}$ is finite or countable then the existence of a choice function follows from ZF.

REMARK 7.2. Assuming ZF, the Axiom of Choice is equivalent to each of the following statements (see [**HJ99**] and [**RR85**] for a much longer list):

(1) *Zorn's lemma:* Suppose that S is a partially ordered set where every totally ordered subset has an upper bound. Then S has a maximal element.

(2) Every vector space has a basis.
(3) Every ideal in a unitary ring is contained in a maximal ideal.
(4) If A is a subset in a topological space X and B is a subset in a topological space Y, the closure of $A \times B$ in $X \times Y$ is equal to the product of the closure of A in X with the closure of B in Y.
(5) *Tychonoff's theorem:* If $(X_i)_{i\in I}$ is a collection of non-empty compact topological spaces, then $\prod_{i\in I} X_i$ is compact.

REMARK 7.3. The following statements require the Axiom of Choice (i.e., are unprovable in ZF, but hold in ZFC), see [**HJ99, RR85**]:

(1) Every union of countably many countable sets is countable.
(2) The Nielsen–Schreier theorem: Every subgroup of a free group is free (Theorem 4.42), to ensure the existence of a maximal subtree. (Note that the Axiom of Choice is needed only for free groups of uncountable rank.)

Note that for finitely generated free groups (and we are mostly interested in these) the Nielsen–Schreier theorem does not require the Axiom of Choice.

In ZF, we have the following irreversible sequence of implications:

Axiom of choice $\Rightarrow$ Ultrafilter lemma $\Rightarrow$ Hahn–Banach extension theorem.

The first implication is easy (see Lemma 7.18), it was proved to be irreversible in [**Hal64**]. Proof of the second implication can be found in [**ŁRN51**], [**Lux62**], [**Lux67**], [**Lux69**], while proofs of its irreversibility is can be found in [**Pin72**] and [**Pin74**].

Thus, the Hahn–Banach extension theorem (see below) can be seen as the analyst's Axiom of Choice, in a weaker form.

THEOREM 7.4 (Hahn–Banach Theorem, see e.g. [**Roy68**]). *Let V be a real vector space, U a subspace of V, and $\varphi : U \to \mathbb{R}$ a linear function. Let $p : V \to \mathbb{R}$ be a map with the following properties:*

$$p(\lambda x) = \lambda p(x) \text{ and } p(x+y) \leqslant p(x) + p(y)\,,\ \forall x, y \in V,\ \lambda \in [0, +\infty)\,,$$

such that $\varphi(x) \leqslant p(x)$ for every $x \in U$. Then there exists a linear extension of φ, $\overline{\varphi} : V \to \mathbb{R}$ such that $\overline{\varphi}(x) \leqslant p(x)$ for every $x \in V$.

In order to state the Ultrafilter Lemma (which we will use to prove existence of ultrafilters and, hence, ultralimits and asymptotic cones), we first define *filters*. We refer the reader to [**Bou65**, §I.6.4] for the basic properties of filters and ultrafilters, and to [**Kei10**] for an in depth survey, including ultraproducts.

DEFINITION 7.5. A *filter* $\mathcal{F}$ on a set I is a collection of subsets of I satisfying the following conditions:

(F_1) $\emptyset \notin \mathcal{F}$.
(F_2) If $A, B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$.
(F_3) If $A \in \mathcal{F}$, $A \subseteq B \subseteq I$, then $B \in \mathcal{F}$.

EXERCISE 7.6. Given an infinite set I, prove that the collection of all complements of finite sets is a filter on I. This filter is called *the Fréchet filter* (or the *cofinite filter*); it is used to define the *cofinite topology* on a topological space.

DEFINITION 7.7. Subsets $A \subset I$ which belong to a filter $\mathcal{F}$ are called *$\mathcal{F}$-large.* We say that a property (P) holds for *$\mathcal{F}$-all i* if (P) is satisfied for all i in some $\mathcal{F}$-large set.

Definition 7.8. A *base of a filter* on a set I is a subset $\mathcal{B}$ of the power set 2^I of I, which satisfies the properties:

(B_1) If $B_i \in \mathcal{B}$, $i = 1, 2$, then $B_1 \cap B_2$ contains an element of $\mathcal{B}$;
(B_2) $\emptyset \notin \mathcal{B}$ and $\mathcal{B}$ is not empty.

As an example, consider a point x in a topological space X. We let $\mathcal{F}_x$ denote the system of neighborhoods of X, i.e., all subsets $N \subset X$ which contain x together with some open neighborhood of x. Then $\mathcal{F}_x$ is a filter. A neighborhood basis of x is an example of a base of a filter. This topological intuition is somewhat useful when thinking about filters on $\mathbb{N}$ (more precisely, *non-principal ultrafilters* defined below): Such a filter can be regarded as a system of *punctured neighborhoods* of ∞ in a topology on $\mathbb{N} \cup \{\infty\}$.

Exercise 7.9. If $\mathcal{B}$ is a base of a filter, then the set $\langle \mathcal{B} \rangle$ of subsets of I containing some $B \in \mathcal{B}$ is a filter.

We will say that $\langle \mathcal{B} \rangle$ is the filter *generated* by $\mathcal{B}$. Thus, one can generate filters using bases in the same fashion one generates a topology using its neighborhood bases.

Given a set I, we let $Filter(I) \subset 2^{2^I}$ denote the set of all filters on I. In particular, $Filter(I)$ has a natural partially order given by the inclusion. If $\mathcal{F}_\alpha, \alpha \in A$, is a (nonempty) collection of filters on I, then the union

$$\mathcal{B} = \bigcup_{\alpha \in A} \mathcal{F}_\alpha$$

is not (in general) a filter on I, but it is a base of a filter. Therefore, $\langle \mathcal{B} \rangle$ is a filter on I.

Exercise 7.10. Use this construction to show that every totally ordered subset A of $Filter(I)$ has an upper bound in $Filter(I)$.

Remark 7.11. The set $Filter(I)$ has the same cardinality as 2^{2^I}, see [**Pos37**].

Definition 7.12. An *ultrafilter* on a set I is a filter $\mathcal{U}$ on I which is a maximal element in the ordered set $Filter(I)$. Equivalently, an ultrafilter can be defined (see [**Bou65**, §I.6.4]) as a collection of subsets of I satisfying the conditions (F_1), (F_2), (F_3) defining a filter and the additional condition:

$$(F_4) \qquad \text{For every } A \subseteq I, \text{ either } A \in \mathcal{U} \text{ or } A^c = I \setminus A \in \mathcal{U}.$$

One direction in this equivalence is clear: If $\mathcal{F}$ satisfies the axioms (F_1)—(F_4), then $\mathcal{F}$ has to be a maximal filter, since every strictly larger filter would have as its members a subset $A \subset I$ as well as A^c, but $A \cap A^c = \emptyset$, contradicting the axioms (F_1) and (F_2).

Exercise 7.13. Given a set I, take a point $x \in I$ and consider the collection $\mathcal{U}_x$ of subsets of I containing x. Prove that $\mathcal{U}_x$ is an ultrafilter on I.

Exercise 7.14. Given the set $\mathbb{Z}$ of integers, prove, using Zorn's lemma, that there exists an ultrafilter containing all the non-trivial subgroups of $\mathbb{Z}$. Such an ultrafilter is called *profinite ultrafilter*. Hint: In the double power set 2^{2^I} consider the partially ordered subset S consisting of all filters which contain all the non-trivial subgroups of $\mathbb{Z}$.

Definition 7.15. An ultrafilter as in Exercise 7.13 is called *a principal (or atomic) ultrafilter*. A filter that cannot be defined in such a way is called a *a non-principal (or* free*) ultrafilter*.

Proposition 7.16. *An ultrafilter on an infinite set I is non-principal if and only if it contains the Fréchet filter.*

Proof. We will prove the equivalence between the negations of the two statements. A principal ultrafilter $\mathcal{U}_x$ on I defined by a point x contains $\{x\}$; hence, by (F_4), it does not contain $I \setminus \{x\}$ which is an element of the Fréchet filter.

Let now $\mathcal{U}$ be an ultrafilter that does not contain the Fréchet filter. This and the axiom (F_4) imply that $\mathcal{U}$ contains a finite subset F of I. If

$$F \cap \bigcap_{A \in \mathcal{U}} A = \emptyset,$$

then there exist $A_1, \ldots, A_n \in \mathcal{U}$ such that

$$F \cap A_1 \cap \cdots \cap A_n = \emptyset.$$

This and the property (F_2) contradict the property (F_1).

It follows that

$$F \cap \bigcap_{A \in \mathcal{U}} A = F' \neq \emptyset,$$

in particular, given an element $x \in F'$, $\mathcal{U}$ is contained in the principal ultrafilter $\mathcal{U}_x$. The maximality of $\mathcal{U}$ implies that $\mathcal{U} = \mathcal{U}_x$. □

Exercise 7.17. (1) Let J be an infinite subset of I. Prove (using Zorn's lemma) that there exists a non-principal ultrafilter $\mathcal{U}$ such that $J \in \mathcal{U}$.

(2) Let $J_1 \supset J_2 \supset J_3 \supset \cdots \supset J_m \supset \ldots$ be an infinite sequence of infinite subsets of I. Prove that there exists a non-principal ultrafilter containing all J_m, $\forall m \in \mathbb{N}$, as its elements.

Lemma 7.18 (The Ultrafilter Lemma). *Every filter on a set I is a subset of some ultrafilter on I.*

Proof. Let $\mathcal{F}$ be the Fréchet filter of I. By Zorn's lemma (cf. Exercise 7.10), there exists a maximal filter $\mathcal{U}$ on I containing $\mathcal{F}$. By maximality, $\mathcal{U}$ is an ultrafilter; $\mathcal{U}$ is non-principal by Proposition 7.16. □

In ZF, the Axiom of Choice is equivalent to Zorn's lemma, and the latter, as we just saw, implies the Ultrafilter Lemma.

Here is an alternative way to define ultrafilters:

Definition 7.19. *An ultrafilter on a set I* is a finitely additive measure ω on the set I, such that ω takes only the values 0 and 1, and such that $\omega(I) = 1$.

We would like to stress that each subset of I is supposed to be measurable with respect to ω, in contrast to the measures (like the Lebesgue measure) that one usually encounters in analysis.

In order to verify equivalence two definitions, consider a measure ω, as Definition 7.19,. This measure defines a subset $\mathcal{U} \subset 2^I$,

$$J \in \mathcal{U} \iff \omega(J) = 1, \quad J \notin \mathcal{U} \iff \omega(J) = 0. \tag{7.1}$$

Conversely, given an ultrafilter $\mathcal{U}$, we define a measure ω on 2^I by the equations (7.1). We leave it to the reader to check that the filter axioms exactly match the finitely additive measure axioms.

Note that for an atomic ultrafilter $\mathcal{U}_x$ defined in Example 7.13, the corresponding measure is the (atomic) Dirac measure δ_x.

DEFINITION 7.20. A *non-principal ultrafilter on a set* I is a finitely additive measure $\omega : 2^I \to \{0,1\}$ such that $\omega(I) = 1$ and $\omega(F) = 0$ for every finite subset F of I.

EXERCISE 7.21. Prove the equivalence between Definitions 7.15 and 7.20.

Thus, in what follows, we will use the terminology *ultrafilter* for both maximal filters on I and finitely additive measures on I as above.

REMARK 7.22. Suppose that ω is an ultrafilter on I. Then:

(1) If $\omega(A_1 \sqcup \cdots \sqcup A_n) = 1$, then there exists $i_0 \in \{1, 2, \ldots, n\}$ such that $\omega(A_{i_0}) = 1$ and $\omega(A_j) = 0$ for every $j \neq i_0$.
(2) If $\omega(A) = 1$ and $\omega(B) = 1$ then $\omega(A \cap B) = 1$.

NOTATION 7.23. Let $(A_i)_{i \in I}$ and $(B_i)_{i \in I}$ be two families of sets or numbers indexed by I, and let $\mathcal{R}$ be a relation which holds for A_i and B_i, for every $i \in I$. We then write $A_i \, \mathcal{R}_\omega \, B_i$ if and only if $A_i \, \mathcal{R} \, B_i$ ω-almost surely, that is

$$\omega\left(\{i \in I \mid A_i \, \mathcal{R} \, B_i\}\right) = 1\,.$$

Examples of such $\mathcal{R}_\omega$'s are: $=_\omega$, $<_\omega$, $\subset_\omega$. For instance, suppose that $(x_n), (y_n)$ are sequences of real numbers and ω is an ultrafilter on $\mathbb{N}$, such that for ω-all $n \in \mathbb{N}$, $x_n < y_n$. Then we will say that

$$x_n <_\omega y_n.$$

Below we explain how existence of non-principal ultrafilters implies the Hahn–Banach in the following special case: V is the real vector space of bounded sequences of real numbers $\mathbf{x} = (x_n)$, $U \subset V$ is the subspace of convergent sequences of real numbers, p is the sup-norm

$$\|\mathbf{x}\|_\infty = \sup_{n \in \mathbb{N}} |x_n|$$

and $\varphi : U \to \mathbb{R}$ is the limit function, i.e.

$$\varphi(\mathbf{x}) = \lim_{n \to \infty} x_n.$$

Instead of the sup-norm, we can as well take

$$p : \mathbf{x} \mapsto \limsup x_n.$$

In other words, we will show how, using a non-principal ultrafilter, one can extend the notion of limit from convergent sequences to bounded sequences. The main tool in this extension is the concept of an *ultralimit*, which we will frequently use in the book.

DEFINITION 7.24. [Ultralimit of a function] Given a function $f : I \to Y$ (where Y is a topological space) define the ω–limit

$$\omega\text{-}\lim_i f(i)$$

to be a point $y \in Y$ such that for every neighborhood U of y, the pre-image $f^{-1}U$ belongs to ω. The point y is called the *ultralimit* of the function f.

Note that, in general, an ultralimit need not be unique. However, it is unique in the case when Y is Hausdorff:

LEMMA 7.25. *1. If Y is compact, then every function $f : I \to Y$ has an ultralimit.*

2. If Y is Hausdorff, then every function $f : I \to Y$ has at most one ultralimit.

PROOF. 1. To prove existence of a limit, assume that there is no point $y \in Y$ satisfying the definition of the ultralimit. Then each point $z \in Y$ possesses a neighborhood U_z such that $f^{-1}U_z \notin \omega$. By compactness, we can cover Y with finitely many of these neighborhoods U_{z_i}, $i = 1, \ldots, n$. Therefore,

$$I = \bigcup_{i=1}^{n} f^{-1}(U_{z_i})$$

and, thus,

$$\emptyset = \bigcap_{i=1}^{n} (I \setminus f^{-1}(U_{z_i})) \in \omega.$$

This contradicts the definition of a filter.

2. The proof of uniqueness of ultralimits is the same as for uniqueness of ordinary limits in Hausdorff spaces. Suppose that $f : I \to Y$ has two ultralimits $y_1 \neq y_2$. Since Y is Hausdorff, the points y_1, y_2 have disjoint neighborhoods U_1, U_2. By the assumption, both sets $f^{-1}(U_1), f^{-1}(U_2)$ are ω-large. However, their intersection is empty since $U_1 \cap U_2$ is empty. This contradicts Axiom (F_2) of filters. (Note that in this part of the proof we did not use the assumption that ω is an ultrafilter, only that it is a filter.) □

EXAMPLE 7.26. Suppose that $I = \mathbb{N}$ and $x_i = (-1)^n$. Then ω-$\lim x_i$ is either -1 or 1, depending on whether the set of odd or even numbers belongs to ω.

Note that the ω–limit satisfies the "usual "calculus properties," e.g., linearity:

$$\omega\text{-}\lim(\lambda f + \mu g) = \lambda\, \omega\text{-}\lim f + \mu\, \omega\text{-}\lim g$$

for all bounded functions $f, g : I \to \mathbb{R}$. (Boundedness is needed to ensure existence of ultralimits.) Now, we can prove Hahn–Banach theorem for the space of convergent sequences U, the space of all bounded sequences V and the functional $\varphi := \lim : U \to \mathbb{R}$. We take

$$\bar{\varphi}((x_i)) = \omega\text{-}\lim x_i.$$

Lemma 7.25 implies that every bounded function $f : I \to \mathbb{R}$ has an ultralimit. In the case when the ordinary limit $\lim_{i\to\infty} x_i$ exists, it equals the ultralimit ω-$\lim x_i$. We leave it to the reader to check the inequality

$$\omega\text{-}\lim x_i \leqslant p((x_i))$$

for both $p((x_i)) = \sup_i |x_i|$ and $p(x_i) = \limsup x_i$. This proves Hahn–Banach theorem (in the special case).

EXERCISE 7.27. Show that the ω–limit of a function $f : I \to Y$ is an accumulation point of the subset $f(I) \subset Y$.

Conversely, if y is an accumulation point of $\{f(i)\}_{i \in I}$, then there is a non-principal ultrafilter ω with $\omega\text{-}\lim f = y$, namely an ultrafilter containing the filter $\mathcal{F}$ on I, which is the preimage of the neighborhood basis of y under f.

Thus, an ultrafilter is a device which selects accumulation points for subsets A in compact Hausdorff spaces Y, in a coherent manner.

Note that when the ultrafilter is principal, that is $\omega = \delta_{i_0}$ for some $i_0 \in I$, and Y is Hausdorff, the δ_{i_0}–limit of a function $f : I \to Y$ is simply the element $f(i_0)$, which is not very interesting. Thus, when considering ω–limits we shall always choose the ultrafilter ω to be non-principal.

Remark 7.28. Recall that when we have a countable collection of sequences

$$\mathbf{x}^{(k)} = \left(x_n^{(k)}\right)_{n \in \mathbb{N}}, k \in \mathbb{N}, x_n^{(k)} \in X,$$

where X is a compact space, we can select a subset of indices $I \subset \mathbb{N}$, such that for every $k \in \mathbb{N}$ the subsequence $\left(x_i^{(k)}\right)_{i \in I}$ converges. This is achieved by the diagonal procedure. The ω–limit allows, in some sense, to do the same for an uncountable collection of (uncountable) sets. Thus, an ultralimit can be seen as an uncountable version of the diagonal procedure.

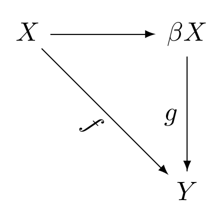
Note also that for applications in Geometric Group Theory, most of the time, one considers only countable index sets I. Thus, in principle, one can avoid using ultrafilters at the expense of getting complicated proofs involving passage to multiple subsequences.

Using ultralimits of maps we will later define ultralimits of sequences of metric spaces; in particular, given metric space (X, dist), we will define an "image of (X, dist) seen from infinitely far away" (an asymptotic cone of (X, dist)). Ultralimits and asymptotic cones will be among key technical tools used in this book.

7.2. Ultrafilters and the Stone–Čech compactification

Let X be a Hausdorff topological space. The *Stone–Čech compactification* of X is a pair consisting of a compact Hausdorff topological space βX and a continuous map $X \to \beta X$ which satisfies the following universal property:

For every continuous map $f : X \to Y$, where Y is a compact Hausdorff space, there exists a unique continuous map $g : X \to Y$, such that the following diagram commutes:



This universal property implies uniqueness of the Stone–Čech compactification in the sense that for any two such compactifications $c_1 : X \to X', c_2 : X \to X''$, there exists a homeomorphism $h : X' \to X''$ such that $c_2 = h \circ c_1$.

Exercise 7.29. Show that $X \to \beta X$ is injective and its image is dense in βX.

In view of this exercise, we will regard X as a subset of βX, so that βX is a compactification of X.

We will now explain how to construct βX using ultrafilters, provided that X has discrete topology, e.g., $X = \mathbb{N}$. We declare βX to be the set of *all ultrafilters* on X. Then, βX is a subset $Filter(X)$, which, in turn, is a subset of the power set

$$2^{2^X}.$$

We equip 2^X and, hence, 2^{2^X}, with the product topology and the subset $\beta X \subset 2^{2^X}$ with the subspace topology.

EXERCISE 7.30. Show that the subset $\beta X \subset 2^{2^X}$ is closed. Thus, by the Tychonoff's theorem, βX is also compact. Since X is Hausdorff, so is 2^X and, hence, 2^{2^X}.

Every $x \in X$ determines the principal ultrafilter δ_x; thus, we obtain an embedding $X \hookrightarrow \beta X$, $x \mapsto \delta_x$. This embedding is continuous since X has discrete topology. Therefore, from now, on we will regard X as a subset of βX.

EXERCISE 7.31. Let $\omega \in \beta X$ be a non-principal ultrafilter. Show that for every neighborhood U of ω in βX, the intersection $X \cap U$ is an ω-large set. Conversely, for every ω-large set $A \subset X$, there exists a neighborhood U of ω in βX such that $A = U \cap X$. In particular, X is dense in βX.

We will now verify the universal property of βX. Let $f : X \to Y$ be a continuous map to a compact Hausdorff space. For every $\omega \in \beta X \setminus X$ we set

$$g(\omega) := \omega\text{-}\lim f.$$

By the definition of the ultralimit of a map, for every point $y \in Y$ and its neighborhood V in Y, the preimage $A = f^{-1}(V)$ is ω-large. Therefore, by Exercise 7.31, there exists a neighborhood U of ω in βX such that $A = U \cap X$. This proves that the map g is continuous. Hence, g is the required continuous extension of f. Uniqueness of g follows from the fact that X is dense in βX.

7.3. Elements of nonstandard algebra

Our discussion of nonstandard algebra mostly follows **[Gol98]**, **[dDW84]**.

Given an ultrafilter ω on I and a collection of sets $X_i, i \in I$, define the *ultraproduct*

$$\prod_{i \in I} X_i/\omega$$

to be the collection of equivalence classes of maps

$$f : I \to \bigcup_{i \in I} X_i$$

with $f(i) \in X_i$ for every $i \in I$, with respect to the equivalence relation $f \sim g$ defined by the property that $f(i) = g(i)$ for ω–all i. Thus, an ultraproduct is a certain quotient of the ordinary product of the sets X_i.

The equivalence class of a map f in the ultraproduct is denoted by f^ω. When the map is given by an indexed family of values $(x_i)_{i \in I}$, where $x_i = f(i)$, we will also use the notation $(x_i)^\omega$ for the equivalence class.

When $X_i = X$ for all $i \in I$, the ultraproduct is called the *ultrapower of* X and denoted by X^ω. Every subset A of X can be embedded into X^ω by

$$a \mapsto \widehat{a} := (a)^\omega.$$

We let $\widehat{A}$ denote the image of A in X^ω.

Note that any algebraic structure on X (group, ring, order, order, etc.) defines the same structure on X^ω, e.g., if G is a group then G^ω is a group, etc. When $X = \mathbb{K}$ is either $\mathbb{N}, \mathbb{Z}$ or $\mathbb{R}$, the ultrapower $\mathbb{K}^\omega$ is sometimes called the *nonstandard extension of* $\mathbb{K}$, and the elements in $\mathbb{K}^\omega \setminus \mathbb{K}$ are called *nonstandard elements.* If X is totally ordered then X^ω is totally ordered as well: $f^\omega \leqslant g^\omega$ (for $f, g \in X^\omega$) if and only if $f(i) \leqslant_\omega g(i)$, with the Notation 7.23. Since ω is an ultrafilter, it follows that $\leqslant_\omega$ is a total order: This is where ultraproducts are superior to the ordinary products, since the ordinary product of totally ordered sets is (in general) only partially ordered.

In particular, we define the ordered semigroup $\mathbb{N}^\omega$ (the nonstandard natural numbers) and the ordered field $\mathbb{R}^\omega$ (the nonstandard real numbers).

Definition 7.32. An element $R \in \mathbb{R}^\omega$ is called *infinitely large* if given any $r \in \mathbb{R} \subset \mathbb{R}^\omega$, one has $R \geqslant \widehat{r}$. Note that given any $R \in \mathbb{R}^\omega$, there exists $n \in \mathbb{N}^\omega$ such that $n > R$.

Exercise 7.33. Prove that $R = (R_i)^\omega \in \mathbb{R}^\omega$ is infinitely large if and only if $\omega\text{-}\lim_i R_i = +\infty$.

Definition 7.34 (Internal subsets). A subset $W^\omega \subset X^\omega$ is called *internal* if "membership in W can be determined by coordinate-wise computation", i.e., if for each $i \in I$ there is a subset $W_i \subset X$ such that for $f \in X^I$

$$f^\omega \in W^\omega \iff f(i) \in_\omega W_i \,.$$

(Recall that the latter means that $f(i) \in W_i$ for ω–all i.) The sets W_i are called *coordinates* of W. We will write $W^\omega = (W_i)^\omega$.

Lemma 7.35.
(1) *If an internal subset A^ω is defined by a family of subsets of bounded cardinality $A_i = \{a_i^1, \ldots, a_i^k\}$, then $A^\omega = \{a_\omega^1, \ldots, a_\omega^k\}$, where $a_\omega^j = \left(a_i^j\right)^\omega$.*
(2) *In particular, if an internal subset A^ω is defined by a constant family of finite subsets $A_i = A \subseteq X$ then $A^\omega = \widehat{A}$.*
(3) *Every finite subset in X^ω is internal.*

Proof. (1) Let $x = (x_i)^\omega \in A^\omega$. The set of indices decomposes as $I = I_1 \sqcup \cdots \sqcup I_k$, where $I_j = \left\{ i \in I \,;\, x_i = a_i^j \right\}$. Then there exists $j \in \{1, \ldots, k\}$ such that $\omega(I_j) = 1$, that is $x_i =_\omega a_i^j$, and $x = a_\omega^j$.

(2) is an immediate consequence of (1).

(3) Let U be a subset in X^ω of cardinality k, and let $x_1, \ldots, x_k$ be its elements. Each element x_r is of the form $(x_i^r)^\omega$ and ω-almost surely $x_i^r \neq x_i^s$ when $r \neq s$. Therefore ω-almost surely the set $A_i = \{x_i^1, \ldots, x_i^k\}$ has cardinality k. It follows that $A^\omega = (A_i)^\omega$ has cardinality k, according to (1), and it contains U. Therefore $U = A^\omega$. □

Lemma 7.36. *If A is an infinite subset in X, then $\widehat{A}$ is not internal.*

Proof. Assume $\widehat{A} = (B_i)^\omega$ for a family $(B_i)_{i\in I}$ of subsets. For every $a \in A$, $\hat{a} \in (B_i)^\omega$, i.e.,

$$a \in B_i \quad \omega - \text{almost surely.} \tag{7.2}$$

Take an infinite sequence $a_1, a_2, \ldots, a_k, \ldots$ of distinct elements in A. Consider the nested sequence of sets

$$I_k = \{i \in I \mid \{a_1, a_2, \ldots, a_k\} \subseteq B_i\}.$$

From (7.2) and Remark 7.22, (2), it follows that $\omega(I_k) = 1$ for every k.

The intersection $J = \bigcap_{n\geqslant 1} I_k$ has ω-measure either 0 or 1. Assume first that $\omega(J) = 0$. Since

$$I_1 = \bigsqcup_{k=1}^{\infty} (I_k \setminus I_{k+1}) \sqcup J,$$

it follows that the set

$$J' = \bigsqcup_{k=1}^{\infty} (I_k \setminus I_{k+1})$$

has $\omega(J') = 1$.

Define the indexed family (x_i) such that $x_i = a_k$ for every $i \in I_k \setminus I_{k+1}$. By the definition, $x_i \in B_i$ for every $i \in J'$. Thus

$$(x_i)^\omega \in (B_i)^\omega = \widehat{A},$$

which implies that $x_i = a$ ω-a.s. for some $a \in A$.

Let $E = \{i \in I \mid x_i = a\}$, $\omega(E) = 1$. Remark 7.22, (2), implies that $E \cap J' \neq \emptyset$, hence, for some $k \in \mathbb{N}$,

$$E \cap (I_k \setminus I_{k+1}) \neq \emptyset.$$

For $i \in E \cap (I_k \setminus I_{k+1})$ we have $x_i = a = a_k$.

The fact that $\omega(I_{k+1}) = 1$ implies that $E \cap I_{k+1} \cap J' \neq \emptyset$. Hence, for some $j \geqslant k+1$,

$$E \cap (I_j \setminus I_{j+1}) \neq \emptyset.$$

For an index i in $E \cap (I_j \setminus I_{j+1})$ we have the equality $x_i = a = a_j$. But as $j > k$, $a_j \neq a_k$, and, thus, we obtain a contradiction.

Assume now that $\omega(J) = 1$. Suppose that this occurs for every sequence (a_k) of distinct elements in A. It follows that ω-almost surely $A \subseteq B_i$. □

Definition 7.37 (internal maps). A map $f^\omega : X^\omega \to Y^\omega$ is *internal* if there exists an indexed family of maps $f_i : X_i \to Y_i$, $i \in I$, such that $f^\omega(x^\omega) = (f_i(x_i))^\omega$.

Note that the range of an internal map is an internal set.

For instance, given a collection of metric spaces (X_i, dist_i), one defines a metric dist^ω on X^ω as the internal function $\text{dist}^\omega : X^\omega \times X^\omega \to \mathbb{R}^\omega$ given by the collection of functions (dist_i), that is $\text{dist}^\omega : X^\omega \times X^\omega \to \mathbb{R}^\omega$,

$$\text{dist}^\omega\left((x_i)^\omega, (y_i)^\omega\right) = \left(\text{dist}_i(x_i, y_i)\right)^\omega. \tag{7.3}$$

The main problem is that dist^ω does not take values in $\mathbb{R}$ but in $\mathbb{R}^\omega$.

Let (Π) be a property of a structure on a set X that can be expressed using elements, subsets, $\in, \subset, \subseteq, =$ and the logical quantifiers $\exists, \forall, \wedge$ (and), $\vee$ (or), $\neg$ (not) and $\Rightarrow$ (implies).

The *non-standard interpretation* $(\Pi)^\omega$ of (Π) is the statement obtained by replacing "$x \in X$" with "$x^\omega \in X^\omega$", and "A subset of X" with "A^ω internal subset of X^ω".

THEOREM 7.38 (Łoś' Theorem, see e.g. [**BS69**], [**Kei76**], Chapter 1, [**dDW84**], p.361). *A property* (Π) *is true in* X *if and only if its non-standard interpretation* $(\Pi)^\omega$ *is true in* X^ω.

We will use the following special cases of this theorem when proving Gromov's theorem on groups of polynomial growth:

COROLLARY 7.39.
(1) *Every non-empty internal subset in* $\mathbb{R}^\omega$ *that is bounded from above (below) has a supremum (infimum).*
(2) *Every non-empty internal subset in* $\mathbb{N}^\omega$ *that is bounded from above (below) has a maximal (minimal) element.*

COROLLARY 7.40 (non-standard induction). *If a non-empty internal subset* A^ω *in* $\mathbb{N}^\omega$ *satisfies the properties:*

- $\widehat{1} \in A^\omega$;
- *for every* $n^\omega \in A^\omega$, $n^\omega + 1 \in A^\omega$,

then $A^\omega = \mathbb{N}^\omega$.

EXERCISE 7.41.
(1) Give a direct proof of Corollary 7.39, (1), for $\mathbb{R}^\omega$.
(2) Deduce Corollary 7.39 from Theorem 7.38.
(3) Deduce Corollary 7.40 from Corollary 7.39.

Suppose we are given $a_n \in \mathbb{R}^\omega$, where $n \in \mathbb{N}^\omega$. Using the nonstandard induction principle on can define the nonstandard products:

$$a_1 \cdots a_n, n \in \mathbb{N}^\omega,$$

using the internal function $f : \mathbb{N}^\omega \to \mathbb{R}^\omega$, given by $f(1) = a_1$, $f(n+1) = f(n)a_{n+1}$.

Various properties of groups can be characterized in terms of ultrapowers, as explained below and in Chapter 16, §16.8.

Ultrapowers and laws in groups.

Suppose that G satisfies a law $w(x_1, \ldots, x_n) = 1$. Then the ordinary product

$$G^I = \prod_{i \in I} G$$

also satisfies this law: For every function $f \in G^I$ and all $i \in I$,

$$w(f_1, \ldots, f_n)(i) = w(f_1(i), \ldots, f_n(i)) = 1.$$

Therefore, being a quotient of G^I, the group G^ω satisfies the law $w(x_1, \ldots, x_n) = 1$ as well.

Moreover:

LEMMA 7.42 (See Lemma 6.15 in [**DS05b**]). *A group* G *satisfies a law if and only if for one (equivalently, every) non-principal ultrafilter* ω *on* $\mathbb{N}$, *the ultrapower* G^ω *does not contain free non-abelian subgroups.*

Proof. For the direct implication note that if G satisfies a law, then G^ω also satisfies the same law. Since a free nonabelian group cannot satisfy a law, the claim follows.

For the converse implication, let ω be an ultrafilter on $\mathbb{N}$, and assume that G does not satisfy any law. Enumerate all the reduced words $u_1, u_2, \ldots$ in two variables x, y and define the sequence of iterated left-commutators:

$$v_1 = u_1, v_2 = [u_1, u_2], v_3 = [v_2, u_3], v_4 = [v_3, u_4], \ldots,$$

$$v_n = [[[u_1, u_2], \ldots, u_{n-1}], u_n], \ldots$$

We will think of v_n's as words in x, y.

Since G does not satisfy any law, for every n there exists a pair (x_n, y_n) of elements in G such that $v_n(x_n, y_n) \neq 1$ in G. Consider the corresponding elements $x = (x_n)^\omega$, $y = (y_n)^\omega$ in the ultrapower G^ω. We claim that the subgroup $F < G^\omega$ generated by x and y is free. Suppose that the subgroup F satisfies a reduced relation. That relation is given by a reduced word u_i for some $i \in \mathbb{N}$. Hence, $u_i(x_n, y_n) = 1$ ω–almost surely. In particular, since ω is a non-principal ultrafilter, for some $n > i$, $u_i(x_n, y_n) = 1$. But then $v_n(x_n, y_n) = 1$ since u_i appears in the iterated commutator v_n, contradicting the choice of x_n, y_n. □

7.4. Ultralimits of sequences of metric spaces

Let $(X_i)_{i \in I}$ be a family of metric spaces parameterized by an infinite set I.

Convention 7.43. From now on, all ultrafilters are non-principal, and we will omit mentioning this property henceforth.

For an ultrafilter ω on I we define the ultralimit

$$X_\omega = \omega\text{-}\lim_i X_i$$

as follows. Let $\prod_i X_i$ be the product of the sets X_i, i.e., it is the set of indexed families of points $(x_i)_{i \in I}$ with $x_i \in X_i$. Define the distance between two points $(x_i), (y_i) \in \prod_i X_i$ by

$$\mathrm{dist}_\omega\big((x_i), (y_i)\big) := \omega\text{-}\lim\big(i \mapsto \mathrm{dist}_{X_i}(x_i, y_i)\big),$$

where we take the ultralimit of the function $i \mapsto \mathrm{dist}_{X_i}(x_i, y_i)$ with values in the compact set $[0, \infty]$. The function dist_ω is a pseudo-distance on $\prod_i X_i$ with values in $[0, \infty]$. Set

$$(X_\omega, \mathrm{dist}_\omega) := (\prod_i X_i, \mathrm{dist}_i)/\sim$$

where we identify points with zero dist_ω–distance. In the case when $X_i = Y$, for all i, the ultralimit $(X_\omega, \mathrm{dist}_\omega)$ is called a *constant ultralimit*.

The reader will notice similarities between this construction and the Cauchy–Bourbaki completion of a metric space. The difference is that we allow distinct metric spaces instead of a single space and, even if $X_i = Y$ for all i, we do not restrict to indexed families of points (x_i) which are Cauchy. The price we have to pay for this is that, at the moment, dist_ω is merely a pseudo-metric, as it takes infinite values (unless the spaces X_i have uniformly bounded diameter).

Given an indexed family of points $(x_i)_{i \in I}$ with $x_i \in X_i$ we denote the equivalence class corresponding to it either by x_ω or by $\omega\text{-}\lim x_i$.

EXERCISE 7.44. If $(X_\omega, \mathrm{dist}_\omega)$ is a constant ultralimit of a sequence of compact metric spaces $X_i = Y$, then X_ω is isometric to Y for all ultrafilters ω.

If the spaces X_i do not have uniformly bounded diameter, then the ultralimit X_ω decomposes into (in general, uncountably many) components consisting of points at mutually finite distance. In order to pick one of these components, we introduce a family of base-points e_i in X_i. The pair (X_i, e_i) is called a *based* or *pointed* metric space. The indexed family (e_i) defines a base-point $\boldsymbol{e} = e_\omega$ in X_ω and we set

$$X_{\omega,\boldsymbol{e}} := \{x_\omega \in X_\omega \mid \mathrm{dist}_\omega(x_\omega, e_\omega) < \infty\}.$$

We define the *based ultralimit* as

$$\omega\text{-}\lim_i (X_i, e_i) := (X_{\omega,\boldsymbol{e}}, e_\omega).$$

By abusing the notation, we will frequently drop $\boldsymbol{e}$ in the notation $X_{\omega,\boldsymbol{e}}$ when the choice of the base-point is clear. Given a family of subsets $A_i \subset X_i$ we let A_ω denote the subset of $X_{\omega,\boldsymbol{e}}$ represented by indexed families $(a_i)_{i\in I}$, $a_i \in A_i$.

EXERCISE 7.45. Let $X = \mathbb{R}^n$ with the Euclidean metric. Then for every sequence $e_i \in X$, $\omega\text{-}\lim(X, e_i) \cong (\mathbb{R}^n, 0)$.

The following theorem relates Gromov-Hausdorff convergence and ultralimits:

THEOREM 7.46 (M. Kapovich and B. Leeb, [**KL95**]). *Suppose that*

$$(X_i, \mathrm{dist}_{X_i}, x_i)_{i\in\mathbb{N}}$$

is a sequence of proper metric spaces Gromov-Hausdorff converging to a pointed proper metric space (X, dist_X, x). *Then for all ultrafilters* ω *there exists an isometry between* ω*-lim*$(X_i, \mathrm{dist}_{X_i}, x_i)$ *and* (X, dist_X, x) *sending* $x_\omega = \omega\text{-}\lim x_n$ *to* x.

PROOF. In view of the properness assumption (and the Arzela-Ascoli theorem), it suffices to show that for each $r > 0$, the Gromov-Hausdorff limit of the sequence pointed of closed metric balls $(\bar{B}(x_i, r), x_i)$ in X_i is isometric to $\omega\text{-}\lim(\bar{B}(x_i, r), \mathrm{dist}_i, x_i)$, where dist_i is the restriction of the distance function dist_{X_i} to $\bar{B}(x_i, r)$. Therefore, the problem reduces to the case when $X_i, i \in \mathbb{N}$, and Y are all compact. We realize Gromov-Hausdorff convergence as Hausdorff convergence in a compact metric space Y, i.e., embed each X_i and X isometrically into Y via isometric maps

$$f_i : X_i \to X'_i := f_i(X_i) \subset Y,$$

such that the Hausdorff limit of the sequence (X'_i) is $X' \cong X$:

$$\lim_{Haus} X'_i = X'.$$

Then the sequence of isometric embeddings there is an isometric embedding

$$f_\omega : X_\omega \to \omega\text{-}\lim Y = Y.$$

Since ω is non-principal, the ω-limit is independent of any finite collections of X_i's and we get:

$$f_\omega(X_\omega) \subset \bigcap_{i_0\in I} \overline{\bigcup_{i\geqslant i_0} X'_i} = X'.$$

On the other hand, $X \subset f_\omega(X_\omega)$ since $f_\omega((x_i)) = x$ whenever $\lim_{i\in I} x_i = x \in Y$. Hence, $X' = f_\omega(X_\omega)$. □

EXAMPLE 7.47. Suppose that X_i is the sequence of spheres of radius $R_i \to \infty$ in $\mathbb{E}^n$ with the induced path-metric. Then

$$X_\omega = \omega\text{-lim}(X_i, x_i) \cong \mathbb{E}^{n-1}$$

for any choice of base-points $x_i \in X_i$. Indeed, for each fixed r, define the sequence of subsets closed r-balls $Y_i(x_i, r) \subset X_i$. Then, since the sequence R_i diverges to infinity, the sequence of spaces $Y_i(x_i, r)$ Gromov-Hausdorff converges to the closed r-ball $\bar{B}(o, r) \subset \mathbb{E}^{n-1}$. Therefore, by the above lemma, for each r there is an isometry $h_r : (\bar{B}(o, r), o) \to \omega\text{-lim}(Y_i(x_i, r), x_i)$. It is clear that

$$X_\omega = \bigcup_{r>0} Y_\omega(r),$$

where each $Y_\omega(r)$ is isometric to $\omega\text{-lim}(Y_i(x_i, r), x_i)$. Composing the isometries h_r and $\omega\text{-lim}(Y_i(x_i, r), x_i) \to (Y_\omega(r), x_r)$ and taking an ultralimit as $r \to \infty$, we obtain the required isometry $\mathbb{E}^{n-1} \to X_\omega$.

LEMMA 7.48 (Functoriality of ultralimits). *1. Let $(X_i, e_i), (X_i', e_i'), i \in I$, be families of pointed metric spaces with ultralimits X_ω, X_ω', respectively. Let $f_i : (X_i, e_i) \to (X_i', e_i')$ be isometric embeddings such that*

$$\omega\text{-}lim\, \text{dist}(f(e_i), e_i') < \infty,$$

i.e.,

$$\text{dist}(f(e_i), e_i') \leqslant Const, \text{ for } \omega\text{-all } i,$$

Then the maps f_i yield an isometric embedding of the ultralimits $f_\omega : X_\omega \to X_\omega'$.

2. If each f_i is an isometry, then so is f_ω.

3. $\Phi_\omega : (f_i) \mapsto f_\omega$ preserves compositions:

$$\Phi_\omega : (g_i \circ f_i) = \Phi_\omega((g_i)) \circ \Phi_\omega((f_i)).$$

PROOF. We define f_ω as

$$f_\omega((x_i)) = (f_i(x_i')).$$

By the definition of distances in X_ω and X_ω',

$$d(f_\omega(x_\omega), f_\omega(y_\omega)) = \omega\text{-lim}\, d(f_i(x_i), f_i(y_i)) = \omega\text{-lim}\, d(x_i, y_i) = d(x_\omega, y_\omega)$$

for any pair of points $x_\omega, y_\omega \in X_\omega$. If each f_i is surjective, then, clearly, f_ω is surjective as well. The composition property is clear as well. ☐

EXERCISE 7.49. Show that injectivity of each f_i does not imply injectivity of f_ω.

The map f_ω defined in this lemma is called the *ultralimit* of the sequence of maps (f_i). An important example illustrating this lemma is the case when each X_i is an interval in $\mathbb{R}$ and, hence, each f_i is a geodesic in Y_i. Then the ultralimit $f_\omega : J_\omega \to X_\omega$ is a geodesic in X_ω (here J_ω is an interval in $\mathbb{R}$).

DEFINITION 7.50. Geodesics $f_\omega : J_\omega \to X_\omega$ are called *limit geodesics* in X_ω.

In general, X_ω contains geodesics which are not limit geodesics. In the extreme case, Y_i may contain only constant geodesics, while Y_ω is a geodesic metric space (containing more than one point). For instance, let $X = \mathbb{Q}$ with the metric induced from $\mathbb{R}$. Of course, $\mathbb{Q}$ contains no nonconstant geodesics, but

$$\omega\text{-lim}(X, 0) \cong (\mathbb{R}, 0),$$

see Exercise 7.61.

LEMMA 7.51. *Each ultralimit (X_ω, e_ω) of a sequence of pointed geodesic metric spaces (X_i, e_i) is again a geodesic metric space.*

PROOF. Let $x_\omega = (x_i), y_\omega = (y_i)$ be points in X_ω. Let $\gamma_i : [0, T_i] \to X_i$ be geodesics connecting x_i to y_i. Clearly,

$$\omega\text{-lim}\, T_i = T = d(x_\omega, y_\omega) = T < \infty.$$

We define the ultralimit γ_ω of the maps γ_i. Then $\gamma_\omega : [0, T] \to X_\omega$ is a geodesic connecting x_ω to y_ω. □

EXERCISE 7.52. Let X be a path-metric space. Then every constant ultralimit of X is a geodesic metric space.

We now return to the discussion of basic properties of ultralimits.

LEMMA 7.53. *Let (X_i, e_i) be pointed $CAT(\kappa_i)$ metric spaces, $\kappa_i \leqslant 0$, and $\kappa = \omega\text{-}lim\, \kappa_i$. Then the ultralimit (X_ω, e_ω) of the sequence (X_i, e_i) is again a pointed $CAT(\kappa)$ space.*

PROOF. It is clear that comparison inequalities for triangles in X_i yield comparison inequalities for limit triangles in X_ω. It remains to show that X_ω is a uniquely geodesic metric space, in which case every geodesic segment in X_ω is a limit geodesic. Suppose that $m_\omega \in X_\omega$ is a point such that

$$d(x_\omega, z_\omega) + d(z_\omega, y_\omega) = d(x_\omega, y_\omega),$$

equivalently, z_ω belongs to *some geodesic* connecting x_ω to y_ω.

Thus, if $z_i \in X_i$ is a sequence representing z_ω, then

$$0 \leqslant d(x_i, z_i) + d(z_i, y_i) = d(x_i, y_i) \leqslant \eta_i, \quad \omega\text{-lim}\, \eta_i = 0.$$

Let us assume that $s_i = d(x_i, z_i) \leqslant d(z_i, y_i)$ and consider the point $q_i \in x_i y_i$ within distance s_i from x_i. Compare the triangle $T_i = T(x_i, y_i, z_i)$ with the Euclidean triangle using the comparison points $p_i = z_i$ and q_i. In the Euclidean comparison triangle $\tilde{T}_i$, we have

$$\omega\text{-lim}\, d(\tilde{z}_i, \tilde{q}_i) = 0$$

(since the constant ultralimit of the sequence of Euclidean planes is the Euclidean plane and, hence, is uniquely geodesic). Since, by the $CAT(0)$-comparison inequality,

$$d(z_i, q_i) \leqslant d(\tilde{z}_i, \tilde{q}_i),$$

we conclude that $(q_i) = z_\omega$ in the space X_ω. Thus, z_ω lies on the limit geodesic connecting x_ω and y_ω. □

LEMMA 7.54 (Ultralimits preserve direct products of metric spaces). *Suppose that $X_i = X_i' \times X_i'', i \in I$, is an indexed family of direct products of metric spaces, i.e., the metrics on X_i are given by the Pythagorean formula* (1.7). *Then for every ω and a family of base-points $e_i \in X_i, e_i = e_i' \times e_i''$ we have an isometry*

$$X_\omega = \omega\text{-}lim(X_i, e_i) \cong \omega\text{-}lim(X_i', e_i') \times \omega\text{-}lim(X_i'', e_i'').$$

Proof. By the definition of an ultralimit, as a set, X_ω splits naturally as a direct product of two ultralimits $X'_\omega = \omega\text{-lim}(X'_i, e'_i)$ and $X''_\omega = \omega\text{-lim}(X''_i, e''_i)$. Consider points $x_\omega = x'_\omega \times x''_\omega$ and $y_\omega = y'_\omega \times y''_\omega$ in X_ω, where $x'_\omega = (x'_i), y'_\omega = (y'_i), x''_\omega = (x''_i), y''_\omega = (y''_i)$. The distance between these points is given by

$$\operatorname{dist}^2(x_\omega, y_\omega) = \omega\text{-lim}\operatorname{dist}^2_{X_i}(x_i, y_i) =$$
$$\omega\text{-lim}\left(\operatorname{dist}^2_{X'_i}(x'_i, y'_i) + \operatorname{dist}^2_{X''_i}(x''_i, y''_i)\right) =$$
$$\omega\text{-lim}\operatorname{dist}^2_{X'_i}(x'_i, y'_i) + \omega\text{-lim}\operatorname{dist}^2_{X''_i}(x''_i, y''_i) = \operatorname{dist}^2_{X'_\omega}(x'_\omega, y'_\omega) + \operatorname{dist}^2_{X''_\omega}(x''_\omega, y''_\omega).$$

Lemma follows. □

7.5. Completeness of ultralimits and incompleteness of ultrafilters

So far, our discussion of ultralimits did not depend on the nature of the set I and the ultrafilter ω (as long as the latter was non-principal). In this section we discuss the question of completeness of ultralimits of families of metric spaces. It turns out that the answer depends on the ultrafilter.

Definition 7.55. An ultrafilter ω is called *countably complete* if it is closed under countable intersections.

Each principal ultrafilter is obviously countably complete. (In fact, a principal ultrafilter is closed under arbitrary intersections.) On the other hand, as we will see soon, any non-principal ultrafilter on a countable set is countably incomplete, and, hence, for the purposes of Geometric Group Theory, countably complete ultrafilters are irrelevant. Existence of countably complete non-principal ultrafilters is unprovable in ZFC, we refer the refer to [**Kei10**] for details and references.

Below is a characterization of complete ultrafilters that we will need.

Lemma 7.56. *The following are equivalent for an ultrafilter ω on a set I:*

1. ω is countably incomplete.

2. There exists a map $\nu : I \to \mathbb{N}$, which sends ω to a non-principal ultrafilter, $\nu(\omega)$, i.e., for each finite subset $S \subset \mathbb{N}$ the preimage of S under ν is ω-large.

Proof. 1. Suppose that ω is countably incomplete and, hence, there exists a sequence (J_n) of ω-large subsets of I with intersection J not in ω. By taking finite intersections of the sets J_n, we can assume that the sequence J_n is strictly decreasing:

$$J_1 \supset J_2 \supset J_3 \supset \dots$$

We define the following function $\nu : I \to \mathbb{N} \cup \{\infty\}$:

For each $i \in I$ we let $\nu(i)$ denote the supremum of the set

$$\{n : i \in J_n\}.$$

If this set is empty, we, of course, have $\nu(i) = 1$; if this set is unbounded, $\nu(i) = \infty$. Clearly, $\nu^{-1}(\infty) = J$ is not ω-large. If there exists a finite subset $[1, n] \subset \mathbb{N}$ such that $K = \nu^{-1}([1, n])$ is ω-large, then K is disjoint from J_{n+1}, which is a contradiction. Hence, $\nu(\omega)$ is a non-principal ultrafilter.

2. Suppose there exists a map $\nu : I \to \mathbb{N}$ which sends ω to a non-principal ultrafilter. Then for each interval $[n, \infty) \subset \mathbb{N}$, the preimage $\nu^{-1}([n, \infty))$ is ω-large. The intersection of these preimages is empty and, hence, ω is countably incomplete. □

COROLLARY 7.57. *Any non-principal ultrafilter on a countable set is countably incomplete.*

Countably complete ultrafilters behave essentially like principal ultrafilters, as far as convergence in metric spaces is concerned:

LEMMA 7.58. *Suppose that X is a 1st countable Hausdorff topological space and $(x_i)_{i\in I}$ is an indexed family in X which ω-converges to some $x \in X$. Then the family $(x_i)_{i\in I}$ is ω-constant:*

$$\omega(\{i \in I : x_i = x\}) = 1.$$

PROOF. Consider a countable basis of topology U_n at the point x. Then for each n,

$$\omega(\{i : x_i \in U_n\}) = 1.$$

Since the intersection J of the sets $\{i : x_i \in U_n\}$ is still ω-large and X is Hausdorff, we conclude that for ω-all i's, $x_i = x$. □

COROLLARY 7.59. *Suppose that ω is countably complete. Then for each family $(X_i, e_i)_{i\in I}$ of pointed metric spaces, the ultralimit ω-lim(X_i, e_i) is isometric to the pointed ultraproduct*

$$\left(\prod_{i\in I}(X_i, e_i)\right)/\omega.$$

In particular, the constant ultralimit $(X, e)_{i\in I}$ of an incomplete metric space X is still incomplete.

Thus, countably complete ultrafilters lead to incomplete ultralimits. We now turn to countably incomplete ultrafilters. The reader interested only in Geometric Group Theory applications, can safely assume here that the index set I is countable.

LEMMA 7.60. *Let $(Y_i)_{i\in I}$ be a family of of metric spaces, and for every i let X_i be a dense subset in Y_i. Then for every countably incomplete ultrafilter ω on I, the natural isometric embedding of the ultralimit ω-lim$_i X_i$ into the ultralimit $Y_\omega = \omega$-lim$_i Y_i$ is surjective. In particular, this holds when $Y_i = \widehat{X}_i$, the metric completion of X_i.*

PROOF. We first give a proof in the case when I is countable, since it is based on a *diagonal subsequence* argument probably familiar to the reader. We will identify I with the set of the natural numbers $\mathbb{N}$ and consider a point $y_\omega \in Y_\omega$ and the corresponding indexed family $(y_i)_{i\in I}$. By density of X_i in Y_i, for each i there exists a sequence $(x_{in})_{n\in\mathbb{N}}$ in X_i whose limit is y_i. For each i we choose n_i such that for $x_i := x_{in_i}$,

$$\mathrm{dist}_{Y_i}(x_i, y_i) < \frac{1}{i}.$$

It follows that

$$\omega\text{-lim}\, \mathrm{dist}_{Y_i}(x_i, y_i) = 0$$

and, hence, $(x_i) = (y_i)$ in Y_ω.

Suppose now that ω is a general countably incomplete ultrafilter and $\nu : I \to \mathbb{N}$ is a mapping which sends ω to a non-principal ultrafilter on $\mathbb{N}$. For each i choose $x_i \in X_i$ such that

$$\mathrm{dist}_{Y_i}(x_i, y_i) < \frac{1}{\nu(i)}.$$

(Here we are using the AC!) Since $\nu(\omega)$ is a non-principal ultrafilter on $\mathbb{N}$,

$$\omega\text{-lim} \frac{1}{\nu(i)} = 0,$$

hence,

$$\omega\text{-lim} \operatorname{dist}_{Y_i}(x_i, y_i) = 0$$

as well. We again conclude that $(x_i) = (y_i)$ in Y_ω. $\square$

EXERCISE 7.61. Let Y be a proper metric space, take a subset $X \subset Y$ equipped with the restriction metric. Then for each countably incomplete ultrafilter ω, the constant ultralimit ω-$\lim(X, e)$ is naturally isometric to $(\bar{X}, e)$, where $\bar{X}$ is the closure of X in Y. Hint: Use Exercise 7.44.

In the next proposition, we make no assumptions about completeness of the ultrafilter ω:

PROPOSITION 7.62. *Every based ultralimit ω-$\lim_i(X_i, e_i)$ of a family of complete metric spaces is a complete metric space.*

PROOF. We will prove that every Cauchy sequence $(x^{(k)})$ in $X_{\omega,\boldsymbol{e}}$ contains a convergent subsequence, this will imply that $(x^{(k)})$ converges as well. We select a subsequence (which we again denote $(x^{(k)})$) such that

$$\operatorname{dist}_\omega \left(x^{(k)}, x^{(k+1)} \right) < \frac{1}{2^k}.$$

Equivalently,

$$\omega\text{-}\lim_i \operatorname{dist}_i \left(x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k},$$

which implies that

$$\operatorname{dist}_i \left(x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k}\,,\ \omega - \text{a.s.},$$

i.e., for every k the following set is ω-large:

$$I_k = \left\{ i \in I \ ;\ \operatorname{dist}_i \left(x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k} \right\}.$$

We can assume that $I_{k+1} \subseteq I_k$, otherwise we replace I_{k+1} with $I_{k+1} \cap I_k$. Thus, we obtain a nested sequence of subsets I_k in I:

$$I_1 \supset I_2 \supset I_3 \supset \dots.$$

Case 1. Assume first that the intersection $J := \bigcap_{k \geqslant 1} I_k$ of these subsets is also ω-large. (This will be always the case is ω is countably complete.)

For every $i \in J$ the sequence $\left(x_i^{(k)} \right)$ is Cauchy, therefore, since the space X_i is complete, this sequence converges to some $y_i \in X_i$. The inequalities

$$\operatorname{dist}_i \left(x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k}\,,\ k \in \mathbb{N}\,,$$

imply that for every $m > k$,

$$\operatorname{dist}_i \left(x_i^{(k)}, x_i^{(m)} \right) < \frac{1}{2^{k-1}}.$$

The latter gives, by taking the limit as $m \to \infty$, that

$$\operatorname{dist}_i \left(x_i^{(k)}, y_i \right) \leqslant \frac{1}{2^{k-1}}\,.$$

Hence,

$$\operatorname{dist}_\omega\left(x^{(k)}\,,\,y_\omega\right) \leqslant \frac{1}{2^{k-1}}\,,$$

for $y_\omega = \omega\text{-}\lim y_i$. We have, thus, obtained a limit y_ω of the sequence $\left(x^{(k)}\right)$.

Case 2. Assume now that $\omega(J) = 0$. Since for every $k \geqslant 1$ we have that

$$I_k = J \sqcup \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1})$$

and $\omega(I_k) = 1$, it follows that

$$\omega\left(\bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1})\right) = 1\,.$$

We define subsets

$$J_k := \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1}) \subset I.$$

We claim that the limit point of the sequence $(x^{(k)})$ is $y_\omega = (y_i) \in Y_\omega$, where $y_i = x_i^{(k)}$ whenever $i \in I_k \setminus I_{k+1}$. This defines y_i for all $i \in J_1$. We extend this definition to the rest of I arbitrarily: Values taken on ω-small sets of indices $i \in I$ do not matter.

For every

$$i \in J_k = \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1})$$

there exists $j \geqslant k$ such that $i \in I_j \setminus I_{j+1}$. By the definition, $y_i = x_i^{(j)}$.

Since

$$i \in I_j \subseteq I_{j-1} \subseteq \cdots \subseteq I_{k+1} \subseteq I_k$$

we may write

$$\operatorname{dist}_i\left(x_i^{(k)}, y_i\right) \leqslant \operatorname{dist}_i\left(x_i^{(k)}, x_i^{(k+1)}\right) + \cdots + \operatorname{dist}_i\left(x_i^{(j-1)}, x_i^{(j)}\right) \leqslant$$
$$\frac{1}{2^k} + \frac{1}{2^{k+1}} + \cdots + \frac{1}{2^{j-1}} \leqslant \frac{1}{2^k}\frac{1}{1-\frac{1}{2}} = \frac{1}{2^{k-1}}\,.$$

Thus, we have

$$\operatorname{dist}_\omega\left(x^{(k)}, y_\omega\right) \leqslant \frac{1}{2^{k-1}}\,,$$

which implies that the sequence $x^{(k)}$ indeed converges to y_ω. □

Corollary 7.63. *Suppose that ω is a countably incomplete ultrafilter (e.g., I is countable). Then:*

1. The ultralimit ω-lim(X_i, e_i) of any family of based metric spaces $(X_i, e_i)_{i\in I}$ is complete.

2. For each family $A_i \subset X_i$ of subsets, the ultralimit $A_\omega \subset X_\omega$ is a closed subset.

Proof. This is a combination of Lemma 7.60 with Proposition 7.62. □

Thus, (countably) incomplete ultrafilters, lead to metric completeness of ultralimits.

EXAMPLE 7.64. Let $(\mathcal{H}_i)_{i\in\mathbb{N}}$ be a sequence of Hilbert spaces and let $Y_i = S(0, R_i) \subset \mathcal{H}_i$ be metric spheres of radii R_i diverging to infinity. Then for each non-principal ultrafilter ω and choice of base-points $y_i \in Y_i$, the ultralimit ω-$\lim(Y_i, y_i)$ is isometric to a Hilbert space. Indeed, in view of the Example 7.47, if we fix n and let Σ_i denote the intersection of Y_i with an n-dimensional subspace in $\mathcal{H}_i$ containing y_i, then ω-$\lim(\Sigma_i, y_i) \cong \mathbb{E}^{n-1}$. It follows that Y_ω is a complete (see Proposition 7.62) geodesic metric space such that:

1. Each finite subset of Y_ω is isometric to a subset of a Euclidean space.

2. For each geodesic segment $xy \subset Y_\omega$ there exists a complete geodesic in Y_ω containing xy.

Combining the first property with Theorem 1.137, we conclude that there exists an isometric embedding $\phi : Y_\omega \to \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space. Without loss of generality, we may asume that $\phi(y_\omega) = 0$. Since Y_ω is a geodesic metric space and $\mathcal{H}$ is uniquely geodesic with geodesics given by line segments, the image $\phi(Y_\omega)$ is a convex subset of $\mathcal{H}$. Furthermore, the second property mentioned above implies with each point $y \neq 0$, the subset $\phi(Y_\omega)$ contains the line $\mathbb{R}y \subset \mathcal{H}$. It follows that $\phi(Y_\omega)$ is a linear subspace V in $\mathcal{H}$. Completeness of V follows from that of Y_ω.

7.6. Asymptotic cones of metric spaces

The concept of an asymptotic cone was first introduced in the geometric group theory by van den Dries and Wilkie in [**dDW84**], although its version for groups of polynomial growth was already used by Gromov in [**Gro81a**], who used Gromov-Hausdorff convergence as a tool. Asymptotic cones (and ultralimits) for general metric spaces were defined by Gromov in [**Gro93**]. The idea is to construct, for a metric space (X, dist), its "image" seen from "infinitely far." More precisely, one defines the notion of a *limit* of a sequence of metric spaces $(X, \varepsilon\mathrm{dist})$, $\varepsilon > 0$, as $\varepsilon \to 0$.

Let (X, d_X) be a metric space and ω be a non-principal ultrafilter on I. For each positive real number λ we define the new metric space $\lambda X = (X, \lambda d_X)$ by rescaling the metric d_X. Suppose that we are given a family $\boldsymbol{\lambda} = (\lambda_i)_{i\in I}$ of positive real numbers indexed by I such that ω-$\lim \lambda_i = 0$ and a family $\boldsymbol{e} = (e_i)_{i\in I}$ of base-points $e_i \in X$ indexed by I. Given this data, the *asymptotic cone* $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ of X is defined as the based ultralimit of rescaled copies of X:

$$\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) := \omega\text{-}\lim_i(\lambda_i \cdot X, e_i).$$

For a family of points $(x_i)_{i\in I}$ in X, the corresponding subset in the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, which is either a one-point set, or the empty set if ω-$\lim \lambda_i \mathrm{dist}(x_i, e_i) = \infty$, is denoted by ω-$\lim x_i$.

The family $\boldsymbol{\lambda} = (\lambda_i)_{i\in I}$ is called the *scaling family*. When either the scaling family or the family of base-points are irrelevant, they are omitted from the notation.

Thus, to each metric space X we attach a collection of metric spaces $\mathrm{Cones}(X)$ consisting of all asymptotic cones $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ of X, that is of all the "images of X seen from infinitely far." The first questions to ask are: How large is the collection $\mathrm{Cones}(X)$ for specific metric spaces X, and what features of X are inherited by the metric spaces in $\mathrm{Cones}(X)$.

An asymptotic cone of a finitely generated group is the asymptotic cone of this group regarded as a metric space, where we use the word metric defined by the given finite generating set. As we will see below, the *bi-Lipschitz homeomorphism class* of such asymptotic cone is independent of the generating set and the choice of base-points, but does depend on the ultrafilter ω and the scaling family $\boldsymbol{\lambda}$.

We begin by noting that the choice of base-points is irrelevant for spaces that are quasihomogeneous:

EXERCISE 7.65. [See also Proposition 7.71.] When the space X is quasihomogeneous, all cones defined by the same fixed ultrafilter ω and sequence of scaling constants $\boldsymbol{\lambda}$, are isometric.

Another simple observation is:

REMARK 7.66. Let α be a positive real number. The map

$$I_\alpha : \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) \to \mathrm{Cone}_\omega\left(X, \boldsymbol{e}, \alpha\boldsymbol{\lambda}\right)\,,\; I_\alpha(\omega\text{-lim}\, x_i) = \omega\text{-lim}\, x_i$$

is a similarity: It multiplies all the distances by the factor α. Thus, for a fixed metric space X, the collection of asymptotic cones $\mathrm{Cones}(X)$ is stable with respect to rescaling of the metric on X.

In particular, since the Euclidean space $\mathbb{E}^n$ is proper, homogeneous and *self-similar* ($\mathbb{E}^n$ is isometric to $\alpha\mathbb{E}^n$ for each $\alpha > 0$), it follows that

$$\mathrm{Cone}_\omega\, \mathbb{E}^n \cong \mathbb{E}^n.$$

The same applies to all finite-dimensional normed vector spaces $(V, \|\cdot\|)$:

$$\mathrm{Cone}_\omega(V, \|\cdot\|) \cong (V, \|\cdot\|).$$

Lemmata 7.54 and 7.51 imply that asymptotic cones preserve direct product decompositions of metric spaces and geodesic metric spaces:

COROLLARY 7.67. (1) $\mathrm{Cone}_\omega(X \times Y) = \mathrm{Cone}_\omega(X) \times \mathrm{Cone}_\omega(Y)$.
(2) *The asymptotic cone of a geodesic space is a geodesic space.*

DEFINITION 7.68. Given a family $(A_i)_{i\in I}$ of subsets of (X, dist), we denote either by $\omega\text{-lim}\, A_i$ or by A_ω the subset of $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ that consists of all the elements $\omega\text{-lim}\, x_i$ such that $x_i \in A_i$ ω–almost surely. We call $\omega\text{-lim}\, A_i$ the *limit set* of the family $(A_i)_{i\in I}$.

Note that if $\omega\text{-lim}\, \lambda_i \mathrm{dist}(e_i, A_i) = \infty$ then the set $\omega\text{-lim}\, A_i$ is empty.

PROPOSITION 7.69 (Van den Dries and Wilkie, Proposition 4.2 in **[dDW84]**). *If ω is countably incomplete (e.g., the index set I is countable) then:*

(1) *Any asymptotic cone (with respect to ω) of a metric space is complete.*
(2) *For each family $A_i \subset X_i$, the limit set ω-lim A_i is a closed subset of* $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$.

PROOF. This is an immediate consequence of Proposition 7.62. □

In Definition 7.50 we introduced the notion of *limit geodesics* in the ultralimit of a sequence of metric spaces. Let $\gamma_i : [a_i, b_i] \to X$ be a family of geodesics with the limit geodesic γ_ω in $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$.

EXERCISE 7.70. Show that the image of γ_ω is the limit set of the sequence of images of the geodesics γ_i.

We saw earlier that geodesics in the ultralimit may fail to be limit geodesics. However, in our example, we took a sequence of metric spaces which were not geodesic. It turns out that, in general, there exist geodesics in $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ that are not limit geodesic, even when X is the Cayley graph of a finitely generated group with a word metric. An example of this can be found in [**Dru09**].

Suppose that X is a metric space and $G \subset \mathrm{Isom}(X)$ is a subgroup. Given a non-principal ultrafilter ω consider the ultraproduct $G^\omega = \prod_{i\in I} G/\omega$. For a family of positive real numbers $\boldsymbol{\lambda} = (\lambda_i)_{i\in I}$ such that $\omega\text{-}\lim \lambda_i = 0$ and a family of base-points $\boldsymbol{e} = (e_i)$ in X, let $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ be the corresponding asymptotic cone. In view of Lemma 7.48, the group G^ω acts isometrically on the ultralimit

$$U := \omega\text{-}\lim(\lambda_i \cdot X).$$

Let $G^\omega_{\boldsymbol{e}} \subset G^\omega$ denote the stabilizer in G^ω of the component $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) \subset U$. In other words,

$$G^\omega_{\boldsymbol{e}} = \{(g_i)^\omega \in G^\omega : \omega\text{-}\lim \lambda_i \mathrm{dist}(g_i(e_i), e_i) < \infty\}.$$

There is a natural homomorphism $G^\omega_{\boldsymbol{e}} \to \mathrm{Isom}(\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}))$. Observe also that if (e_i) is a bounded family in X then the group G has a diagonal embedding into $G^\omega_{\boldsymbol{e}}$.

PROPOSITION 7.71. *Suppose that $G \subset \mathrm{Isom}(X)$ and the action $G \curvearrowright X$ is cobounded. Then for every asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ the action $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is transitive. In particular, $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is a homogeneous metric space.*

PROOF. Let $D < \infty$ be such that $G \cdot x$ is a D-net in X. Given two indexed families $(x_i), (y_i)$ of points in X, there exists an indexed family (g_i) of elements of G such that

$$\mathrm{dist}(g_i(x_i), y_i) \leqslant 2D.$$

Therefore, if $g_\omega := (g_i)^\omega \in G^\omega$, then $g_\omega(\omega\text{-}\lim_i x_i) = \omega\text{-}\lim_i y_i$. Hence the action

$$G^\omega \curvearrowright U = \omega\text{-}\lim_i(\lambda_i \cdot X)$$

is transitive. It follows that the action $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is transitive as well. □

EXERCISE 7.72. 1. Construct an example of a metric space X, a bounded sequence (e_i) and an asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ so that for the isometry group $G = \mathrm{Isom}(X)$ the action $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is not effective, i.e., the homomorphism

$$G^\omega_{\boldsymbol{e}} \to \mathrm{Isom}\left(\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})\right)$$

has nontrivial kernel. Construct an example when the kernel of the above homomorphism contains the entire group G embedded diagonally in $G^\omega_{\boldsymbol{e}}$.

2. Show that $\mathrm{Ker}(G \to QI(X))$ is contained in $\mathrm{Ker}(G \to \mathrm{Isom}(X_\omega))$.

Suppose that X admits a cocompact discrete action of a subgroup $G < \mathrm{Isom}(X)$. The problem of how large the class of spaces $\mathrm{Cones}(X)$ can be, that is the problem of the dependence of the topological/metric type of $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ on the ultrafilter ω and the scaling sequence $\boldsymbol{\lambda}$, is, in general, quite hard. In some special cases, it is related to the Continuum Hypothesis (the hypothesis stating that there is no cardinal number between $\aleph_0$ and $2^{\aleph_0}$). Consider, for concreteness, the group $SL(n,\mathbb{R})$, $n \geqslant 3$, equipped with a fixed left-invariant metric.

Kramer, Shelah, Tent and Thomas have shown in [**KSTT05**] that:

(1) If the Continuum Hypothesis (CH) is not true then the group $SL(n,\mathbb{R})$, $n \geqslant 3$, has $2^{2^{\aleph_0}}$ non-homeomorphic asymptotic cones.
(2) If the CH is true then all asymptotic cones of $SL(n,\mathbb{R})$, $n \geqslant 3$, are isometric. Moreover, under the same assumption, a finitely generated group (with a fixed finite generating set) has at most continuum of non-isometric asymptotic cones.

Moreover, according to Theorem 1.4 in Blake Thornthon's PhD thesis [**Tho02**]:

<u>Theorem</u> 7.73. *Assuming CH, if X is a nonpositively curved symmetric space then all asymptotic cones of X are isometric to each other, i.e., up to isometry asymptotic cones are independent of the scaling sequence and the choice of a non-principal ultrafilter.*

The case of $SL(2,\mathbb{R})$ was settled independently of the CH by A. Dyubina–Erschler and I. Polterovich (see Theorem 9.165).

Chronologically, the first non-trivial example of metric space X such that the set $\mathrm{Cones}(X)$ contains very few elements (up to bilipschitz homeomorphisms) is that of virtually nilpotent groups, and is due to P. Pansu, see Theorem 14.28.

C. Druţu and M. Sapir constructed in [**DS05b**] an example of two-generated and recursively presented (but not finitely presented) group with continuum of pairwise non-homeomorphic asymptotic cones. The construction is independent of the Continuum Hypothesis. The example can be adapted so that at least one asymptotic cone is a real tree.

Note that if a finitely presented group G has *one* asymptotic cone which is a real tree, then the group is hyperbolic and hence *every* asymptotic cone of G is a real tree, see Theorem 9.161.

Historical remarks. The first instance (that we are aware of) where asymptotic cones of metric spaces were defined is the 1966 paper [**BDCK66**], where this is done in the context of normed vector spaces. Their definition, though, works for all metric spaces.

On the other hand, Gromov introduced the modified Hausdorff distance (see §5.1 for a definition) and the corresponding limits of sequences of pointed metric spaces in his work on groups of polynomial growth [**Gro81a**]. This approach is no longer appropriate in the case of more general metric spaces, as we will explain below.

Firstly, the modified Hausdorff distance does not distinguish between a space its dense subset, therefore in order to have a well defined limit one has to require *a priori* for the limit be complete.

Secondly, if a pointed sequence of proper geodesic metric spaces $(X_n, \mathrm{dist}_n, x_n)$ converges to a complete geodesic metric space (X, dist, x) in the modified Hausdorff distance, then the limit space X is proper. Indeed given a ball $B(x,R)$ in X, for every ϵ there exists an n such that $B(x,R)$ is at Hausdorff distance at most ϵ from the ball $B(x_n,R)$ in X_n. From this and the fact that all spaces X_n are proper it follows that for every sequence (y_n) in $B(x,R)$ and every ε there exists a subsequence of (y_n) of diameter $\leqslant \varepsilon$. A diagonal argument and completeness of

X allow to conclude that (y_n) has a convergent subsequence, and therefore that $B(x,R)$ is compact.

In view of Theorem 7.46, for a proper geodesic metric space (X, dist), the existence of a sequence of pointed metric spaces of the form $(X, \lambda_n \mathrm{dist}, e_n)$ convergent in the modified Hausdorff metric, implies the existence of proper asymptotic cones. On the other hand, if X is, for instance, the hyperbolic plane or a non-elementary hyperbolic group, no asymptotic cone of X is proper, see Theorem 9.165. Therefore, in such a case, the sequence $(X, \frac{1}{n}\mathrm{dist})$ has no subsequence convergent with respect to the modified Hausdorff metric.

7.7. Ultralimits of asymptotic cones are asymptotic cones

In this section we show that ultralimits of asymptotic cones are asymptotic cones, following [**DS05b**]. To this end, we first describe a construction of ultrafilters on Cartesian products that generalizes the standard notion of product of ultrafilters, as defined in [**She78**, Definition 3.2 in Chapter VI]. In what follows, we view ultrafilters as in Definition 7.19. Throughout the section, ω will denote an ultrafilter on a set I and $\mu = (\mu_i)_{i \in I}$ a family, indexed by I, of ultrafilters on a set J.

DEFINITION 7.74. We define a new ultrafilter $\omega\mu$ on $I \times J$ such that for every subset A in $I \times J$, $\omega\mu(A)$ is equal to the ω–measure of the set of all $i \in I$ such that $\mu_i(A \cap (\{i\} \times J)) = 1$.

LEMMA 7.75. *$\omega\mu$ is an ultrafilter over $I \times J$.*

PROOF. It suffices to prove that $\omega\mu$ is finitely additive and that it takes the zero value on finite sets.

We first prove that $\omega\mu$ is finitely additive, using the fact that ω and μ_i are finitely additive. Let A and B be two disjoint subsets of $I \times J$. Fix an arbitrary $i \in I$. The sets $A \cap (\{i\} \times J)$ and $B \cap (\{i\} \times J)$ are disjoint, hence

$$\mu_i((A \cup B) \cap (\{i\} \times J)) = \mu_i(A \cap (\{i\} \times J)) + \mu_i(B \cap (\{i\} \times J)).$$

The finite additivity of ω implies that

$$\omega\mu(A \sqcup B) = \omega\mu(A) + \omega\mu(B).$$

Also, given a finite subset A of $I \times J$, $\omega\mu(A) = 0$. Indeed, since the set of i's for which $\mu_i(A \cap (\{i\} \times J)) = 1$ is empty, $\omega\mu(A) = 0$ by definition. □

LEMMA 7.76 (double ultralimit of real numbers). *For every doubly indexed family of real numbers α_{ij}, $i \in I, j \in J$ we have that*

$$\omega\mu\text{-}\lim \alpha_{ij} = \omega\text{-}\lim \left(\mu_i\text{-}\lim_j \alpha_{ij}\right), \tag{7.4}$$

where the second limit on the right hand side is taken with respect to $j \in J$.

PROOF. Let a be the limit $\omega\mu$-limα_{ij}. For every neighborhood U of a,

$$\omega\mu \{(i,j) \mid \alpha_{ij} \in U\} = 1 \Leftrightarrow$$
$$\omega \{i \in I \mid \mu_i \{j \mid \alpha_{ij} \in U\} = 1\} = 1\,.$$

This implies that

$$\omega \left\{ i \in I \mid \mu_i\text{-}\lim_j \alpha_{ij} \in \overline{U} \right\} = 1\,,$$

which, in turn, implies that

$$\omega\text{-}\lim(\mu_i\text{-}\lim_j \alpha_{ij}) \in \overline{U}\,.$$

This holds for every neighborhood U of $a \in \mathbb{R} \cup \{\pm\infty\}$. Therefore, we conclude that

$$\omega\text{-lim}(\mu_i\text{-lim } \alpha_{ij}) = a.$$

□

Lemma 7.76 implies a similar result for ultralimits of spaces.

PROPOSITION 7.77 (double ultralimit of spaces). *Let $(X_{ij}, \text{dist}_{ij})$ be a doubly indexed sequence of metric spaces, $(i, j) \in I \times J$, and let $e = (e_{ij})$ be a doubly indexed sequence of points $e_{ij} \in X_{ij}$. We denote by e_i the sequence $(e_{ij})_{j \in J}$.*

Then the map

$$\omega\mu\text{-lim}(x_{ij}) \mapsto \omega\text{-lim}(\mu_i\text{-lim} x_{ij}) \ , \tag{7.5}$$

is an isometry from

$$\omega\mu\text{- lim}(X_{ij}, e_{ij})$$

onto

$$\omega\text{- lim}(\mu_i\text{- lim}(X_{ij}, e_{ij}) \ , \ e_i')$$

where, $e_i' = \mu_i\text{- lim}\, e_{ij}$.

COROLLARY 7.78 (ultralimits of asymptotic cones are asymptotic cones). *Let X be a metric space. Consider double indexed families of points $\boldsymbol{e} = (e_{ij})_{(i,j) \in I \times J}$ in X and of positive real numbers $\boldsymbol{\lambda} = (\lambda_{ij})_{(i,j) \in I \times J}$ such that*

$$\mu_i\text{-lim}_j \lambda_{ij} = 0$$

for every $i \in I$. Let $\text{Cone}_{\mu_i}(X, (e_{ij}), (\lambda_{ij}))$ be the corresponding asymptotic cone of X. The map

$$\omega\mu\text{-lim}(x_{ij}) \mapsto \omega\text{-lim}(\mu_i\text{-lim}(x_{ij}))\,, \tag{7.6}$$

is an isometry from $\text{Cone}_{\omega\mu}(X, \boldsymbol{e}, \boldsymbol{\lambda})$ onto

$$\omega\text{-lim}\left(\text{Cone}_{\mu_i}(X, (e_{ij}), (\lambda_{ij}))\,, \mu_i\text{-lim} e_{ij}\right) \ .$$

PROOF. The statement follows from Proposition 7.77. The only thing to be proved here is that

$$\omega\mu\text{-lim}\lambda_{ij} = 0$$

Let $\varepsilon > 0$. For every $i \in I$ we have that

$$\mu_i\text{-lim}\lambda_{ij} = 0,$$

whence,

$$\mu_i \{j \in I \mid \lambda_{ij} < \varepsilon\} = 1 \ .$$

It follows that

$$\{i \in I \mid \mu_i \{j \in I \mid \lambda_{ij} < \varepsilon\} = 1\} = I,$$

therefore, the ω-measure of this set is 1. We conclude that

$$\omega\mu \{(i, j) \in I \times J \mid \lambda_{ij} < \varepsilon\} = 1. \quad □$$

COROLLARY 7.79. *Let X be a metric space. The collection of all asymptotic cones of X is stable with respect to rescaling, ultralimits and taking asymptotic cones.*

PROOF. It is an immediate consequence of Corollary 7.78 and Remark 7.66. □

COROLLARY 7.80. *Let X, Y be metric spaces such that all asymptotic cones of X are isometric to Y. Then all asymptotic cones of Y are isometric to Y.*

This, in particular, implies that the following are *examples of metric spaces isometric to all their asymptotic cones.*

EXAMPLES 7.81.
(1) The $2^{\aleph_0}$–universal real tree T_C, according to Theorem 9.165.
(2) A non-discrete Euclidean building that is the asymptotic cone of $SL(n,\mathbb{R})$, $n \geqslant 3$, under the Continuum Hypothesis, according to **[KSTT05]** and **[KL98b]**.
(3) A graded nilpotent Lie group with a Carnot–Caratheodory metric, according to Theorem 14.28 of P. Pansu.

7.8. Asymptotic cones and quasiisometries

The following simple lemma shows why asymptotic cones are useful in studying quasiisometries, since they become bi-Lipschitz maps of asymptotic cones, and the latter maps are much easier to handle. It is a direct generalization of Lemma 7.48 on functoriality of ultralimits with respect to isometries.

LEMMA 7.82. *Let $(X, e_i), (X', e'_i)$ be pointed metric spaces, and let $\boldsymbol{\lambda} = (\lambda_i)$ be a scaling family. Define the asymptotic cones*

$$X_\omega = \mathrm{Cone}_\omega(X, (e_i), \boldsymbol{\lambda}), \quad X'_\omega = \mathrm{Cone}_\omega(X', (e'_i), \boldsymbol{\lambda}).$$

Then the following holds for every family of (L, A)-coarse Lipschitz maps $f_i : X \to X'$, satisfying

$$\omega\text{-}lim\, d(f_i(e_i), e'_i) < \infty :$$

1. The ultralimit $f_\omega : X_\omega \to X'_\omega$ of the family (f_i),

$$f_\omega((x_i)) := (f_i(x_i)),$$

is L-Lipschitz.

2. If f_i is an (L, A)-quasiisometric embedding, then f_ω is an L-bi-Lipschitz embedding.

3. The correspondence $\Phi_\omega : (f_i) \mapsto f_\omega$ is functorial:

$$\Phi_\omega : (g_i \circ f_i) \mapsto g_\omega \circ f_\omega.$$

4. If $X = X'$ and f_i's have uniformly bounded displacement, i.e., for ω-all i,

$$\mathrm{dist}(f_i(x), x)) \leqslant A, \quad \forall x \in X,$$

then $f_\omega = \mathrm{id}_X$.

5. If each f_i is an (L, A)-quasiisometry, then f_ω is an L-bi-Lipschitz homeomorphism.

PROOF. 1. We have the inequalities:

$$\frac{1}{\lambda_i}\mathrm{dist}(f_i(x_i), f_i(y_i)) \leqslant L\frac{1}{\lambda_i}\mathrm{dist}(x_i, y_i) + \frac{A}{\lambda_i}.$$

Passing to the ω–limit, we obtain

$$\mathrm{dist}_\omega(f_\omega(x_\omega), f_\omega(y_\omega)) \leqslant L\ \mathrm{dist}_\omega(x_\omega, y_\omega).$$

where $x_\omega = (x_i), y_\omega = (y_i)$. Thus, f_ω is L-Lipschitz.

2. In this case we also have the inequalities

$$L^{-1}\frac{1}{\lambda_i}\text{dist}(x_i, y_i) - \frac{A}{\lambda_i} \leqslant \frac{1}{\lambda_i}\text{dist}(f_i(x_i), f_i(y_i)),$$

which, after passing to the ultralimit, become

$$L^{-1}\text{dist}_\omega(x_\omega, y_\omega) \leqslant \text{dist}_\omega(f_\omega(x_\omega), f_\omega(y_\omega)).$$

Thus, f_ω is an L-bi-Lipschitz embedding.

Parts 3 and 4 are clear. Part 5 follows from 1, 3 and 4. □

One may ask if a converse to this lemma is true, for instance: Does the existence of a (coarse Lipschitz) map between metric spaces that induces bi-Lipschitz maps between asymptotic cones imply quasiisometry of the original metric spaces? We say that two spaces are *asymptotically bi-Lipschitz* if the latter holds. (This notion is introduced in [**dC09**].) See Remark 14.29 for an example of asymptotically bi-Lipschitz spaces which are not quasiisometric to each other.

Here is an example of application of asymptotic cones to the study of quasi-isometries.

Lemma 7.83. *Suppose that $X = \mathbb{E}^n$ or $\mathbb{R}_+$ and $f : X \to X$ is an (L, A)–quasiisometric embedding. Then f is a quasiisometry, furthermore, $\mathcal{N}_C(f(X)) = X$, for some $C = C(L, A)$.*

Proof. We will give a proof in the case of $\mathbb{E}^n$ as the other case is analogous. Suppose that the assertion is false, i.e., there is a sequence of (L, A)–quasiisometric embeddings $f_i : \mathbb{E}^n \to \mathbb{E}^n$, sequence of real numbers r_i diverging to infinity and points $y_i \in \mathbb{E}^n$ such that $\text{dist}(y_i, f(\mathbb{E}^n)) = r_i$. Let $x_i \in \mathbb{E}^n$ be a point such that $\text{dist}(f(x_i), y_i) \leqslant r_i + 1$. Using x_i, y_i as base-points on the domain and range for f_i, rescale the metrics on the domain and the range by $\lambda_i = \frac{1}{r_i}$ and take the corresponding ultralimits. In the limit we get a bi-Lipschitz embedding

$$f_\omega : \mathbb{E}^n \to \mathbb{E}^n,$$

whose image misses the point $y_\omega \in \mathbb{E}^n$. However each bi-Lipschitz embedding of Euclidean spaces is necessarily proper, therefore, by the invariance of domain theorem, the image of f_ω is both closed and open. Contradiction. □

Remark 7.84. Alternatively, one can prove the above lemma (without using ultralimits) as follows: Approximate f by a continuous mapping g. Then, since g is proper, it has to be onto.

Corollary 7.85. *$\mathbb{E}^n$ is quasiisometric to $\mathbb{E}^m$ if and only if $n = m$.*

On the other hand, one cannot use ultralimits (at least directly) to prove that hyperbolic spaces of different dimensions are not quasiisometric to each other: All their ultralimits are isometric to the same universal real tree.

CHAPTER 8

Hyperbolic Space

The real hyperbolic space is the oldest and easiest example of hyperbolic spaces, which will be discussed in detail in Chapter 9. The real hyperbolic space has its origin in the following classical question that has challenged the geometers for nearly 2000 years:

QUESTION. *Does Euclid's fifth postulate follow from the rest of the axioms of Euclidean geometry? (The fifth postulate is equivalent to the statement that given a line L and a point P in the plane, there exists exactly one line through P parallel to L.)*

After a long history of unsuccessful attempts to establish a positive answer to this question, N.I. Lobachevski, J. Bolyai and C.F. Gauss independently (in the early 19th century) developed a theory of non-Euclidean geometry (which we now call "hyperbolic geometry"), where Euclid's fifth postulate is replaced by the axiom:

"For every point P which does not belong to L, there are infinitely many lines through P parallel to L."

Independence of the 5th postulate from the rest of the Euclidean axioms was proved by E. Beltrami in 1868, *via* a construction of a model of the hyperbolic geometry. In this chapter we will use the unit ball and the upper half-space models of hyperbolic geometry, the latter of which is due to H. Poincaré.

Given the classical nature of the subject, there are many books about real hyperbolic spaces, for instance, [**And05**], [**Bea83**], [**BP92**], [**Rat06**], [**Thu97**]. Our treatment of hyperbolic spaces is not meant to be comprehensive, we only cover the material needed elsewhere in the book. The purpose of this chapter is threefold:

1. It motivates many ideas and constructions in more general *Gromov-hyperbolic spaces*, which appear in the next chapter.

2. It provides the necessary geometric background for *lattices* in the isometry group $PO(n, 1)$ of the hyperbolic n-space. This background will be needed in the proof of various rigidity theorems for such lattices, which are due to Mostow, Tukia and Schwartz (Chapters 21 and 22).

3. We will use some basic hyperbolic geometry as a technical tool in proofs of a purely group-theoretic theorem, Stalling's theorem on ends of groups: Hyperbolic geometry appears in both proofs of this theorem given in the book, Chapters 18 and 19.

8.1. Moebius transformations

We will think of the sphere $\mathbb{S}^n$ as the 1-point compactification of the Euclidean n-space $\mathbb{E}^n$,

$$\mathbb{S}^n = \widehat{\mathbb{E}^n} = \mathbb{E}^n \cup \{\infty\}$$

Accordingly, we will regard the 1-point compactification of a hyperplane in $\mathbb{E}^n$ as a *round sphere* (of infinite radius) and the 1-point compactification of a line in $\mathbb{E}^n$ as a *round circle*. Another way to justify this treatment of hyperplanes and lines is that hyperplanes in $\mathbb{E}^n$ appear as Chabauty–limits of round spheres: Consider a sequence of round spheres $S(\mathbf{a}_i, R_i)$ in $\mathbb{E}^n$ passing through the origin ($|\mathbf{a}_i| = R_i$) with the sequence R_i diverging to infinity.

EXERCISE 8.1. Every sequence of spheres as above subconverges to a linear hyperplane in $\mathbb{R}^n$.

The *inversion* in the radius r sphere $\Sigma = S(\mathbf{0}, r) = \{\mathbf{x} : |\mathbf{x}| = r\}$ is the map

$$J_\Sigma : \mathbf{x} \mapsto r^2 \frac{\mathbf{x}}{|\mathbf{x}|^2}, \quad J_\Sigma(\mathbf{0}) = \infty, \quad J_\Sigma(\infty) = \mathbf{0}.$$

One defines the inversion J_Σ in the sphere $\Sigma = S(\mathbf{a}, r) = \{\mathbf{x} : |\mathbf{x} - \mathbf{a}| = r\}$ by the formula

$$J_\Sigma = T_\mathbf{a} \circ J_{S(\mathbf{0},r)} \circ T_{-\mathbf{a}}, \quad J_\Sigma(\mathbf{x}) = r^2 \frac{(\mathbf{x} - \mathbf{a})}{|\mathbf{x} - \mathbf{a}|^2} + \mathbf{a},$$

where $T_\mathbf{a}$ is the translation by the vector $\mathbf{a}$. Inversions map round spheres to round spheres and round circles to circles; inversions also preserve the Euclidean angles. We will regard the reflection in a Euclidean hyperplane as an inversion (this inversion fixes the point ∞). This is justified by:

EXERCISE 8.2. Suppose that the sequence of spheres $\Sigma_i = S(\mathbf{a}_i, r_i)$ converges to a linear hyperplane Π in $\mathbb{R}^n$. Show that the sequence of inversions J_i in the spheres Σ_i converges uniformly on compacts in $\mathbb{R}^n$ to the reflection in Π.

DEFINITION 8.3. A *Moebius transformation* of $\mathbb{E}^n$ (or, more precisely, of $\mathbb{S}^n$) is a composition of finitely many inversions in $\mathbb{E}^n$. The group of all Moebius transformations of $\mathbb{E}^n$ is denoted $Mob(\mathbb{E}^n)$ or $Mob(\mathbb{S}^n)$.

In particular, Moebius transformations preserve angles, send circles to circles and spheres to spheres.

For instance, every translation is a Moebius transformation, since it is the composition of two reflections in parallel hyperplanes. Every rotation in $\mathbb{E}^n$ is the composition of at most n inversions (reflections), since every rotation in $\mathbb{E}^2$ is the composition of two reflections. Every dilation $\mathbf{x} \mapsto \lambda\mathbf{x}, \lambda > 0$ is the composition of two inversions in spheres centered at 0. Thus, the group of *Euclidean similarities*

$$Sym(\mathbb{E}^n) = \{g : g(\mathbf{x}) = \lambda A\mathbf{x} + \mathbf{b}, \lambda > 0, A \in O(n), \mathbf{b} \in \mathbb{R}^n\},$$

is a subgroup of $Mob(\mathbb{S}^n)$.

The *cross-ratio* of a quadruple of points in $\mathbb{S}^n$ is defined as:

$$[\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}] := \frac{|\mathbf{x} - \mathbf{y}| \cdot |\mathbf{z} - \mathbf{w}|}{|\mathbf{y} - \mathbf{z}| \cdot |\mathbf{w} - \mathbf{x}|}.$$

Here and in what follows we assume, by default, that $\mathbf{y} \neq \mathbf{z}, \mathbf{x} \neq \mathbf{w}$. In the formula for the cross-ratio we use the chordal distance on the sphere (defined via the standard embedding of $\mathbb{S}^n$ in $\mathbb{E}^{n+1}$). Instead, we identify, via the stereographic projection, $\mathbb{S}^n$ with the extended Euclidean space $\widehat{\mathbb{E}^n} = \mathbb{E}^n \cup \{\infty\}$ and use Euclidean distances, provided that the points $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$ are not equal to the point ∞. Even if one of these points is ∞, we can define the cross-ratio by declaring that the two infinities appearing in the fraction defining the cross-ratio cancel each other.

This cross-ratio turns out to be equal to the one defined via the chordal metric (since the stereographic projection is the restriction of a Moebius transformation, see Example 8.6).

THEOREM 8.4. *1. A map $g : \mathbb{S}^n \to \mathbb{S}^n$ is a Moebius transformation if and only if it preserves cross-ratios of quadruples of points in $\mathbb{S}^n$.*

2. If a Moebius transformation g fixes the point ∞ in $\widehat{\mathbb{E}^n}$, then g is a Euclidean similarity.

We refer the reader to [**Rat06**, Theorems 4.3.1, 4.3.2] for a proof.

This theorem has an immediate corollary:

COROLLARY 8.5. *The subgroup $Mob(\mathbb{S}^n)$ is closed in the topological group $Homeo(\mathbb{S}^n)$, equipped with the topology of pointwise convergence.*

EXAMPLE 8.6. Let us construct a Moebius transformation σ sending the open unit ball $\mathbf{B}^n = B(0,1) \subset \mathbb{E}^n$ to the upper half-space $\mathbf{U}^n$,

$$\mathbf{U}^n = \{\mathbf{x} = (x_1, ...x_n) : x_n > 0\}.$$

We take σ to be the composition of translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{e}_n$, where $\mathbf{e}_n = (0, ..., 0, 1)$, inversion J_Σ, where $\Sigma = \partial \mathbf{B}^n$, translation $\mathbf{x} \mapsto \mathbf{x} - \frac{1}{2}\mathbf{e}_n$ and, lastly, the dilation $\mathbf{x} \to 2\mathbf{x}$. The reader will notice that the restriction of σ to the boundary sphere Σ of $\mathbf{B}^n$ is nothing but the stereographic projection with the pole at $-\mathbf{e}_n$.

Note that the map σ sends the origin $0 \in \mathbf{B}^n$ to the point $\mathbf{e}_n \in \mathbf{U}^n$.

Given a subset $A \subset \mathbb{S}^n$, we will use the notation $Mob(A)$ for the stabilizer of A in $Mob(\mathbb{S}^n)$.

EXERCISE 8.7. Each Moebius transformation $g \in Mob(\mathbf{B}^n)$ commutes with the inversion J in the boundary sphere of $\mathbf{B}^n$.

Low-dimensional Moebius transformations. Suppose now that $n = 2$. The group $SL(2, \mathbb{C})$ acts on the extended complex plane $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ by *linear-fractional transformations*:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}. \tag{8.1}$$

Note that the matrix $-I$ is in the kernel of this action, thus, the action factors through the group $PSL(2, \mathbb{C}) = SL(2, \mathbb{C})/\pm I$. If we identify the complex-projective line $\mathbb{CP}^1$ with the sphere $\mathbb{S}^2 = \mathbb{C} \cup \infty$ *via* the map $[z : w] \mapsto z/w$, this action of $SL(2, \mathbb{C})$ on $\mathbb{S}^2$ is nothing but the action of $SL(2, \mathbb{C})$ on $\mathbb{CP}^1$ obtained *via* projection of the linear action of $SL(2, \mathbb{C})$ on $\mathbb{C}^2 \setminus 0$.

EXERCISE 8.8. Show the group $PSL(2, \mathbb{C})$ acts faithfully on $\mathbb{S}^2$.

EXERCISE 8.9. Prove that the subgroup $SL(2, \mathbb{R}) \subset SL(2, \mathbb{C})$ preserves the upper half-plane $\mathbf{U}^2 = \{z : Im(z) > 0\}$. Moreover, $SL(2, \mathbb{R})$ is the stabilizer of $\mathbf{U}^2$ in $SL(2, \mathbb{C})$.

EXERCISE 8.10. Prove that each matrix in $SL(2, \mathbb{C})$ is either of the form

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$$

or it can be written as a product

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

Hint: If a matrix is not of the first type then it is a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $c \neq 0$. Use this information and multiplications on the left and on the right by matrices

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

to create zeroes on the diagonal in the matrix.

LEMMA 8.11. *$PSL(2, \mathbb{C})$ is the subgroup $Mob_+(\mathbb{S}^2)$ of Moebius transformations of $\mathbb{S}^2$ which preserve orientation.*

PROOF. 1. Every linear-fractional transformation is a composition of

$$j : z \mapsto z^{-1},$$

translations, dilations and rotations (see Exercise 8.10). Note that $j(z)$ is the composition of the complex conjugation with the inversion in the unit circle. Thus, $PSL(2, \mathbb{C}) \subset Mob_+(\mathbb{S}^2)$. Conversely, let $g \in Mob(\mathbb{S}^2)$ and $z_0 := g(\infty)$. Then $h = j \circ \tau \circ g$ fixes the point ∞, where $\tau_0(z) = z - z_0$. Let $z_1 = h(0)$. Then composition f of h with the translation $\tau_1 : z \mapsto z - z_1$ has the property that $f(\infty) = \infty, f(0) = 0$. Thus, $f \in CO(2)$ and h preserves orientation. It follows that f has the form $f(z) = \lambda z$, for some $\lambda \in \mathbb{C} \setminus 0$. Since f, $\tau_0, \tau{-}1, j$ are linear-fractional transformation, it follows that g is also linear-fractional. ☐

EXERCISE 8.12. Show that the group $Mob(\mathbb{S}^1)$ equals the group of real-linear fractional transformations

$$x \mapsto \frac{ax + b}{cx + d},$$

$ad - bc \neq 0, a, b, c, d \in \mathbb{R}$.

8.2. Real hyperbolic space

The easiest way to introduce the real-hyperbolic n-space $\mathbb{H}^n$ is by using its models: Upper half-space, unit ball and the projectivization of the two-sheeted hyperboloid in the Lorentzian model. Different features of $\mathbb{H}^n$ are best visible in different models.

Upper half-space model. We equip $\mathbf{U}^n$ with the Riemannian metric

$$ds^2 = \frac{d\mathbf{x}^2}{x_n^2} = \frac{dx_1^2 + ... + dx_n^2}{x_n^2} \tag{8.2}$$

The Riemannian manifold $(\mathbf{U}^n, ds^2)$ is called the *n-dimensional hyperbolic space* and denoted $\mathbb{H}^n$. This space is also frequently called the *real-hyperbolic* space, in order to distinguish it from other spaces also called *hyperbolic* (e.g., complex-hyperbolic space, quaternionic-hyperbolic space, Gromov-hyperbolic space, etc.). We will use the terminology *hyperbolic space* for $\mathbb{H}^n$ and add adjective *real* in case when other notions of hyperbolicity are involved in the discussion. In case $n = 2$, we identify $\mathbb{R}^2$ with the complex plane, so that $\mathbf{U}^2 = \{z|Im(z) > 0\}$, $z = x + iy$, and

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

Note that the hyperbolic Riemannian metric ds^2 on $\mathbf{U}^n$ is conformally-Euclidean, hence, hyperbolic angles are equal to the Euclidean angles. One computes hyperbolic volumes of solids in $\mathbb{H}^n$ by the formula

$$Vol(\Omega) = \int_\Omega \frac{dx_1 ... dx_n}{x_n^n}$$

Consider the projection to the x_n-axis in $\mathbf{U}^n$ given by the formula

$$\pi : (x_1, ..., x_n) \mapsto (0, ..., 0, x_n).$$

EXERCISE 8.13. 1. Verify that $d_x\pi$ does not increase the length of tangent vectors $\mathbf{v} \in T_x\mathbb{H}^n$ for every $x \in \mathbb{H}^n$.

2. Verify that for a unit vector $\mathbf{v} \in T_x\mathbb{H}^n$, $\|d_x\pi(\mathbf{v})\| = 1$ if and only if $\mathbf{v}$ is "vertical", i.e., it has the form $(0, ..., 0, v_n)$.

Here and in what follows, the norm $\|\cdot\|$ is the one with respect to the hyperbolic Riemannian metric on the tangent spaces to $\mathbb{H}^n$; the notation $|\cdot|$ is reserved for the Euclidean norm.

EXERCISE 8.14. Suppose that $\mathbf{p} = a\mathbf{e}_n, \mathbf{q} = b\mathbf{e}_n$, where $0 < a < b$. Let α be the vertical path $\alpha(t) = (1-t)\mathbf{p} + t\mathbf{q}$, $t \in [0,1]$ connecting $\mathbf{p}$ to $\mathbf{q}$. Show that α is the shortest path (with respect to the hyperbolic metric) connecting $\mathbf{p}$ to $\mathbf{q}$ in $\mathbb{H}^n$. In particular, α is a hyperbolic geodesic and

$$d(\mathbf{p}, \mathbf{q}) = \log(b/a).$$

Hint: Use the previous exercise.

We note that the metric ds^2 on $\mathbb{H}^n$ is clearly invariant under the "horizontal" Euclidean translations $\mathbf{x} \mapsto \mathbf{x} + \mathbf{v}$, where $\mathbf{v} = (v_1, ..., v_{n-1}, 0)$ (since they preserve the Euclidean metric and the x_n-coordinate). Similarly, ds^2 is invariant under the dilations

$$h : \mathbf{x} \mapsto \lambda\mathbf{x}, \lambda > 0$$

since h scales both numerator and denominator in (8.2) by λ^2. Lastly, ds^2 is invariant under Euclidean rotations which fix the x_n-axis (since they preserve the x_n-coordinate). Clearly, the group generated by such isometries of $\mathbb{H}^n$ act transitively on $\mathbb{H}^n$, which means that $\mathbb{H}^n$ is a *homogeneous Riemannian manifold.*

EXERCISE 8.15. Show that $\mathbb{H}^n$ is a complete Riemannian manifold. You can either use homogeneity of $\mathbb{H}^n$ or show directly that every Cauchy sequence in $\mathbb{H}^n$ lies in a compact subset of $\mathbb{H}^n$.

EXERCISE 8.16. Show that the inversion $J = J_\Sigma$ in the unit sphere Σ centered at the origin, is an isometry of $\mathbb{H}^n$. The proof is easy but (somewhat) tedious calculation, which is best done using *calculus* interpretation of the pull-back Riemannian metric.

EXERCISE 8.17. Show that every inversion preserving $\mathbb{H}^n$ is an isometry of $\mathbb{H}^n$. To prove this, use compositions of the inversion J_Σ in the unit sphere with translations and dilations.

In order to see clearly other isometries of $\mathbb{H}^n$, it is useful to consider the *unit ball model* of the hyperbolic space.

Unit ball model. Consider the open unit Euclidean n-ball

$$\mathbf{B}^n := \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < 1\}$$

in $\mathbb{E}^n$. We equip $\mathbf{B}^n$ with the Riemannian metric

$$ds^2_{\mathbf{B}} = 4\frac{dx_1^2 + ... + dx_n^2}{(1 - |\mathbf{x}|^2)^2}.$$

The Riemannian manifold $(\mathbf{B}^n, ds^2)$ is called the *unit ball model* of the hyperbolic n-space. What is clear in this model is that the group $O(n)$ of orthogonal transformations of $\mathbb{R}^n$ preserves $ds^2_{\mathbf{B}}$ (since its elements preserve $|x|$ and, hence, the denominator of $ds^2_{\mathbf{B}}$). The two models of the hyperbolic space are related by the Moebius transformation $\sigma : \mathbf{B}^n \to \mathbf{U}^n$ defined in the previous section.

Exercise 8.18. Show that $ds^2_{\mathbf{B}} = \sigma^*(ds^2)$. The proof is again a straightforward calculation similar to the Exercise 8.16. Namely, first, pull-back ds^2 *via* the dilation $\mathbf{x} \to 2\mathbf{x}$, then apply pull-back *via* the translation $\mathbf{x} \mapsto \mathbf{x} - \frac{1}{2}\mathbf{e}_n$, etc. Thus, σ is an isometry of the Riemannian manifolds $(\mathbf{B}^n, ds^2_{\mathbf{B}}), (\mathbf{U}^n, ds^2)$.

Lemma 8.19. *The group $O(n)$ is the stabilizer of $\mathbf{0}$ in the group of isometries of $(\mathbf{B}^n, ds^2_{\mathbf{B}})$.*

Proof. Note that if $g \in \mathrm{Isom}(\mathbf{B}^n)$ fixes $\mathbf{0}$, then its derivative at the origin $dg_{\mathbf{0}}$ is an orthogonal transformation u. Thus, the derivative (at the origin) of the composition $h = u^{-1}g \in \mathrm{Isom}(\mathbf{B}^n)$ is the identity. Therefore, for every geodesic γ in $\mathbb{H}^n$ such that $\gamma(0) = 0$, $D_0 h(\gamma'(0)) = \gamma'(0)$. Since each geodesic in a Riemannian manifold is uniquely determined by its initial point and initial velocity, we conclude that $h(\gamma(t)) = \gamma(t)$ for every t. By completeness of $\mathbb{H}^n$, for every $q \in \mathbf{B}^n$ there exists a geodesic γ connecting p to q. It follows that $h(q) = q$ and, therefore, $g = u \in O(n)$. □

Corollary 8.20. *The stabilizer of the point $\mathbf{e}_n \in \mathbf{U}^n$ in the group $\mathrm{Isom}(\mathbb{H}^n)$ is contained in the group of Moebius transformations.*

Proof. Note that σ sends $\mathbf{0} \in \mathbf{B}^n$ to $\mathbf{e}_n \in \mathbf{U}^n$, and σ is Moebius. Thus, $\sigma : \mathbf{B}^n \to \mathbf{U}^n$ conjugates the stabilizer $O(n)$ of $\mathbf{0}$ in $\mathrm{Isom}(\mathbf{B}^n, ds^2_{\mathbf{B}})$ to the stabilizer $K = \sigma^{-1}O(n)\sigma$ of $\mathbf{e}_n$ in $\mathrm{Isom}(\mathbf{U}^n, ds^2)$. Since $O(n) \subset Mob(\mathbb{S}^n), \sigma \in Mob(\mathbb{S}^n)$, the claim follows. □

Corollary 8.21. *a.* $\mathrm{Isom}(\mathbb{H}^n)$ *equals the group $Mob(\mathbb{H}^n)$ of Moebius transformations of $\mathbb{S}^n$ preserving $\mathbb{H}^n$.*

b. $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on the unit tangent bundle $U\mathbb{H}^n$ of $\mathbb{H}^n$.*

Proof. a. Since the two models of $\mathbb{H}^n$ differ by a Moebius transformation, it suffices to work with $\mathbf{U}^n$.

1. We already know that the $\mathrm{Isom}(\mathbb{H}^n) \cap Mob(\mathbb{H}^n)$ contains a subgroup acting transitively on $\mathbb{H}^n$. We also know, that the stabilizer K of p in $\mathrm{Isom}(\mathbb{H}^n)$ is contained in $Mob(\mathbb{H}^n)$. Thus, given $g \in \mathrm{Isom}(\mathbb{H}^n)$ we first find

$$h \in Mob(\mathbb{H}^n) \cap \mathrm{Isom}(\mathbb{H}^n)$$

such that $k = h \circ g(p) = p$. Since $k \in Mob(\mathbb{H}^n)$, we conclude that $\mathrm{Isom}(\mathbb{H}^n) \subset Mob(\mathbb{H}^n)$.

2. We leave it to the reader to verify that the restriction homomorphism $Mob(\mathbb{H}^n) \to Mob(\mathbb{S}^{n-1})$ is injective. Every $g \in Mob(\mathbb{S}^{n-1})$ extends to a composition of inversions preserving $\mathbb{H}^n$. Thus, the above restriction map is a group isomorphism. We already know that inversions $J \in Mob(\mathbb{H}^n)$ are hyperbolic isometries. Thus, $Mob(\mathbb{H}^n) \subset \mathrm{Isom}(\mathbb{H}^n)$.

b. Transitivity of the action of $\mathrm{Isom}(\mathbb{H}^n)$ on $U\mathbb{H}^n$ follows from the fact that this group acts transitively on $\mathbb{H}^n$ and that the stabilizer of p acts transitively on the set of unit vectors in $T_p\mathbb{H}^n$. □

For the next lemma we recall that we treat straight lines as circles.

LEMMA 8.22. *Geodesics in $\mathbb{H}^n$ are arcs of circles orthogonal to the boundary sphere of $\mathbb{H}^n$. Furthermore, for every such arc α in $\mathbf{U}^n$, there exists an isometry of $\mathbb{H}^n$ which carries α to a segment of the x_n-axis.*

PROOF. It suffices to consider complete hyperbolic geodesics $\alpha : \mathbb{R} \to \mathbb{H}^n$. Since $\sigma : \mathbf{B}^n \to \mathbf{U}^n$ sends circles to circles and preserves angles, it again suffices to work with the upper half-space model. Let α be a hyperbolic geodesic in $\mathbf{U}^n$. Since $\mathrm{Isom}(\mathbb{H}^n)$ acts transitively on $U\mathbb{H}^n$, there exists a hyperbolic isometry g such that the hyperbolic geodesic $\beta = g \circ \alpha$ satisfies: $\beta(0) = p = \mathbf{e}_n$ and the vector $\beta'(0)$ has the form $\mathbf{e}_n = (0, ..., 0, 1)$. We already know that the curve

$$\gamma : t \mapsto e^t \mathbf{e}_n$$

is a hyperbolic geodesic, see Exercise 8.14. Furthermore, $\gamma'(0) = \mathbf{e}_n$ and $\gamma(0) = p$. Thus, $\beta = \gamma$ is a circle orthogonal to the boundary of $\mathbb{H}^n$. Since $\mathrm{Isom}(\mathbb{H}^n) = Mob(\mathbb{H}^n)$ and Moebius transformations map circles to circles and preserve angles, lemma follows. □

COROLLARY 8.23. *The space $\mathbb{H}^n$ is* uniquely geodesic*, i.e., for every pair of points in $\mathbb{H}^n$ there exists a unique unit speed geodesic segment connecting these points.*

PROOF. By the above lemma, it suffices to consider points p, q on the x_n-axis. But, according to Exercise 8.14, the vertical segment is the unique length-minimizing path between such p and q. □

COROLLARY 8.24. *Let $H \subset \mathbb{H}^n$ be the intersection of $\mathbb{H}^n$ with a round k-sphere orthogonal to the boundary of $\mathbb{H}^n$. Then H is a* totally-geodesic *subspace of $\mathbb{H}^n$, i.e., for every pair of points $p, q \in H$, the unique hyperbolic geodesic γ connecting p and q in $\mathbb{H}^n$, is contained in H. Furthermore, if $\iota : H \to \mathbb{H}^n$ is the embedding, then the Riemannian manifold $(H, \iota^* ds^2)$ is isometric to $\mathbb{H}^k$.*

PROOF. The first assertion follows from the description of geodesics in $\mathbb{H}^n$. To prove the second assertion, by applying an appropriate isometry of $\mathbb{H}^n$, it suffices to consider the case when H is contained in a coordinate k-dimensional subspace in $\mathbb{R}^n$:

$$H = \{(0, ..., 0, x_{n-k+1}, .., x_n) : x_n > 0\}.$$

Then

$$\iota^* ds^2 = \frac{dx_{n-k+1}^2 + ... + dx_n^2}{x_n^2}$$

is isometric to the hyperbolic metric on $\mathbb{H}^k$ (by relabeling the coordinates). □

We will refer to the submanifolds $H \subset \mathbb{H}^n$ as *hyperbolic subspaces.*

EXERCISE 8.25. Show that the hyperbolic plane violates the 5th Euclidean postulate: For every (geodesic) line $L \subset \mathbb{H}^2$ and every point $P \notin L$, there are infinitely many lines through P which are parallel to L (i.e., disjoint from L).

EXERCISE 8.26. Prove that:

- The unit sphere $\mathbb{S}^{n-1}$ (with its standard topology) is the ideal boundary (in the sense of Definition 2.73) of the hyperbolic space $\mathbb{H}^n$ in the unit ball model.
- The extended Euclidean space $\widehat{\mathbb{E}^n} = \mathbb{S}^n$ is the ideal boundary of the hyperbolic space $\mathbb{H}^{n+1}$ in the upper half-space model.

Note that the Moebius transformation $\sigma : \mathbf{B}^n \to \mathbf{U}^n$ carries the ideal boundary of $\mathbf{B}^n$ to the ideal boundary of $\mathbf{U}^n$. Observe also that all Moebius transformations which preserve $\mathbb{H}^n$ in either model, induce Moebius transformations of the ideal boundary of $\mathbb{H}^n$.

It follows from Corollaries 8.24 and 8.53 that $\mathbb{H}^n$ has sectional curvature -1, therefore all the considerations in §2.11.1, in particular those concerning the ideal boundary, apply to it. Later on, in §9.11 (Chapter 9), we will give yet another definition of ideal boundaries, for hyperbolic metric spaces in the sense of Gromov.

Lorentzian model of $\mathbb{H}^n$. We refer the reader to [**Rat06**] and [**Thu97**] for the material below.

Consider the *Lorentzian space* $\mathbb{R}^{n,1}$, which is $\mathbb{R}^{n+1}$ equipped with the indefinite nondegenerate quadratic form

$$Q(\mathbf{x}) = x_1^2 + \ldots + x_n^2 - x_{n+1}^2,$$

which is the quadratic form of the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i - x_{n+1} y_{n+1}.$$

Let H denote the upper sheet of the 2-sheeted hyperboloid in $\mathbb{R}^{n,1}$:

$$Q(\mathbf{x}) = -1, \quad x_{n+1} > 0.$$

The restriction of Q to the tangent bundle of H is positive-definite and, hence, defines a Riemannian metric ds^2 on H. We identify the unit ball $\mathbf{B}^n$ in $\mathbb{R}^n$ with the ball

$$\{(x_1, \ldots, x_n, 0) : x_1^+ \ldots + x_n^2 < 1\} \subset \mathbb{R}^{n+1}$$

via the inclusion $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$,

$$(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, 0).$$

Let $\pi : H \to \mathbf{B}^n$ denote the radial projection from the point $-\mathbf{e}_{n+1}$:

$$\pi(\mathbf{x}) = tx - (1-t)\mathbf{e}_{n+1}, \quad t = \frac{1}{x_{n+1} + 1}.$$

One then verifies that

$$\pi : (H, ds^2) \to \mathbb{H}^n = \left(\mathbf{B}^n, \frac{4d\mathbf{x}^2}{(1-|x|^2)^2}\right)$$

is an isometry. Accordingly, intersections of H with k-dimensional linear subspaces of $\mathbb{R}^{n+1}$ are k-dimensional hyperbolic subspaces of $\mathbb{H}^n$.

Instead of working with the upper sheet H of the hyperboloid $\{Q = -1\}$ it is sometimes convenient to work with the projectivization of this hyperboloid or, equivalently, of the open cone

$$\{Q(\mathbf{x}) < 0\}.$$

Then the stabilizer $O(n,1)^+$ of H in $O(n,1)$ is naturally isomorphic to the quotient $PO(n,1) = O(n,1)/\pm I$. The stabilizer of H in $O(n,1)$ acts isometrically on H. Furthermore, this stabilizer is the entire isometry group of (H, ds^2).

Thus, $\mathrm{Isom}(\mathbb{H}^n) \cong PO(n,1) \cong O(n,1)^+ < O(n,1)$; in particular, the Lie group $\mathrm{Isom}(\mathbb{H}^n)$ is linear.

EXERCISE 8.27. $O(n,1)^+$ is isomorphic to the group $SO(n,1) = O(n,1) \cap SL(n+1,\mathbb{R})$.

The distance function in $\mathbb{H}^n$ in terms of the Lorentzian inner product is given by the formula:

$$\cosh d(\mathbf{x},\mathbf{y}) = -\langle \mathbf{x},\mathbf{y}\rangle\,, \tag{8.3}$$

which is a direct analogue of the familiar formula for the angular metric on the unit sphere in terms of the Euclidean inner product. In order to see this, it suffices to consider the 1-dimensional hyperbolic space $\mathbb{H}^1$ identified with the hyperbola $x_1^2 - x_2^2 = -1, x_2 > 0$, in $\mathbb{R}^{1,1}$. Thus hyperbola is parameterized as

$$\mathbf{x}(t) = (\sinh(t), \cosh(t)), \quad t \in \mathbb{R}.$$

It is immediate from the definition of the induced Riemannian metric on $\mathbb{H}^1$ that this is an isometric parameterization of $\mathbb{H}^1$ and, hence,

$$t = \mathrm{dist}(\mathbf{e}_2, \mathbf{x}), \quad \mathbf{x} = \mathbf{x}(t).$$

Lastly,

$$\langle \mathbf{e}_2, \mathbf{x}\rangle = -\cosh(t).$$

The general case follows from transitivity of the isometry group of $\mathbb{H}^n$.

EXERCISE 8.28 (Rigidity of n-point configurations). Every n-tuple of points $(p_1, \ldots, p_n)$ in $\mathbb{H}^n$ is uniquely determined, up to an isometry of $\mathbb{H}^n$, by their mutual distances

$$\mathrm{dist}(p_i, p_j), \quad i < j.$$

In particular, a geodesic triangle in $\mathbb{H}^n$ is uniquely determined (up to congruence) by its side-lengths. Hint: Use the distance formula (8.3) and the fact that a quadratic form is uniquely determined (up to an isometry) by its Gramm matrix.

The Lorentzian model of $\mathbb{H}^n$ is a luxury one has in studying real-hyperbolic spaces, as the unit ball and the upper half-space models work just fine. However, *linear models* become a necessity when dealing with other hyperbolic spaces, complex-hyperbolic and quaternionic ones (see §8.9), as the unit ball and upper half-spaces models for such spaces become much more awkward to use.

8.3. Classification of isometries

Every continuous map of a closed disk to itself has a fixed point. Since every isometry of $\mathbb{H}^n$ (in the unit ball model) extends to a Moebius transformation of the closed ball $\mathbb{D}^n$, isometries of the hyperbolic space are classified by their fixed points in $\mathbb{D}^n$.

Definition 8.29. An isometry $g \in \mathrm{Isom}(\mathbb{H}^n)$ is *elliptic* if it fixes a point $x \in \mathbb{H}^n$.

Conjugating an elliptic isometry g (fixing $x \in \mathbb{H}^n$) by an isometry $h \in \mathrm{Isom}(\mathbb{H}^n)$, sending x to the center of the ball $\mathbf{B}^n$, we obtain another elliptic isometry

$$f = hgh^{-1}$$

which fixes the center of $\mathbf{B}^n$. Since f commutes with the inversion J in the unit sphere $\mathbb{S}^{n-1}$, we obtain:

$$f(\infty) = JfJ(\infty) = Jf(0) = J(0) = \infty.$$

Therefore, in view of Theorem 8.4, we conclude that f has to be a Euclidean similarity fixing the origin and preserving the unit ball $\mathbf{B}^n$. Such f is necessarily an orthogonal transformation, an element of the orthogonal group $O(n)$. We obtain:

Lemma 8.30. *An element $g \in \mathrm{Isom}(\mathbb{H}^n) = Mob(\mathbf{B}^n)$ is elliptic if and only if g is conjugate in $\mathrm{Isom}(\mathbb{H}^n)$ to an orthogonal transformation.*

Suppose that a Moebius transformation g of the boundary sphere $\mathbb{S}^{n-1}$ fixes three distinct points $\xi_1, \xi_2, \xi_3 \in \mathbb{S}^{n-1}$. Let C denote the unique round circle through these three points. The circle C appears as the boundary circle of a unique hyperbolic plane $\mathbb{H}^2 \subset \mathbb{H}^n$. Since g fixes the points ξ_1, ξ_2, ξ_3, it has to preserve C and, hence, $\mathbb{H}^2$. Furthermore, g preserves the hyperbolic geodesic $\gamma \subset \mathbb{H}^2$ asymptotic to ξ_1, ξ_2. There exists a unique horoball $B \subset \mathbb{H}^2$ centered at ξ_3, whose boundary touches the geodesic γ; we let $x \in \gamma$ denote this point of tangency. By combining these observations, we conclude that g fixes the point x and is, therefore, elliptic. Moreover, we also see that g fixes two linearly independent vectors $v_1, v_3 \in T_x\mathbb{H}^2$: These are the initial velocity vectors of the geodesic rays ρ_1, ρ_3 emanating from x and asymptotic to ξ_1, ξ_2 respectively. Therefore, g fixes x and acts as the identity map on the tangent plane $T_x\mathbb{H}^2$.

Exercise 8.31. Use these facts to conclude that the isometry g fixes the hyperbolic plane $\mathbb{H}^2$ and the circle C pointwise. Alternatively, argue that a linear–fractional transformation fixing three points in C is the identity map.

We, thus, obtain:

Lemma 8.32. *Each isometry of $\mathbb{H}^n$ fixing at least three points in the boundary sphere $\mathbb{S}^{n-1}$ is elliptic and, moreover, fixes pointwise a hyperbolic plane in $\mathbb{H}^n$.*

Of course, elliptic isometries need not fix any points in $\mathbb{S}^{n-1}$, for instance, the antipodal map

$$\mathbf{x} \mapsto -\mathbf{x}, \quad \mathbf{x} \in \mathbf{B}^n$$

is an elliptic isometry which has unique fixed point in $\mathbb{D}^n$. Another example to keep in mind is that each rotation $g \in SO(3)$ is an elliptic isometry of $\mathbb{H}^3 = \mathbf{B}^3$, which has exactly two fixed points in the boundary sphere.

In view of Lemma 8.32, in order to classify non-elliptic isometries, we have to consider isometries with one or two fixed points in $\mathbb{S}^{n-1}$.

Definition 8.33. An isometry g of $\mathbb{H}^n$ is called *parabolic* if it has exactly one fixed point in the boundary sphere $\mathbb{S}^{n-1}$.

Note that such an isometry cannot be elliptic, since a fixed point $x \in \mathbb{H}^n$ together with a fixed point $\xi \in \mathbb{S}^{n-1}$ determine a unique geodesic $\gamma \subset \mathbb{H}^n$ through x asymptotic to ξ. Therefore, an isometry g fixing both x and ξ also fixes the entire geodesic γ and, hence, the second ideal boundary point $\hat{\xi} \in \mathbb{S}^{n-1}$ of γ.

It is now convenient to switch from the unit ball model to the upper half-space model $\mathbf{U}^n$; we choose a Moebius transformation $h : \mathbf{B}^n \to \mathbf{U}^n$ which sends the fixed point ξ of g to the point ∞ in $\widehat{E^n}$. Conjugating g via h, we obtain a parabolic isometry

$$f = ghg^{-1}$$

whose unique fixed point is ∞. Such f has to act as a Euclidean similarity on $\mathbb{E}^{n-1}$ which has no fixed points in $\mathbb{E}^{n-1}$.

EXERCISE 8.34. Suppose that $f \in Sim(\mathbb{E}^{n-1})$ has no fixed points in $\mathbb{E}^{n-1}$. Then f has the form

$$f(\mathbf{x}) = A\mathbf{x} + \mathbf{b},$$

with $A \in O(n-1)$.

A Euclidean isometry $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ is called a *Euclidean skew motion* with the translational component b if the vector b is nonzero and is fixed by A. Note that we allow Euclidean translations as special cases as skew motions (with the identity orthogonal component A).

EXERCISE 8.35. 1. Suppose that $f(x) = A\mathbf{x}+\mathbf{b}$ is a Euclidean isometry without fixed points in $\mathbb{E}^{n-1}$. Then f is conjugate by a translation in $\mathbb{R}^n$ to a Euclidean skew motion.

2. Conversely, Euclidean skew motions have no fixed points in $\mathbb{E}^{n-1}$.

To summarize:

LEMMA 8.36. *An isometry of $\mathbb{H}^n$ is parabolic if and only if it is conjugate in $Mob(\mathbb{S}^n)$ to a Euclidean skew motion.*

The last class of isometries of $\mathbb{H}^n$ consists of *hyperbolic isometries*. Each hyperbolic isometry g has exactly two fixed points $\xi, \hat{\xi}$ in the boundary sphere $\mathbb{S}^{n-1}$. In order to distinguish such isometries from elliptic isometries, consider the unique geodesic γ in $\mathbb{H}^n$ asymptotic to the points $\xi, \hat{\xi}$. This geodesic has to be preserved by g. Therefore, g induces an isometry $\gamma \to \gamma$. The isometry group of $\mathbb{R}$ consists of three types of elements:

1. The identity map.
2. Reflections, $R_a : x \mapsto a - x$, $a \in \mathbb{R}$.
3. Nontrivial translations $x \mapsto x + b$, $b \in \mathbb{R} \setminus \{0\}$.

It is clear that if g induces an isometry of type 1 or 2 of the geodesic γ, then g is necessarily elliptic. This leads to:

DEFINITION 8.37. An isometry $g \in \mathrm{Isom}(\mathbb{H}^n)$ is *hyperbolic* if it preserves a geodesic $\gamma \subset \mathbb{H}^n$ and acts on this geodesic as a nonzero Euclidean translation $x \mapsto x + b$. The number b is called the *translation number* τ_g of g. The geodesic γ is called the *axis* of g.

EXERCISE 8.38. Show that each hyperbolic isometry has unique axis. Hint: Assuming that g has two distinct axes, consider the action of g on their ideal boundary points.

Note that g, of course, fixes the ideal points $\xi, \hat{\xi} \in \mathbb{S}^{n-1}$ of its γ. One can distinguish g from an elliptic isometry fixing these points by noting that g is hyperbolic if and only it its derivative at these points *is not an orthogonal transformation*.

EXERCISE 8.39. Prove this characterization of hyperbolic isometries in terms of their derivatives. Hint: First consider the case when g fixes 0 and ∞, and consider the derivative at the origin. Then reduce the general case to this one.

As with the elliptic and parabolic isometries, we can conjugate each hyperbolic isometry g to a Euclidean similarity, by sending (via a Moebius transformation $h : \mathbf{B}^n \to \mathbf{U}^n$) the fixed points $\xi, \hat{\xi}$ to 0 and ∞ respectively. The conjugate Moebius transformation

$$f = hgh^{-1}$$

has the form

$$f(\mathbf{x}) = \lambda A\mathbf{x}, \quad A \in O(n-1), \quad \lambda > 0, \quad \lambda \neq 1.$$

The translation number τ_g equals

$$\tau_g = |\log(\lambda)|,$$

since

$$\mathrm{dist}(\mathbf{e}_n, \lambda\mathbf{e}_n) = |\log(\lambda)|.$$

In the case when $n = 3$ and we can identify $\mathrm{Isom}_+(\mathbb{H}^n)$ with the group $PSL(2, \mathbb{C})$, one can give a simple *numerical* characterization of elliptic, parabolic and hyperbolic isometries:

Suppose that g is an orientation-preserving Moebius transformation of $\mathbb{C}$, represented by the matrices $\pm A$,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL(2, \mathbb{C}).$$

We assume that $A \neq \pm I$, i.e., g is not the identity map (in which case, g is, of course, elliptic).

EXERCISE 8.40. 1. g is elliptic iff $\mathrm{tr}(A) \in (-2, 2) \subset \mathbb{C}$.
2. g is parabolic iff $\mathrm{tr}(A) = \pm 2$.
3. g is hyperbolic iff $\mathrm{tr}(A) \notin [-2, 2]$.
Hint: Use the fact that each $g \in PSL(2, \mathbb{C})$ is conjugate to a Euclidean similarity.

Lastly, we note that the elliptic-parabolic-hyperbolic classification of isometries can be generalized in the context of $CAT(-1)$ spaces X. Instead of relying upon the (unavailable) fixed-point theorem for general continuous maps, one classifies isometries g of X according to their *translation numbers*:

$$\tau_g = \inf_{x \in X} d(x, gx).$$

- An isometry g is elliptic if $\tau_g = 0$ and the infimum in the definition of τ_g is realized in X.
- An isometry g is elliptic if $\tau_g = 0$ and the infimum in the definition of τ_g is not realized in X.
- An isometry g is hyperbolic if $\tau_g > 0$.

EXERCISE 8.41. Show that the classification of isometries of $\mathbb{H}^n$ described in this section is equivalent to their classification via translation numbers.

8.4. Hyperbolic trigonometry

In this section we consider geometry of triangles in the hyperbolic plane. We refer to [**Bea83, Rat06, Thu97**] for the proofs of the hyperbolic trigonometric formulae introduced in this section. Recall that a (geodesic) triangle $T = T(A, B, C)$ as a 1-*dimensional object*. From the Euclidean viewpoint, a hyperbolic triangle T is a concatenations of circular arcs connecting points A, B, C in $\mathbb{H}^2$, where the circles containing the arcs are orthogonal to the boundary of $\mathbb{H}^2$. Besides such "conventional" triangles, it is useful to consider *generalized* hyperbolic triangles where some vertices are *ideal*, i.e., they belong to the (ideal) boundary circle of $\mathbb{H}^2$. Such triangles are easiest to introduce by using the Euclidean interpretation of hyperbolic triangles: One simply allows some (or, even all) vertices A, B, C to be points on the boundary circle of $\mathbb{H}^2$, the rest of the definition is exactly the same. *However, we no longer allow two vertices which belong to the boundary circle* $\mathbb{S}^1$ *to be the same.* More intrinsically, an triangle $T(A, B, C)$, where, say, B and C are in $\mathbb{H}^2$ and $A \in \mathbb{S}^1$ is the concatenation of the geodesic arc BC and geodesic rays CA and BA (although, the natural orientation of the latter is from A to B).

The vertices of T which happen to be points of the boundary circle $\mathbb{S}^1$ are called the *ideal vertices* of T. The *angle* of T at its ideal vertex A is just the Euclidean angle, which has to be zero, since both sides of T at A are orthogonal to the ideal boundary circle $\mathbb{S}^1$.

In general, we will use the notation $\alpha = \angle_A(B, C)$ to denote the angle of T at A. From now on, a *hyperbolic triangle* means either a usual triangle or a triangle where some vertices are ideal. We still refer to such triangles as *triangles in* $\mathbb{H}^2$, even though, some of the vertices could lie on the ideal boundary, so, strictly speaking, an ideal hyperbolic triangle in $\mathbb{H}^2$ is not a subset of $\mathbb{H}^2$. An *ideal* hyperbolic triangle, is a triangle where all the vertices are distinct ideal points in $\mathbb{H}^2$. The same conventions will be used for hyperbolic triangles in $\mathbb{H}^n$.

1. General triangles. Consider hyperbolic triangles T in $\mathbb{H}^2$ with the side-lengths a, b, c and the opposite angles α, β, γ, see Figure 8.1.

a. Hyperbolic Sine Law:

$$\frac{\sinh(a)}{\sin(\alpha)} = \frac{\sinh(b)}{\sin(\beta)} = \frac{\sinh(c)}{\sin(\gamma)}. \tag{8.4}$$

b. Hyperbolic Cosine Law:

$$\cosh(c) = \cosh(a)\cosh(b) - \sinh(a)\sinh(b)\cos(\gamma) \tag{8.5}$$

c. Dual Hyperbolic Cosine Law:

$$\cos(\gamma) = -\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)\cosh(c) \tag{8.6}$$

2. Right triangles. Consider a right-angled hyperbolic triangle with the hypotenuse c, the other side-lengths a, b and the opposite angles α, β. Then, hyperbolic cosine laws become:

$$\cosh(c) = \cosh(a)\cosh(b), \tag{8.7}$$

$$\cos(\alpha) = \sin(\beta)\cosh(a), \tag{8.8}$$

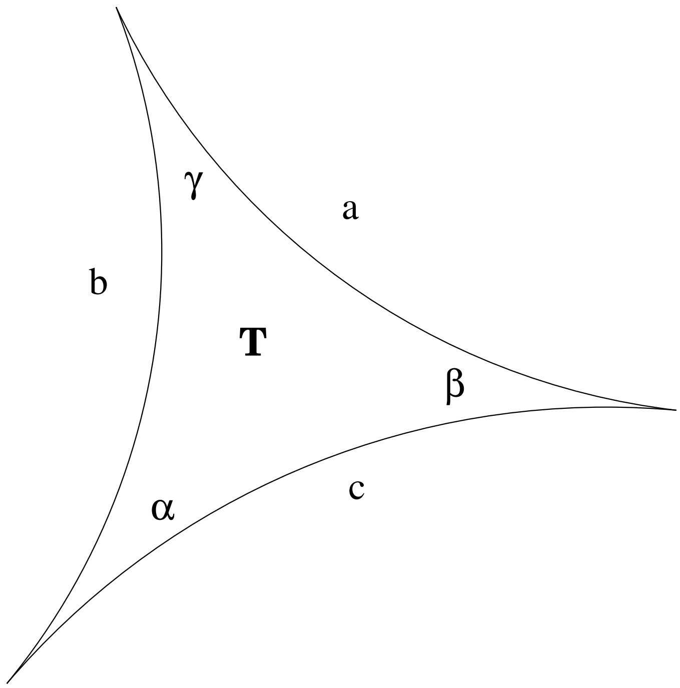


FIGURE 8.1. Geometry of a general hyperbolic triangle.

$$\cos(\alpha) = \frac{\tanh b}{\tanh c} \tag{8.9}$$

In particular,

$$\cos(\alpha) = \frac{\cosh(a)\sinh(b)}{\sinh(c)}. \tag{8.10}$$

3. First variation formula for right triangles. We now hold the side a fixed and vary the hypotenuse in the above right-angled triangle. By combining (8.7) and (8.5) we obtain the *First Variation Formula*:

$$c'(0) = \frac{\cosh(a)\sinh(b)}{\sinh(c)} b'(0) = \cos(\alpha) b'(0). \tag{8.11}$$

The equation $c'(0) = \cos(\alpha)b'(0)$ is a special case of the *First Variation Formula* in Riemannian geometry, which applies to general Riemannian manifolds.

As an application of the first variation formula, consider a hyperbolic triangle with vertices A, B, C, side-lengths a, b, c and the angles β, γ opposite to the sides b, c. Then

LEMMA 8.42. *$a + b - c \geqslant ma$, where*

$$m = \min\{|1 - \cos(\beta)|, |1 - \cos(\gamma)|\}.$$

PROOF. We let $g(t)$ denote the unit speed parameterizations of the segment BC, such that $g(0) = C, g(a) = B$. Let $c(t)$ denote the distance $\mathrm{dist}(A, g(t))$ (such that $b = c(0), c = c(a)$) and let $\beta(t)$ denote the angle $\angle Ag(t)B$. We leave it to the reader to verify that

$$|1 - \cos(\beta(t))| \geqslant m.$$

Consider the function

$$f(t) = t + b - c(t), \quad f(0) = 0, \quad f(a) = a + b - c.$$

By the 1st variation formula,

$$c'(t) = \cos(\beta(t))$$

and, hence,

$$f'(t) = 1 - \cos(\beta(t)) \geqslant m$$

Thus,

$$a + b - c = f(a) \geqslant ma \quad \square$$

EXERCISE 8.43. [Monotonicity of the hyperbolic distance] Let $T_i, i = 1, 2$ be right hyperbolic triangles with vertices A_i, B_i, C_i (where A_i or B_i could be ideal vertices) so that $A = A_1 = A_2$, $A_1B_1 \subset A_2B_2$, $\alpha_1 = \alpha_2$ and $\gamma_1 = \gamma_2 = \pi/2$. See Figure 8.2. Then $a_1 \leqslant a_2$. Hint: Use (8.9).

In other words, if $\sigma(t), \tau(t)$ are hyperbolic geodesic with unit speed parameterizations, so that $\sigma(0) = \tau(0) = A \in \mathbb{H}^2$, then the distance $d(\sigma(t), \tau)$ from the point $\sigma(t)$ to the geodesic τ, is a monotonically increasing function of t.

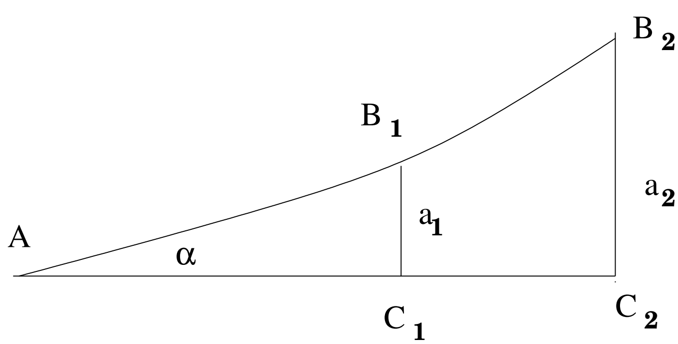


FIGURE 8.2. *Monotonicity of distance.*

8.5. Triangles and curvature of $\mathbb{H}^n$

Given points $A, B, C \in \mathbb{H}^n$ we define the *hyperbolic triangle* $T = [A, B, C] = \Delta ABC$ with vertices A, B, C. We topologize the set $Tri(\mathbb{H}^n)$ of hyperbolic triangles T in $\mathbb{H}^n$ by using topology on triples of vertices of T, i.e., a subset topology in $(\bar{\mathbf{B}}^n)^3$.

EXERCISE 8.44. Angles of hyperbolic triangles are continuous functions on $Tri(\mathbb{H}^n)$.

EXERCISE 8.45. Every hyperbolic triangle T in $\mathbb{H}^n$ is contained in (the compactification of) a 2-dimensional hyperbolic subspace $\mathbb{H}^2 \subset \mathbb{H}^n$. Hint: Consider a triangle $T = [A, B, C]$, where A, B belong to a common vertical line.

So far, we considered only geodesic hyperbolic triangles, we now introduce their 2-dimensional counterparts. First, let $T = T(A, B, C)$ be a generalized hyperbolic triangle in $\mathbb{H}^2$. We will assume that T is *nondegenerate*, i.e., is not contained in a hyperbolic geodesic. Such triangle T cuts $\mathbb{H}^2$ into several connected components, exactly one of which is a convex region with the boundary equal to T itself. (For instance, if all vertices of T are points in $\mathbb{H}^2$, then $\mathbb{H}^2 \setminus T$ consists of two components, while if T is an ideal triangle, then $\mathbb{H}^2 \setminus T$ is a disjoint union of four convex regions.) The closure of this region is called *solid* (generalized) hyperbolic triangle and denoted $\blacktriangle = \blacktriangle(A, B, C)$. It T is degenerate, we set $\blacktriangle := T$. More generally, if $T \subset \mathbb{H}^n$ is a hyperbolic triangle, then the *solid triangle* bounded by T is the solid triangle bounded by T in the hyperbolic plane $\mathbb{H}^2 \subset \mathbb{H}^n$ containing T. We will retain the notation $\blacktriangle$ for solid triangles in $\mathbb{H}^n$.

EXERCISE 8.46. Let S be a hyperbolic triangle with the sides $\sigma_i, i = 1, 2, 3$. Then there exists an ideal hyperbolic triangle T in $\mathbb{H}^2$ with the sides $\tau_i, i = 1, 2, 3$, bounding solid triangle $\blacktriangle$, so that $S \subset \blacktriangle$ and σ_1 is contained in the side τ_1 of T. See Figure 8.3.

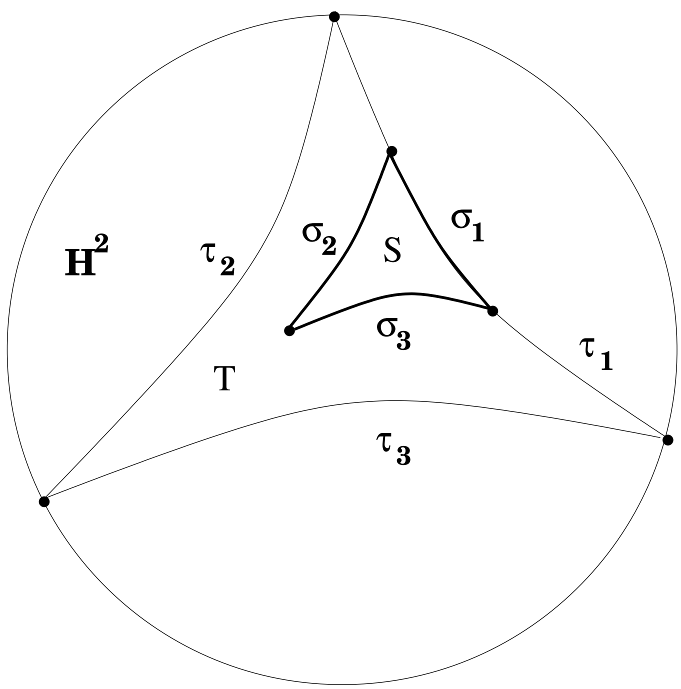


FIGURE 8.3. *T*riangles in the hyperbolic plane.

LEMMA 8.47. $\text{Isom}(\mathbb{H}^2)$ *acts transitively on the set of ordered triples of pairwise distinct points in* $\mathbb{H}^2$.

PROOF. Let $a, b, c \in \mathbb{R} \cup \infty$ be distinct points. By applying inversion we send a to ∞, so we can assume $a = \infty$. By applying a translation in $\mathbb{R}$ we get $b = 0$. Lastly, composing a map of the type $x \to \lambda x$, $\lambda \in \mathbb{R} \setminus 0$, we send c to 1. The composition of the above maps is a Moebius transformation of $\mathbb{S}^1$ and, hence, equals to the restriction of an isometry of $\mathbb{H}^2$. $\square$

COROLLARY 8.48. *All ideal hyperbolic triangles are congruent to each other.*

EXERCISE 8.49. Generalize the above corollary to: Every hyperbolic triangle is uniquely determined by its angles. Hint: Use hyperbolic trigonometry.

We will use the notation $T_{\alpha,\beta,\gamma}$ to denote unique (up to congruence) triangle with the angles α, β, γ.

EXERCISE 8.50. The group $Mob(\mathbb{S}^n)$ acts transitively on 3-point subsets of $\mathbb{S}^n$. (Hint: Use the fact that any triple of points in $\mathbb{S}^n$ is contained in a round circle; then apply Lemma 8.47.)

LEMMA 8.51. *Suppose that $(x, y, z), (x', y', z'), (x_i, y_i, z_i), i \in \mathbb{N}$ are triples of distinct points in $\mathbb{S}^n$ and*

$$\lim_{i\to\infty} (x_i, y_i, z_i) = (x', y', z'). \tag{8.12}$$

Assume that $\gamma_i \in Mob(\mathbb{S}^n)$ are such that

$$\gamma_i(x, y, z) = (x_i, y_i, z_i).$$

Then the sequence (g_i) belongs to a compact subset of $Mob(\mathbb{S}^n)$.

PROOF. We let $T, T', T_i \subset \mathbb{H}^{n+1}$ denote the (unique) ideal triangles with the vertices $(x, y, z), (x', y', z'), (x_i, y_i, z_i)$ respectively. Then each g_i sends T to T_i and maps the center c ot T to the center c_i of T_i. The limit (8.12) implies that

$$\lim_{i\to\infty} c_i = c',$$

where c' is the center of T'. The Arzela-Ascoli theorem now implies precompactness of the sequence (g_i) in $\mathrm{Isom}(\mathbb{H}^{n+1})$ and, hence, in $Mob(\mathbb{S}^n)$. $\square$

We now return to the study of geometry of hyperbolic triangles.

Given a hyperbolic triangle T bounding a solid triangle $\blacktriangle$, the *area of* T is the area of $\blacktriangle$

$$Area(T) = \iint_{\blacktriangle} \frac{dxdy}{y^2}.$$

Area of a degenerate hyperbolic triangle is, of course, zero. Here is an example of the area calculation. Consider the triangle $T = T_{0,\alpha,\pi/2}$ (which has angles $\pi/2, 0, \alpha$). We can realize T as the triangle with the vertices $i, \infty, e^{i\alpha}$. Computing hyperbolic area of this triangle (and using the substitution $x = \cos(t), \alpha \leqslant t \leqslant \pi/2$), we obtain

$$Area(T) = \iint_{\blacktriangle} \frac{dxdy}{y^2} = \frac{\pi}{2} - \alpha.$$

For $T = T_{0,0,\alpha}$, we subdivide T in two right triangles congruent to $T_{0,\alpha/2,\pi/2}$ and, thus, obtain

$$Area(T_{0,0,\alpha}) = \pi - \alpha. \tag{8.13}$$

In particular, area of the ideal triangle equals π.

LEMMA 8.52. $Area(T_{\alpha,\beta,\gamma}) = \pi - (\alpha + \beta + \gamma)$.

PROOF. The proof given here is due to Gauss, it appears in the letter from Gauss to Bolyai, see [**Gau73**]. We realize $T = T_{\alpha,\beta,\gamma}$ as a part of the subdivision of an ideal triangle $T_{0,0,0}$ in four triangles, the rest of which are $T_{0,0,\alpha'}, T_{0,0,\beta'}, T_{0,0,\gamma'}$, where $\theta' = \pi - \theta$ is the complementary angle. See Figure 8.4. Using additivity of area and equation (8.13), we obtain the area formula for T. □

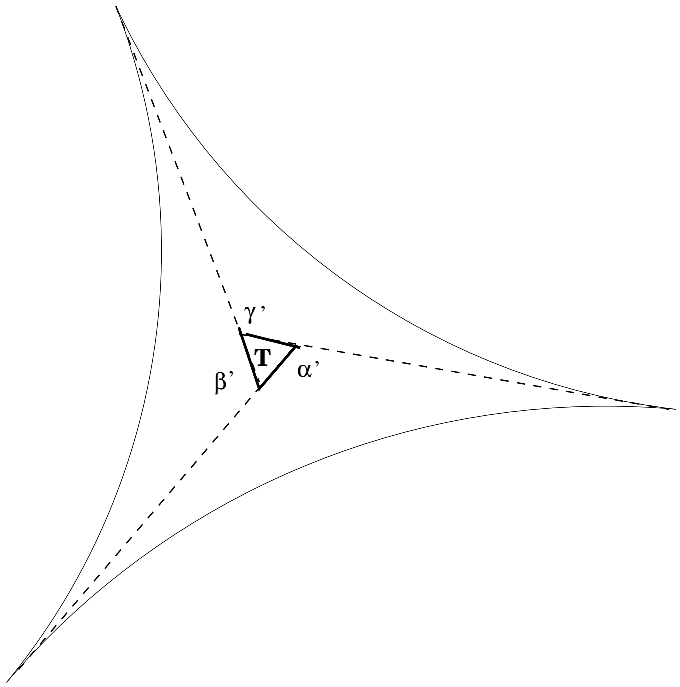


FIGURE 8.4. *C*omputation of area of the triangle T.

Curvature computation. Our next goal is to compute sectional curvature of $\mathbb{H}^n$. Since $\mathrm{Isom}(\mathbb{H}^n)$ acts transitively on pairs (p, P), where $P \subset T_pM$ is a 2-dimensional subspace, it follows that $\mathbb{H}^n$ has *constant* sectional curvature κ (see Section 2.6). Since $\mathbb{H}^2 \subset \mathbb{H}^n$ is totally geodesic and isometrically embedded (in the sense of Riemannian geometry), κ is the same for $\mathbb{H}^n$ as for $\mathbb{H}^2$.

COROLLARY 8.53. *The Gaussian curvature κ of $\mathbb{H}^2$ equals -1.*

PROOF. Instead of computing curvature tensor (see e.g. [**dC92**] for the computation), we will use Gauss-Bonnet formula. Comparing the area computation given in Lemma 8.52 with Gauss-Bonnet formula (Theorem 2.22) we conclude that $\kappa = -1$. □

Note that scaling properties of the sectional curvature (see §2.6) imply that the sectional curvature of

$$\left(\mathbf{U}^n, a^2 \frac{dx^2}{x_n^2}\right)$$

equals $-a^{-2}$ for every $a \neq 0$.

8.6. Distance function on $\mathbb{H}^n$

We begin by defining the following quantities:

$$\operatorname{dist}(z,w) = \operatorname{arccosh}\left(1 + \frac{|z-w|^2}{2\operatorname{Im} z \operatorname{Im} w}\right) \quad z, w \in \mathbf{U}^2 \tag{8.14}$$

and, more generally,

$$\operatorname{dist}(\mathbf{p},\mathbf{q}) = \operatorname{arccosh}\left(1 + \frac{|\mathbf{p}-\mathbf{q}|^2}{2p_n q_n}\right) \quad \mathbf{p}, \mathbf{q} \in \mathbf{U}^n \tag{8.15}$$

It is immediate that $dist(p,q) = dist(q,p)$ and that $dist(p,q) = 0$ if and only if $p = q$. However, it is, *a priori*, far from clear that dist satisfies the triangle inequality.

LEMMA 8.54. dist *is invariant under* $\operatorname{Isom}(\mathbb{H}^n) = Mob(\mathbf{U}^n)$.

PROOF. First, it is clear that dist is invariant under the group $Euc(\mathbf{U}^n)$ of Euclidean isometries which preserve $\mathbf{U}^n$. Next, any two points in $\mathbf{U}^n$ belong to a vertical half-plane in $\mathbf{U}^n$. Applying elements of $Euc(\mathbf{U}^n)$ to this half-plane, we can transform it to the coordinate half-plane $\mathbf{U}^2 \subset \mathbf{U}^n$. Thus, the problem reduces to the case $n = 2$ and orientation-preserving Moebius transformations of $\mathbb{H}^2$. We leave it to the reader as an exercise to show that the map $z \mapsto -\frac{1}{z}$ (which is an element of $PSL(2,\mathbb{R})$) preserves the quantity

$$\frac{|z-w|^2}{\operatorname{Im} z \operatorname{Im} w}$$

and, hence, the function dist. Now, the assertion follows from Exercise 8.10 and Lemma 8.11. □

Recall that $d(p,q)$ denotes the hyperbolic distance between points $p, q \in \mathbf{U}^n$.

PROPOSITION 8.55. $\operatorname{dist}(p,q) = d(p,q)$ *for all points* $p, q \in \mathbb{H}^n$. *In particular, the function* dist *is indeed a metric on* $\mathbb{H}^n$.

PROOF. As in the above lemma, it suffices to consider the case $n = 2$. We can also assume that $p \neq q$. First, suppose that $p = i$ and $q = ib$, $b > 1$. Then, by Exercise 8.14,

$$\operatorname{dist}(p,q) = \int_1^b \frac{dt}{t} = \log(b), \quad \exp(d(p,q)) = b.$$

On the other hand, the formula (8.14) yields:

$$\operatorname{dist}(p,q) = \operatorname{arccosh}\left(1 + \frac{(b-1)^2}{2b}\right).$$

Hence,

$$\cosh(\operatorname{dist}(p,q)) = \frac{e^{\operatorname{dist}(p,q)} + e^{-\operatorname{dist}(p,q)}}{2} = 1 + \frac{(b-1)^2}{2b}.$$

Now, the equality $\operatorname{dist}(p,q) = d(p,q)$ follows from the identity

$$1 + \frac{(b-1)^2}{2b} = \frac{b+b^{-1}}{2}.$$

For general points p, q in $\mathbb{H}^2$, by Lemma 8.22, there exists a hyperbolic isometry which sends p to i and q to a point of the form $ib, b \geqslant 1$. We already know that both hyperbolic distance d and the quantity dist are invariant under the action of $\mathrm{Isom}(\mathbb{H}^2)$. Thus, the equality $d(p,q) = \mathrm{dist}(p,q)$ follows from the special case of points on the y-axis. □

Exercise 8.56. 1. Deduce from (8.14) that

$$\log\left(1+\frac{|z-w|^2}{2\,\mathrm{Im}\,z\,\mathrm{Im}\,w}\right) \leqslant d(z,w) \leqslant \log\left(1+\frac{|z-w|^2}{2\,\mathrm{Im}\,z\,\mathrm{Im}\,w}\right)+\ln 2$$

for all points $z, w \in \mathbf{U}^2$.

2. Suppose that $\hat{A}, \hat{B}$ are distinct points in $\mathbb{S}^1$ and A, B are points which belong to the geodesic in $\mathbb{H}^2$ connecting $\hat{A}$ to $\hat{B}$. Show that

$$\mathrm{dist}(A,B) = \log[A : B : \hat{B} : \hat{A}].$$

Hint: First do the computation when $\hat{A} = 0, \hat{B} = \infty$ in the upper half-plane model.

8.7. Hyperbolic balls and spheres

Pick a point $p \in \mathbb{H}^n$ and a positive real number R. Then the *hyperbolic sphere* of radius R centered at p is the set

$$S(p,R) = \{x \in \mathbb{H}^n : d(x,p) = R\}.$$

Exercise 8.57. 1. Prove that $S(\mathbf{e}_n, R) \subset \mathbb{H}^n = \mathbf{U}^n$ equals the Euclidean sphere of center $\cosh(R)\mathbf{e}_n$ and radius $\sinh(R)$. *Hint.* It follows immediately from the distance formula (8.14).

2. Suppose that $S \subset \mathbf{U}^n$ is the Euclidean sphere with Euclidean radius R and the center x so that $x_n = a$. Then $S = S(p,r)$, where the hyperbolic radius r equals

$$\frac{1}{2}\left(\log(a+R) - \log(a-R)\right).$$

Since the group of Euclidean similarities acts transitively on $\mathbf{U}^n$, it follows that every hyperbolic sphere is also a Euclidean sphere. A non-computational proof of this fact is as follows: Since the hyperbolic metric $ds^2_{\mathbf{B}}$ on $\mathbf{B}^n$ is invariant under $O(n)$, it follows that hyperbolic spheres centered at 0 in $\mathbf{B}^n$ are also Euclidean spheres. The general case follows from transitivity of $\mathrm{Isom}(\mathbb{H}^n)$ and the fact that isometries of $\mathbb{H}^n$ are Moebius transformations, which, therefore, send Euclidean spheres to Euclidean spheres.

Lemma 8.58. *If $B(x_1, R_1) \subset B(x_2, R_2)$ are hyperbolic balls, then $R_1 \leqslant R_2$.*

Proof. It follows from the triangle inequality that the diameter of a metric ball $B(x,R)$ is the longest geodesic segment contained in $B(x,R)$. Therefore, let $\gamma \subset B(x_1,R_1)$ be a diameter. Then γ is contained in $B(x_2,R_2)$ and, hence, its length is $\leqslant 2R_2$. However, the length of γ is $2R_1$, therefore, $R_1 \leqslant R_2$. □

Exercise 8.59. Show that this lemma fails for general metric spaces.

8.8. Horoballs and horospheres in $\mathbb{H}^n$

Horoballs and horospheres play prominent role in the theory of discrete groups of isometries of hyperbolic n-space, primarily due to the *thick-thin decomposition*, which we will discuss in detail in §10.6.3. Later on, in Chapter 22 we will deal with families of disjoint horoballs in $\mathbb{H}^n$, while proving Schwartz' theorem on quasi-isometric rigidity of nonuniform lattices.

Consider a geodesic ray $r = x\xi$ in $\mathbb{H}^n = \mathbf{B}^n$, connecting a point $x \in \mathbb{H}^n$ to a boundary point $\xi \in \mathbb{S}^{n-1}$. We let b_r denote the Busemann function on $\mathbb{H}^n$ for the ray r ($b_r(x) = 0$). By Lemma 2.81, the open horoball $B(\xi)$ defined by the inequality $b_r < 0$, equals the union of open balls

$$B(\xi) = \bigcup_{t \geqslant 0} B(r(t), t)\,.$$

As we saw in §8.7, in particular Exercise 8.57, each ball $B(r(t), t)$ is a Euclidean ball centered in a point $r(T_t)$ with $T_t > t$. Therefore, this union of hyperbolic balls is the open Euclidean ball with boundary tangent to $\mathbb{S}^{n-1}$ at ξ, and containing the point x. According to Lemma 2.83, the closed horoball and the horosphere defined by $b_r \leqslant 0$ and $b_r = 0$, respectively, are the closed Euclidean ball and its boundary sphere, both with the point ξ removed.

EXERCISE 8.60. The isometry group of $\mathbb{H}^n$ acts transitively on the set of open horoballs in $\mathbb{H}^n$.

We conclude that the set of horoballs (closed or open) with center ξ is the same as the set of Euclidean balls in $\mathbf{B}^n$ (closed or open) tangent to $\mathbb{S}^{n-1}$ at ξ, with the point ξ removed.

Applying the map $\sigma : \mathbf{B}^n \to \mathbf{U}^n$ to horoballs and horospheres in $\mathbf{B}^n$, we obtain horoballs and horospheres in the upper-half space model $\mathbf{U}^n$ of $\mathbb{H}^n$. Being a Moebius transformation, σ carries Euclidean spheres to Euclidean spheres (recall that a compactified Euclidean hyperplane is also regarded as a Euclidean sphere). Recall that $\sigma(-\mathbf{e}_n) = \infty$. Therefore, every horosphere in $\mathbf{B}^n$ centered at $-\mathbf{e}_n$ is sent by σ to an $n-1$-dimensional Euclidean subspace E of $\mathbf{U}^n$ whose compactification contains the point ∞. Hence, E has to be a horizontal Euclidean subspace, i.e., a subspace of the form

$$\{\mathbf{x} \in \mathbf{U}^n : x_n = t\}$$

for some fixed $t > 0$. Restricting the metric ds^2 to such E we obtain the Euclidean metric rescaled by t^{-2}. Thus, the restriction of the Riemannian metric ds^2 to every horosphere is isometric to the Euclidean $n-1$ space $\mathbb{E}^{n-1}$. When working with horoballs and horospheres we will frequently use their identification with Euclidean half-spaces and hyperplanes in $\mathbf{U}^n$.

On the other hand, the restriction of the hyperbolic distance function to a horosphere is very far from the Euclidean metric: It follows from Exercise 8.56 that as the distance D between points x, y in a fixed horosphere Σ tends to infinity, the distance $\mathrm{dist}(x, y)$ in $\mathbb{H}^n$ also tends to infinity, but logarithmically slower:

$$\mathrm{dist}(x, y) \asymp \log(D).$$

Thus, horospheres in $\mathbb{H}^n$ are *exponentially distorted*, see §5.9.

We next consider intersections of horoballs $B(\xi_1) \cap B(\xi_2)$. If $\xi_1 = \xi_2$ then this intersection is either $B(\xi_1)$ or $B(\xi_2)$, whichever of these horoballs is smaller. Suppose now that $\xi_1 \neq \xi_2$. The horoballs $B(\xi_1)$, $B(\xi_2)$ are said to be *opposite* in this case. Using the upper half-space model, we find an isometry of $\mathbb{H}^n$ sending ξ_2 to ∞ and $B(\xi_1)$ to $\{x_n > 1\}$. After applying this isometry, we can assume that $B(\xi_1) = \{x_n > 1\}$ and $B(\xi_1)$ is a Euclidean round ball. Then the intersection of the horoballs is clearly bounded and, furthermore, the intersection

$$B(\xi_2) \cap \{x_n = 1\}$$

is either empty or is a round Euclidean ball. This proves:

LEMMA 8.61. *1. The intersection of two horoballs with the same center is another horoball with the same center.*

2. The intersection of two opposite horoballs is always bounded.

3. The intersection of a horoball with the horosphere Σ bounding an opposite horoball is either empty or is a metric ball with respect to the intrinsic (flat) Riemannian metric of Σ.

EXERCISE 8.62. Consider the upper half-space model for the hyperbolic space $\mathbb{H}^n$ and the vertical geodesic ray r in $\mathbb{H}^n$:

$$r = \{(0, \dots, 0, x_n) : x_n \geqslant 1\}.$$

Show that the Busemann function b_r for the ray r is given by

$$b_r(x_1, \dots, x_n) = -\log(x_n).$$

Consider the boundary horosphere $H \subset \mathbb{H}^n$ of the horoball

$$B = \{(x_1, \dots, x_n) : x_n > 1\}.$$

Define the projection

$$\pi : B^c := \mathbb{H}^n \setminus B \to H, \quad \pi(x_1, \dots, x_{n-1}, x_n) = (x_1, \dots, x_{n-1}, 1).$$

EXERCISE 8.63. For $x \in B^c$, the norm (computed with respect to the hyperbolic metric) of the derivative $d\pi_x$ equals x_n. In particular, $\|d\pi_x\| \leq 1$ with equality iff $x \in H$.

We now switch to the Lorentzian model. In view of the projection $\pi : H \to \mathbf{B}^n = \mathbb{H}^n$, can identify the ideal boundary points $\xi \in \mathbb{S}^{n-1}$ of $\mathbb{H}^n$ with lines in the *future light cone*

$$C^{\uparrow} = \{\mathbf{x} \in \mathbb{R}^{n+1} : x_{n+1} > 0, Q(\mathbf{x}) = 0\}.$$

EXERCISE 8.64. Given a point $\xi \in C^{\uparrow}$, show that the corresponding Busemann function b_ξ on $\mathbb{H}^n$ (up to constant) equals

$$-\log(-\langle \mathbf{x}, \xi \rangle).$$

Accordingly, horospheres in $\mathbb{H}^n$ are projections of intersections of affine hyperplanes $\{\langle \mathbf{x}, \xi \rangle = a\} \cap H$, where $a < 0$. Similarly, show that open horoballs are projections of the intersections

$$\{\langle \mathbf{x}, \xi \rangle > a\} \cap H, \quad a < 0.$$

8.9. $\mathbb{H}^n$ is a symmetric space

A *symmetric space* is a complete simply connected Riemannian manifold X such that for every point p there exists a global isometry of X which is a geodesic symmetry σ_p with respect to p, that is, for every geodesic γ through p, $\sigma_p(\gamma(t)) = \gamma(-t)$. We will discuss general symmetric spaces and their discrete groups of isometries in more details in Chapter 10; we also refer the reader to [**BH99**, II.10], [**Ebe72**] and [**Hel01**] for the detailed treatment.

Let us verify that each symmetric space X is a homogeneous Riemannian manifold. Indeed, given points $p, q \in X$, let m denote the midpoint of a geodesic connecting p to q. Then $\sigma_m(p) = q$. Thus, X can be naturally identified with the quotient G/K, where G is a Lie group (acting transitively and isometrically on X) and $K < G$ is a compact subgroup. In the case of symmetric spaces of nonpositive curvature we are interested in, the group G is *semisimple* and K is its maximal compact subgroup. Another important subgroup, in the nonpositively curved case, is the *Borel subgroup* $B < G$, it is a minimal subgroup of G such that the quotient G/B is compact. Geometrically speaking, the quotient G/B is identified with the *Fürstenberg boundary* of X. In the case of negatively curved symmetric spaces, G/B is the ideal boundary $\partial_\infty X$ of X in the sense of §2.11.3. The solvable group B has a further decomposition as the semidirect product

$$N \rtimes T,$$

where the group T is abelian and the subgroup N is nilpotent. The subgroup T is a *maximal (split) torus* of G. Both groups play important role in geometry of symmetric spaces. The dimension of T is the *rank* of X (and of G). A symmetric space is negatively curved if and only if it has rank 1. In this situation, the group N acts simply-transitively on a horosphere in X. Accordingly, $\partial_\infty X$ can be identified with a one-point compactification of N. This algebraic description of $\partial_\infty X$ plays an important role in proofs of rigidity theorems for rank 1 symmetric spaces.

In this section we describe how the real-hyperbolic space fits into the general framework of symmetric spaces. We will also discuss briefly other negatively curved symmetric spaces, as it turns out that besides real-hyperbolic spaces $\mathbb{H}^n$, there are three other families of negatively curved symmetric spaces: $\mathbf{C}\mathbb{H}^n, n \geqslant 2$ (complex-hyperbolic spaces), $\mathbf{H}\mathbb{H}^n, n \geqslant 2$ (quaternionic hyperbolic spaces) and $\mathbf{O}\mathbb{H}^2$ (octonionic hyperbolic plane).

Generalities of negatively curved symmetric spaces $\mathbb{H}^n, \mathbf{C}\mathbb{H}^n, \mathbf{H}\mathbb{H}^n, \mathbf{O}\mathbb{H}^2$. All four classes symmetric spaces can be described via a "linear algebra" model, generalizing the Lorentzian model of $\mathbb{H}^n$, although things become quite complicated in the case of $\mathbf{O}\mathbb{H}^2$ due to lack of associativity.

In the first three cases, the symmetric space X appears as a projectivization of a certain open cone V_- in a vector space (or a module in the case of $\mathbf{H}\mathbb{H}^n$), equipped with a hermitian form $\langle \cdot, \cdot \rangle$. The distance function in X is given by the formula:

$$\cosh^2(\text{dist}(\mathbf{p}, \mathbf{q})) = \frac{\langle \mathbf{p}, \mathbf{q} \rangle \langle \mathbf{q}, \mathbf{p} \rangle}{\langle \mathbf{p}, \mathbf{p} \rangle \langle \mathbf{q}, \mathbf{q} \rangle}, \tag{8.16}$$

where $\mathbf{p}, \mathbf{q} \in V_-$ represent points in X.

In the case of all negatively curved symmetric spaces, the maximal torus T isn isomorphic to $\mathbb{R}_+$, while the group N is *2-step nilpotent*. Accordingly, the Lie algebra of N splits (as a vector space) as a direct sum

$$\mathfrak{n} = \mathfrak{n}_1 \oplus \mathfrak{n}_2$$

and this decomposition is T-invariant (one of these Lie algebras is trivial in the real-hyperbolic case). The subalgebra $\mathfrak{n}_2$ is the Lie algebra of the center of N. Each element $t \in T$ acts on $\mathfrak{n}$ with two distinct eigenvalues λ_1, λ_2, which are evaluations on t of two homomorphisms $\lambda_1, \lambda_2 : T \to \mathbb{R}_+$, called *characters*.

Special features of rank 1 symmetric spaces. The rank one symmetric spaces X are also characterized among symmetric spaces by the property that any two segments of the same length are congruent in X, i.e., the subgroup $K < G$ (the stabilizer of a point $p \in X$) acts transitively on each R-sphere $S(p, R)$ centered at p. Another distinguishing characteristic of negatively curved symmetric spaces X

is that their horospheres are exponentially distorted in X (cf. §8.8), while for all other nonpositively curved symmetric spaces, horospheres are quasi-isometrically embedded. Furthermore, two horoballs with distinct centers in negatively curved symmetric spaces have bounded intersection, while it is not the case for the rest of the symmetric spaces.

Real-hyperbolic spaces $\mathbb{H}^n$. We note that in the unit ball model of $\mathbb{H}^n$ we clearly have the symmetry σ_p with respect to the origin $p = 0$, namely, $\sigma_0 : \mathbf{x} \mapsto -\mathbf{x}$. Since $\mathbb{H}^n$ is homogeneous, it follows that it has a symmetry at every point. Thus, $\mathbb{H}^n$ is a symmetric space.

EXERCISE 8.65. Prove that the linear-fractional transformation $\sigma_i \in PSL(2, \mathbb{R})$ defined by $\pm S_i$, where

$$S_i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

fixes i and is a symmetry with respect to i.

We proved in §8.5 that $\mathbb{H}^n$ has negative curvature -1. In particular, it contains no totally-geodesic Euclidean subspaces of dimension $\geqslant 2$ and, thus, $\mathbb{H}^n$ has rank 1.

The isometry group of $\mathbb{H}^n$ is $PO(n, 1)$, its maximal compact subgroup is $K \cong O(n)$, its Borel subgroup is the semidirect product $\mathbb{R}^{n-1} \rtimes \mathbb{R}_+ = N \rtimes T$. In the upper half-space model, the group N consists of Euclidean translations in $\mathbb{R}^{n-1}$, while T consists of dilations $\mathbf{x} \mapsto t\mathbf{x}, t > 0$.

There are many properties which distinguish the real-hyperbolic space among other rank 1 symmetric spaces, for instance, the fact that the subgroup N is abelian, which, geometrically, reflects flatness of the intrinsic Riemannian metric of the horospheres in $\mathbb{H}^n$. Another example is the fact that only in the real-hyperbolic space triangles are uniquely determined by their side-lengths: This is false for other hyperbolic spaces.

Complex-hyperbolic spaces. Consider the complex vector space $V = \mathbb{C}^{n+1}$ equipped with the Hermitian bilinear form

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{k=1}^{n} v_k \bar{w}_k - v_{n+1} \bar{w}_{n+1}.$$

The group $U(n, 1)$ is the group of complex-linear automorphisms of $\mathbb{C}^{n+1}$ preserving this bilinear form. Consider the negative cone

$$V_- = \{\mathbf{v} : \langle \mathbf{v}, \mathbf{v} \rangle < 0\} \subset \mathbb{C}^{n+1}.$$

Then the *complex-hyperbolic space* $\mathbb{CH}^n$ is the projectivization of V_-. The group $PU(n, 1)$ acts naturally on $X = \mathbb{CH}^n$. One can describe the Riemannian metric on $\mathbb{CH}^n$ as follows. Let $\mathbf{p} \in V_-$ be a vector such that $\langle \mathbf{p}, \mathbf{p} \rangle = 1$; the tangent space $T_{[\mathbf{p}]}X$ of X at the projection $[\mathbf{p}]$ of $\mathbf{p}$, is the projection of the orthogonal complement $\mathbf{p}^\perp$ in $\mathbb{C}^{n+1}$. Let $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{n+1}$ be vectors orthogonal to p, i.e.,

$$\langle \mathbf{p}, \mathbf{v} \rangle = \langle \mathbf{p}, \mathbf{w} \rangle = 0.$$

Then define

$$(\mathbf{v}, \mathbf{w})_p := -Im \langle \mathbf{v}, \mathbf{w} \rangle .$$

This determines a $PU(n,1)$-invariant Riemannian metric on X. The corresponding distance function (8.16) is $PU(n,1)$-invariant. The Cartan involution fixing the point $[\mathbf{e}_{n+1}]$ is the projectivization of the diagonal matrix $Diag(-1,\ldots,-1,1)$.

The maximal compact subgroup of $PU(n,1)$ is $U(n)$, the nilpotent subgroup $N < B < G$ is the *Heisenberg group*, its Lie algebra splits as

$$\mathbb{C}^n \oplus \mathbb{R},$$

where one should think of $\mathbb{R}$ as the set of imaginary complex numbers (the reason for this will become clear shortly).

An important special feature of complex-hyperbolic spaces is the fact that they are *Kähler manifolds*: The $PU(n,1)$-invariant complex structure on $\mathbb{CH}^n$ is the restriction of the complex structure on the ambient complex-projective space. The corresponding *almost complex structure* on the tangent bundle of $\mathbb{CH}^n$ is given by the multiplication by i:

$$J(\mathbf{v}) = i\mathbf{v}, \quad \mathbf{v} \in T_p\mathbb{CH}^n.$$

This complex structure is *hermitian*, i.e., J preserves the Riemannian metric on $\mathbb{CH}^n$. Furthermore, J and $(v,w)_p$ together define a $PU(n,1)$-invariant *symplectic structure* on $\mathbb{CH}^n$ (a closed nondegenerate 2-form), given by

$$\omega(\mathbf{v},\mathbf{w}) = (\mathbf{v}, J\mathbf{w}).$$

This Kähler nature of $\mathbb{CH}^n$ means that one can use tools of complex analysis and complex differential geometry in order to study complex-hyperbolic spaces and their quotients by discrete isometry groups.

As we noted earlier, geodesic triangles $T \subset \mathbb{CH}^n$ are not uniquely determined by their side-lengths. The additional invariant which determines geodesic triangles is their *symplectic area*, which is defined as the integral

$$\int_S \omega$$

of the symplectic form ω on $\mathbb{CH}^n$ over any surface $S \subset \mathbb{CH}^n$ bounded by T. (The fact that the area is independent of the choice of S follows from Stokes Theorem, since the form ω is closed.)

Quaternionic-hyperbolic spaces. Consider the ring $\mathbf{H}$ of *quaternions*; the elements of the quaternion ring have the form

$$q = x + iy + jz + kw, \quad x,y,z,w \in \mathbb{R}.$$

The quaternionic conjugation is given by

$$\bar{q} = x - iy - jz - kw$$

and

$$|q| = (q\bar{q})^{1/2} \in \mathbb{R}_+$$

is the quaternionic norm. A *unit quaternions* is a quaternion of the unit norm. Let V be a left $n+1$-dimensional free left module over $\mathbf{H}$:

$$V = \{\mathbf{q} = (q_1,\ldots,q_{n+1}) : q_m \in \mathbf{H}\}.$$

Consider the quaternionic-hermitian inner product of signature $(n,1)$:

$$\langle \mathbf{p},\mathbf{q}\rangle = \sum_{m=1}^{n} p_m\bar{q}_m - p_{n+1}\bar{q}_{n+1}.$$

Then the group $G = Sp(n, 1)$ is the group of automorphisms of the module V preserving this inner product. The quotient of V by the group of nonzero quaternions $\mathbf{H}^\times$ (with respect to the left multiplication action) is the n-dimensional *quaternionic-projective space* PV. Analogously to the case of real and complex hyperbolic spaces, consider the negative cone

$$V_- = \{\mathbf{q} \in V : \langle \mathbf{q}, \mathbf{q} \rangle < 0\}.$$

The group G acts naturally on $PV_- \subset PV$ through the group $PSp(n, 1)$, the quotient of G by the subgroup of unit quaternions embedded in the subgroup of diagonal matrices in G,

$$q \mapsto qI.$$

The space PV_- is called the n-dimensional *quaternionic-hyperbolic space* $\mathbf{H}\mathbb{H}^n$. As in the real and complex cases, the Cartan involution fixing the point $[\mathbf{e}_{n+1}]$ is the projectivization of the diagonal matrix $Diag(-1, \ldots, -1, 1)$.

The maximal compact subgroup of G is $Sp(n)$, the Lie algebra of the nilpotent subgroup $N < B$ splits as a real vector space as

$$\mathfrak{n}_1 \oplus \mathfrak{n}_2 = \mathbb{H}^n \oplus Im(\mathbf{H}),$$

where $Im(\mathbf{H})$ is the 3-dimensional real vector space of imaginary quaternions.

The octonionic-hyperbolic plane. One defines *octonionic-hyperbolic plane* $\mathbf{O}\mathbb{H}^2$ analogously to $\mathbf{H}\mathbb{H}^n$, only using the algebra $\mathbf{O}$ of Cayley octonions instead of quaternions. An extra complication comes from the fact that the algebra $\mathbf{O}$ is not associative, which means that one cannot talk about free $\mathbf{O}$-modules.

The space $\mathbf{O}\mathbb{H}^2$ has dimension 16. It is identified with the quotient G/K, where G is a real form of the exceptional Lie group F_4 and the maximal compact subgroup $K < G$ is isomorphic to $Spin(9)$, the 2-fold cover of the orthogonal group $SO(9)$. The Lie algebra of the nilpotent subgroup $N < B < G$ has dimension 15; it splits as a real vector space as

$$\mathfrak{n}_1 \oplus \mathfrak{n}_2 = \mathbf{O} \oplus Im(\mathbf{O}),$$

where $Im(\mathbf{O})$ is the 7-dimensional vector space consisting of imaginary octonions.

We refer to Mostow's book [**Mos73**] and Parker's survey [**Par08**] for a more detailed discussion of negatively curved symmetric spaces.

8.10. Inscribed radius and thinness of hyperbolic triangles

Suppose that T is a hyperbolic triangle in the hyperbolic plane $\mathbb{H}^2$ with the sides $\tau_i, i = 1, 2, 3$, so that T bounds the solid triangle $\blacktriangle$. For a point $x \in \blacktriangle$ define the quantities

$$\Delta_x(T) := \max_{i=1,2,3} d(x, \tau_i).$$

and

$$\Delta(T) := \inf_{x \in \blacktriangle} \Delta_x(T).$$

The goal of this section is to estimate $\Delta(T)$ from above. It is immediate that the infimum in the definition of $\Delta(T)$ is realized by a point $x_o \in \blacktriangle$ which is equidistant from all the three sides of T, i.e., by the intersection point of the angle bisectors.

Define the *inscribed radius* $inrad(T)$ of T is the supremum of radii of hyperbolic disks contained in $\blacktriangle$.

LEMMA 8.66. *$\Delta(T) = inrad(T)$.*

PROOF. Suppose that $D = B(X, R) \subset \blacktriangle$ is a hyperbolic disk. Unless D touches two sides of T, there exists a disk $D' = B(X', R') \subset \blacktriangle$ which contains D and, hence, has larger radius, see Lemma 8.58. Suppose, therefore, that $D \subset \blacktriangle$ touches two boundary edges of T, hence, center X of D belongs to the bisector σ of the corner ABC of T. Unless D touches all three sides of T, we can move the center X of D along the bisector σ away from the vertex B so that the resulting disk $D' = B(X', R')$ still touches only the sides AB, BC of T. We claim that the (radius R' of D' is larger than the radius R of D. In order to prove this, consider hyperbolic triangles $[X, Y, B]$ and $[X', Y', B']$, where Y, Y' are the points of tangency between D, D' and the side BA. These right-angled triangles have the common angle $\angle_b xy$ and satisfy

$$d(B, X) \leqslant d(B, X').$$

Thus, the inequality $R \leqslant R'$ follows from the Exercise 8.43. □

Thus, we need to estimate inradius of hyperbolic triangles from above. Recall that by Exercise 8.46, for every hyperbolic triangle S in $\mathbb{H}^2$ there exists an ideal hyperbolic triangle T, so that $S \subset \blacktriangle$. Clearly, $inrad(S) \leqslant inrad(T)$. Since all ideal hyperbolic triangles are congruent, it suffices to consider the ideal hyperbolic triangle T in $\mathbb{U}^2$ with the vertices $-1, 1, \infty$. The inscribed circle C in T has Euclidean center $(0, 2)$ and Euclidean radius 1. Therefore, by Exercise 8.57, its hyperbolic radius equals $\log(3)/2$. By combining these observations with Exercise 8.45, we obtain

PROPOSITION 8.67. *For every hyperbolic triangle T, $\Delta(T) = inrad(T) \leqslant \frac{\log(3)}{2}$. In particular, for every hyperbolic triangle in $\mathbb{H}^n$, there exists a point $p \in H^n$ so that distance from p to all three sides of T is $\leqslant \frac{\log(3)}{2}$.*

Another way to measure thinness of a hyperbolic triangle T is to compute distance from points of one side of T to the union of the two other sides. Let T be a hyperbolic triangle with sides $\tau_j, j = 1, 2, 3$. Define

$$\delta(T) := \max_j \sup_{p \in \tau_j} d(p, \tau_{j+1} \cup \tau_{j+2}),$$

where indices of the sides of T are taken modulo 3. In other words, if $\delta = \delta(T)$ then each side of T is contained in the δ-neighborhood of the union of the other two sides.

PROPOSITION 8.68. *For every geodesic triangle S in $\mathbb{H}^n$, $\delta(S) \leqslant \operatorname{arccosh}(\sqrt{2})$.*

PROOF. First of all, as above, it suffices to consider the case $n = 2$. Let $\sigma_j, j = 1, 2, 3$ denote the edges of S. We will estimate $d(p, \sigma_2 \cup \sigma_3)$ (from above) for $p \in \sigma_1$. We enlarge the hyperbolic triangle S to an ideal hyperbolic triangle T as in Figure 8.5. For every $p \in \sigma_1$, every geodesic segment g connecting p to a point of $\tau_2 \cup \tau_3$ has to cross $\sigma_2 \cup \sigma_3$. In particular,

$$d(p, \sigma_2 \cup \sigma_3) \leqslant d(p, \tau_2 \cup \tau_3).$$

Thus, it suffices to show that $\delta(T) \leqslant \operatorname{arccosh}(\sqrt{2})$ for the ideal triangle T as above. We realize T as the triangle with the (ideal) vertices $A_1 = \infty, A_2 = -1, A_3 = 1$ in $\partial_\infty \mathbb{H}^2$. We parameterize the sides $\tau_i = A_{j-1}A_{j+1}, j = 1, 2, 3$ modulo 3, according to their orientation. Then, by the Exercise 8.43, for every i,

$$d(\tau_j(t), \tau_{j-1})$$

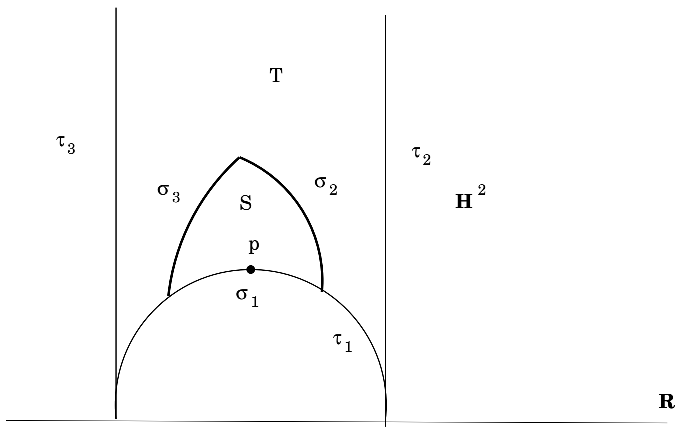


FIGURE 8.5. Enlarging the hyperbolic triangle S.

is monotonically increasing. Thus,

$$\sup_t d(\tau_1(t), \tau_2 \cup \tau_3)$$

is achieved at the point $p = \tau_1(t) = i = \sqrt{-1}$ and equals $d(p,q)$, where $q = -1+\sqrt{2}i$. Then, using formula 8.15, we get $d(p,q) = \operatorname{arccosh}(\sqrt{2})$. Note that alternatively, one can get the formula for $d(p,q)$ from (8.8) by considering the right triangle $[p,q,-1]$ where the angle at p equals $\pi/4$. $\square$

As we will see in Section 9.1, the above propositions mean that *all hyperbolic triangles are uniformly thin.*

COROLLARY 8.69.

$$\sup_{T\in Tri(\mathbb{H}^n)} \delta(T) = \operatorname{arccosh}(\sqrt{2}).$$

8.11. Existence-uniqueness theorem for triangles

Proof of Lemma 2.53. We will prove this result for the hyperbolic plane $\mathbb{H}^2$, this will imply the lemma for all $\kappa < 0$ by rescaling the metric on $\mathbb{H}^2$. We leave the cases $\kappa \geqslant 0$ to the reader as the proof is similar. The proof below is goes back to Euclid (in the case of $\mathbb{E}^2$). Let c denote the largest of the numbers a, b, c. Draw a geodesic $\gamma \subset \mathbb{H}^2$ through points x, y such that $d(x,y) = c$. Then

$$\gamma = \gamma_x \cup xy \cap \gamma_y,$$

where γ_x, γ_y are geodesic rays emanating from x and y respectively. Now, consider the hyperbolic circles $S(x,b)$ and $S(y,a)$ centered at x, y and having radii b, a respectively. Since $c \geq \max(a,b)$,

$$\gamma_x \cap S(y,a) \subset \{x\}, \quad \gamma_y \cap S(x,b) \subset \{y\},$$

while

$$S(x,b) \cap xy = p, \quad S(y,a) \cap xy = y.$$

By the triangle inequality on $c \leq a+b$, p separates q from y (and q separates x from p). Therefore, both the ball $B(x,b)$ and its complement contain points of the circle $S(y,a)$, which (by connectivity) implies that $S(x,b) \cap S(y,a) \neq \emptyset$. Therefore, the triangle with the side-lengths a, b, c exists. Uniqueness (up to congruence) of this triangle follows from Exercise 8.28; alternatively it can be derived from the hyperbolic cosine law. $\square$

CHAPTER 9

Gromov-hyperbolic spaces and groups

The goal of this chapter is to review basic properties of δ-hyperbolic spaces and word-hyperbolic groups, which are far-reaching generalizations of the real-hyperbolic space $\mathbb{H}^n$ and of groups acting geometrically on $\mathbb{H}^n$. The advantage of δ-hyperbolicity is that it can be defined in the context of arbitrary metric spaces which need not even be geodesic. These spaces were introduced in the seminal essay by Mikhael Gromov on hyperbolic groups [**Gro87**], although ideas of *combinatorial* curvature and (in retrospect) hyperbolic properties of finitely-generated groups are much older. These ideas go back to work of Max Dehn (Dehn algorithm for the word problem in hyperbolic surface groups), Martin Grindlinger (small cancelation theory), Alexandr Olʹshanskiĭ (who used what we now call *relative hyperbolicity* in order to construct finitely-generated groups with exotic properties) and many others.

9.1. Hyperbolicity according to Rips

We begin our discussion of δ-hyperbolic spaces with the notion of hyperbolicity in the context of geodesic metric spaces, which (according to Gromov) is due to Ilya (Eliyahu) Rips. This definitions will be then applied to Cayley graphs of groups, leading to the concept of *hyperbolic groups* discussed later in this chapter. Rips notion of hyperbolicity is based on the thinness properties of hyperbolic triangles which are established in §8.10.

Let (X, d) be a geodesic metric space. As in §8.5, a geodesic triangle T in X is a concatenation of three geodesic segments τ_1, τ_2, τ_3 connecting the points A_1, A_2, A_3 (vertices of T) in the natural cyclic order. Unlike the real-hyperbolic space, we no longer have uniqueness of geodesics, thus T is not (in general) determined by its vertices. We define a measure of the thinness of T similar to the one in §8.10 of Chapter 8.

Definition 9.1. The *thinness radius* of the geodesic triangle T is the number

$$\delta(T) := \max_{j=1,2,3} \left(\sup_{p \in \tau_j} d(p, \tau_{j+1} \cup \tau_{j+2}) \right),$$

A triangle T is called δ*-thin* if $\delta(T) \leqslant \delta$.

Definition 9.2 (Rips' definition of hyperbolicity). A geodesic hyperbolic space X is called δ-hyperbolic (in the sense of Rips) if every geodesic triangle T in X is δ-thin. The infimum of all δ's for which X is δ-hyperbolic is called the *hyperbolicity constant* of X.

A space X which is δ-hyperbolic for some $\delta < \infty$ is called *Rips–hyperbolic.* In what follows, we will refer to δ–hyperbolic spaces in the sense of Rips simply as being δ–hyperbolic.

Below are several simple but important geometric features of δ-hyperbolic spaces.

First of all, note that general Rips–hyperbolic metric spaces X are by no means uniquely geodesics; the notation xy used in what follows will mean that xy is *some* geodesic connecting x to y. The next lemma shows that geodesics in X between the given pair of points are "almost unique" and justifies, to some extent, the abuse of notation that we are committing.

LEMMA 9.3 (Thin bigon property). *If X is δ–hyperbolic, then all geodesics xy, zx with $d(y,z) \leqslant D$ are at Hausdorff distance at most $D+\delta$ from each other. In particular, if α, β are geodesics connecting points $x, y \in X$, then* $\mathrm{dist}_{Haus}(\alpha,\beta) \leqslant \delta$.

PROOF. Every point p on xy is, either at distance at distance at most δ from xz, or at distance at most δ from yz; in the latter case p is at distance at most $D+\delta$ from xz. □

Lemma 9.4 below is the *fellow-traveling property of hyperbolic geodesics*, which sharpens the conclusion of Lemma 9.3.

LEMMA 9.4 (Fellow-traveling property). *Let $\alpha(t), \beta(t)$ be geodesics in a δ-hyperbolic space X, such that $\alpha(0) = \beta(0) = o$ and $d(\alpha(t_0), \beta(t_0)) \leqslant D$ for some $t_0 \geqslant 0$. Then for all $t \in [0, t_0]$,*

$$d(\alpha(t), \beta(t)) \leqslant 2(D+\delta).$$

PROOF. By the previous lemma, for every $t \in [0, t_0]$ there exists $s \in [0, t_0]$ so that

$$d(\beta(t), \alpha(s)) \leqslant c = \delta + D.$$

Applying the triangle inequality, we see that

$$|t-s| \leqslant c,$$

hence, $d(\alpha(t), \beta(t)) \leqslant 2c = 2(\delta + D)$. □

The notion of thin triangles generalizes naturally to the concept of *thin polygons*. A *geodesic n-gon* P in a metric space X is a concatenation of geodesic segments $\sigma_i, i = 1, \ldots, n$, connecting points $p_i, i = 1, \ldots, n$, in the natural cyclic order. The points p_i are called the *vertices* of the polygon P and the geodesics σ_i are called the *sides* or the *edges* of P. A polygon P is called η-thin if every side of P is contained in the η-neighborhood of the union of the other sides.

EXERCISE 9.5. Suppose that X is a δ-hyperbolic metric space. Show that every n-gon in X is $\delta(n-2)$-thin. Hint: Triangulate an n-gon P by $n-3$ diagonals emanating from a single vertex. Now, use δ-thinness of triangles in X inductively.

We next improve the estimate provided by this exercise.

LEMMA 9.6 (thin polygons). *If X is δ–hyperbolic then, for $n \geqslant 2$, every geodesic $n+1$-gon in X is η_n-thin with*

$$\eta_n = \delta \lceil \log_2 n \rceil.$$

PROOF. First of all, it suffices to consider the case $n = 2^k$, for otherwise we add to the polygon edges of zero length until the number of sides reaches $2^k + 1$. We prove the estimate on thinness of (2^k+1)-gons by induction on k. For $k = 1$ the statement amounts to the δ-thinness of triangles. Suppose that k is at least 2 and that the thinness estimate holds for all $(2^{k-1}+1)$-gons. Take a geodesic $n+1$-gon P with the sides $\tau_i = p_i p_{i+1}$, $i = 0, \ldots, n-1$, $\tau = p_n p_0$. and consider the edge $\tau = \tau_n$ of P.

We subdivide P into three pieces by introducing the diagonals $p_0 p_m$ and $p_m p_n$, where $m = 2^{k-1}$. These pieces are two $2^m + 1$-gons and one triangle:

$$P' = p_0 p_1 \ldots p_m, \quad P'' = p_m p_{m+1} \ldots p_n, \quad T = p_0 p_m p_n.$$

By the induction hypothesis, the polygons P', P'' are $\delta(k-1)$-thin, while the triangle T is δ-thin. Therefore, τ is contained in the $\delta(k-1+1) = \delta k$-neighborhood of the union of the other sides of P. □

We now give some examples of Rips–hyperbolic metric spaces.

EXAMPLE 9.7. (1) Proposition 8.67 implies that $\mathbb{H}^n$ is δ-hyperbolic for $\delta = \operatorname{arccos}(\sqrt{2})$.

(2) Suppose that (X, d) is δ-hyperbolic and $a > 0$. Then the metric space $(X, a \cdot d)$ is $a\delta$-hyperbolic. Indeed, distances in $(X, a \cdot d)$ are obtained from distances in (X, d) by multiplication by a. Therefore, the same is true for distances between the edges of geodesic triangles.

(3) Let X_κ is the model surface of curvature $\kappa < 0$ as in §2.11.1. Then X_κ is δ-hyperbolic for

$$\delta_\kappa = |\kappa|^{-1/4} \operatorname{arccos}(\sqrt{2}).$$

Indeed, the Riemannian metric on X_κ is obtained by multiplying the Riemannian metric on $\mathbb{H}^2$ by $|\kappa|^{-1/2}$. This has the effect of multiplying all distances in $\mathbb{H}^2$ by $|\kappa|^{-1/4}$. Hence, if d is the distance function on $\mathbb{H}^2$ then $|\kappa|^{-1/4} d$ is the distance function on X_κ.

(4) Suppose that X is a $CAT(\kappa)$-space where $\kappa < 0$, see §2.11.1. Then X is δ_κ-hyperbolic. Indeed, all triangles in X are thinner then triangles in X_κ. Therefore, given a geodesic triangle T with the edges $\tau_i, i = 1, 2, 3$, and a point $p_1 \in \tau_1$ we take the comparison triangle $\tilde{T} \subset X_\kappa$ and the comparison point $\tilde{p}_1 \in \tilde{\tau}_1 \subset \tilde{T}$. Since $\tilde{T}$ is δ_κ-thin, there exists a point $\tilde{p}_i \in \tilde{\tau}_i, i = 2$ or $i = 3$, souch that $d(\tilde{p}_1, \tilde{p}_i) \leqslant \delta_\kappa$. Let $p_i \in \tau_i$ be the comparison point of $\tilde{p}_i$. By the comparison inequality,

$$d(p_1, p_i) \leqslant d(\tilde{p}_1, \tilde{p}_i) \leqslant \delta_\kappa,$$

and, hence, T is δ_κ-thin. In particular, if X is a simply-connected complete Riemannian manifold of sectional curvature $\leqslant \kappa < 0$, then X is δ_κ-hyperbolic.

(5) Let X be a simplicial tree, and d be a path-metric on X. Then, by the Exercise 2.58, X is $CAT(-\infty)$. Thus, by (4), X is δ_κ-hyperbolic for every $\delta_\kappa = |\kappa|^{-1/4} \operatorname{arccos}(\sqrt{2})$. Since

$$\inf_\kappa \delta_\kappa = 0,$$

it follows that X is 0-hyperbolic. Of course, this fact one can easily see directly by observing that every triangle in X is a tripod.

(6) Every geodesic metric space of diameter $\leqslant \delta < \infty$ is δ-hyperbolic.

EXERCISE 9.8. Let X be the circle of radius R in $\mathbb{R}^2$ with the induced path-metric d. Thus, (X, d) has diameter πR. Show that X is $\pi R/2$-hyperbolic and is not δ-hyperbolic for any $\delta < \pi R/2$.

Not every geodesic metric space is hyperbolic:

EXAMPLE 9.9. For instance, let us verify that the Euclidean plane $\mathbb{E}^2$ is not δ-hyperbolic for any δ. Pick a nondegenerate triangle $T \subset \mathbb{R}^2$. Then $\delta(T) = k > 0$ for some k. Therefore, if we scale T by a positive constant c, then $\delta(cT) = ck$. Sending $c \to \infty$, shows that $\mathbb{E}^2$ is not δ-hyperbolic for any $\delta > 0$. More generally, if a metric space X contains an isometrically embedded copy of $\mathbb{E}^2$, then X is not hyperbolic.

Here is an example of a metric space which is not hyperbolic, but does not contain a quasiisometrically embedded copy of $\mathbb{E}^2$ either. Consider the wedge X of countably many circles C_i each given with a path-metric of the overall length $2\pi i$, $i \in \mathbb{N}$. We equip X with the path-metric such that each C_i is isometrically embedded. Exercise 9.8 shows that X is not hyperbolic.

EXERCISE 9.10. Show that X contains no quasiisometrically embedded copy of $\mathbb{E}^2$. Hint: Use coarse topology.

More interesting examples of non-hyperbolic spaces containing no quasi-isometrically embedded copies of $\mathbb{E}^2$ are given by various solvable groups, e.g. the solvable Lie group Sol_3 and the Cayley graph of the Baumslag–Solitar group $BS(n, 1)$, see [**Bur99**].

Below we describe briefly another measure of thinness of triangles which can be used as an alternative definition of Rips–hyperbolicity. It is also related to the minimal size of triangles, described in Definition 5.107, consequently it is related to the filling area of the triangle *via* a Besikovitch–type inequality as described in Proposition 5.110.

DEFINITION 9.11. For a geodesic triangle $T \subset X$ with the sides τ_1, τ_2, τ_3, define the *inradius* of T to be

$$\Delta(T) := \inf_{x \in X} \max_{i=1,2,3} d(x, \tau_i).$$

In the case of the real-hyperbolic plane, as we saw in Lemma 8.66, this definition coincides with the radius of the largest circle inscribed in T. Clearly, $\Delta(T) \leqslant \delta(T)$ and

$$\Delta(T) \leqslant \mathrm{minsize}(T)$$

We next show that

$$\mathrm{minsize}(T) \leqslant 2\delta(T). \tag{9.1}$$

Indeed, suppose that $T = T(x_1, x_2, x_3)$ and the side τ_i of T connects x_i to x_{i+1} (i is taken modulo 3). Let a denote the length of τ_1.

Define the continuous function

$$f(t) = d(\tau_1(t), \tau_2) - d(\tau_1(t), \tau_3),$$

it takes the value $d(x_1, x_2) = a$ at $t = 0$ and the value $-a$ at $t = a$. Therefore, by the intermediate value theorem, there exists $t_1 \in [0, a]$ such that

$$d(\tau_1(t_1), \tau_2) = d(\tau_1(t_1), \tau_3) \leqslant \delta(T).$$

Taking $p_1 = \tau_1(t_1)$ and $p_i \in \tau_i, i = 2, 3$ to be the points closest to p_1, we get

$$d(p_1, p_2) \leqslant \delta, d(p_1, p_3) \leqslant \delta(T),$$

hence,

$$\text{minsize}(T) \leqslant 2\delta(T).$$

Hyperbolicity and combings. One might criticize Rips definition of hyperbolicity by observing that it is difficult to verify that the given geodesic metric space is hyperbolic, as Rips' definition requires one to identify geodesic segments in the space. The notion of *thin bicombing* below can be used to circumvent this problem; for instance, it was used successfully to verify hyperbolicity of the *curve complex* by Bowditch in [**Bow06b**].

Let Γ be a connected graph with the standard metric. A *combing* of Γ is a map c which associates to every pair of vertices u, v in Γ an edge-path p_{uv} in Γ connecting u to v. A combing c is called a *bicombing* if p_{uv} equals p_{vu} run in the reverse. A combing c is said to be *consistent* if for every vertices u, v in Γ and every integer subinterval $[t, s]$ in the domain of p_{uv}, the restriction of p_{uv} to $[t, s]$ equals $p_{u',v'}$, where $u' = p_{uv}(t), v' = p_{uv}(s)$. A combing is called *proper* if there is a constant C such that whenever $d(u, v) \leqslant 1$, we also have

$$\text{length}(p_{uv}) \leqslant C.$$

A bicombing is called *thin* it is consistent, proper and there exists a constant δ such that for every triple of vertices u, v, w in Γ there exists a vertex x within distance $\leqslant \delta$ from the images of all three paths

$$p_{uv}, \quad p_{vw}, \quad p_{wu}.$$

More generally, one defines the notion of a thin bicombing for a general metric space X by assuming that the paths p_{uv} connecting points of X are 1-Lipschitz and repeating the rest of the definition. It is now immediate that every Rips-hyperbolic metric space admits a thin bicombing, namely, the one given by geodesics. Conversely:

<u>Theorem</u> 9.12. *(U. Hamenstädt,* [**Ham07**, Proposition 3.5]*) If a geodesic metric space X (or a connected graph with the standard metric) admits a thin bicombing, then X is Rips-hyperbolic. Furthermore, the paths p_{uv} are $(L, 0)$-quasigeodesic for some L.*

9.2. Geometry and topology of real trees

In this section we consider in more detail a special class of hyperbolic spaces, the *real trees.* In view the Definition 2.59, a geodesic metric space is a real tree if and only if it is 0–hyperbolic.

Lemma 9.13. *If X is a real tree then any two points in X are connected by a unique topological arc in X.*

Proof. Let $D = d(x, y)$. Consider a topological arc, i.e., a continuous injective map $\alpha : [0, 1] \to X$, $x = \alpha(0), y = \alpha(1)$. Let $\alpha^* = xy, \alpha^* : [0, D] \to X$ be a geodesic connecting x to y. (This geodesic is unique by 0-hyperbolicity of X.) We claim that the image of α contains the image of α^*. Indeed, we can approximate α by piecewise-geodesic (nonembedded!) arcs

$$\alpha_n = x_0 x_1 \cup \ldots \cup x_{n-1} x_n, \quad x_0 = x, \quad x_n = y.$$

Since the $n + 1$-gon P in X, which is the concatenation of α_n with yx is 0-thin, $\alpha^* \subset \alpha_n$, cf. Lemma 9.6. Therefore, the image of α also contains the image of α^*.

Consider the continuous map $(\alpha^*)^{-1} \circ \alpha : [0, D] \to [0, D]$. Applying the intermediate value theorem to this function, we see that the images of α and α^* are equal. $\square$

EXERCISE 9.14. Prove the converse to this lemma: If X is a path-metric space where any two points are connected by a unique topological arc, then X is isometric to a real tree. In particular, if X is a path-metric space homeomorphic to a tree, then X is isometric to a tree.

We refer the reader to [**Bow91a**] for further discussion of characterizations of metric trees.

DEFINITION 9.15. Let T be a real tree and p be a point in T. The *space of directions* at p, denoted Σ_p, is defined as the space of germs of geodesics in T emanating from p, i.e., the quotient $\Sigma_p := \Re_p / \sim$, where

$$\Re_p = \{r : [0, a) \to T \mid a > 0,\ r \text{ is isometric, } r(0) = p\}$$

and

$$r_1 \sim r_2 \iff \exists\, \varepsilon > 0 \text{ such that } r_1|_{[0,\varepsilon)} \equiv r_2|_{[0,\varepsilon)}.$$

By Lemma 9.13, for every topological arc $c : [a, b] \to T$ in a tree, the image of c coincides with the geodesic segment $c(a)c(b)$. It follows that we may also define Σ_p as the space of germs of topological arcs instead of geodesic arcs.

DEFINITION 9.16. Define the *valence* $val(p)$ of a point p in a real tree T to be the cardinality of the set Σ_p. A *branch-point* of T is a point p of valence $\geqslant 3$. The *valence* of T is the supremum of valences of points in T.

EXERCISE 9.17. Show that $val(p)$ equals the number of connected components of $T \setminus \{p\}$.

DEFINITION 9.18. A real tree T is called *α–universal* if every real tree with valence at most α can be isometrically embedded into T.

We refer the reader to [**MNLGO92**] for a study of universal trees. In particular, the following holds:

THEOREM 9.19 ([**MNLGO92**]). *For every cardinal number $\alpha > 2$ there exists an α–universal tree, and it is unique up to isometry.*

9.3. Gromov hyperbolicity

One drawback of the Rips definition of hyperbolicity is that it uses geodesics. Below is an alternative definition of hyperbolicity, due to Gromov, where one needs to verify certain inequalities only for quadruples of points in a metric space (which need not be geodesic). Gromov's definition is less intuitive than the one of Rips, but, as we will see, it is more suitable in certain situations.

Let (X, dist) be a metric space (*which is no longer required to be geodesic*). Pick a base-point $p \in X$. For each $x \in X$ set $|x|_p := \text{dist}(x, p)$ and define the *Gromov product*

$$(x, y)_p := \frac{1}{2} \left(|x|_p + |y|_p - \text{dist}(x, y) \right).$$

Note that the triangle inequality immediately implies that $(x, y)_p \geqslant 0$ for all x, y, p; the Gromov product measures how far the triangle inequality for the points x, y, p is from being an equality.

REMARK 9.20. The Gromov product is a generalization of the inner product in vector spaces with p serving as the origin. For instance, suppose that $X = \mathbb{R}^n$ with the usual inner product, $p = 0$ and $|v|_p := \|v\|$ for $v \in \mathbb{R}^n$. Then

$$\frac{1}{2}\left(|x|_p^2 + |y|_p^2 - \|x-y\|^2\right) = x \cdot y.$$

EXERCISE 9.21. Suppose that X is a metric tree. Then $(x,y)_p$ is the distance $\mathrm{dist}(p,\gamma)$ from p to the geodesic segment $\gamma = xy$.

For general metric spaces general, a direct calculation using triangle inequalities shows that all points $p, x, y, z \in X$ satisfy the inequality

$$(p,x)_z + (p,y)_z \leqslant |z|_p - (x,y)_p$$

with the equality

$$(p,x)_z + (p,y)_z = |z|_p - (x,y)_p. \tag{9.2}$$

if and only $d(x,z) + d(z,y) = d(x,y)$. Thus, for every $z \in \gamma = xy$,

$$(x,y)_p = d(z,p) - (p,x)_z - (p,y)_z \leqslant d(z,p).$$

In particular, $(x,y)_p \leqslant \mathrm{dist}(p,\gamma)$.

LEMMA 9.22. *Suppose that X is δ–hyperbolic in the sense of Rips. Then the Gromov product in X is "comparable" to* $\mathrm{dist}(p,xy)$*: For every $x, y, p \in X$ and geodesic xy,*

$$(x,y)_p \leqslant \mathrm{dist}(p,xy) \leqslant (x,y)_p + 2\delta.$$

PROOF. The inequality $(x,y)_p \leqslant \mathrm{dist}(p,xy)$ was proved above; thus, we have to establish the other inequality. Note that since the triangle $T(p,x,y)$ is δ–thin, for each point $z \in xy$ we have

$$\min\{(x,p)_z, (y,p)_z\} \leqslant \min\{\mathrm{dist}(z,px), \mathrm{dist}(z,py)\} \leqslant \delta.$$

By continuity of the distance function, there exists a point $z \in xy$ such that $(x,p)_z, (y,p)_z \leqslant \delta$. By applying the equality (9.2) we get:

$$|z|_p - (x,y)_p = (p,x)_z + (p,y)_z \leqslant 2\delta.$$

Since $|z|_p \leqslant \mathrm{dist}(p,xy)$, we conclude that $\mathrm{dist}(p,xy) \leqslant (x,y)_p + 2\delta$. □

For each pointed metric space (X,p) we define its *Gromov-hyperbolicity constant* $\delta_p = \delta_p(X) \in [0,\infty]$ as

$$\delta_p := \sup\{\min((x,z)_p, (y,z)_p) - (x,y)_p\},$$

where the supremum is taken over all triples of points $x, y, z \in X$.

EXERCISE 9.23. If $\delta_p \leqslant \delta$ then $\delta_q \leqslant 2\delta$ for all $q \in X$.

DEFINITION 9.24. A metric space X is said to be *δ–hyperbolic in the sense of Gromov*, if $\delta_p \leqslant \delta < \infty$ for all $p \in X$. In other words, for every quadruple $x, y, z, p \in X$, we have

$$(x,y)_p \geqslant \min((x,z)_p, (y,z)_p) - \delta.$$

EXERCISE 9.25. The real line with the usual metric is 0-hyperbolic in the sense of Gromov.

EXERCISE 9.26. Each δ-hyperbolic space in the sense of Gromov satisfies

$$(x,u)_p \geqslant \min\{(x,y)_p, (x,z)_p, (z,u)_p\} - 2\delta$$

for all $x, y, z, u, p \in X$.

Computing Gromov-hyperbolicity constant for a given metric space is, typically, not an easy task. We will see that all real trees are 0-hyperbolic in Gromov's sense. It was recently proven by Nica and Spakula [**NS14**] that the Gromov-hyperbolicity constant for the hyperbolic plane $\mathbb{H}^2$ is $\log(2)$.

We next compare the two notions of hyperbolicity introduced so far.

LEMMA 9.27. *If a metric space X is δ–hyperbolic in the sense of Rips, then it is 3δ–hyperbolic in the sense of Gromov.*

PROOF. Consider points $x, y, z, p \in X$ and the geodesic triangle $T(x, y, z) \subset X$ with vertices x, y, z. Let $m \in xy$ be the point nearest to p. Then, since the triangle $T(x, y, z)$ is δ–thin, there exists a point $n \in xz \cup yz$ such that $\text{dist}(n, m) \leqslant \delta$. Assume that $n \in yz$. Then, by Lemma 9.22,

$$(y,z)_p \leqslant \text{dist}(p, yz) \leqslant \text{dist}(p, xy) + \delta.$$

On the other hand, by the same Lemma 9.22,

$$\text{dist}(p, xy) \leqslant (x,y)_p - 2\delta.$$

Combining these two inequalities, we obtain

$$(y,z)_p \leqslant (x,y)_p - 3\delta.$$

Therefore,

$$(x,y)_p \geqslant \min\left((x,z)_p, (y,z)_p\right) - 3\delta.$$

□

We now prove a "converse" to this lemma:

LEMMA 9.28. *If X is a geodesic metric space which is δ–hyperbolic in the sense Gromov, then X is 2δ–hyperbolic in the sense of Rips.*

PROOF. 1. We first show that in such space geodesics connecting any pair of points are "almost" unique, i.e., if α is a geodesic connecting x to y and p is a point in X such that

$$\text{dist}(x,p) + \text{dist}(p,y) \leqslant \text{dist}(x,y) + 2\delta$$

then $\text{dist}(p, \alpha) \leqslant 2\delta$. We suppose that $\text{dist}(p, x) \leqslant \text{dist}(p, y)$. If $\text{dist}(p, x) \geqslant \text{dist}(x, y)$ then $\text{dist}(x, y) \leqslant 2\delta$ and thus

$$\min(\text{dist}(p,x), p(y)) \leqslant 2\delta,$$

and we are done.

Therefore, assume that $\text{dist}(p, x) < \text{dist}(x, y)$ and let $z \in \alpha$ be such that $\text{dist}(z, y) = \text{dist}(p, y)$. Since X is δ–hyperbolic in the sense Gromov,

$$(x,y)_p \geqslant \min((x,z)_p, (y,z)_p) - \delta.$$

Thus we can assume that $(x, y)_p \geqslant (x, z)_p$. Then

$$\text{dist}(y,p) - \text{dist}(x,y) \geqslant \text{dist}(z,p) - \text{dist}(x,z) - 2\delta \iff$$

$$\text{dist}(z,p) \leqslant 2\delta.$$

We conclude that $\text{dist}(p, \alpha) \leqslant 2\delta$.

2. Consider now a geodesic triangle $T(x,y,p) \subset X$ and let $z \in xy$. Our goal is to show that z belongs to $\mathcal{N}_{4\delta}(px \cup py)$. We have:

$$(x,y)_p \geqslant \min((x,z)_p, (y,z)_p) - \delta.$$

Assume that $(x,y)_p \geqslant (x,z)_p - \delta$. Set $\alpha := py$. We will show that $z \in \mathcal{N}_{2\delta}(\alpha)$.

By combining $\mathrm{dist}(x,z) + \mathrm{dist}(y,z) = \mathrm{dist}(x,y)$ and $(x,y)_p \geqslant (x,z)_p - \delta$, we obtain

$$\mathrm{dist}(y,p) \geqslant \mathrm{dist}(y,z) + \mathrm{dist}(z,p) - 2\delta.$$

Therefore, by Part 1, $z \in \mathcal{N}_{2\delta}(\alpha)$ and hence the triangle $T(x,y,z)$ is 2δ–thin. □

Corollary 9.29 (M. Gromov, **[Gro87]**, section 6.3C.). *For geodesic metric spaces, Gromov–hyperbolicity is equivalent to Rips–hyperbolicity.*

Another corollary of the Lemmata 9.27 and 9.28 is:

Corollary 9.30. *A geodesic metric space is a real tree if and only if it is 0-hyperbolic in the sense of Gromov.*

This corollary has a "converse" (see e.g. **[Dre84]** or **[GdlH90**, Ch. 2, Proposition 6**]**):

Theorem 9.31. *Every 0-hyperbolic metric space (in the sense of Gromov) admits an isometric embedding into a tree.*

Furthermore:

Theorem 9.32 (M. Bonk, O. Schramm **[BS00]**). *Every δ-hyperbolic metric space (in the sense of Gromov) admits an isometric embedding into a geodesic metric space which is also δ-hyperbolic.*

Question 9.33. Does there exist a $\aleph$-*quasiuniversal* δ-hyperbolic space, i.e., a Gromov-hyperbolic metric space X such that every δ-hyperbolic metric space Y of cardinality $\leqslant \aleph$, admits an (L,A) quasiisometric embedding into X, with L and A depending only on δ?

A partial positive answer to this question is provided by a universality theorem of Bonk and Schramm **[BS00]**, see Theorem 9.207.

We next consider behavior of hyperbolicity under quasiisometries.

Exercise 9.34. Gromov–hyperbolicity is invariant under $(1,A)$-quasiisometries.

Exercise 9.35. Let X be a metric space and $N \subset X$ be an R-net. Show that the embedding $N \hookrightarrow X$ is an $(1,R)$-quasiisometry. Thus, X is Gromov–hyperbolic if and only if N is Gromov–hyperbolic. In particular, a group (G, d_S) with word metric d_S is Gromov–hyperbolic if and only if the Cayley graph $\Gamma_{G,S}$ of G is Rips–hyperbolic.

The drawback is that for general nongeodesic metric spaces, Gromov–hyperbolicity fails to be QI invariant:

Example 9.36 (Gromov–hyperbolicity is not QI invariant). This example is taken from **[Väi05]**. Consider the graph X of the function $y = |x|$, where the metric on X is the restriction of the metric on $\mathbb{R}^2$. (This is not a path-metric!) Then the map $f : \mathbb{R} \to X, f(x) = (x, |x|)$ is a quasiisometry:

$$|x - x'| \leqslant d(f(x), f(x')) \leqslant \sqrt{2}|x - x'|.$$

Let $p = (0, 0)$ be the base-point in X and for $t > 0$ we let $x := (2t, 2t)$, $y := (-2t, 2t)$ and $z := (t, t)$. The reader will verify that

$$\min((x, z)_p, (y, z)_p) - (x, y)_p) = t\left(\frac{7\sqrt{2}}{2} - 3\right) > t.$$

Therefore, the quantity $\min((x, z)_p, (y, z)_p) - (x, y)_p)$ is unbounded from above as $t \to \infty$ and hence X is not δ-hyperbolic for any $\delta < \infty$. In particular, X is QI to a Gromov–hyperbolic space $\mathbb{R}$, but is not Gromov–hyperbolic itself. We will see, as a corollary of Morse Lemma (Corollary 9.43), that in the context of geodesic spaces, hyperbolicity is a QI invariant.

9.4. Ultralimits and stability of geodesics in Rips–hyperbolic spaces

In this section we will see that every hyperbolic geodesic metric spaces X *asymptotically* resembles a tree. This property will be used to prove *Morse Lemma*, which establishes that quasigeodesics in δ-hyperbolic spaces are uniformly close to geodesics.

Lemma 9.37. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of geodesic δ_i–hyperbolic spaces with δ_i tending to 0. Then for every non-principal ultrafilter ω each component of the ultralimit X_ω is a metric tree.*

Proof. First, according to Lemma 7.51, ultralimit of geodesic metric spaces is again a geodesic metric space. Thus, in view of Lemma 9.28, it suffices to verify that X_ω is 0-hyperbolic in the sense of Gromov (since it will be 0-hyperbolic in the sense of Rips and, hence, a metric tree). This is one of the few cases where Gromov–hyperbolicity is superior to Rips–hyperbolicity: It suffices to check 0-hyperbolicity condition only for quadruples of points.

We know that for every quadruple x_i, y_i, z_i, p_i in X_i,

$$(x_i, y_i)_{p_i} \geqslant \min((x_i, z_i)_{p_i}, (y_i, z_i)_{p_i}) - \delta_i.$$

By taking the ultralimit of this inequality, we obtain (for every quadruple of points x_ω, y_ω, z_ω, p_ω in X_ω):

$$(x_\omega, y_\omega)_{p_\omega} \geqslant \min((x_\omega, z_\omega)_{p_\omega}, (y_\omega, z_\omega)_{p_\omega}),$$

since $\omega\text{-}\lim \delta_i = 0$. Thus, X_ω is 0-hyperbolic. □

Corollary 9.38. *Every geodesic in the tree X_ω is a limit geodesic.*

Proof. 1. Suppose first that $x_\omega y_\omega$ is a geodesic segment in X_ω, $x_\omega = (x_i), y_\omega = (y_i)$. Then the ultralimit of geodesic segments $x_i y_i \subset X_i$ is a geodesic segment connecting x_ω to y_ω. Since each component of X_ω is 0-hyperbolic, it is uniquely geodesic, i.e., there exists a unique geodesic segment connecting x_ω to y_ω.

2. We consider the case of biinfinite geodesics in X_ω and leave the proof for geodesic rays to the reader. Let $l_\omega \subset X_\omega$ be a biinfinite geodesic parameterized by the isometric embedding $\gamma_\omega : \mathbb{R} \to X_\omega$. Take the points $x_{\omega,n} := \gamma_\omega(n), y_{\omega,n} = \gamma_\omega(-n)$. For each n the finite geodesic segment $x_{\omega,n}y_{\omega,n}$ is the ultralimit of geodesic segments $x_{i,n}y_{i,n} \subset X_i$. Then the entire l_ω is the ultralimit of the sequence of geodesic segments $x_{i,i}y_{i,i}$. □

EXERCISE 9.39. Find a flaw in the following "proof" of Lemma 9.37: Since X_i is δ_i-hyperbolic, it follows that every geodesic triangle T_i in X_i is δ_i-thin. Suppose that $\omega\text{-lim}\, d(x_i, e_i) < \infty$, $\omega\text{-lim}\, d(p_i, e_i) < \infty$. Taking the limit in the definition of thinness of triangles, we conclude that the ultralimit of triangles $T_\omega = \omega\text{-lim}\, T_i \subset X_\pm$ is 0-thin. Therefore, every geodesic triangle in X_ω is 0-thin.

The following fundamental theorem in the theory of hyperbolic spaces is called *Morse Lemma* or *stability of hyperbolic geodesics*.

THEOREM 9.40 (Morse Lemma). *There exists a function $D = D(L, A, \delta)$, such that the following holds. If X be a δ–hyperbolic geodesic space, then for every (L, A)–quasigeodesic $f : [a, b] \to X$, the Hausdorff distance between the image of f and a geodesic segment $f(a)f(b) \subset X$ is at most D.*

PROOF. Set $c = d(f(a), f(b))$. Given a quasigeodesic f and $f^* : [0, c] \to X$ parameterizing the geodesic $f(a)f(b)$, we define two numbers:

$$D_f = \sup_{t\in[a,b]} d(f(t), Im(f^*))$$

and

$$D_f^* = \sup_{t\in[0,c]} d(f^*(t), Im(f)).$$

Then

$$\text{dist}_{Haus}(Im(f), Im(f^*)) = \max(D_f, D_f^*).$$

We will prove that D_f is uniformly bounded in terms of L, A, δ, since the proof for D_f^* is completely analogous.

Suppose that the quantities D_f are not uniformly bounded, that is, exists a sequence of (L, A)–quasigeodesics $f_n : [-n, n] \to X_n$ in δ-hyperbolic geodesic metric spaces X_n, such that

$$\lim_{n\to\infty} D_n = \infty,$$

where $D_n = D_{f_n}$. Pick points $t_n \in [-n, n]$ such that for $\gamma_n^* = f_n^*([-n, n])$ and $x_n := f_n(t_n)$, we have:

$$|\text{dist}(x_n, \gamma_n^*) - D_n| \leqslant 1.$$

In other words, the points x_n "almost" realize the maximal distance between the points of $f_n([-n, n])$ and the geodesic γ_n^*.

Define the sequence $\boldsymbol{\lambda}$ of scaling factors

$$\lambda_n = \frac{1}{D_n}.$$

As in Lemma 7.82, we consider two sequences of pointed metric spaces

$$(\lambda_n X_n, x_n), \quad (\lambda_n[-n, n], t_n).$$

Note that the ultralimit $\omega\text{-lim}\, \frac{n}{D_n}$ could be infinite, however, it cannot be zero. Let

$$(X_\omega, x_\omega) = \omega\text{-lim}\, (\lambda_n X_n, x_n)$$

and

$$(Y, y) := \omega\text{-lim}\, (\lambda_n[-n, n], t_n).$$

The metric space Y is either a nondegenerate segment in $\mathbb{R}$ or a closed geodesic ray in $\mathbb{R}$ or the whole real line. Note that the distance from the points of the image of f_n to γ_n^* in the rescaled metric space $\lambda_n X_n$ is at most $1 + \lambda_n$. Each map

$$f_n : Y_n \to \lambda_n X_n$$

is an $(L, A\lambda_n)$–quasigeodesic. Therefore, the ultralimit

$$f_\omega = \omega\text{-lim}\, f_n : (Y, y) \to (X_\omega, x_\omega)$$

is an L-bi-Lipschitz map (cf. Lemma 7.82). In particular this map is a continuous embedding and the image of f_ω is a geodesic γ in X_ω, see Lemma 9.13.

On the other hand, the sequence of geodesic segments $\gamma_n^* \subset \lambda_n X_n$ also ω–converges to a geodesic $\gamma^* \subset X_\omega$, this geodesic is either a finite geodesic segment or a geodesic ray or a complete geodesic. In any case, by our choice of the points x_n, γ is contained in the 1-neighborhood of the geodesic γ^* and, at the same time, $\gamma \neq \gamma^*$ since $x_\omega \in \gamma \setminus \gamma^*$.

The last step of the proof is to get a contradiction with the fact that X_ω is a real tree. If γ^* is a finite geodesic, it connects the end-points of the geodesic γ, thereby contradicting the fact that each metric tree is uniquely geodesic. Suppose that γ^* is a complete geodesic. We then pick two points $y_\omega, z_\omega \in \gamma$ such that x_ω is the midpoint of the geodesic segment $y_\omega z_\omega$, while the distance between y_ω and z_ω is sufficiently high, say, larger than 4. We next find points $y'_\omega \in \gamma^*, z'_\omega \in \gamma^*$ within distance $\leqslant 1$ from y_ω and z_ω respectively. Since

$$d(y_\omega, x_\omega) = d(z_\omega, x_\omega) \geqslant 2,$$

the point x_ω does not belong to union

$$y_\omega y'_\omega \cup z_\omega z'_\omega.$$

At the same time, x_ω does not belong to the geodesic segment $y'_\omega z'_\omega$ since the latter is contained in the geodesic γ^*. Therefore, the side $y_\omega z_\omega$ of the geodesic quadrilateral Q with the vertices

$$y_\omega,\ z_\omega,\ z'_\omega,\ y'_\omega$$

is not contained in the union of the three other sides. This contradicts the fact that Q is 0-thin. We proof in the remaining case, when γ is a geodesic ray is similar and is left to the reader. □

Historical Remark 9.41. The first version of this theorem was proven by Morse in [**Mor24**] in the following setting. Consider a compact surface S equipped with two Riemannian metrics g_1, g_2 of negative curvature. Now, lift, the metrics g_1, g_2 to the universal cover of S. Then each geodesic with respect to the lift $\tilde{g}_1$ of g_1 is a (uniform) quasigeodesic with respect to the lift $\tilde{g}_2$ of g_2. Morse proved that all geodesics with respect to $\tilde{g}_1$ are uniformly close to the geodesics with respect to $\tilde{g}_2$, as long as their end-points are the same. Later on, Busemann, [**Bus65**], proved a version of this lemma in the case of $\mathbb{H}^n$, where metrics in question were not necessarily Riemannian. A version in terms of quasigeodesics is due to Mostow [**Mos73**], in the context of negatively curved symmetric spaces, although his proof is general. The first proof for general δ-hyperbolic geodesic metric spaces is due to Gromov, see [**Gro87**, 7.2.A]. Of course, neither Morse, nor Busemann, nor Mostow, nor Gromov used ultralimits: Their proofs were based on an analysis of nearest-point projections to geodesics. We will give an effective proof of the Morse Lemma in Section 9.10.

Remark 9.42. Stability of geodesics fails in the Eucldiean plane $\mathbb{E}^2$ and, hence, for general $CAT(0)$ spaces. Nevertheless, some versions of the Morse Lemma remain true for non-hyperbolic $CAT(0)$ spaces, see [**Sul14**] and [**MK**].

COROLLARY 9.43 (QI invariance of hyperbolicity). *Suppose that X, X' are quasi-isometric geodesic metric spaces and X' is hyperbolic. Then X is also hyperbolic.*

PROOF. Suppose that X' is δ'-hyperbolic and $f : X \to X'$ is an (L, A)–quasiisometry and $f' : X' \to X$ is its quasiinverse. Pick a geodesic triangle $T \subset X$. Its image under f is a quasigeodesic triangle S in X' whose sides are (L, A)–quasigeodesic. Therefore each of the quasigeodesic sides σ_i of S is within distance $\leqslant D = D(L, A, \delta')$ from a geodesic σ_i^* connecting the end-points of this side. See Figure 9.1. The geodesic triangle S^* formed by the segments $\sigma_1^*, \sigma_2^*, \sigma_3^*$ is δ'-thin. Therefore, the quasigeodesic triangle $f'(S^*) \subset X$ is $\epsilon := L\delta' + A$–thin, i.e. each quasigeodesic $\tau_i' := f'(\sigma_i^*)$ is within distance $\leqslant \epsilon$ from the union τ_{i-1}', τ_{i+1}'. However,

$$dist_{Haus}(\tau_i, \tau_i') \leqslant LD + 2A.$$

Putting this all together, we conclude that the triangle T is δ-thin with

$$\delta = 2(LD + 2A) + \epsilon = 2(LD + 2A) + L\delta' + A. \quad \square$$

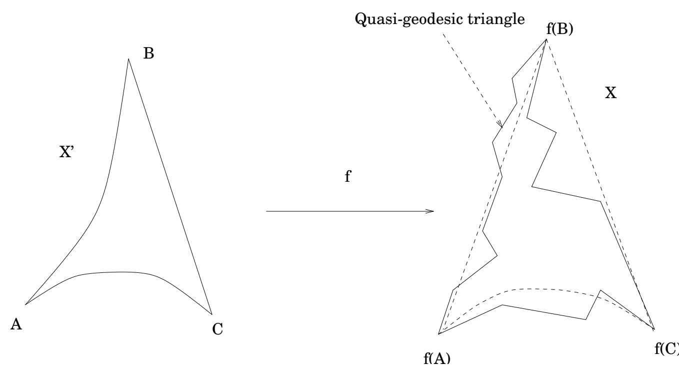


FIGURE 9.1. Quasiisometric image of a geodesic triangle.

Observe that in Morse Lemma, we are not claiming, of course, that the distance $d(f(t), f^*(t))$ is uniformly bounded, only that for every t there exist s and s^* such that

$$d(f(t), f^*(s)) \leqslant D,$$

and

$$d(f^*(t), f(s^*)) \leqslant D.$$

Here $s = s(t), s^* = s^*(t)$. However, applying triangle inequalities we get for $B = A + D$ the following estimates:

$$L^{-1}t - B \leqslant s \leqslant Lt + B \tag{9.3}$$

and

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B) \tag{9.4}$$

Lastly, we note that Proposition 9.158 proven later on, which characterizes hyperbolic geodesic metric spaces as the ones for which every asymptotic cone is a tree, provides an alternative proof of QI invariance of hyperbolicity: If two quasiisometric metric spaces have bi-Lipschitz homeomorphic asymptotic cones and one cone is a tree, then the other cone is also a tree.

9.5. Local geodesics in hyperbolic spaces

A map $p : I \to X$ of an interval $I \subset \mathbb{R}$ into a metric space X, is called a *k-local geodesic* if the restriction of p to each length k subinterval $I' \subset I$ is an isometric embedding. The notion of local geodesics is in line with the concept of geodesics in Riemannian geometry: (Unit speed) Riemannian geodesics are not required to be isometric embeddings, but *locally* they always are. If M is a Riemannian manifold with injectivity radius $\geqslant \epsilon > 0$, then every unit speed Riemannian geodesic in M is an ϵ-local geodesic in the metric sense.

EXERCISE 9.44. Suppose that X is a real tree and k is a positive number. Then each k-local geodesic in X is a geodesic.

A *coarse version* of this exercise works for general hyperbolic spaces as well, it is due to Gromov [**Gro87**, ???]:

THEOREM 9.45 (Local geodesics are uniform quasigeodesics). *Suppose that X is a δ-hyperbolic geodesic metric space in the sense of Rips, $\delta > 0$. Then for $k = 6\delta$, each k-local geodesic in X is a $(3, 4\delta)$-quasigeodesic.*

PROOF. For each pair of points $x, y \in X$ we partition X in two *half-spaces*

$$\mathcal{D}(x, y) = \{z \in X : d(x, z) \leqslant d(z, y)\}, \quad \mathcal{D}(y, x) = \{z \in X : d(y, z) \leqslant d(z, x)\}.$$

The intersection of these half-spaces is the *bisector* $Bis(x, y)$ of the pair (x, y), consisting of all points equidistant from x and y.

FIGURE 9.2. Bisector and half-spaces.

The key to the proof is the following:

LEMMA 9.46. *Consider three points $x_0, x_1, x_2 \in X$ such that*

$$\epsilon = 3\delta = d(x_0, x_1) = d(x_1, x_2) = \frac{1}{2} d(x_0, x_2).$$

Then:

1. $\mathrm{dist}(x_0, \mathcal{D}(x_1, x_0)) \geqslant \delta, \quad \mathrm{dist}(x_2, \mathcal{D}(x_1, x_2)) \geqslant \delta.$
2. *The half-space $\mathcal{D}(x_1, x_0)$ contains $\mathcal{D}(x_2, x_1)$ and, moreover,*

$$\mathrm{dist}(\mathcal{D}(x_0, x_1), \mathcal{D}(x_2, x_1)) \geqslant \delta/2.$$

PROOF. 1. Let y be a point in the bisector $Bis(x_0, x_1)$ nearest to x_0 and let $m \in x_0x_1$ be the midpoint of a geodesic segment x_0x_1 connecting x_0 to x_1. Since the geodesic triangle $\Delta(x_0, y, x_1)$ is δ-thin, the distance from m to one of the two sides x_0y, yx_1 of this triangle does not exceed δ. We will assume that this side is x_0y, since the other case is obtained by relabeling.

FIGURE 9.3. Nested half-spaces.

Let $z \in x_0y$ be the point closest to m. Then, by the triangle inequality,

$$d(x_0, y) \geqslant d(x_0, z) \geqslant d(x_0, m) - \delta = \frac{\epsilon}{2} - \delta.$$

Since $\epsilon = 3\delta$, we obtain

$$\text{dist}(x, Bis(x_0, x_1)) = d(x_0, y) \geqslant \delta/2.$$

2. Take points $y_1 \in \mathcal{D}(x_0, x_1), \quad y_2 \in \mathcal{D}(x_2, x_1)$, i.e.,

$$D_0 = d(x_0, y_1) \leqslant D_1 = d(y_1, x_1), \quad D_3 = d(x_2, y_2) \leqslant D_2 = d(y_2, x_1).$$

Our goal is to estimate the distance

$$\eta = d(y_1, y_2)$$

from below, this will provide a lower bound on the distance between the half-spaces. We let u denote the midpoint of a geodesic segment y_1y_2. Then

$$d(x_0, y_1) \leqslant D_0 + \frac{\eta}{2}, \quad d(x_2, y_2) \leqslant D_3 + \frac{\eta}{2}.$$

We also have

$$D = d(u, x_1) \geqslant \max(D_1 - \frac{\eta}{2}, D_2 - \frac{\eta}{2}).$$

FIGURE 9.4. Estimating distance between half-spaces.

Since the triangle $\Delta(x_0, u, x_2)$ is δ-thin, and, by the hypothesis of Lemma, x_1 lies on the geodesic segment x_0x_2, the distance from x_1 to one of the sides x_0u, x_2u of this triangle is at most δ. We will assume that this side is x_0u. Let $v \in x_0u$ denote the point closest to x_1; this point divides the segment x_0u into two subsegments vu, vx_0 of the lengths D_1', D_1'' respectively. We have

$$D_1' \geqslant D - \delta \geqslant D_1 - \delta - \frac{\eta}{2} \geqslant D_0 - \delta - \frac{\eta}{2},$$

which implies

$$D_1'' = d(x_0, u) - D_1' \leqslant (D_0 + \frac{\eta}{2}) - (D_0 - \delta - \frac{\eta}{2}) = \eta + \delta.$$

Combining this with the triangle inequality for $\Delta(x_0, v, x_1)$, we obtain:

$$\epsilon \leqslant D_1'' + \delta \leqslant \eta + 2\delta,$$

implying

$$\eta \geqslant \epsilon - 2\delta = 3\delta - 2\delta = \delta. \quad \square$$

We now can prove the theorem. Let q be a 6δ-local geodesic in X. We first consider the special case when q has length $n\epsilon$, $n \in \mathbb{N}$ (where, as before, $\epsilon = 3\delta$) and estimate from below the distance between the end-points of q in terms of the length of q. We subdivide q into n subsegments

$$x_0x_1, x_1x_2, \ldots, x_{n-1}x_n$$

of length ϵ. Since q is a $k = 2\epsilon$-local geodesic, the unions

$$x_{i-1}x_i \cup x_ix_{i+1}, \quad i = 1, \ldots, n-1$$

are geodesic segments in X. Therefore, by Lemma 9.46,

$$\text{dist}(\mathcal{D}(x_{i-1}, x_i), \mathcal{D}(x_{i+1}, x_i)) \geqslant \delta$$

for each i. Furthermore, the distances from x_0 to $\mathcal{D}(x_2, x_1)$ and from x_n to $\mathcal{D}(x_{n-1}, x_n)$ are at least $\delta/2$. Thus, every path q' connecting x_0 to x_n has length at least

$$\frac{\delta}{2} + (n-1)\delta + \frac{\delta}{2} = n\delta,$$

which is the length of q divided by 3.

FIGURE 9.5. Estimating the distance between the end-points of a local geodesic.

Consider now the general case. Suppose that p is a k-local geodesic in X, x_0, x are points on p such that the length ℓ of p between them equals $n\epsilon + \sigma$, $0 < \sigma < \epsilon$. We represent the portion of p between x_0, x as the concatenation of two sub-paths: q (of length $n\epsilon$, connecting x_0 to x_n) and r (of length σ, connecting x_n to x). Then (by the special case considered above)

$$3d(x_0, x_n) \geqslant \text{length}(q) \geqslant \ell - \epsilon,$$

$$3d(x_0, x) > 3d(x_0, x_n) - 3\epsilon \geqslant \ell - 4\epsilon,$$

$$\frac{1}{3}\ell - \frac{4}{3}\epsilon = \frac{1}{3}\ell - 4\delta < d(x_0, x).$$

Hence, p is a $(3, 4\delta)$-quasigeodesic in X. $\square$

9.6. Quasiconvexity in hyperbolic spaces

The usual notion of convexity is not particularly useful in the context of hyperbolic geodesic metric spaces, it is replaced with the one of *quasiconvexity*.

DEFINITION 9.47. Let X be a geodesic metric space and $Y \subset X$. Then the *quasiconvex hull* $H(Y)$ of Y in X is the union of all geodesics $y_1 y_2 \subset X$, with the end-points y_1, y_2 contained in Y.

Accordingly, a subset $Y \subset X$ is called R-*quasiconvex* if $H(Y) \subset \mathcal{N}_R(Y)$. A subset Y is called *quasiconvex* if it is quasiconvex for some $R < \infty$.

Let X be a δ-hyperbolic geodesic metric space.

The thin triangle property immediately implies that subsets of a δ-hyperbolic geodesic metric space satisfy:

1. Every metric ball $B(x, R)$ in is δ-quasiconvex.

2. Suppose that $Y_i \subset X$ is R_i-quasiconvex, $i = 1, 2$, and $Y_1 \cap Y_2 \neq \emptyset$. Then $Y_1 \cup Y_2$ is $R_1 + R_2 + \delta$-quasiconvex.

Thus, quasiconvex subsets behave somewhat differently from the convex ones, since the union of convex sets (with nonempty intersection) need not be convex.

An example of a non-quasiconvex subset is a horosphere in $\mathbb{H}^n$: Its quasiconvex hull is the horoball bounded by this horosphere.

EXERCISE 9.48. The quasiconvex hull of any subset $Y \subset X$ of a δ-hyperbolic geodesic metric space, is 2δ-quasiconvex in X. Hint: Use the fact that quadrilaterals in X are 2δ-thin.

Thus, quasiconvex hulls are quasiconvex.

The following results connect quasiconvexity and quasiisometry for subsets of Gromov–hyperbolic geodesic metric spaces.

THEOREM 9.49. *Let X, Y be geodesic metric spaces, such that X is δ-hyperbolic geodesic metric space. Then for every quasiisometric embedding $f : Y \to X$, the image $f(Y)$ is quasiconvex in X.*

PROOF. Let $y_1, y_2 \in Y$ and $\alpha = y_1 y_2 \subset Y$ be a geodesic connecting y_1 to y_2. Since f is an (L, A) quasiisometric embedding, $\beta = f(\alpha)$ is an (L, A) quasigeodesic in X. By the Morse Lemma,

$$dist_{Haus}(\beta, \beta^*) \leqslant R = D(L, A, \delta),$$

where β^* is any geodesic in X connecting $x_1 = f(y_1)$ to $x_2 = f(y_2)$. Therefore, $\beta^* \subset \mathcal{N}_R(f(Y)$, and $f(Y)$ is R-quasiconvex. □

The map $f : Y \to f(Y)$ is a quasiisometry, where we use the restriction of the metric from X to define a metric on $f(Y)$. Of course, $f(Y)$ is not a geodesic metric space, but it is quasiconvex; thus, applying the same arguments as in the proof of Theorem 9.43, we conclude that Y is also hyperbolic.

Conversely, let $Y \subset X$ be a coarsely connected subset, i.e., there exists a constant $c < \infty$ such that the complex $\mathrm{Rips}_C(Y)$ is connected for all $C \geqslant c$, where we again use the restriction of the metric d from X to Y to define the Rips complex. Then we define a path-metric $d_{Y,C}$ on Y by looking at infima of lengths of edge-paths in $Rips_C(Y)$ connecting points of Y. The following is a converse to Theorem 9.49:

THEOREM 9.50. *Suppose that $Y \subset X$ is coarsely connected and Y is quasiconvex in X. Then the identity map $f : (Y, d_{Y,C}) \to (X, \mathrm{dist}_X)$ is a quasiisometric embedding for all $C \geqslant 2c + 1$.*

PROOF. Let C be such that $H(Y) \subset \mathcal{N}_C(Y)$. First, if $d_Y(y, y') \leqslant C$ then $\mathrm{dist}_X(y, y') \leqslant C$ as well. Hence, f is coarsely Lipschitz. Let $y, y' \in Y$ and γ is a geodesic in X of length L connecting y, y'. Subdivide γ into $n = \lfloor L \rfloor$ of unit subsegments and a subsegment of the length $L - n$:

$$z_0 z_1, \ldots, z_{n-1} z_n, \quad z_n, z_{n+1},$$

where $z_0 = y, z_{n+1} = y'$. Since each z_i belongs to $\mathcal{N}_c(Y)$, there exist points $y_i \in Y$ such that $\mathrm{dist}_X(y_i, z_i) \leqslant c$, where we take $y_0 = z_0, y_{n+1} = z_{n+1}$. Then

$$\mathrm{dist}_X(z_i, z_{i+1}) \leqslant 2c + 1 \leqslant C$$

and, hence, z_i, z_{i+1} are connected by an edge (of length C) in $\mathrm{Rips}_C(Y)$. Now it is clear that

$$d_{Y,C}(y, y') \leqslant C(n + 1) \leqslant C\mathrm{dist}_X(y, y') + C. \quad \square$$

REMARK 9.51. It is proven in [**Bow94**] that in the context of subsets of negatively pinched complete simply-connected Riemannian manifolds X, quasiconvex hulls $Hull(Y)$ are essentially the same as convex hulls $H(Y)$:

There exists a function $L = L(C)$ such that for every C-quasiconvex subset $Y \subset X$,

$$H(Y) \subset Hull(Y) \subset \mathcal{N}_{L(C)}(Y).$$

9.7. Nearest-point projections

In general, nearest-point projections to geodesics in δ-hyperbolic geodesic spaces are not well defined. The following lemma shows, nevertheless, that they are *coarsely-well defined*:

Let γ be a geodesic in δ-hyperbolic geodesic space X. For a point $x \in X$ let $p = \pi_\gamma(x)$ be a point nearest to x.

LEMMA 9.52. *Let $p' \in \gamma$ be such that $d(x, p') < d(x, p) + R$. Then*

$$d(p, p') \leqslant 2(R + 2\delta).$$

In particular, if $p, p' \in \gamma$ are both nearest to x then

$$d(p, p') \leqslant 4\delta.$$

PROOF. Consider the geodesics α, α' connecting x to p and p' respectively. Let $q' \in \alpha'$ be the point within distance $\delta + R$ from p' (this point exists unless $d(x, p) < \delta + R$ in which case $d(p, p') \leqslant 2(\delta + R)$ by the triangle inequality). Since the triangle $\Delta(x, p, p')$ is δ-thin, there exists a point

$$q \in xp \cup pp' \subset xp \cup \gamma$$

within distance δ from q. If $q \in \gamma$, we obtain a contradiction with the fact that the point p is nearest to x on γ (the point q will be closer). Thus, $q \in xp$. By the triangle inequality

$$d(x, p') - (R + \delta) = d(x, q') \leqslant d(x, q) + \delta \leqslant d(x, p) - d(q, p) + \delta.$$

Thus,

$$d(q, p) \leqslant d(x, p) - d(x, p') + R + 2\delta \leqslant R + 2\delta.$$

Since $d(p', q) \leqslant R + 2\delta$, we obtain $d(p', p) \leqslant 2(R + 2\delta)$. □

This lemma can be strengthened, we now show that the nearest-point projection to a quasigeodesic subspace in a hyperbolic space is coarse Lipschitz:

LEMMA 9.53. *Let $X' \subset X$ be an R-quasiconvex subset. Then the nearest-point projection $\pi = \pi_{X'} : X \to X'$ is $(2, 2R + 9\delta)$-coarse Lipschitz.*

PROOF. Suppose that $x, y \in X$ such that $d(x, y) = D$. Let $x' = \pi(x), y' = \pi(y)$. Consider the quadrilateral formed by geodesic segments $xy \cup [y, y'], [y', x'] \cup [x', x]$. Since this quadrilateral is 2δ-thin, there exists a point $q \in x'y'$ which is within distance $\leqslant 2\delta$ from $x'x \cup xy$ and $xy \cup yy$.

Case 1. We first assume that there are points $x'' \in xx', y'' \in yy$ such that

$$d(q, x'') \leqslant 2\delta, d(q, y'') \leqslant 2\delta.$$

Let $q' \in X'$ be a point within distance $\leqslant R$ from q. By considering the paths

$$xx'' \cup x''q \cup qq', \quad yy'' \cup y''q \cup qq'$$

and using the fact that $x' = \pi(x), y' = \pi(y)$, we conclude that

$$d(x', x'') \leqslant R + 2\delta, \quad d(y', y'') \leqslant R + 2\delta.$$

Therefore,

$$d(x', y') \leqslant 2R + 9\delta.$$

Case 2. Suppose that there exists a point $q'' \in xy$ such that $d(q, q'') \leqslant 2\delta$. Setting $D_1 = d(x, q''), D_2 = d(y, q'')$, we obtain

$$d(x, x') \leqslant d(x, q') \leqslant D_1 + R + 2\delta, d(y, y') \leqslant d(y, q') \leqslant D_2 + R + 2\delta$$

which implies that

$$d(x', y') \leqslant 2D + 2R + 4\delta.$$

In either case, $d(x', y') \leqslant 2d(x, y) + 2R + 9\delta$. □

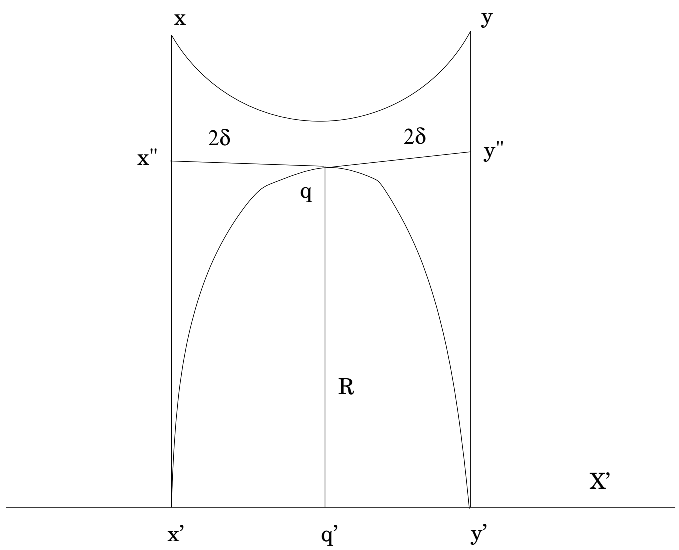


FIGURE 9.6. Projection to a quasiconvex subset.

9.8. Geometry of triangles in Rips–hyperbolic spaces

In the case of real-hyperbolic space we relied upon hyperbolic trigonometry in order to study geodesic triangles. Trigonometry no longer makes sense in the context of Rips–hyperbolic spaces X, so instead one compares geodesic triangles in X to geodesic triangles in real trees, i.e., to tripods, in the manner similar to the comparison theorems for $CAT(\kappa)$-spaces. In this section we describe comparison maps to tripods, called *collapsing maps*. We will see that such maps are $(1, 14\delta)$-quasiisometries. We will use the collapsing maps in order to get a detailed information about geometry of triangles in X.

A *tripod* $\tilde{T}$ is a metric graph, which, as a graph, is isomorphic to the 3-pod, see Example 1.32. We will use the notation o for the center of the tripod. By abusing the notation, we will regard a tripod $\tilde{T}$ as a geodesic triangle whose vertices are the extreme points (leaves) $\tilde{x}_i$ of $\tilde{T}$; hence, we will use the notation $\mathcal{T} = \tilde{T} = T(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$. Accordingly, the *side-lengths* of a tripod are lengths of the sides of the corresponding triangle.

REMARK 9.54. Using the symbol $\sim$ in the notation for a tripod is motivated by the comparison geometry, as we will compare geodesic triangles in δ-hyperbolic spaces with the tripods $\tilde{T}$: This is analogous to comparing geodesic triangles in metric spaces to geodesic triangles in constant curvature spaces, see Definition 2.55.

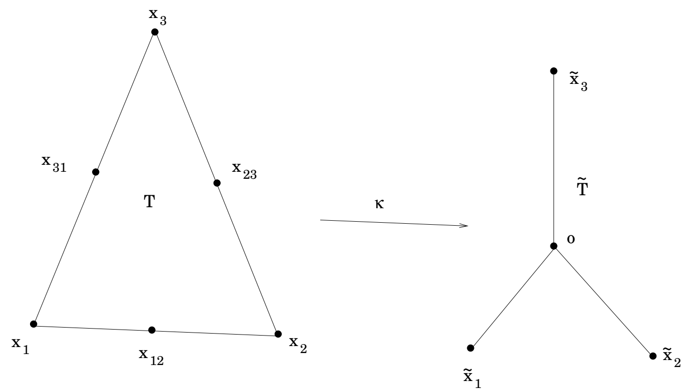


FIGURE 9.7. Collapsing map of a triangle to a tripod.

EXERCISE 9.55. For any three numbers $a_i \in \mathbb{R}_+, i = 1, 2, 3$ satisfying the triangle inequalities $a_i \leqslant a_j + a_k$ ($\{1, 2, 3\} = \{i, j, k\}$), there exists a unique (up to isometry) tripod $\tilde{T} = \mathcal{T}_{a_1,a_2,a_3}$ with the side-lengths a_1, a_2, a_3.

Now, given a geodesic triangle $T = T(x_1, x_2, x_3)$ with side-lengths $a_i, i = 1, 2, 3$ in a metric space X, there exists a unique (possibly up to postcomposition with an isometry $\tilde{T} \to \tilde{T}$) map κ to the *comparison* tripod $\tilde{T}$,

$$\kappa : T \to \tilde{T} = \mathcal{T}_{a_1,a_2,a_3},$$

which restricts to an isometry on every edge of T: The map κ sends the vertices x_i of T to the leaves $\tilde{x}_i$ of the tripod $\tilde{T}$. The map κ is called the *collapsing map* for T. We say that the points $x, y \in T$ are *dual* to each other if $\kappa(x) = \kappa(y)$.

EXERCISE 9.56. The collapsing map κ is 1-Lipschitz and preserves the Gromov-products $(x_i, x_j)_{x_k}$.

Then,

$$(x_i, x_j)_{x_k} = d(\tilde{x}_k, [\tilde{x}_i, \tilde{x}_j]) = d(\tilde{x}_k, o).$$

By taking the preimage of $o \in \tilde{T}$ under the maps $\kappa|(x_i x_j)$ we obtain points

$$x_{ij} \in x_i x_j$$

called the *central points* of the triangle T:

$$d(x_i, x_{ij}) = (x_j, x_k)_{x_i}.$$

LEMMA 9.57 (Approximation of triangles by tripods). *Assume that a geodesic metric space X is δ–hyperbolic in the sense of Rips, and consider an arbitrary geodesic triangle $T = \Delta(x_1, x_2, x_3)$ with the central points $x_{ij} \in x_i x_j$. Then for every $\{i, j, k\} = \{1, 2, 3\}$ we have:*

1. $d(x_{ij}, x_{jk}) \leqslant 6\delta$.

2. $d_{Haus}(x_j x_{ji}, x_j x_{kj}) \leqslant 7\delta$.

3. Distances between dual points in T are $\leqslant 14\delta$. In detail: Suppose that $\alpha_{ji}, \alpha_{jk} : [0, t_j] \to X$ ($t_j = d(x_j, x_{ij}) = d(x_j, x_{jk})$) are unit speed parameterizations of geodesic segments $x_j x_{ji}, x_j x_{jk}$. Then

$$d(\alpha_{ji}(t), \alpha_{jk}(t)) \leqslant 14\delta$$

for all $t \in [0, t_j]$.

Proof. The geodesic $x_i x_j$ is covered by the closed subsets $\overline{\mathcal{N}}_\delta(x_i x_k)$ and $\overline{\mathcal{N}}_\delta(x_j x_k)$, hence by connectedness there exists a point p on $x_i x_j$ at distance at most δ from both $x_i x_k$ and $x_j x_k$. Let $p' \in x_i x_k$ and $p'' \in x_j x_k$ be points at distance at most δ from p. The inequality

$$(x_j, x_k)_{x_i} = \frac{1}{2}\left[d(x_i, p) + d(p, x_j) + d(x_i, p') + d(p', x_k) - d(x_j, p'') - d(p'', x_k)\right]$$

combined with the triangle inequality implies that

$$|(x_j, x_k)_{x_i} - d(x_i, p)| \leqslant 2\delta,$$

and, hence $d(x_{ij}, p) \leqslant 2\delta$. Then $d(x_{ik}, p') \leqslant 3\delta$, whence $d(x_{ij}, x_{ik}) \leqslant 6\delta$. It remains to apply Lemma 9.3 to obtain 2 and Lemma 9.4 to obtain 3. □

We thus obtain

Proposition 9.58. *κ is a $(1, 14\delta)$-quasiisometry.*

Proof. The map κ is a surjective 1-Lipschitz map. On the other hand, Part 3 of the above lemma implies that

$$d(x, y) - 14\delta \leqslant d(\kappa(x), \kappa(y))$$

for all $x, y \in T$. □

Proposition 9.58 allows one to reduce (up to a uniformly bounded error) study of geodesic triangles in δ-hyperbolic spaces to study of tripods. For instance suppose that $m_{ij} \in x_i x_j$ are points such that

$$d(m_{ij}, m_{jk}) \leqslant r$$

for all i, j, k. We already know that this property holds for the central points x_{ij} of T (with $r = 6\delta$). Next result shows that points m_{ij} have to be uniformly close to the central points:

Corollary 9.59. *Under the above assumptions, $d(m_{ij}, x_{ij}) \leqslant r + 14\delta$.*

Proof. Since κ is 1-Lipschitz,

$$d(\kappa(m_{ik}), \kappa(m_{jk})) \leqslant r$$

for all i, j, k. By definition of the map κ, all three points $\kappa(m_{ij})$ cannot lie in the same leg of the tripod $\tilde{T}$, except when one of them is the center o of the tripod. Therefore, $d(\kappa(m_{ij}), o) \leqslant r$ for all i, j. Since κ is $(1, 14\delta)$-quasiisometry,

$$d(m_{ij}, x_{ij}) \leqslant d(\kappa(m_{ik}), \kappa(m_{jk})) + 14\delta \leqslant r + 14\delta.$$

Definition 9.60. We say that a point $p \in X$ is an R-centroid of a triangle $T \subset X$ if distances from p to all three sides of T are $\leqslant R$.

Corollary 9.61. *Every two R-centroids of T are within distance at most $\phi(R) = 4R + 28\delta$ from each other.*

Proof. Given an R-centroid p, let $m_{ij} \in x_i x_j$ be the nearest points to p. Then

$$d(m_{ij}, m_{jk}) \leqslant 2R$$

for all i, j, k. By previous corollary,

$$d(m_{ij}, x_{ij}) \leqslant 2R + 14\delta.$$

Thus, triangle inequalities imply that every two centroids are within distance at most $2(2R + 14\delta)$ from each other. □

Let $p_3 \in \gamma_{12} = x_1x_2$ be a point closest to x_3. Taking $R = 2\delta$ and combining Lemma 9.22 with Lemma 9.52, we obtain:

COROLLARY 9.62. $d(p_3, x_{12}) \leqslant 2(2\delta + 2\delta) = 6\delta$.

We now can define a continuous quasiinverse $\bar{\kappa}$ to κ as follows: We map the geodesic segment $\tilde{x}_1\tilde{x}_2 \subset \tilde{T}$ isometrically to a geodesic x_1x_2. We send $o\tilde{x}_3$ onto a geodesic $x_{12}x_3$ by an affine map. Since

$$d(x_{12}, x_{32}) \leqslant 6\delta$$

and

$$d(x_3, x_{32}) = d(\tilde{x}_3, 0),$$

we conclude that the map $\bar{\kappa}$ is $(1, 6\delta)$-Lipschitz.

EXERCISE 9.63.

$$d(\bar{\kappa} \circ \kappa, Id) \leqslant 32\delta.$$

9.9. Divergence of geodesics in hyperbolic metric spaces

Another important feature of hyperbolic spaces is the *exponential divergence* of its geodesic rays. This can be deduced from the thinness of polygons described in Lemma 9.6, as shown below. Our arguments are inspired by those in [**Pap03**].

LEMMA 9.64. *Let X be a geodesic metric space, δ–hyperbolic in the sense of Rips. If xy is a geodesic of length $2r$ and m is its midpoint, then every path joining x, y outside the ball $B(m, r)$ has length at least $2^{\frac{r-1}{\delta}}$.*

PROOF. Consider such a path $\mathfrak{p}$, of length ℓ. We divide this path first into two arcs of length $\frac{\ell}{2}$, then into four arcs of length $\frac{\ell}{4}$ etc., until we obtain $n = 2^k$ arcs of length $\frac{\ell}{2^k} \leqslant 2$. The minimal k satisfying this condition equals $\lfloor \log_2 \ell \rfloor$. Let $x_0 = x, x_1, ..., x_n = y$ be the consecutive subdivision points on $\mathfrak{p}$ obtained after this procedure. Lemma 9.6 applied to a geodesic polygon $x_0x_1 \dots x_n$ implies that m is contained in the $k\delta$-neighborhood of the broken geodesic

$$\mathfrak{q} = \bigcup_{i=0}^{n} x_ix_{i+1}.$$

Let $p \in \mathfrak{q}$ be the point closest to m. Since $d(x_i, m) \geqslant r$ for each i and $d(x_i, p) \leqslant 1$ for some i, we conclude that

$$r \leqslant k\delta + 1$$

and, hence,

$$r - 1 \leqslant \delta \log_2(\ell), \quad 2^{\frac{r-1}{\delta}} \leqslant \ell. \quad \square$$

The content of the next two lemmas can be described by saying that *geodesic rays in hyperbolic spaces diverge (at least) exponentially fast.*

LEMMA 9.65. *Let X be a geodesic metric space, δ–hyperbolic in the sense of Rips, and let x and y be two points on the sphere $S(o, R)$ such that* $\mathrm{dist}(x, y) = 2r$. *Then every path joining x and y outside $\overline{B}(o, R)$ has length at least $\psi(r) = 2^{\frac{r-1}{\delta}-3} - 12\delta$.*

PROOF. Let $m \in xy$ be the midpoint. Since $d(o,x) = d(o,y)$, it follows that m is also one of the centroids of the triangle $T(x,y,o)$ in the sense of §9.8. Then, by using Lemma 9.57 (Part 1), we see that $d(m,o) \leqslant (R-r) + 6\delta$. Therefore, the closed ball $\overline{B}(m, r-6\delta)$ is contained in $\overline{B}(o,R)$. Let $\mathfrak{p}$ be a path joining x and y outside $\overline{B}(o,R)$, and let xx' and $y'y$ be subsegments of xy of length 6δ. Lemma 9.64 implies that the path $x'x \cup \mathfrak{p} \cup yy'$ has length at least

$$2^{\frac{r-6\delta-1}{\delta}}$$

whence $\mathfrak{p}$ has length at least

$$2^{\frac{r-1}{\delta}-3} - 12\delta.$$

□

LEMMA 9.66. *Let X be a δ–hyperbolic in the sense of Rips, and let x and y be two points on the sphere $S(o, r_1 + r_2)$ such that there exist two geodesics xo and yo intersecting the sphere $S(o, r_1)$ in two points x', y' at distance larger than 14δ. Then every path joining x and y outside $B(o, r_1 + r_2)$ has length at least $\psi(r_2 - 15\delta) = 2^{\frac{r_2-1}{\delta}-18} - 12\delta$.*

PROOF. Let m be the midpoint m of xy; since $T(x,y,o)$ is isosceles, m is one of the centroids of this triangle. Since $d(x',y') > 14\delta$, they cannot be dual points on $\Delta(x,y,o)$ in the sense of §9.8. Let $x'', y'' \in xy$ be dual to x', y'. Thus (by Lemma 9.57 (Part 3)),

$$d(o,x'') \leqslant r_1 + 14\delta, d(o,x'') \leqslant r_1 + 14\delta.$$

Furthermore, by the definition of dual points, since m is a centroid of $\Delta(x,y,o)$, m belongs to the segment $x''y'' \subset xy$. Thus, by quasiconvexity of metric balls, see §9.6,

$$d(m,o) \leqslant r_1 + 14\delta + \delta = r_1 + 15\delta.$$

By the triangle inequality,

$$r_1 + r_2 = d(x,o) \leqslant r + d(m,o) \leqslant r + r_1 + 15\delta, \quad r_2 - 15\delta \leqslant r.$$

Since the function ψ in Lemma 9.65 is increasing,

$$\psi(r_2 - 15\delta) \leqslant \psi(r).$$

Combining this with Lemma 9.65 (where we take $R = r_1 + r_2$), we obtain the required inequality. □

COROLLARY 9.67. *Suppose that $\rho, \rho' \in Ray_p(X)$ are inequivalent rays. Then for every sequence t_n diverging to ∞,*

$$\lim_{i\to\infty} d(\rho(t_i), \rho'(t_i)) = \infty.$$

PROOF. Suppose to the contrary, there exists a divergent sequence t_i such that $d(\rho(t_i), \rho'(t_i)) \leqslant D$. Then, by Lemma 9.4, for every $t \leqslant t_i$,

$$d(\rho(t), \rho'(t)) \leqslant 2(D + \delta).$$

Since $\lim_{i\to\infty} t_i = \infty$, it follows that $\rho \sim \rho'$. A contradiction. □

We now promote the conclusion of Lemmas 9.64 and 9.66 to the notion of *divergence* of geodesics and spaces. In both definitions, X is a 1-ended geodesic metric space (which need not be hyperbolic).

Definition 9.68 (Divergence of a geodesic). Let $\gamma : \mathbb{R} \to X$ be a biinfinite geodesic in X. This geodesic is said to have *divergence* $\geqslant \zeta(r)$ if for the points $x = \gamma(-r), y = \gamma(r)$, the infimum of lengths of paths $\mathfrak{p}$ connecting x to y outside of the ball $B(\gamma(0), r)$ is $\geqslant \zeta(r)$.

The assumption that X is 1-ended in this definition is needed to ensure that the paths $\mathfrak{p}$ connecting x to y exist.

Exercise 9.69. Suppose that X is the Cayley graph of a finitely-generated group. Show that each geodesic in X has at most exponential divergence.

In view of this definition, lemma 9.64 says that in every δ-hyperbolic geodesic metric space, every biinfinite geodesic has divergence $\geqslant 2^{\frac{r-1}{2\delta}}$: Geodesics in Rips-hyperbolic spaces have at least exponential divergence. Similarly to the definition of divergence of geodesics, one defines divergence of quasigeodesics.

We refer the reader to [**Ger94, KL98a, Mac, DR09, BC12, AK11, Sul14**] for a more detailed treatment of divergence of geodesics in metric spaces.

Lemma 9.66 suggests a notion of divergence of a space based on *uniform divergence of pairs of geodesic rays* rather than of geodesics.

Definition 9.70 (Uniform divergence of a space). A continuous function $\eta : \mathbb{R}_+ \to \mathbb{R}$ is called a *uniform divergence function* for X if for every point $o \in X$ and geodesic segments $\alpha = ox, \beta = oy$ in X, for all $r, R \in \mathbb{R}_+$ satisfying

$$R + r \leq \min(d(o,x), d(o,y)), \quad d(\alpha(r), \beta(r)) \geq \eta(0)$$

and every path $\mathfrak{p}$ in $X \setminus B(o, r+R)$ connecting $\alpha(R+r)$ to $\beta(R+r)$, we have

$$\text{length}(\mathfrak{p}) \geqslant \eta(R).$$

For instance, in view of Lemma 9.66, every hyperbolic geodesic metric space has an exponential uniform divergence function.

Theorem 9.71 (P. Papasoglu, [**Pap95c**]). *If X is a geodesic metric space with proper uniform divergence function, then X is hyperbolic.*

9.10. Morse Lemma revisited

In this section we use Lemma 9.64 to give another proof of the Morse Lemma, this time with an explicit bound on the distance between quasigeodesic and geodesic paths. We note that a more refined (and sharp) estimate in the Morse Lemma is established by V. Schur in [**Sch13**].

Theorem 9.72. *For every (L, A)-quasigeodesic $\mathfrak{q} : [0, T] \to X$ in a δ-hyperbolic geodesic space X, the image of $\mathfrak{q}$ is within Hausdorff distance $\leqslant D = D(L, A, \delta)$ from every geodesic $xy \subset X$ connecting the endpoints of $\mathfrak{q}$. The function D can be estimated from above as*

$$D \leqslant L(A + 1 + 2R_*)(L + A),$$

where

$$R_* = R_*(L, A, \delta) \leqslant \frac{L + 2A}{6} + 2\delta \log_2(7L(L + A)) + 2\delta \log_2(\delta)$$

provided that δ is at least 2.

Proof. Let $\mathfrak{q} : [0,T] \to X$ be an (L,A)-quasigeodesic path in X. We let $x = \mathfrak{q}(0), y = \mathfrak{q}(T)$. Then the set

$$N = \mathfrak{q}([0,T] \cap \mathbb{Z}) \cup \{y\}$$

is an $(L+A)/2$-net in the image of $\mathfrak{q}$. Note that for $i, i+1 \in [0,T] \cap \mathbb{Z}$,

$$d(x_i, x_{i+1}) \leqslant (L+A)$$

and

$$d(x_n, y) \leqslant L + A,$$

where $n = \lfloor T \rfloor$.

Take a geodesic segment xy connecting the endpoints x, y of $\mathfrak{q}$. We let $m \in xy$ denote the point within the largest distance (denoted R) from the net N. We parameterize the segment xy via an isometric map $\gamma : [a,b] \to X$ such that $\gamma(a) = x, \gamma(b) = y$, $\gamma(0) = m$. We first consider the *generic* case when $a \leqslant -2R$, $2R \leqslant b$. We mark four points

$$z = \gamma(-2R), x' = \gamma(-R), y' = \gamma(R), w = \gamma(2R)$$

in the segment xy and let $x_i = \mathfrak{q}(i), x_j = \mathfrak{q}(j)$ denote the points in the net N closest to z and w respectively. Due to our choice of m to be the point in xy farthest from N, we obtain the bound

$$\max\left(d(z, x_i), d(w, x_j)\right) \leqslant R.$$

We will consider the case $i \leqslant j$ for convenience of the notation and leave the case $i \geqslant j$ to the reader. Then we have a broken geodesic β connecting x_i and x_j:

$$\beta = x_i x_{i+1} \cdots x_{j-1} x_j.$$

Since $\mathfrak{q}$ is an (L,A)-quasigeodesic, the length $\ell(\beta)$ of the path β is at most

$$(L+A)(j-i) \leqslant (L+A)L(d(x_i, x_j) + A).$$

Since $d(x_i, x_j) \leqslant 6R + (L+A)$, we obtain the bound

$$\ell(\beta) \leqslant (L+A)L(6R + (L+A) + A) = (L+A)L(6R + L + 2A).$$

Since none of the points $x_k, k = i, \ldots, j$ belongs to the open ball $B(m,R)$, we conclude that the piecewise-geodesic concatenation

$$\alpha = x'z \star zx_i \star \beta \star x_j w \star wy'$$

is disjoint from the open ball $B(m,R)$. Therefore, in view of Lemma 9.64, the length of α is at least

$$2^{\frac{R-1}{\delta}}$$

which is a superlinear function of R. On the other hand, α has length at most

$$4R + \ell(\beta) \leqslant 4R + (L+A)L(6R + L + 2A)$$

which is a linear function of R. Therefore, the inequality

$$2^{\frac{R-1}{\delta}} \leqslant 4R + (L+A)L(6R + L + 2A) \tag{9.5}$$

forces $R \leqslant R_*$ for some $R_* = R_*(L, A, \delta)$. Below we will estimate R_* (from above) explicitly.

Figure 9.8. The path α.

Exercise 9.73. For $\delta \geqslant 2, c_1 > 0, c_2 > 0$, if

$$2^{\frac{R-1}{\delta}} \leqslant c_1 R + c_2$$

then

$$R \leqslant \frac{c_2}{c_1} + 2\delta \log_2(c_1) + 2\delta \log_2(\delta)$$

Therefore, we obtain the estimate:

$$R_* \leqslant \frac{L(L+A)(L+2A)}{4+6L(L+A)} + 2\delta \log_2(4+6L(L+A)) + 2\delta \log_2(\delta) \leqslant$$

$$\frac{L+2A}{6} + 2\delta \log_2(7L(L+A)) + 2\delta \log_2(\delta).$$

We next consider the *nongeneric* cases, i.e., when $d(m,y) < 2R$ or $d(x,m) < 2R$. There are several subcases to analyze, we will deal with the case

$$d(x,m) \geqslant 2R, \quad R \leqslant d(m,y) < 2R$$

and leave the other two possibilities to the reader since they are similar. We define the points x', z and x_i as before, but now use the point y to play the role of x_j. The broken path β above will be replaced with the broken path

$$\beta := x_i x_{i+1} \cdots x_n y$$

and we will use the concatenation $x'z \star zx_i \star \beta \star yy'$. With this modification, the same inequality (9.5) still holds and we again conclude that $R \leqslant R_*$, where R_* is the same function as above.

So far, we proved that the geodesic segment xy is contained in the R_*-neighborhood of the image of $\mathfrak{q}$. More precisely, we proved that for each $t \in [a,b]$ there exists $s = s(t) \in \{0, 1, \ldots, n, T\}$ such that $d(\mathfrak{q}(s), \gamma(t)) \leqslant R_*$. This choice of $s(t)$ need not be unique, but we will use $s(a) = 0, s(b) = T$ since the respective distances in X will be zero in this situation. It is convenient at this point to reparameterize the geodesic xy so that $a = 0$. The function $t \mapsto s$, of course, is not continuous, but it is coarse Lipschitz:

$$|s(t) - s(t+1)| \leqslant L(A + 1 + 2R_*)$$

for all $t, t+1 \in [0,b]$. We will replace this function with a piecewise-linear function as follows. For every $t \in [0,b] \cap \mathbb{Z}$, we set $f(t) := s(t)$. We extend this function linearly over each unit interval contained in $[0,b]$ and having the form $[i, i+1]$, $i \in \mathbb{Z}$ or $[\lfloor b \rfloor, b]$. By abusing the terminology, we will refer to these intervals as *integer intervals*. The resulting function f is continuous on the interval $[0,b]$ and maps it onto the interval $[0,T]$. Moreover, since $\mathfrak{q}$ is (L,A)-quasigeodesic, the function f maps each integer interval to an interval of the length at most $C = L(A+1+2R_*)$.

Now, for $s \in [0,T]$ find t such that $f(t) = s$. Then t belongs to one of the integer subintervals $[i, i+1] \subset [0,b]$ (or the interval $[i,b]$, $i = \lfloor b \rfloor$). We have $d(\gamma(i), \mathfrak{q}(s(i))) \leqslant R_*$ and, furthermore,

$$d(\mathfrak{q}(s), \mathfrak{q}(s(i))) \leqslant C(L+A).$$

Therefore, $\mathfrak{q}(s)$ is within distance $\leqslant C(L+A) = L(A+1+2R_*)(L+A)$ from the geodesic xy. Since

$$R_* \leqslant D := L(A+1+2R_*)(L+A)$$

(as $L \geqslant 1$), we conclude that the Hausdorff distance between the geodesic xy and the quasigeodesic $\mathfrak{q}$ is at most D. $\square$

9.11. Ideal boundaries

We consider the general notion of the *ideal boundary* defined in §2.11.3, in the special case when X is geodesic, δ–hyperbolic and locally compact (equivalently, proper). We start by proving an analogue of Lemma 9.74 in the context of hyperbolic spaces.

LEMMA 9.74. *Suppose that X is geodesic, δ–hyperbolic and locally compact (equivalently, proper). Then for each $x \in X$ and $\xi \in \partial_\infty X$, there exists a geodesic ray ρ with the initial point x, asymptotic to ξ.*

PROOF. Let ρ' be a geodesic ray asymptotic to ξ, with the initial point x_0. Consider a sequence of geodesic segments $\gamma_n : [0, D_n] \to X$, connecting p to $x_n = \rho'(n)$, where $D_n = d(x, \rho'(n))$. The δ-hyperbolicity of X implies that the image of γ_n is at Hausdorff distance at most $\delta + \operatorname{dist}(x, x_0)$ from x_0x_n, where x_0x_n is the initial subsegment of ρ'. Combining properness of X with the Arzela-Ascoli theorem, we see that the maps γ_n subconverge to a geodesic ray ρ, $\rho(0) = p$. Clearly, the image of ρ is at Hausdorff distance at most $\delta + \operatorname{dist}(x, x_0)$ from the image of ρ. In particular, ρ is asymptotic to ρ'. □

In view of Lemma 9.74, in order to understand $\partial_\infty X$ it suffices to restrict to the set $Ray_x(X)$ of geodesic rays in X emanating from $x \in X$. The important difference between this lemma and the one for $CAT(0)$ spaces (Lemma 9.74) is that the ray ρ now may not be unique. Nevertheless we will continue to use the notation $x\xi$, which now means that $x\xi$ is *one* of the geodesic rays with the initial point x and asymptotic to ξ. This abuse of notation is harmless in view of the following lemma which generalizes the *fellow–travelling property* for geodesic segments in hyperbolic spaces.

LEMMA 9.75 (Asymptotic rays are uniformly close). *Let ρ_1, ρ_2 be asymptotic geodesic rays in X with the common initial point $\rho_1(0) = \rho_2(0) = x$. Then for each t,*

$$d(\rho_1(t), \rho_2(t)) \leqslant 2\delta.$$

PROOF. Suppose that the rays ρ_1, ρ_2 are within distance $\leqslant C$ from each other. Take T much larger than t. Then (since the rays are asymptotic) there exists $S \in \mathbb{R}_+$ such that

$$d(\rho_1(T), \rho_2(S)) \leqslant C.$$

By δ–thinness of the triangle $\Delta(p, \rho_1(T), \rho_2(S))$, the point $\rho_1(t)$ is within distance $\leqslant \delta$ from a point either on $p\rho_2(S)$ or on $\rho_1(T)\rho_2(S)$. Since the length of $\rho_1(T)\rho_2(S)$ is $\leqslant C$ and T is much larger than t, it follows that there exists t' such that

$$\operatorname{dist}(\rho_1(t), \rho_2(t')) \leqslant \delta.$$

By the triangle inequality, $|t - t'| \leqslant \delta$ and, hence, $\operatorname{dist}(\rho_1(t), \rho_2(t)) \leqslant 2\delta$. □

Our next goal is to extend the topology τ defined on $\partial_\infty X$ (i.e., the quotient topology of the compact-open topology on the set of all rays in X, see §2.11.3) to a topology on the union $\bar{X} = X \cup \partial_\infty X$. There are several natural ways to do so, all resulting in the same topology, which is a compactification of X by its ideal boundary $\partial_\infty X$.

Shadow topology $\mathcal{T}_{x,k}$ on $\bar{X}$. Our next goal is to topologize $\bar{X}$ and to describe some basic properties of this topology. We fix a point $x \in X$ and a number $k \geqslant 3\delta$. For each ideal boundary point $\xi \in \partial_\infty X$ we fix a geodesic ray $\rho = x\xi$ asymptotic to ξ. We define the topology $\mathcal{T}_{x,k}$ on $\bar{X}$ by declaring that its basis at points $z \in X$ consists of open metric balls $B(z, r)$, $r > 0$, and defining basic neighborhoods $U_y = U_{x,y,k}(\xi)$ at points $\xi \in \partial_\infty X$ as

$$U_y = \{z \in \bar{X} : \forall xz, xz \cap B(y, k) \neq \emptyset\},$$

where $y = \rho(t)$, $t \geqslant 0$.

FIGURE 9.9. Shadow topology.

Note that the requirement in this definition is that *each* geodesic segment or a ray from x to z intersects the open ball $B(y, r)$. We need to check that the collection of basic sets we defined is indeed a basis of topology. It follows from Lemma 9.75 that ξ belongs to $U_y(\xi)$ for each $y = \rho(t)$. Furthermore, δ-hyperbolicity of X implies that for every $t \geqslant 0$

$$U_{\rho(t')} \subset U_{\rho(t)}, \tag{9.6}$$

provided that t' is at least $t + k + \delta$. Therefore,

$$U_{y_3} \subset U_{y_1} \cap U_{y_2}, y_i = \rho(t_i), t_3 = \max(t_1, t_2) + k + \delta.$$

We next have to verify that each basic set is open, more precisely, each point $u \in U_y$ is contained in a basic set $U_z \subset U_y$. Suppose that $u \in U_y \cap X$ and $u_n \in X$ is a sequence converging to u. Assume for a moment that for each n there exists a geodesic segment xu_n disjoint from the open ball $B(y, k)$. Then a subsequential limit of xu_n would connect x to u and also avoid the ball $B(y, k)$. (Here we are using properness of X.) However, this would imply that $u \notin U_y$, which is a contradiction.

The proof for boundary points is similar. Suppose that $\xi \in U_y(\eta) \cap \partial_\infty X$; let $\rho = x\xi$ be a ray connecting x to ξ. We again assume that there is a sequence t_n diverging to infinity, points $u_n \in B(\rho(t_n), k)$ and geodesic segments $\gamma_n = xu_n$ which all avoid the open ball $B(y, r)$. By δ-thinness of geodesic triangles, each segment γ_n is contained in the $k + \delta$-neighborhood of the ray ρ. Therefore, after passing to a subsequence, we obtain a limit ray γ (of the segments γ_n), which is asymptotic to ξ and which also avoids the ball $B(y, k)$. However, this contradicts the assumption that ξ belongs to U_y.

To summarize, we now have a topology $\mathcal{T}_{x,k}$ on the set $\bar{X}$.

LEMMA 9.76. *With respect to the topology $\mathcal{T}_{x,k}$:*
1. X is an open and dense subset of $\bar{X}$.
2. $\bar{X}$ is 1st countable.
3. $\bar{X}$ is Hausdorff.

PROOF. 1. Openness of X is clear. Density of X follows from the fact that for each $\xi \in \partial_\infty X$ the sequence $(\rho(n))_{n\in\mathbb{N}}$ converges to ξ, where $\rho = x\xi$.

2. The 1st countability is clear at the points $x \in X$; at the points $\xi \in \partial_\infty X$, the 1st countability follows from the fact that the sets

$$U_{\rho(n)}, n \in \mathbb{N},$$

form a basis of topology at ξ, see (9.6).

3. We will check that any two distinct points $\xi_1, \xi_2 \in \partial_\infty X$ have disjoint neighborhoods and leave the other cases as an exercise to the reader. Since the geodesic rays $x\xi_i$, $i = 1, 2$, diverge, there exist points $y_i \in x\xi_i$ such that

$$\mathrm{dist}(y_i, x\xi_{3-i}) > k + \delta, \quad i = 1, 2.$$

We claim that the basic neighborhoods U_{y_1}, U_{y_2} are disjoint. Otherwise, there exists $z \in xu \cap B(y_2, k)$ and a geodesic xz which intersects $B(y, k)$. Since the triangle $\Delta(x, y_2, z)$ is δ-thin, it follows that the point $xu \cap B(y, k)$ is within distance $\leqslant \delta$ form the subsegment xy_2 of $x\xi_2$. However, this contradicts the assumption that the minimal distance from y_1 to $x\xi_2$ is $> k + \delta$. Interchanging the roles of the points y_1, y_2, we conclude that $U_{y_1} \cap U_{y_2} = \emptyset$. □

Consider the set $Geo_x(X)$ consisting of geodesics in X (finite or half-infinite) emanating from x. In order to ensure that all maps in $Geo_x(X)$ have the same domain, we extend each geodesic segment $\gamma : [0, T] \to X$ by the constant map to the half-line $[T, \infty)$. We quip $Geo_x(X)$ with the compact-open topology (equivalently, the topology of uniform convergence on compacts). The space $Geo_x(X)$ is compact by the Arzela–Ascoli theorem. Since X is Hausdorff, so is $Geo_x(X)$. There is a natural quotient map $\epsilon : Geo_x(X) \to \bar{X}$ which sends a finite geodesic or a geodesic ray emanating from x to its terminal point in $\bar{X}$:

$$\epsilon : xy \mapsto y, \quad y \in \bar{X}.$$

Lemma 9.77. *The map $\epsilon : Geo_x(X) \to \bar{X}$ is continuous.*

Proof. The statement is clear for finite geodesic segments. Let $\xi \in \partial_\infty X$ be an ideal boundary point and let γ_n denote a sequence of geodesic segments/rays, $\gamma_n = xx_n$, $x_n \in \bar{X}$, such that

$$\lim_{n\to\infty} \gamma_n = \gamma,$$

where γ is a ray asymptotic to ξ. We claim that the sequence $\epsilon(ga_n) = x_n$ converges to ξ in the topology $\mathcal{T}_{x,k}$. Pick a point y on the geodesic ray $x\xi$ (which could be different from γ). Since the rays γ and ρ are within distance $\leqslant 2\delta$ from each other, and the convergence $\gamma_n \to \gamma$ is uniform on compacts, for every there exists N such that for all $n > N$, the intersection

$$\gamma_n \cap B(y, 3\delta) \subset \gamma_n \cap B(y, k)$$

is nonempty. Hence, x_n belongs to U_y. □

Corollary 9.78. *1. ϵ is a closed map. In particular, $\epsilon : Geo_x(X) \to (\bar{X}, \mathcal{T}_{x,k})$ is a quotient map.*

2. $(\bar{X}, \mathcal{T}_{x,k})$ is a compact topological space.

3. $(\bar{X}, \mathcal{T}_{x,k})$ is a compactification of X.

Proof. 1. The statement follows from the fact that $Geo_x(X)$ is compact and $(\bar{X}, \mathcal{T}_{x,k})$ is Hausdorff.

2. Continuous image of a compact topological space is again compact.

3. This part follows from openness and density of X in $\bar{X}$ combined with compactness of $\bar{X}$. □

Corollary 9.79. *1. For all $k_1, k_2 \geqslant 3\delta$, the topologies $\mathcal{T}_{x,k_1}, \mathcal{T}_{x,k_2}$ are equal.*

2. The topology $\mathcal{T}_{x,k}$ is independent of the choice of geodesic rays $x\xi$ used to define $\mathcal{T}_{x,k}$

PROOF. For all different choices of k's and the rays, the topologies are the quotient topologies of $Geo_x(X)$ with respect to the same quotient map ϵ. $\square$

The topology on $\bar{X}$ is independent of the choice of a base-point x:

LEMMA 9.80. *For all $x_1, x_2 \in X$, $\mathcal{T}_{x_1,k} = \mathcal{T}_{x_2,k}$.*

PROOF. The equality of two topologies is clear at the points of X. Consider, therefore, an ideal boundary point $\xi \in \partial_\infty X$. We use a geodesic segment x_1x_2 and geodesic rays $x_i\xi$ to form a generalized geodesic triangle $\Delta(x_1, x_2, \xi)$ in X. Since this triangle is 2δ-thin, we pick points $y_i \in x_i\xi$ within distance $\leqslant 2\delta$ from each other. Then each $u \in B(y_1, k)$ is contained in the ball $B(y_2, k+\delta)$. Therefore, each basic neighborhood $U_{y_2,k}$ of ξ in the topology $\mathcal{T}_{x_2,k}$ is contained in the basic neighborhood $U_{y_2,k+\delta}$ of ξ in the topology $\mathcal{T}_{x_2,k+\delta}$. Hence, the topology $\mathcal{T}_{x_2,k}$ is finer than $\mathcal{T}_{x_1,k+\delta} = \mathcal{T}_{x_1,k}$. Switching the roles of x_1 and x_2, we conclude that $\mathcal{T}_{x_1,k} = \mathcal{T}_{x_2,k}$. $\square$

In view of these basic results, from now on, we will omit the subscripts in the notation for the topology on $\bar{X}$. Using the identification of $\bar{X}$ with the quotient space of $Geo_x(X)$, we can also give an alternative description of converging sequences in $\bar{X}$.

LEMMA 9.81. *For a sequence (x_n) in $\bar{X}$ and a point $\xi \in \partial_\infty X$, the following are equivalent:*

1. $\lim_{n\to\infty} x_n = \xi$.

2. Every convergent subsequence in xx_n converges to a ray asymptotic to ξ.

PROOF. (1) $\Rightarrow$ (2). Suppose first that $\lim_{n\to\infty} x_n = x$. Let $\gamma_m = xx_{n_m}$ be a convergent sequence of segments/rays and let γ be their limiting ray in $Geo_x(X)$. Pick an arbitrary point $y = \rho(t)$ on the ray $\rho = x\xi$. Since $\lim_{n\to\infty} x_n = \xi$, the intersections $\gamma_n \cap B(y, k) \neq \emptyset$ for all sufficiently large n. Hence, each subsegment $\gamma_n([0, T])$ is contained in the $k+\delta$-neighborhood of the ray $x\xi$. Since this holds for all T, we conclude that the limit ray γ is also contained in the $k+\delta$-neighborhood of $x\xi$. It follows that the ray γ is asymptotic to ξ.

(2) $\Rightarrow$ (1). After passing to subsequences, we can assume that the sequence xx_n converges to a ray $\rho = x\xi$ and that the sequence (x_n) converges to an ideal boundary point η. Continuity of the map ϵ now implies that $\epsilon(\rho) = \xi$. $\square$

We owe the following remark to Bernhard Leeb:

REMARK 9.82. Even if a sequence (x_n) converges, this does not imply that there exists a convergent sequence of geodesic rays xx_n.

We compute two examples of compactifications and ideal boundaries of hyperbolic spaces.

1. Suppose that $X = \mathbb{H}^n$ is the real-hyperbolic space. We claim that $\bar{X}$ is naturally homeomorphic to the closed ball $\mathbb{D}^n$, where we use the unit ball model of $\mathbb{H}^n$. Let o denote the center of the unit ball $\mathbf{B}^n$. The map $\epsilon : Geo_o(X) \to \mathbb{D}^n$ is a bijection. The fact that the map ϵ restricts to a homeomorphism $X \to \mathbb{H}^n$ is clear. Bicontinuity of this map at the points of $\partial_\infty X$ follows from the fact that a sequence $\gamma_n \in Geo_o(X)$ converges to a geodesic ray ρ iff

$$\lim_{n\to\infty} \gamma'(0) = \rho'(0).$$

2. Suppose that X is a simplicial tree of finite constant valence $val(X) \geqslant 3$, equipped with the standard metric. Fix a vertex $p \in X$. Since X is a $CAT(-\infty)$-space, the map $\epsilon : Geo_p(X) \to \bar{X}$ is a bijection, hence, a homeomorphism (in view of the quotient topology on $\bar{X}$).

We that $\partial_\infty X$ is homeomorphic to the Cantor set. Since we know that $\partial_\infty X$ is compact and Hausdorff, it suffices to verify that $\partial_\infty X$ is totally disconnected and contains no isolated points. Let $\rho \in Ray_p(X)$ be a ray. For each n pick a ray $\rho_n \in Ray_p(X)$ which coincides with ρ on $[0, n]$, but $\rho_n(t) \neq \rho(t)$ for all $t > n$ (this is where we use the assumption that $val(X) \geqslant 3$). It is then clear that

$$\lim_{n\to\infty} \rho_n = \rho$$

uniformly on compacts. Hence, $\partial_\infty X$ has no isolated points. Recall that for $k = \frac{1}{2}$, we have open sets $U_{n,k}(\rho)$ forming a basis of neighborhoods of ρ. We also note that each $U_{n,k}(\rho)$ is also closed, since (for a tree X as in our example) it is also given by

$$\{\rho' : \rho(t) = \rho'(t), t \in [0, n]\}.$$

Therefore, $\partial_\infty X$ is totally-disconnected as for any pair of distinct points $\rho, \rho' \in Ray_p(X)$, they have open, closed and disjoint neighborhoods $U_{n,k}(\rho), U_{n,k}(\rho')$. Thus, $\partial_\infty X$ is compact, Hausdorff, perfect, consists of at least 2 points and is totally-disconnected. Therefore, $\partial_\infty X$ is homeomorphic to the Cantor set.

We now return to the discussion of ideal boundaries of arbitrary proper geodesic hyperbolic spaces.

LEMMA 9.83 (The visibility property). *Let X be a proper geodesic Gromov–hyperbolic space. Then for each pair of distinct points $\xi, \eta \in \partial_\infty X$ there exists a geodesic γ in X which is asymptotic to both ξ and η.*

PROOF. Consider geodesic rays ρ, ρ' emanating from the same point $p \in X$ and asymptotic to ξ, η respectively. Since $\xi \neq \eta$ (Corollary 9.67), for each $R < \infty$ the set

$$K(R) := \{x \in X : \text{dist}(x, \rho) \leqslant R, \text{dist}(x, \rho') \leqslant R\}$$

is compact. Consider the sequences $x_n := \rho(n), x'_n := \rho'(n)$ on ρ, ρ' respectively. Since the triangles $T(p, x_n, x'_n)$ are δ–thin, each segment $\gamma_n := x_n x'_n$ contains a point within distance $\leqslant \delta$ from both px_n, px'_n, i.e., $\gamma_n \cap K(\delta) \neq \emptyset$. Therefore, by the Arzela–Ascoli theorem, the sequence of geodesic segments γ_n subconverges to a complete geodesic γ in X. Since

$$\gamma \subset \mathcal{N}_\delta(\rho \cup \rho'),$$

it follows that γ is asymptotic to ξ and to η. □

EXERCISE 9.84. Suppose that X is δ-hyperbolic. Show that there are no complete geodesics γ in X such that

$$\lim_{n\to\infty} \gamma(-n) = \lim_{n\to\infty} \gamma(n).$$

Hint: Use the fact that geodesic bigons in X are δ-thin.

EXERCISE 9.85 (Ideal bigons are 2δ-thin). Suppose that $\alpha, \beta : \mathbb{R} \to X$ are geodesics in X which are both asymptotic to points $\xi, \eta \in \partial_\infty X$. Then

$$\text{dist}_{Haus}(\alpha, \beta) \leqslant 2\delta.$$

Hint: For $n \in \mathbb{N}$ define

$$z_n, w_n \in \beta(\mathbb{R})$$

to be the nearest points to $x_n = \alpha(n), y_n = \alpha(-n)$. Let $x_n y_n, z_n w_n$ be the subsegments of α, β between x_n, y_n and y_n, z_n respectively. Now use the fact that the quadrilateral

$$x_n y_n \cup y_n w_n \cup w_n z_n \cup z_n x_n$$

is 2δ-thin.

Triangles in $\bar{X}$. We now generalize (geodesic) triangles in X to triangles with some vertices in $\partial_\infty X$, similarly to the definitions made in §8.4. Namely a (generalized) triangle in $\bar{X}$ is a concatenation of geodesics connecting three points A, B, C in $\bar{X}$; geodesics are now allowed to be finite, half-infinite and infinite. The points A, B, C are called the vertices of the triangle. As in the case of $\mathbb{H}^n$, *we do not allow two ideal vertices of a triangle T to coincide.* By abusing the terminology, we will again refer to such generalized triangles as *hyperbolic triangles*. As with hyperbolic geodesics, we continue to use the notation $T(A, B, C)$ and $\Delta(A, B, C)$ for geodesic triangles with the vertices A, B and C, even though geodesics connecting the vertices are not unique.

An ideal triangle is a triangle where all three vertices are in $\partial_\infty X$. We topologize the set $Tri(X)$ of hyperbolic triangles in X by the compact-open topology on the set of their geodesic edges. Given a hyperbolic triangle $T = T(A, B, C)$ in X, we find a sequence of finite triangles $T_i \subset X$ whose vertices converge to the respective vertices of T. Passing to a subsequence if necessary and taking the limit of the sides of the triangles T_i, we obtain geodesics connecting the vertices A, B, C of T. The resulting triangle T', of course, need not be equal to T (since geodesics connecting points in $\bar{X}$ are not, in general, not be unique), however, in view of Exercise 9.85, sides of T' are within distance $\leqslant 2\delta$ from the respective sides of T. We will say that the sequence of triangles T_i *coarsely converges* to the triangle T (cf. Definition 5.31).

EXERCISE 9.86. Every (generalized) hyperbolic triangle T in X is 5δ-thin. In particular,

$$\text{minsize}(T) \leqslant 4\delta.$$

Hint: Use a sequence of finite triangles coarsely converging to T and the fact that finite triangles are δ-thin.

Centroids of triangles with ideal vertices. We now return to the discussion of proper geometric metric spaces X which are δ-hyperbolic in the sense of Rips. Exercise 9.86 allows one to define *centroids* of triangles T in $\bar{X}$. As in Definition 9.60 we say that a point $p \in X$ is an *R-centroid of T* if p is within distance $\leqslant R$ from all three sides of T. Furthermore, we will say that p is a *centroid* of T if

$$d(p, \tau_i) \leqslant 5\delta, i = 1, 2, 3.$$

LEMMA 9.87. *The distance between any two R-centroids of a triangle T is at most*

$$r(R, \delta) = 4R + 32\delta.$$

PROOF. Let p, q be R-centroids of T. We coarsely approximate T by a sequence of finite triangles $T_i \subset X$. Then for every $\epsilon > 0$, for all sufficiently large i, the points p, q are $R + 2\delta + \epsilon$-centroids of T_i. Therefore, by Corollary 9.61 applied to the triangles T_i,

$$d(p, q) \leqslant \phi(R + 2\delta + \epsilon) = 4(R + 2\delta + \epsilon) + 28\delta = 4R + 32\delta + 2\epsilon$$

Since this holds for every $\epsilon > 0$, we conclude that $d(p, q) \leqslant 4R + 32\delta$. □

NOTATION 9.88. Given a topological space Z, we let $\mathrm{Trip}(Z)$ denote the set of ordered triples of pairwise distinct elements of Z, equipped with the subspace topology induced from Z^3.

We define the correspondence

$$center : \mathrm{Trip}(\partial_\infty X) \to X$$

which sends every triple of distinct points in $\partial_\infty X$ first to the set of ideal triangles T that they span and then to the set of centroids of these ideal triangles. Lemma 9.87 implies:

COROLLARY 9.89. *For every* $\xi \in \mathrm{Trip}(\partial_\infty X)$,

$$\mathrm{diam}(center(\xi)) \leqslant r(7\delta, \delta) = 60\delta.$$

EXERCISE 9.90. Suppose that γ_n are geodesics in X asymptotic to points $\zeta_n, \eta_n \in \partial_\infty X$ and such that

$$\lim_{n\to\infty} \zeta_n = \zeta, \quad \lim_{n\to\infty} \eta_n = \eta, \quad \eta \neq \zeta.$$

Show that the sequence (γ_n) subconverges to a geodesic asymptotic to both ξ and η.

Use this exercise to conclude:

EXERCISE 9.91. If $K \subset \mathrm{Trip}(\partial_\infty X)$ is a compact subset, then $center(K)$ is a bounded subset of X.

Conversely, prove:

EXERCISE 9.92. Let $B \subset X$ be a bounded subset and $K \subset \mathrm{Trip}(\partial_\infty X)$ be a subset such that $center(K) \subset B$. Show that K is relatively compact in $\mathrm{Trip}(\partial_\infty X)$. Hint: For every $\xi \in K$, every ideal edge of a triangle spanned by ξ intersects the 5δ-neighborhood of B. Now, use the Arzela-Ascoli theorem.

Loosely speaking, the two exercises show that the correspondence *center* is coarsely continuous (the image of a compact is bounded) and coarsely proper (the preimage of a bounded subset is relatively compact).

Cone topology. Suppose that X is a proper hyperbolic geodesic metric space. Later on, it will be convenient to use another topology on $\bar{X}$, called *the cone (or, radial) topology.* This topology is not equivalent to the topology τ: With few exceptions, $\bar{X}$ is noncompact with respect to this topology (even if $X = \mathbb{H}^n, n \geqslant 2$).

DEFINITION 9.93. Fix a base point $p \in X$. We use the metric topology on X and will say that a sequence $x_i \in X$ *conically converges* to a point $\xi \in \partial_\infty X$ if there is a constant R such that $x_i \in \mathcal{N}_R(p\xi)$ and

$$\lim_{i\to\infty} d(p, x_i) = \infty.$$

A subset $C \subset \bar{X}$ is closed in the conical topology if its intersection with X is closed in the metric topology of X and $C \cap \partial_\infty X$ contains conical limits of sequences in $C \cap X$. We will refer to the resulting topology as the *cone topology* on $\bar{X}$.

EXERCISE 9.94. If a sequence (x_i) converges to $\xi \in \partial_\infty X$ in the cone topology, then it also converges to ξ in the topology τ on $\bar{X}$.

As an example, consider $X = \mathbb{H}^n$ in the upper half-space model, $\xi = \mathbf{0} \in \mathbb{R}^{n-1}$, and let L be any vertical hyperbolic geodesic asymptotic to ξ. Then a sequence $x_i \in X$ converges ξ in the cone topology if and only if all the points x_i belong to some Euclidean cone with the axis L, while the Euclidean distance from x_i to $\mathbf{0}$ tends to zero. See Figure 9.10. This explains the name *cone topology.*

EXERCISE 9.95. Suppose that a sequence (x_i) converges to a point $\xi \in \partial_\infty \mathbb{H}^n$ along a horosphere centered at ξ. Show that the sequence (x_i) contains no convergent subsequences in the cone topology on $\bar{X}$.

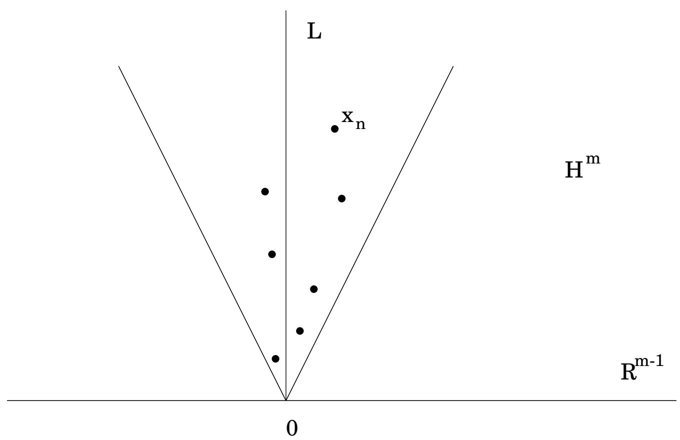


FIGURE 9.10. *C*onvergence in the cone topology.

9.12. Gromov boridification of Gromov-hyperbolic spaces

The definition of $\bar{X}$ and its topology, used in the previous section, worked fine for geodesic hyperbolic metric spaces. Gromov extended this definition to the case when X is an arbitrary (nonempty) δ-hyperbolic metric space.

Pick a base-point $p \in X$. A sequence (x_n) in X is said to *converge at infinity* if

$$\lim_{(m,n)\to\infty} (x_m, x_n)_p = \infty.$$

In particular, for such a sequence,

$$\lim_{n\to\infty} d(p, x_n) = \infty.$$

Define the relation $\sim$ on sequences converging at infinity by

$$(x_n) \sim (y_n) \iff \lim_{n\to\infty} (x_n, y_n)_p = \infty.$$

EXERCISE 9.96. 1. Show that $\sim$ is an equivalence relation using Definition 9.24.

2. Show that each sequence (x_n) converging at infinity is equivalent to each subsequence in (x_n).

3. Show that if $(x_n), (y_n)$ are inequivalent sequences converging at infinity, then

$$\sup_{m,n}(x_m, y_n)_p < \infty.$$

4. Show that for two sequences $(x_m), (y_n)$,

$$\lim_{(m,n)\to\infty}(x_m, y_n)_p = \infty$$

if and only if

$$\lim_{(m,n)\to\infty}(x_m, y_n)_q = \infty,$$

$q \in X$.

5. Suppose that $\mathbb{N} \to X, n \mapsto x_n$ is an isometric embedding. Show that the sequence (x_n) converges at infinity.

DEFINITION 9.97. The *Gromov boundary* $\partial_{Gromov}X$ of X consists of equivalence classes of sequences converging at infinity. Given a sequence (x_n) converging at infinity, we will use the notation $[x_n]$ for the eauivalence class of this sequences. The union $X \cup \partial_{Gromov}X$ is the *Gromov bordification* of X.

It will be convenient to extend the notation $[x_n]$ for sequences (x_n) which converge in X; we set

$$[x_n] = \lim_{n\to\infty} x_n \in X$$

for such sequences.

The Gromov product extends to $X \cup \partial_{Gromov}X$ by taking limits of Gromov products in X:

1.

$$(\xi, \eta)_p = \inf \lim \inf_{(m,n)\to\infty}(x_m, y_n)_x,$$

where the infimum is taken over all sequences $(x_m), (y_n)$ representing ξ and η respectively.

2.

$$(\xi, y)_p = \inf \lim \inf_{n\to\infty}(x_n, y)_p,$$

where the sequence (x_n) represents ξ. The infimum in this definition is again taken over all sequences (x_m) representing ξ.

REMARK 9.98. Taking the infimum and lim inf in this definition is, by no means, the only choice. However, all four possible choices in the definition of $(\xi, \eta)_p$ and $(\xi, y)_p$ differ by $\leqslant 2\delta$, see **[Väi05]**.

EXERCISE 9.99. For points $x, y \in X \cup \partial_{Gromov}X$ we have

$$(x, y)_p = \inf\{\lim \inf_{i\to\infty}(x_i, y_i)_p\}$$

where the infimum is taken over all sequences $(x_i), (y_i)$ in X such that $x = [x_i], y = [y_i]$.

LEMMA 9.100. *For all points $p \in X, x, y, z \in X \cup \partial_{Gromov} X$ we have the inequality*

$$(x, y)_p \geqslant \min\{(y, z)_p, (z, x)_p\} + \delta.$$

PROOF. For each point x, y, z we consider sequences $(x_i), (y_i), (z_i)$ in X such that $x = [x_i], y = [y_i], z = [z_i]$. We assume that $(x_i), (y_i)$ are chosen so that

$$(x, y)_p = \lim_{i\to\infty} (x_i, y_i)_p = (x, y)_p.$$

Then for each i we have

$$(x_i, y_i)_p \geqslant \min\{(y_i, z_i)_p, (z_i, x_i)_p\} + \delta.$$

Then

$$(x, y)_p = \lim_{i\to\infty} (x_i, y_i)_p \geqslant \lim\inf_{i\to\infty} \min\{(y_i, z_i)_p, (z_i, x_i)_p\} + \delta \geqslant$$
$$\min\{\lim\inf_{i\to\infty} (y_i, z_i)_p, \lim\inf_{i\to\infty} (z_i, x_i)_p\} + \delta$$

Now the claim follows from the Exercise 9.99. □

The *Gromov topology* on

$$\bar{X} = X \cup \partial_{Gromov} X$$

is the metric topology on X, while a basis of topology at $\xi \in \partial_{Gromov} X$ consists of the sets

$$\mathcal{U}_{\xi,R} = \{x \in \bar{X} : (\xi, x)_p > R\}.$$

Thus, a sequence (x_n) in $\bar{X}$ converges to $\xi \in \partial_{Gromov} X$ if and only if

$$\lim_{n\to\infty} (x_n, \xi)_p = \infty.$$

LEMMA 9.101. *A sequence (x_n) converges to $\xi \in \partial_{Gromov} X$ if and only if (x_n) converges at infinity and $[x_n] = \xi$.*

PROOF. 1. Suppose that (x_n) converges to $\xi \in \partial_{Gromov} X$. Then

$$\lim_{m\to\infty} (x_m, \xi)_p = \lim_{m\to\infty} (x_m, \xi)_p = \infty.$$

By Lemma 9.100, we have

$$(x_m, x_n) \geqslant \min\{(x_m, \xi)_p, (x_n, \xi)_p\} - \delta \to \infty_{n,m\to\infty}.$$

Hence, the sequence (x_n) converges at infinity. Let (y_n) be a sequence in X representing ξ. We claim that $(x_n) \sim (y_n)$. Indeed,

$$(x_n, y_n) \geqslant \min\{(x_n, \xi)_p, (y_n, \xi)_p\} - \delta \to \infty_{n\to\infty}.$$

Thus, $[x_n] = \xi$.

2. Suppose that $[x_n] = \xi$. For each n we pick a sequence $(y_{nm})_{m\in\mathbb{N}}$ representing ξ such that

$$(x_n, \xi) = \lim_{m\to\infty} (x_n, y_{nm})_p.$$

Taking a diagonal subsequence $(y_{n_k,m_k})_{k\in\mathbb{N}}$ which represents ξ, we obtain

$$\lim_{k\to\infty} (x_{n_k}, \xi)_p = \lim_{k\to\infty} (x_{n_k}, y_{n_k,m_k})_p = \infty.$$

Since

$$\lim_{k\to\infty} (x_{n_k}, x_k)_p = \infty,$$

Lemma 9.100 applied to the points x_{n_k}, x_n and ξ implies that

$$\lim_{k\to\infty} (x_x, \xi)_p = \infty. \quad \square$$

Suppose now that X is a geodesic metric space which is a δ_1-hyperbolic (in the sense of Rips) and δ_2-hyperbolic (in Gromov's sense). We define a map

$$h : X \cup \partial_\infty X \to X \cup \partial_{Gromov} X$$

which is the identity on X and sends $\xi = [\rho]$ in $\partial_\infty X$ to the equivalence class of the sequence $(\rho(n))$.

EXERCISE 9.102. The map h is well-defined, i.e.:

1. If $\rho : \mathbb{R}_+ \to X$ is a geodesics ray then the sequence $(\rho(n))$ converges at infinity.
2. If two rays ρ_1, ρ_2 are asymptotic then $[\rho_1(n)] = [\rho_2(n)]$.

LEMMA 9.103. *If X is a proper geodesic metric space then the map h is a bijection.*

PROOF. 1. Injectivity of h. Suppose that ρ_1, ρ_2 are rays in X emanating from $p \in X$, such that $[\rho_1(n)] = [\rho_2(n)]$. Set $x_n = \rho_1(n), y_n = \rho_2(n)$ and for each n choose a geodesic $x_n y_n$ in X. For each n let $z_n \in x_n y_n$ be a point within distance $\leqslant \delta_1$ from both sides px_n, py_n of the geodesic trian gle $T(px_n y_n)$. Let $x'_n \in px_n, y'_n \in py_n$ be points within distance $\leqslant \delta_1$ from z_n. Since $[x_n] = [y_n]$,

$$\lim_{n\to\infty} d(p, z_n) = \infty,$$

which implies that

$$\lim_{n\to\infty} d(p, x'_n) = \lim_{n\to\infty} d(p, y'_n) = \infty.$$

Since $d(x'_n, y'_n) \leqslant 2\delta_1$, in view of δ_1-thinness of the triangle $T(px'_n y'_n)$, we have that the Hausdorff distance between the geodesic segments px'_n, py'_n is $\leqslant 2\delta_1$. We conclude that the rays ρ_1, ρ_2 are Hausdorff-close to each other. Hence, the map h is injective.

2. Surjectivity of h. Let (x_n) be a sequence in X converging at infinity. Since X is proper, the sequence of geodesic segmanets px_n contains a subsequence px_{n_k} which converges (uniformly on compacts) to a geodesic ray $p\xi$ in X. For each k let $x'_{n_k} \in px_{n_k}$ denote a point within distance $\leqslant \delta$ from $p\xi$, and such that the distance $d(p, x'_{n_k})$ is maximal among all such points. Then

$$[x_n] = [x_{n_k}] = [x'_{n_k}].$$

Let $y_{n_k} \in p\xi$ denote a point within distance $\leqslant \delta_1$ from x'_{n_k}. Then

$$\lim_{n\to\infty} d(p, x'_{n_k}) = \infty$$

and $[x'_{n_k}] = [y'_{n_k}]$. It follows that h sends the equivalence class of the ray $p\xi$ to the equivalence class of the sequences $[x_n]$. □

THEOREM 9.104. *The map h is a homeomorphism.*

PROOF. We will verify continuity of h and h^{-1} at each point $\xi \in \partial_\infty X$. We fix $k \geqslant 10\delta$ and consider the topology $\mathcal{T}_{p,k}$ of shadow-convergence in $\bar{X}$.

1. Pick a ray $p\xi = \rho(\mathbb{R}_+)$ asymptotic to ξ. Suppose that (x_n) is a sequence in $\bar{X}$ which shadow-converges to $\xi \in \partial_\infty X$. Then there exists a sequence $t_n \in \mathbb{R}_+$ diverging to infinity, such that for $y_n = \rho(t_n)$, the segment (or a ray) px_n intersects the ball $B(y_n, k)$ at a point x'_n. Since

$$[x'_n] = [y_n] = \xi$$

$$\lim_{n\to\infty}(x'_n, \xi) = \infty,$$

see Lemma 9.101. On the other hand,

$$(x_n, x'_n)_p = d(p, x'_n) \to \infty.$$

The inequality

$$(x_n, \xi) \geqslant \min\{(x_n, x'_n)_p, (x'_n, \xi)_p\} + \delta_2$$

now implies that

$$\lim_{n\to\infty}(x_n, \xi) = \infty.$$

Therefore, the map h is continuous.

2. Suppose that $x_n \in X$ is a sequence converging at infinity, $[x_n] = \eta \in \partial_{Gromov} X$. We let $x\xi$ denote a geodesic ray in X with $\xi = h^{-1}(\eta)$. We will show that (x_n) shadow-converges to ξ. For each n pick a geodesic ray $x_n\xi$. Since

$$(x_n, \xi)_p \to \infty,$$

the minimal distance from p to $x_n\xi$ diverges to ∞. Let $z_n \in x_n\xi$ be a point within distance $\leqslant 4\delta_1$ from

$$x_n p \cup p\xi.$$

Let $y_n \in p\xi$, $u_n \in px_n$ be points within distance $\leqslant 4\delta_1$ from z_n. Since

$$\lim_{n\to\infty} d(p, z_n) = \lim_{n\to\infty} d(p, y_n) = \infty$$

and $u_n \in B(y_n, 8\delta_1)$, we conclude that the sequence (x_n) shadow-converges to ξ. The proof in the case when x_n is a sequence in $\partial_{Gromov} X$ converging to η is similar and is left to the reader. (Alternatively, continuity of h^{-1} follows from Lemma 1.16.) □

In view of this theorem, we will be identifying the visual and Gromov ideal boundaries of X.

9.13. Boundary extension of quasiisometries of hyperbolic spaces

The goal of this section is to explain how quasiisometries of Rips–hyperbolic spaces extend to their ideal boundaries.

9.13.1. Extended Morse Lemma. We first extend the Morse lemma to the case of quasigeodesic rays and complete geodesics.

LEMMA 9.105 (Extended Morse Lemma). *Suppose that X is a proper δ–hyperbolic geodesic space. Let ρ be an (L, A)–quasigeodesic ray or a complete (L, A)–quasigeodesic. Then there is ρ^*, which is either a geodesic ray or a complete geodesic in X, such that the Hausdorff distance between the images of ρ and ρ^* is at most $D(L, A, \delta)$. Here D is the function which appears in the Morse lemma.*

Moreover, there are two functions $s = s(t), s^* = s^*(t)$ such that:

$$L^{-1}t - B \leqslant s \leqslant Lt + B, \tag{9.7}$$

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B), \tag{9.8}$$

and for every t,

$$d(\rho(t), \rho^*(s)) \leqslant D, \quad d(\rho^*(t), \rho(s^*)) \leqslant D.$$

Here $B = A + D$.

Proof. We will consider only the case of quasigeodesic rays $\rho : [0, \infty) \to X$ as the other case is similar. For each i we define the finite quasigeodesic

$$\rho_i := \rho\big|_{[0,i]}$$

and the geodesic segment $\rho_i^* = px_i$, connecting the points $p = \rho(0), x_i = \rho(i)$. According to the Morse lemma,

$$\mathrm{dist}_{Haus}(\rho_i, \rho_i^*) \leqslant D(L, A, \delta).$$

By properness of X, the sequence of geodesic segments ρ_i^* subconverges to a complete geodesic ray ρ^*. It is clear that

$$\mathrm{dist}_{Haus}(\rho, \rho^*) \leqslant D(L, A, \delta).$$

The estimates (9.7) and (9.8) follow from the inequalities (9.3) and (9.4) in the case of finite geodesic segments. □

Corollary 9.106. *If ρ is a quasigeodesic ray as in the above lemma, there exists a point $\xi \in \partial_\infty X$ such that $\lim_{t\to\infty} \rho(t) = \xi$.*

Proof. According to Lemma 9.105, the quasigeodesic ray ρ is close to a geodesic ray $\rho^* = p\xi$. Since $d(\rho(t), p\xi) \leqslant D$ for all t, it follows that

$$\lim_{t\to\infty} \rho(t) = \xi. \quad \square$$

We will refer to the point η as $\rho(\infty)$. Note that if ρ' is another quasigeodesic ray Hausdorff-close to ρ, then $\rho(\infty) = \rho'(\infty)$.

Below is another useful application of the Extended Morse Lemma. Given a geodesic γ in X we let $\pi_\gamma : X \to \gamma$ denote a nearest-point projection.

Proposition 9.107 (Quasiisometries commute with projections). *There exists $C = C(L, A, \delta)$ such that the following holds. Let X, X' be proper δ-hyperbolic geodesic metric spaces and let $f : X \to X'$ be an (L, A)-quasiisometry. Let α be a (finite or infinite) geodesic in X, and let $\beta \subset X'$ be a geodesic which is $D(L, A, \delta)$-close to $f(\alpha)$. Then the map f* almost commutes *with the nearest-point projections π_α, π_β:*

$$d(f\pi_\alpha(x), \pi_\beta f(x)) \leqslant C, \quad \forall x \in X.$$

Proof. For each (finite or infinite) geodesic $\gamma \subset X$ consider the triangle $\Delta = \Delta_{x,\gamma}$ where one side of Δ is γ and x is a vertex: The other two sides of Δ are geodesics connecting x to the (finite or ideal) end-points of γ. We use the same definition for triangles in X'.

Let $c = center(\Delta) \in \gamma$ denote a centroid of Δ: The distance from c to each side of Δ is $\leqslant 6\delta$. By Corollary 9.62,

$$d(c, \pi_\gamma(x)) \leqslant 21\delta$$

for all $x \in X$. Consider now a geodesic $\alpha \subset X$, a point $x \in X$ and its image $y = f(x)$ in X'. We let β be a geodesic in X' within distance $\leqslant D(L, A, \delta)$ from $f(\alpha)$.

Figure 9.11. Quasiisometries almost commute with projections.

Applying f to the centroid $c_{x,\alpha} = c(\Delta_{x,\alpha})$, we obtain a point $a \in X'$ whose distance to each side of the quasigeodesic triangle $f(\Delta_{x,\alpha})$ is $\leqslant 2\delta L + A$. Hence, the distance from a to each side of the geodesic triangle $\Delta_{y,\beta}$ is at most $R := 2\delta L + A + D(L, A, \delta)$. Hence, a is an R-centroid of $\Delta_{y,\beta}$. By Lemma 9.87, it follows that the distance from a to the centroid $c_{y,\beta} = center(\Delta_{y,\beta})$ is at most $8R + 32\delta$. Since $d(\pi_\beta(y), c(\Delta_{y,\beta})) \leqslant 21\delta$, we obtain:

$$d(f(\pi_\alpha(x)), \pi_\beta f(x)) \leqslant C := 21\delta + 8R + 27\delta + 21\delta L + A. \quad \square$$

9.13.2. The extension theorem. We are now ready to prove the main theorem of this section, which is a fundamental fact of the theory of hyperbolic spaces:

THEOREM 9.108 (Extension Theorem). *Suppose that $f : X \to X'$ is a quasi-isometry between two Rips–hyperbolic proper metric spaces. Then f admits a homeomorphic extension $f_\infty : \partial_\infty X \to \partial_\infty X'$. This extension is such that the map $\bar{f} = f \cup f_\infty$ is continuous at each point $\eta \in \partial_\infty X$. The extension satisfies the following functoriality properties:*

1. For every pair of quasiisometries $f_i : X_i \to X_{i+1}, i = 1, 2$, we have

$$(f_2 \circ f_1)_\infty = (f_2)_\infty \circ (f_1)_\infty.$$

2. For every pair of quasiisometries $f_1, f_2 : X \to X'$ satisfying $\mathrm{dist}(f_1, f_2) < \infty$*, we have*

$$(f_2)_\infty = (f_1)_\infty.$$

PROOF. First, we construct the extension f_∞.

Given $\xi \in \partial_\infty X$, we pick a sequence (x_n) in X representing ξ. We claim that the sequence (y_n), $y_n = f(x_n)$, converges at infinity in Y. Indeed, according to Lemma 9.22, for any pair of indices m, n, we have

$$(x_m, x_n)_p \leqslant \mathrm{dist}(p, x_m x_n) \leqslant L\mathrm{dist}(f(p), f(x_m x_n)) + LA \leqslant$$

$$LD\mathrm{dist}(f(p), y_m y_n) + LA \leqslant LD((y_m, y_n)_{f(p)} + 2\delta) + LA$$

where $D = D(L, A, \delta)$ is the constant from the Morse Lemma. Therefore,

$$\lim_{m,n\to\infty} (y_m, y_n)_{f(p)} = \infty.$$

The same argument shows that if $[x_n] = [x'_n] = \xi$ then $[f(x_n)] = [f(x'_n)]$. Therefore, we set $f_\infty(\xi) := [y_n]$. We next show that for each $\xi \in \partial_\infty X$ the restriction of $\bar{f}$ to $X \cup \{\xi\}$ is continuous at ξ. Since $\bar{X}$ and $\bar{Y}$ are 1st countable, it suffices to verify sequential continuity. We have that $x_n \in X$ converges to ξ if and only if $[x_n] = \xi$. Since $\xi' = f_\infty(\xi) = [f(x_n)]$, it follows that the sequence $(f(x_n))$ converges to ξ'. Therefore, the restriction of $\bar{f}$ to $X \cup \{\xi\}$ is continuous at ξ. Since $\bar{X}$ is compact and Hausdorff, it is regular. Part 2 of Lemma 1.16 implies continuity of $\bar{f} : \bar{X} \to \bar{Y}$ at each $\xi \in \partial_\infty X$.

We next check the functoriality properties (1) and (2) of the extension maps. Suppose that $\xi = [x_n] \in \partial_\infty X$ (where (x_n) is a sequence in X converging at infinity), $\eta = (f_1)_\infty(\xi)$. Then

$$\eta = [f_1(x_n)], \quad (f_2)_\infty(\eta) = [f_2 \circ f_1(x_n)] = (f_2 \circ f_1)_\infty(\xi).$$

This implies Property 1. To verify Property 2, note that for each $\xi = [x_n]$ is also represented by the sequence (y_n), $y_n = f(x_n)$, since the distances $d(x_n, y_n)$ are uniformly bounded.

Lastly, we verify that f_∞ is a homeomorphism. Let g be a quasiinverse of $f : X \to X'$; this inverse also has a continuous extension

$$g_\infty : \partial_\infty X' \to \partial_\infty X.$$

Since $\mathrm{dist}(g \circ f, \mathrm{id}_X) < \infty$ and $\mathrm{dist}(f \circ g, \mathrm{id}_{X'}) < \infty$ by the functoriality properties, we obtain:

$$\mathrm{id}_{X'} = (f \circ g)_\infty = f_\infty \circ g_\infty,$$

$$\mathrm{id}_X = (g \circ f)_\infty = g_\infty \circ f_\infty.$$

Hence, g_∞ is the continuous inverse of f_∞, and

$$f_\infty : \partial_\infty X \to \partial_\infty X'$$

is a homeomorphism. □

EXERCISE 9.109. Suppose that f is merely a QI embedding $X \to X'$. Show that the continuous extension f_∞ given by $[x_n] \mapsto [f(x_n)]$ is injective.

HISTORICAL REMARK 9.110. The above extension theorem was first proven by Efremovich and Tikhomirova in [**ET64**] for the real-hyperbolic space and, soon afterwards, reproved by Mostow [**Mos73**]. We will see in Chapter 20 that the homeomorphisms f_∞ are *quasisymmetric*, in particular, they enjoy certain regularity properties which are critical for proving QI rigidity theorems in the context of hyperbolic groups and spaces.

The next lemma is a simple but useful corollary of Theorem 9.108 (the functoriality part):

COROLLARY 9.111. *Suppose that f, g, h are quasiisometries of X such that $\mathrm{dist}(h, g \circ f) < \infty$. Then*

$$h_\infty = g_\infty \circ f_\infty.$$

In particular, if $f : X \to X'$ is a quasiisometry quasiequivariant with respect to isometric group actions $G \curvearrowright X, G \curvearrowright X'$, then f_∞ is also G-equivariant.

We thus obtained a functor from quasiisometries between Rips–hyperbolic spaces to homeomorphisms between their boundaries.

The following lemma is a "converse" to the 2nd functoriality property in Theorem 9.108:

LEMMA 9.112. *Let X and Y be proper geodesic δ–hyperbolic spaces. In addition we assume that there exists $R < \infty$ such that every $x \in X$ is an R-centroid of an ideal triangle in X. Then quasiisometries $f, f' : X \to Y$ with equal extension maps $f_\infty = f'_\infty$ are uniformly close to each other. More precisely, there exists $D(L, A, R, \delta)$ such that each pair of (L, A)–quasiisometries $f, f' : X \to Y$ with $f_\infty = f'_\infty$, satisfies:*

$$\mathrm{dist}(f, f') \leqslant D(L, A, R, \delta).$$

PROOF. By Lemma 9.105, for each $x \in X$ the points $y = f(x), y' = f'(x)$ are C-centroids of an ideal geodesic triangle $S \subset Y$ whose ideal vertices are the images of the ideal vertices of T under f_∞. Here $C = LR + A + D(L, A, \delta)$. Lemma 9.87 implies that C-centroids of ideal triangles are uniformly close to each other:

$$d(y, y') \leqslant r(C, \delta).$$

We conclude that

$$d(f(x), f'(x)) \leqslant D(L, A, R, \delta) = 2(LR + A) + r(C, \delta). \quad \square$$

Suppose that X is hyperbolic and $\partial_\infty X$ contains at least 3 points. Then X has at least one ideal triangle and, hence, at least one centroid of an ideal triangle. If, in addition, X is quasihomogeneous, then, for some $R < \infty$, every $x \in X$ is an R-centroid of an ideal triangles in X. Thus, the above lemma applies to the real-hyperbolic space, more generally, all negatively curved symmetric space, and, as we will sees soon, all non-elementary hyperbolic groups.

EXERCISE 9.113. Suppose that X is a complete simply-connected Riemannian manifold of dimension > 1 and sectional curvature $\leqslant a < 0$. Show that every point $x \in X$ is an R-centroid of an ideal triangle, for some uniform R.

EXAMPLE 9.114. The line $X = \mathbb{R}$ is 0-hyperbolic, its ideal boundary consists of two points. Take a translation $f : X \to X$, $f(x) = x + a$. Then f_∞ is the identity map of $\{-\infty, \infty\}$ but there is no bound on the distance from f to the identity.

Here is an important corollary of Theorem 9.108 and Lemma 9.112:

COROLLARY 9.115. *Let X be a Rips–hyperbolic space. Then the map $f \mapsto f_\infty$, sending quasiisometries of X to homeomorphisms of $\partial_\infty X$, descends to a homomorphism $QI(X) \to Homeo(X)$. Furthermore, under the hypothesis of Lemma 9.112, this homomorphism is injective.*

In §20.5 we will identify the image of this homomorphism in the case of the real-hyperbolic space $\mathbb{H}^n$, it will be the subgroup of $Homeo(\mathbb{S}^{n-1})$ consisting of *quasimoebius* homeomorphisms.

9.13.3. Boundary extension and quasiactions. In view of Corollary 9.115, we have

COROLLARY 9.116. *Suppose that X is a Rips-hyperbolic space. Then every quasiaction ϕ of a group G on X extends (by $g \mapsto \phi(g)_\infty$) to an action ϕ_∞ of G on $\partial_\infty X$ by homeomorphisms.*

LEMMA 9.117. *Suppose that X satisfies the hypothesis of Lemma 9.112 and $G \curvearrowright X$ is a properly discontinuous quasiaction. Then the kernel for the associated boundary action ϕ_∞ is finite.*

PROOF. The kernel K of ϕ_∞ consists of the elements $g \in G$ such that the distance from $\phi(g)$ to the identity is finite. Since $\phi(g)$ is an (L, A)-quasiisometry of X, it follows from Lemma 9.112, that

$$\mathrm{dist}(\phi(g), \mathrm{id}_X) \leqslant D(L, A, R, \delta).$$

As ϕ was properly discontinuous, the subgroup K is finite. $\square$

9.13.4. Conical limit points of quasiactions. Suppose that ϕ is a quasiaction of a group G on a Rips–hyperbolic space X. A point $\xi \in \partial_\infty X$ is called a *conical limit point* for the quasiaction ϕ if there exists a sequence $g_i \in G$ such that $\phi(g_i)(x)$ converges to ξ in the conical topology. In other words, for some (equivalently every) geodesic ray $\gamma \subset X$ asymptotic to ξ, and some (equivalently every) point $x \in X$, there exists a constant $R < \infty$ such that:

- $\lim_{i \to \infty} \phi(g_i)(x) = \xi$.

- $d(\phi(g_i)(x), \gamma) \leqslant R$ for all i.

LEMMA 9.118. *Suppose that $\psi : G \curvearrowright X$ is a cobounded quasiaction. Then every point of the ideal boundary $\partial_\infty X$ is a conical limit point for ψ.*

PROOF. Let $\xi \in \partial_\infty X$ and let $x_i \in X$ be a sequence converging to ξ in the conical topology (e.g., we can take $x_i = \gamma(i)$, where γ is a geodesic ray in X asymptotic to ξ). Fix a point $x \in X$ and R such that for every $x' \in X$ there exists $g \in G$ satisfying

$$d(x', \phi(g)(x)) \leqslant R.$$

Then, by coboundedness of the quasiaction ψ, there exists a sequence $g_i \in G$ for which

$$d(x_i, \phi(g_i)(x)) \leqslant R.$$

It follows that ξ is a conical limit point of the quasiaction ψ. □

COROLLARY 9.119. *Suppose that G is a finitely-generated group, $f : X \to G$ is a quasiisometry and $G \curvearrowright G$ is the isometric action by the left multiplications. Let $\psi : G \curvearrowright X$ be the quasiaction, obtained by conjugating $G \curvearrowright G$ via f. Then every point of $\partial_\infty X$ is a conical limit point for the quasiaction ψ.*

PROOF. The action $G \curvearrowright G$ by the left multiplications is cobounded, hence, the conjugate quasiaction $\psi : G \curvearrowright X$ is also cobounded. □

If ϕ_∞ is a topological action of a group G on $\partial_\infty X$, obtained by the extension of a quasiaction ϕ of G on X, then *conical limit points* of the action $G \curvearrowright \partial_\infty X$ are defined as the conical limit points for the quasiaction $G \curvearrowright X$.

9.14. Hyperbolic groups

We now come to the *raison d'être* for δ-hyperbolic spaces, namely, *hyperbolic groups*.

DEFINITION 9.120. A finitely-generated group G is called *Gromov–hyperbolic* or *word-hyperbolic*, or simply *hyperbolic*, if one of its Cayley graphs is hyperbolic. A hyperbolic group is called *elementary* if it is virtually cyclic. A hyperbolic group is called non-elementary otherwise.

Immediate examples of hyperbolic groups are:

EXAMPLE 9.121. 1. Trivially, finite groups are hyperbolic.

2. Every finitely-generated free group is hyperbolic: Taking the Cayley graph corresponding to a free generating set, we obtain a simplicial tree, which is 0-hyperbolic.

We will see more examples of hyperbolic groups below.

Many examples of hyperbolic groups can be constructed *via the small cancelation* theory, see e.g. [**GdlH90, GS90, IS98**]. For instance, let G be a 1-relator group with the presentation

$$\langle x_1, \ldots, x_n | w^m \rangle ,$$

where $m \geqslant 2$ and w is a cyclically reduced word in the generators x_i. Then G is hyperbolic. (This was proven by B. B. Newman in [**New68**, Theorem 3] before the notion of hyperbolic groups was introduced; Newman proved that for such groups G the *Dehn's algorithm* applies, which is equivalent to hyperbolicity, see §9.16.)

Below is a combinatorial characterization of hyperbolic groups among Coxeter groups. Let Γ be a finite Coxeter graph and $G = C_\Gamma$ the corresponding Coxeter group. A *parabolic subgroup* of Γ is the Coxeter subgroup defined by a full subgraph Λ of Γ. It is clear that every parabolic subgroup of G admits a natural homomorphism to G, sending the generators $g_v, v \in V(\Lambda)$, to the generators g_v of G. As it turns out that such homomorphisms are always injective, see e.g. [**Hum97**], page 113.

<u>Theorem</u> 9.122 (G. Moussong [**Mou88**]). *A Coxeter group G is Gromov–hyperbolic if and only if the following condition holds:*

No parabolic subgroup of G is virtually isomorphic to the direct product of two infinite groups.

In particular, a Coxeter group is hyperbolic if and only if it contains no free abelian subgroups of rank 2.

Problem 9.123. Is there a similar characterization of Gromov–hyperbolic groups among Shephard groups and generalized von Dyck groups?

Another outstanding open problem of the same flavor is:

Problem 9.124 (flat closing problem). Suppose that G is a $CAT(0)$ group. Is it true that G contains a subgroup isomorphic to $\mathbb{Z}^2$?

This problem is open even for fundamental groups of closed Riemannian manifolds of nonpositive curvature and for 2-dimensional $CAT(0)$ groups.

Since changing generating sets does not affect the quasiisometry type of the Cayley graph and Rips–hyperbolicity is invariant under quasiisometries (Corollary 9.43), we conclude that a group G is hyperbolic if and only if all its Cayley graphs are hyperbolic. Furthermore, if groups G, G' are quasiisometric, then G is hyperbolic if and only if G' is hyperbolic. In particular, if G, G' are virtually isomorphic, then G is hyperbolic if and only if G' is hyperbolic. For instance, all virtually free groups are hyperbolic.

In view of the Milnor–Schwarz lemma:

Observation 9.125. If G is a group acting geometrically on a Rips–hyperbolic metric space, then G is also hyperbolic.

Definition 9.126. A group G is called $CAT(\kappa)$ if it admits a geometric action on a $CAT(\kappa)$ space.

Thus, every $CAT(-1)$ group is hyperbolic. In particular, fundamental groups of compact Riemannian manifolds of negative curvature are hyperbolic. If S is a compact connected surface then $\pi_1(S)$ is hyperbolic if and only if S is neither the torus nor the Klein bottle.

The following is an outstanding open problem in geometric group theory:

Problem 9.127. Construct a hyperbolic group G which is not a $CAT(-1)$ group.

Here are some examples of non-hyperbolic groups:

1. $\mathbb{Z}^n$ is not hyperbolic for every $n \geqslant 2$. Indeed, $\mathbb{Z}^n$ is QI to $\mathbb{R}^n$ and $\mathbb{R}^n$ is not hyperbolic (see Example 9.9).

2. A deeper fact is that hyperbolic groups cannot contain subgroups isomorphic to $\mathbb{Z}^2$.

3. More generally, if G contains a solvable subgroup S, then G is not hyperbolic unless S is virtually cyclic.

4. Even more generally, for every subgroup S of a hyperbolic group G, the group S is either elementary hyperbolic or contains a nonabelian free subgroup. In particular, every amenable subgroup of a hyperbolic group is virtually cyclic.

5. Furthermore, if $C \lhd G$ is an infinite cyclic normal subgroup of a hyperbolic group, then either C is finite, or G/C is finite.

We refer the reader to [**BH99**] for the proofs of 2, 3, 4 and 5.

Remark 9.128. There are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups, see Theorem 9.147. A subgroup $H \leqslant G$ of a hyperbolic group G is called *quasiconvex* if it is a quasiconvex subset of a Cayley graph of G. If $H \leqslant G$ is a quasiconvex subgroup , then, according to Theorem 9.50, H is quasiisometrically embedded in G and, hence, is hyperbolic itself.

Examples of quasiconvex subgroups are given by finite subgroups (which is clear) and (less obviously) infinite cyclic subgroups. Let G be a hyperbolic group with a word metric d. Define the *translation length* of $g \in G$ as

$$\|g\| := \lim_{n\to\infty} \frac{d(g^n, e)}{n}.$$

It is clear that $\|g\| = 0$ if g has finite order. On the other hand, every cyclic subgroup $\langle g\rangle \subset G$ is quasiconvex and $\|g\| > 0$ for every g of infinite order, see Chapter III.Γ, Propositions 3.10, 3.15 of [**BH99**].

Obstructions to hyperbolicity.

If a finitely-generated group G satisfies one of the following, then G is not hyperbolic:

(1) G contains an amenable subgroup which is not virtually cyclic.
(2) G contains an infinite cyclic subgroup which is not quasiisometrically embedded, i.e., an infinite order element g such that $\|g\| = 0$.
(3) G has infinite cohomological dimension over $\mathbb{Q}$.
(4) G does not contain a free nonabelian subgroup.
(5) G does not have the type $\mathbf{F}_\infty$.
(6) G contains infinitely many conjugacy classes of finite subgroups.
(7) G has unbounded torsion.
(8) G does not admit a uniformly proper map to a Hilbert space.
(9) G is not hopfian.

Proofs of 1—7 can be found in [**BH99**], while 8 and 9 are proven by Z. Sela in [**Sel92**] and [**Sel99**].

Strangely, all known examples of groups of the type $\mathbf{F}_3$ contain either $\mathbb{Z}^2$ or a solvable Baumslag–Solitar subgroup $BS(p, 1)$, cf. [**Bra99**]. In some cases, e.g., fundamental groups of compact 3-dimensional manifolds or free-by-cyclic groups (see [**Bri00**]), absence of such subgroups implies hyperbolicity. In the case of 3-dimensional manifolds, this result is a corollary of Perelman's Geometrization Theorem (it suffices to rule out free abelian subgroups of rank 2 in this case). The following is a well-known open problem:

Problem 9.129. Construct an example of a non-hyperbolic group of the type $\mathbf{F}_\infty$ which contains no Baumslag–Solitar subgroups.

9.15. Ideal boundaries of hyperbolic groups

We define the *ideal boundary* $\partial_\infty G$ of a hyperbolic group G as the ideal boundary of some (every) Cayley graph of G: It follows from Theorem 9.108, that boundaries of different Cayley graphs are equivariantly homeomorphic. Here are two simple examples of ideal boundaries of hyperbolic groups.

Since $\partial_\infty \mathbb{H}^n = \mathbb{S}^{n-1}$, we conclude that for the fundamental groups G of closed hyperbolic n-manifolds, $\partial_\infty G$ is homeomorphic to the sphere $\mathbb{S}^{n-1}$. The same applies to the fundamental groups of compact negatively curved n-dimensional Riemannian manifolds. Similarly, if $G = F_n$ is the free group of rank $n \geqslant 2$, then free generating set S of G yields the Cayley graph $X = \Gamma_{G,S}$, which is a simplicial tree of constant valence > 2. Therefore, as we saw in §9.11, $\partial_\infty X$ is homeomorphic to the Cantor set. Thus, $\partial_\infty F_n$ is the Cantor set as well.

LEMMA 9.130. *Let G be a hyperbolic group and $Z = \partial_\infty G$. Then Z consists of 0, 2 or continuum of points; in the latter case Z is perfect. In the first two cases, G is elementary, otherwise G is non-elementary. In the latter case, the kernel of the action $G \curvearrowright Z$ is the unique maximal finite normal subgroup of G.*

PROOF. Let X be a Cayley graph of G. If G is finite, then X is bounded and, hence $Z = \emptyset$. Thus, we assume that G is infinite. By Exercise 4.85, X contains a complete geodesic γ, thus, Z has at least two distinct points, the limit points of γ. If $\mathrm{dist}_{Haus}(\gamma, X) < \infty$, X is quasiisometric to $\mathbb{R}$ and, hence, G is 2-ended. Therefore, G is virtually cyclic by Part 3 of Theorem 6.22.

We assume, therefore, that $\mathrm{dist}_{Haus}(\gamma, X) = \infty$. Then there exists a sequence of vertices $x_n \in X$ satisfying $\lim \mathrm{dist}(x_n, \gamma) = \infty$. Let $y_n \in \gamma$ be a nearest vertex to x_n and $g_n \in G$ be such that $g_n(y_n) = e \in G$. Then applying g_n to the union of geodesics

$$x_n y_n \cup \gamma$$

and taking the limit as $n \to \infty$, we obtain a complete geodesic $\beta \subset X$ (the limit of a subsequence $g_n(\gamma)$) and a geodesic ray ρ meeting β at e, such that for every $x \in \rho$, e is a nearest point on γ to x. Therefore, $\rho(\infty)$ is a point different from $\gamma(\pm\infty)$, and Z contains at least three distinct points. Let p be a centroid of a corresponding ideal triangle. Then $G \cdot o$ is a 1-net in X and, we are, therefore, in the situation of Lemma 9.112. Let K denote the kernel of the action $G \curvearrowright Z$. Then every $k \in K$ moves every point in X by $\leqslant D(1, 0, 1, \delta)$, where D is the function defined in Lemma 9.112. It follows that K is a finite group. Since G is infinite, Z is also infinite. Suppose that $F \lhd G$ is a finite normal subgroup. Since the quotient map $G \to \bar{G} = G/F$ is a quasiisometry, it induces a G-equivariant homeomorphism $\partial_\infty G \to \partial_\infty \bar{G}$. Since F acts trivially on $\partial_\infty \bar{G}$, it acts trivially on $\partial_\infty G$. It follows that K is the unique maximal finite normal subgroup of G.

Let $\xi \in Z$ and let ρ be a ray in X asymptotic to ξ. Then, there exists a sequence $g_n \in G$ for which $g_n(e) = x_n \in \rho$. Let $\gamma \subset X$ be a complete geodesic asymptotic to points η, ζ different from ξ. We leave it to the reader to verify that either

$$\lim_n g_n(\eta) = \xi,$$

or

$$\lim_n g_n(\zeta) = \xi.$$

Since Z is infinite, we can choose ξ, η such that their images under the given sequence g_n are not all equal to ξ. Thus, ξ is an accumulation point of Z and, hence,

Z is a perfect topological space. Since Z is 2nd countable infinite, compact and Hausdorff, it follows that Z has the cardinality of continuum. $\square$

We next describe some dynamical properties of the actions of hyperbolic groups on their ideal boundaries.

DEFINITION 9.131. Let $G \subset Homeo(Z)$ be a group of homeomorphisms of a compact Hausdorff space Z. The group G is said to be a *convergence group* if G acts properly discontinuously on $\mathrm{Trip}(Z)$, where $\mathrm{Trip}(Z)$ is the set of triples of distinct elements of Z. A convergence group G is said to be a *uniform* if $\mathrm{Trip}(Z)/G$ is compact.

THEOREM 9.132 (P. Tukia, [**Tuk94**]). *Suppose that X is a proper δ-hyperbolic geodesic metric space with the ideal boundary $Z = \partial_\infty X$ consisting of at least three points. Let $G \curvearrowright X$ be an isometric action and $G \curvearrowright Z$ be the corresponding topological action. Then the action $G \curvearrowright X$ is geometric if and only if $G \curvearrowright Z$ is a uniform convergence action.*

PROOF. Recall that we have a correspondence $center : \mathrm{Trip}(Z) \to X$ sending each triple of distinct points in Z to the set of centroids of the corresponding ideal triangles. Furthermore, by Corollary 9.89, for every $\xi \in \mathrm{Trip}(Z)$,

$$\mathrm{diam}(center(\xi)) \leqslant 60\delta.$$

Clearly, the correspondence *center* is G-equivariant. Moreover, the image of every compact K in $\mathrm{Trip}(Z)$ under *center* is bounded (see Exercise 9.91).

Assume now that the action $G \curvearrowright X$ is geometric. Given a compact subset $K \subset \mathrm{Trip}(Z)$, suppose for a moment that the set

$$G_K := \{g \in G | gK \cap K \neq \emptyset\}$$

is infinite. Then there exists a sequence $\xi_n \in K$ and an infinite sequence $g_n \in G, g_0 = e \in G$, $g_n(\xi_n) \in K$ for all $n > 0$. The diameter of the set

$$E = \left(\bigcup_n center(g_n(\xi_n)) \right) \subset X$$

is bounded and each g_n sends some $p_n \in E$ to an element of E. This, however, contradicts proper discontinuity of the action of G on X. Thus, the action $G \curvearrowright \mathrm{Trip}(Z)$ is properly discontinuous.

Similarly, since the action $G \curvearrowright X$ is cobounded, the G-orbit of some metric ball $B(p, R)$ covers the entire X. Thus, using equivariance of *center*, for every $\xi \in \mathrm{Trip}(Z)$, there exists $g \in G$ such that

$$center(g\xi) \subset B = B(x, R + 60\delta).$$

Since $center^{-1}(B)$ is relatively compact in $\mathrm{Trip}(Z)$ (see Exercise 9.92), we conclude that G acts cocompactly on $\mathrm{Trip}(Z)$. We conclude that $G \subset Homeo(Z)$ is a uniform convergence group.

The proof of the converse is essentially the same argument run in the reverse. Let $K \subset \mathrm{Trip}(Z)$ be a compact whose G-orbit is the entire $\mathrm{Trip}(Z)$. Then the set $center(K)$, which is the union of sets of centroids of points $\xi' \in K$, is a bounded subset $B \subset X$. By equivariance of the correspondence *center*, it follows that the G-orbit of B is the entire X. Hence, the action $G \curvearrowright X$ is cobounded. The argument for proper discontinuity of the action $G \curvearrowright \mathrm{Trip}(Z)$ is similar, we use the fact that the preimage of a sufficiently large metric ball $B \subset X$ under the correspondence *center* is nonempty and relatively compact in $\mathrm{Trip}(Z)$. Then proper discontinuity of the action $G \curvearrowright X$ follows from proper discontinuity of $G \curvearrowright \mathrm{Trip}(Z)$. $\square$

COROLLARY 9.133. *Suppose that G is a nonelementary hyperbolic group. Then the image $\bar{G}$ of G in $Homeo(\partial_\infty G)$ is a uniform convergence group.*

The converse to Theorem 9.132 is a deep theorem of B. Bowditch [**Bow98b**]:

THEOREM 9.134. *Let Z be a perfect compact Hausdorff space consisting of more than one point. Suppose that $G \subset Homeo(Z)$ is a uniform convergence group. Then G is hyperbolic and, moreover, there exists an equivariant homeomorphism $Z \to \partial_\infty G$.*

Note that in the proof of Part 1 of Theorem 9.132 we did not really need the property that the action of G on itself was isometric, a geometric quasiaction (see Definition 5.58) suffices:

THEOREM 9.135. *Suppose that X is a δ-hyperbolic proper geodesic metric space. Assume that there exists R such that every point in X is an R-centroid of an ideal triangle in X. Let $\phi : G \curvearrowright X$ be a geometric quasiaction. Then the extension $\phi_\infty : G \to Homeo(Z)$, $Z = \partial_\infty X$, of the quasiaction ϕ to a topological action of G on Z, is a uniform convergence action.*

PROOF. The proof of this result closely follows the proof of Theorem 9.132; the only difference is that ideal triangles $T \subset X$ are not mapped to ideal triangles by quasiisometries $\phi(g), g \in G$. However, ideal quasigeodesic triangles $\phi(g)(T)$ are uniformly close to ideal triangles which suffices for the proof. $\square$

The next theorem related ends and ideal boundaries of hyperbolic spaces:

THEOREM 9.136. *Suppose that X is a Rips-hyperbolic proper metric space. Then there exists a continuous surjection*

$$\eta : \partial_\infty X \to \epsilon(X)$$

such that the preimages $\eta^{-1}(\xi)$ are connected components of $\partial_\infty X$. Moreover, the map η is equivariant with respect to the isometry group of X.

We refer the reader to [**GdlH90**, Chapter 7, Proposition 17] for a proof. We also refer the reader to [**KB02**] for the more detailed discussion of ideal boundaries of hyperbolic groups.

9.16. Linear isoperimetric inequality and Dehn algorithm for hyperbolic groups

Let G be a hyperbolic group, we suppose that Γ is a δ-hyperbolic Cayley graph of G. We will assume that $\delta \geqslant 2$ is a natural number. Recall that a *loop* in Γ is required to be a closed edge-path. Since the group G acts transitively on the vertices of Γ, the number of G-orbits of loops of length $\leqslant 12\delta$ in Γ is bounded. We attach a 2-cell along every such loop. Let X denote the resulting cell complex; the action of G on Γ extends naturally to a cellular action on X. Recall that for a loop γ in Γ, $\ell(\gamma)$ denotes the length of γ and $A(\gamma)$ the least combinatorial area of a disk in X bounding γ, see §4.10.

Our goal is to show that X is simply-connected and satisfies a linear isoperimetric inequality. We will prove a somewhat stronger statement. Namely, suppose that

X is a connected two-dimensional cell complex whose 1-skeleton $X^{(1)}$ (equipped with the standard metric) is δ-hyperbolic (with δ a natural number) and such that for every loop γ of length $\leqslant 12\delta$ in X, $A(\gamma) \leqslant K < \infty$. The following theorem was first proven by Gromov in Section 2.3 of **[Gro87]**:

THEOREM 9.137 (Hyperbolicity implies linear isoperimetric inequality). *Under the above assumptions, for every loop $\gamma \subset X$,*

$$A(\gamma) \leqslant K\ell(\gamma). \tag{9.9}$$

Since the argument in the proof of the theorem is by induction on the length of γ, the following proposition is the key. In the proposition, by saying that a loop γ based at a vertex $v \in X^{(1)}$ is a product of two loops γ_1, γ_2 in $X^{(1)}$, we mean that γ_1, γ_2 are also based at v and that γ represents the same element of $\pi_1(X^{(1)}, v)$ as the (concatenation) product $\gamma_1 \star \gamma_2$. Furthermore, d is the standard metric on the graph $X^{(1)}$.

PROPOSITION 9.138. *Every loop γ in $X^{(1)}$ of length larger than 12δ is a product of two loops, one of the length $\leqslant 12\delta$ and another one of the length $< \ell(\gamma)$.*

PROOF. We assume that γ is parameterized by its arc-length, and that $\ell(\gamma) = n$.

Case 1. Assume that there exists a vertex $u = \gamma(t)$ such that the vertex $v = \gamma(t + 6\delta)$ satisfies $d(u, v) < 6\delta$. After reparameterizing γ, we may assume that $t = 0$. Let p denote a geodesic vu in $X^{(1)}$ and $-p$ the same geodesic run in the reverse. Then γ is the product of the loops

$$\gamma_1 = \gamma([0, 6\delta]) \star p$$

and

$$\gamma_2 = (-p) \star \gamma([6\delta, n]).$$

(Here $u = \gamma(0)$ serves as a base-point.) Since $\ell(p) < \ell(\gamma([0, 6\delta]))$, we have $\ell(\gamma_1) \leqslant 12\delta$ and $\ell(\gamma_2) < \ell(\gamma_1)$. Thus, the statement of the proposition holds in the Case 1.

FIGURE 9.12. Case 1.

Case 2. Assume now that for every integer t, $d(\gamma(t), \gamma(t + 6\delta)) = 6\delta$, where $t + 6\delta$ is considered modulo n. In other words, every subarc of γ of length 6δ is a geodesic segment in $X^{(1)}$.

Set $v_0 = \gamma(0)$ and let $v = \gamma(t)$ denote a vertex whose distance $d(v, v_0)$ to v_0 is the largest possible, in particular it is at least 6δ.

FIGURE 9.13. Case 2.

Define the vertices $v_\pm = \gamma(t \pm 3\delta)$ on γ and consider the geodesic triangle $T = v_0 v_- v_+$ with the edge $v_- v_+$ equal to the geodesic subarc of γ between these vertices. Since the triangle T is δ-thin, the point $v \in v_- v_+$ is within distance $\leqslant \delta$ either from the side $v_0 v_-$ or from $v_0 v_+$. After reparameterizing γ in the reverse direction if necessary, we may assume that there exists a vertex $u \in v_0v_+$ within distance $\leqslant \delta$ from v. Set

$$r = d(v_0, u), \quad s = d(u, v_+).$$

Then, by the triangle inequalities, $d(v_0, v) \leqslant r+\delta$, while $s \geqslant 3\delta - \delta = 2\delta$. Therefore,

$$d(v_0, v) = r + s \geqslant r + 2\delta > r + \delta \geqslant d(v_0, v).$$

This contradicts our choice of v as the point in $X^{(0)}$ on γ with the largest distance to v_0. We, thus, conclude that the Case 2 cannot occur. □

Proof of Theorem 9.137. The proof of the inequality (9.9) is by induction on the length of γ.

1. If $\ell(\gamma) \leqslant 12\delta$ then $A(\gamma) \leqslant K \leqslant K\ell(\gamma)$.

2. Suppose that the inequality holds for $\ell(\gamma) \leqslant n$, $n \geqslant 12\delta$. If $\ell(\gamma) = n+1$, then γ is the product of loops γ_1, γ_2 as in Proposition 9.138: $\ell(\gamma_2) < \ell(\gamma)$, $\ell(\gamma_1) \leqslant 12\delta$. Then, inductively,

$$A(\gamma_2) \leqslant K\ell(\gamma_2), \quad A(\gamma_1) \leqslant K,$$

and, thus,

$$A(\gamma) \leqslant A(\gamma_2) + A(\gamma_1) \leqslant K\ell(\gamma_2) + K \leqslant K\ell(\gamma). \quad \square$$

Below are two corollaries of Proposition 9.138, which was the key to the proof of the linear isoperimetric inequality.

Corollary 9.139 (M. Gromov, [**Gro87**]). *Every hyperbolic group is finitely-presented.*

Proof. Proposition 9.138 means that every loop in the Cayley graph of Γ is a product of loops of length $\leqslant 12\delta$. Attaching 2-cells to Γ along the G-images of these loops we obtain a simply-connected complex X on which G acts geometrically. Thus, G is finitely-presented. □

Corollary 9.140 (M. Gromov, [**Gro87**], section 6.8N). *Let Y be a coarsely connected Rips–hyperbolic metric space. Then X satisfies the linear isoperimetric inequality:*

$$Ar_\mu(\mathfrak{c}) \leqslant K\ell(\mathfrak{c})$$

for all sufficiently large μ and for appropriate $K = K(\mu)$.

Proof. Quasiisometry invariance of isoperimetric functions implies that it suffices to prove the assertion for Γ, the 1-skeleton of a connected R-Rips complex $Rips_R(X)$ of X. By Proposition 9.138, every loop γ in Γ is a product of $\leqslant \ell(\gamma)$ loops of length $\leqslant 12\delta$, where Γ is δ-hyperbolic in the sense of Rips. Therefore, for any $\mu \geqslant 12\delta$, we get

$$Ar_\mu(\gamma) \leqslant \ell(\gamma). \quad \square$$

Dehn algorithm. A (finite) presentation $\langle X|R\rangle$ is called *Dehn* if for every nontrivial word w representing $1 \in G$, the word w contains more than half of a defining relator. A word w is called *Dehn-reduced* if it contains no more than half of any relator. Given a word w, we can inductively reduce the length of w by replacing subwords u in w with u' such that $u'u^{-1}$ is a relator and $|u'| < |u|$. This, of course, does not change the element g of G represented by w. As the length of w decreases on each step, eventually, we get a Dehn-reduced word v representing $g \in G$. Since the presentation $\langle X|R\rangle$ is Dehn, either $v = 1$ (in which case $g = 1$)

or $v \neq 1$ in which case $g \neq 1$. This algorithm is, probably, the simplest way to solve the word problem in groups. It is also, historically, the oldest: Max Dehn introduced it in order to solve the word problem for hyperbolic surface groups.

Geometrically, Dehn reduction represents a based homotopy of the path in X represented by the word w (the base-point is $1 \in G$). Similarly, one defines the *cyclic* Dehn reduction, where the reduction is applied to (unbased) loops represented by w and the *cyclically Dehn* presentation: If w is a null–homotopic loop in X then this loop contains a subarc which is more than half of a relator. Again, if G admits a cyclically Dehn presentation then the word problem in G is solvable.

LEMMA 9.141. *Each δ-hyperbolic group G admits a finite (cyclically) Dehn presentation.*

PROOF. Start with an arbitrary finite presentation of G. Then add to the list of relators all the words of length $\leqslant 12\delta$ representing the identity in G. Since the set of such words is finite, we obtain a new finite presentation of the group G. The fact that the new presentation is (cyclically) Dehn is just the induction step of the proof of Proposition 9.137. □

Note, however, that the construction of a (cyclically) Dehn presentation requires solvability of the word problem for G (or, rather, for the words of the length $\leqslant 12\delta$) and, hence, is not *a priori* algorithmic. Nevertheless, we will see below that a Dehn presentation for δ-hyperbolic groups (with known δ) is algorithmically computable.

The converse of Proposition 9.137 is true as well, i.e., if a finitely-presented group satisfies a linear isoperimetric inequality then it is hyperbolic. We shall discuss this in §9.22.

9.17. The small cancellation theory

As we noted earlier, one of the origins of the theory of hyperbolic groups is the *small cancellation theory.* In this section we briefly discuss one class of *small cancellation conditions*, namely, $C'(r)$. We refer the reader to the books by Lyndon and Schupp [**LS77**], Ol′shanskiĭ [**Ol′91a**], and the appendix by Strebel to [**GdlH90**], for details.

Consider a presentation $P = \langle X|R\rangle$. We define a new set of relators R^* by first symmetrizing R (adding the relator R_k^{-1} for each relator $R_k \in R$) and then cyclically conjugating each relator by generators $x_i \in X$. The new set of relators R^* is symmetric ($R = R^{-1}$) and is invariant under conjugation via generators $x_i \in X$ and their inverses.

DEFINITION 9.142. The presentation $P^* = \langle X|R^*\rangle$ is *reduced* if each relator in R^* is reduced and no relator is repeated.

A (nonempty) word w in $X \cup X^{-1}$ is called a *piece* with respect to the presentation P^* if w appears as a common prefix in two distinct elements R_i, R_j of R^*. The *relative length* of a piece w is

$$\ell_{rel}(w) = \max_{R_i} \frac{|w|}{|R_i|},$$

where the maximum is taken over all relators $R_i \in R^*$ containing w as a prefix.

Definition 9.143. For $\lambda > 0$, a presentation P is said to satisfy the *small cancellation condition* $C'(\lambda)$ if each piece of $\ell_{rel}(w) \leqslant \lambda$ for each piece w.

Theorem 9.144. *For each presentation P satisfying the condition $C'(1/7)$, the presentation P^* is cyclically Dehn.*

Corollary 9.145. *If G a group admits a finite presentation satisfying the condition $C'(1/7)$, then G is hyperbolic.*

This argument is a typical example of the small cancelation theory, see . Rips in his paper [**Rip82**], did not use the language of hyperbolic groups, but the language of the small cancelation theory.

Theorem 9.146 (S. Gersten, [**Ger87**]). *The presentation complex of a presentation P satisfying the condition $C'(1/6)$ is aspherical.*

9.18. The Rips construction

The goal of this section is to describe the *Rips construction*, which associates a hyperbolic group to an arbitrary finite presentation of an arbitrary group.

Theorem 9.147 (The Rips Construction, E. Rips [**Rip82**]). *Let Q be a group with a finite presentation $\langle A|R\rangle$. Then, this presentations gives rise to a short exact sequence*

$$1 \to K \to G \to Q \to 1$$

where G is hyperbolic and K is finitely generated. Furthermore, the group K in this construction is finitely-presentable if and only if Q is finite.

Proof. Let $A = \{a_1, ..., a_m\}$, $R = \{R_1, ..., R_n\}$. For $i = 1, ..., m, j = 1, 2$, pick even natural numbers $r_i < s_i$, $p_{ij} < q_{ij}$, $u_{ij} < v_{ij}$, such that all the intervals

$$[r_i, s_i], \quad [p_{ij}, q_{ij}] \quad [u_{ij}, v_{ij}], i = 1, ..., m, j = 1, 2$$

are pairwise disjoint and all the numbers $r_i, s_i, p_{ij}, q_{ij}, u_{ij}, v_{ij}$ are at least 10 times larger than the lengths of the words in R. Define the group G by the presentation P, where the generators are $a_1, ..., a_m, b_1, b_2$, and the relators are:

$$R_i b_1 b_2^{r_2} b_1 b_2^{r_i+1} \cdots b_1 b_2^{s_i}, \quad i = 1, ..., n. \tag{9.10}$$

$$a_i^{-1} b_j a_i b_1 b_2^{u_{ij}} b_1 b_2^{u_{ij}+1} \cdots b_1 b_2^{v_{ij}}, \quad i = 1, ..., m, j = 1, 2. \tag{9.11}$$

$$a_i b_j a_i^{-1} b_1 b_2^{p_{ij}} b_1 b_2^{p_{ij}+1} \cdots b_1 b_2^{q_{ij}}, \quad i = 1, ..., m, j = 1, 2. \tag{9.12}$$

Now, define the map

$$\tilde{\phi}(a_i) = a_i, i = 1, ..., m, \quad \phi(b_j) = 1, \quad j = 1, 2.$$

The map $\tilde{\phi}$ extends to a epimorphism $F_{m+2} \to Q$ which sends all the relators R_k to $1 \in Q$; therefore, it descends to an epimorphism $\phi : G \to Q$. We claim that the kernel K of ϕ is generated by b_1, b_2. First, the kernel, of course, contains b_1, b_2. The subgroup generated by b_1, b_2 is clearly normal in G, because of the relators (9.11) and (9.12). Thus, indeed, b_1, b_2 generate K.

EXERCISE 9.148. The presentation P satisfies the small cancellation condition $C'(1/7)$. Hint: Show that the product of generators b_1, b_2 appearing at the end of each relator cannot get cancelled when we multiply conjugates of the relators in P and their inverses.

In particular, the group G is hyperbolic. In view of Theorem 9.146, the presentation complex of the presentation P is aspherical. Therefore, G has cohomological dimension $\leqslant 2$.

Lastly, we will verify that K cannot be finitely-presentable, unless Q is finite. R. Bieri proved in [**Bie76b**, Theorem B] that if G is a group of cohomological dimension $\leqslant 2$ and $H \lhd G$ is a finitely-presentable normal subgroup of infinite index, then H is free.

Suppose that the subgroup K is free. Then rank of K is at most 2 since K is 2-generated. The elements $a_1, a_2 \in G$ act on K as automorphisms (by conjugation). However, considering the action of a_1, a_2 on the abelianization, we see that because p_{ij}, q_{ij} are even, the images of the generators b_1, b_2 cannot generate the abelianization of K. A similar argument shows that K cannot be cyclic; therefore, K is trivial and, hence, $b_1 = b_2 = 1$ in G. However, this clearly contradicts the fact that the presentation (9.10) — (9.12) is a Dehn presentation. □

The Rips construction shows that there are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups. Furthermore,

COROLLARY 9.149. *Some hyperbolic groups have unsolvable membership problem.*

PROOF. Indeed, start with a finitely-presented group Q with unsolvable word problem and apply the Rips construction to Q. Then $g \in G$ belongs to the normal subgroup $K \lhd G$ if and only if g maps to the identity in Q. Since Q has unsolvable word problem, the problem of membership of g in K is unsolvable as well. □

On the other hand, the membership problem is solvable for quasiconvex subgroups, see Theorem 9.203.

9.19. Central co-extensions of hyperbolic groups and quasiisometries

We now consider a central co-extension

$$1 \to A \to \tilde{G} \xrightarrow{r} G \to 1 \tag{9.13}$$

with A a finitely-generated abelian group and G hyperbolic. The main result of this section is:

THEOREM 9.150 (W. Neumann, L. Reeves, [**NR97**]). *The group $\tilde{G}$ is QI to $A \times G$.*

PROOF. In the case when $A \cong \mathbb{Z}$, the first published proof belongs to S. Gersten [**Ger92**], although, it appears that D.B.A. Epstein and G. Mess also knew this result. Our proof follows the one in [**NR97**].

First of all, since an epimorphism with finite kernel is a quasiisometry, it suffices to consider the case when A is free abelian of finite rank. Our main goal is to construct a Lipschitz section (which is not a homomorphism!) $s : G \to \tilde{G}$ of the sequence (9.13). We first consider the case when A is infinite cyclic. Each fiber $r^{-1}(g), g \in G$, admits a canonical bijection to $\mathbb{Z}$:

$$ga \mapsto a \in A.$$

This defines a natural order $\leq$ on $r^{-1}(g)$. We let ι denote the embedding

$$\mathbb{Z} \cong A \hookrightarrow \tilde{G}.$$

Fix $\mathcal{X}$, a symmetric generating set of $\tilde{G}$; we will use the same name for its image under r. We let $\langle \mathcal{X} | \mathcal{R} \rangle$ be a finite presentation of G. We will use the notation $|w|$ for the word length with respect to this generating set, $w \in \mathcal{X}^*$, where $\mathcal{X}^*$ is the set of all words in $\mathcal{X}$, as in §4.2. Lastly, let $\tilde{w}$ and $\bar{w}$ denote the elements of $\tilde{G}$ and G respectively, represented by $w \in \mathcal{X}^*$.

LEMMA 9.151. *There is $C \in \mathbb{N}$ such that for every $g \in G$ the subsets*

$$\{\tilde{w}\iota(-C|w|) : w \in \mathcal{X}^*, \bar{w} = g\} \subset r^{-1}(g)$$

are bounded from above with respect to the order $\leq$.

PROOF. We will use the fact that G satisfies the linear isoperimetric inequality

$$Area(\alpha) \leqslant K|\alpha|$$

for every $\alpha \in \mathcal{X}^*$ representing the identity in G. We will assume that $K \in \mathbb{N}$. For each $R \in \mathcal{X}^*$ such that $R^{\pm 1}$ is a defining relator for G, the word R represents some $\tilde{R} \in A$. Therefore, since G is finitely-presented, we define a natural number T for which

$$\iota(T) = \max\{\tilde{R} : R^{\pm 1} \text{ is a defining relator of } G\}.$$

We claim that for each $u \in \mathcal{X}^*$ representing the identity in G,

$$\iota(T Area(u)) \geq \tilde{u} \in A. \tag{9.14}$$

Since general relators u of G are products of words of the form hRh^{-1}, $R \in \mathcal{R}$, (where $Area(u)$ is at most the number of these terms in the product) it suffices to verify that for $w = h^{-1}Rh$,

$$\tilde{w} \leq \iota(T),$$

where R is a defining relator of G and $h \in \mathcal{X}^*$. The latter inequality follows from the fact that the multiplications by $\bar{h}$ (resp. $\bar{h}^{-1}$) determine an order isomorphism (resp. its inverse) between $r^{-1}(1)$ and $r^{-1}(\bar{h})$.

Set $C := TK$. We are now ready to prove lemma. Let w, v be in $\mathcal{X}^*$ representing the same element $g \in G$. Set $u := v^{-1}$. Then $q = wu$ represents the identity and, hence, by (9.14),

$$\tilde{q} = \tilde{w}\tilde{u} \leq \iota(C|q|) = \iota(C|w|) + \iota(C|u|).$$

We now switch to the addition notation for $A \cong \mathbb{Z}$. Then,

$$w - v \leq \iota(C|w|) + \iota(C|v|),$$

and

$$w - \iota(C|w|) \leq v + \iota(C|v|).$$

Therefore, taking v to be a fixed word representing g, we conclude that all the differences $w - \iota(C|w|)$ are bounded from above. $\square$

In view of this lemma, we define a section $s : G \to \tilde{G}$

$$s(g) := \max\{\tilde{w}\iota(-C|w|) : w \in \mathcal{X}^*, \bar{w} = g\}$$

of the exact sequence (9.13). The unique word $w = w_g$ realizing the maximum in the definition of s is called *maximizing*. The section s, of course, need not be a group homomorphism. We will see, nevertheless, that it is not far from being one. Define the cocycle

$$\sigma(g_1, g_2) := s(g_1)s(g_2) - s(g_1g_2),$$

where the difference is taking place in $r^{-1}(g_1g_2)$. The next lemma does not use hyperbolicity of G, only the definition of s.

LEMMA 9.152. *The set $\sigma(G, X)$ is finite.*

PROOF. Let $x \in \mathcal{X}$, $g \in G$. We have to estimate the difference

$$s(g)x - s(gx).$$

Let w_1 and w_2 denote maximizing words for g and gx respectively. Note that the word w_1x also represents gx. Therefore, by the definition of s,

$$\widetilde{w_1x}\iota(-C(|w_1| + 1)) \le \tilde{w}_2\iota(-C|w_2|).$$

Hence, there exists $a \in A, a \geqslant 0$, satisfying

$$\widetilde{w_1}\iota(-C(|w_1|)\widetilde{x}\iota(-C)a = \tilde{w}_2\iota(-C|w_2|)$$

and, thus

$$s(g)\widetilde{x}\iota(-C)a = s(gx). \tag{9.15}$$

Since w_2x^{-1} represents g, we similarly obtain

$$s(gx)\widetilde{x}^{-1}\iota(-C)b = s(g), \quad b \ge 0, b \in A. \tag{9.16}$$

By combining equations (9.15) and (9.16), and switching to the additive notation for the group operation in A we get

$$a + b = \iota(2C).$$

Since $a \ge 0, b \ge 0$, we conclude that $-\iota(C) \le a - \iota(C) \le \iota(C)$. Therefore, (9.15) implies that

$$|s(g)x - s(gx)| \leqslant C.$$

Since the finite interval $[-\iota(C), \iota(C)]$ in A is a finite set, lemma follows. □

REMARK 9.153. Actually, more is true: The image of $\sigma : G \times G \to A$ is a finite set; in other words, the map $s : G \to \tilde{G}$ is a *quasihomomorphism* and the extension class of the central coextension (9.13) is a bounded cohomology class.

Moreover, all (degree $d \geqslant 2$) cohomology classes of hyperbolic groups are *bounded*: The natural homomorphism

$$H_b^d(G, A) \to H^d(G, A)$$

is surjective, see §3.9.3 for the definition of the bounded cohomology groups H_b^*. However, the proof is more difficult; we refer the reader to **[Min01]** for the details.

Letting L denote the maximum of the word lengths (with respect to the generating set $\mathcal{X}$) of the elements in the sets $\sigma(G, \mathcal{X}), \sigma(\mathcal{X}, G)$, we conclude (in view of Lemma 9.152) that the map $s : G \to \tilde{G}$ is $(L + 1)$–Lipschitz. Given the section $s : G \to \tilde{G}$, we define the projection $\phi = \phi_s : \tilde{G} \to A$ by

$$\phi(\tilde{g}) = \tilde{g} - s \circ r(\tilde{g}). \tag{9.17}$$

It is immediate that ϕ is Lipschitz since s is Lipschitz.

We now extend this construction to the case of central co-extensions with free abelian kernel of finite rank. Let $A = \prod_{i=1}^n A_i, A_i \cong \mathbb{Z}$. Consider the central co-extension (9.13). The homomorphisms $A \to A_i$ induce quotient maps $\eta_i : \tilde{G} \to \tilde{G}_i$ with the kernels $\prod_{j \neq i} A_j$. Each $\tilde{G}_i$, in turn, is a central co-extension

$$1 \to A_i \to \tilde{G}_i \xrightarrow{r_i} G \to 1. \tag{9.18}$$

Assuming that each central co-extension (9.18) has a Lipschitz section s_i, we obtain the corresponding Lipschitz projection $\phi_i : \tilde{G}_i \to A_i$ given by (9.17). This yields a Lipschitz projection

$$\Phi : \tilde{G} \to A, \Phi = (\phi_1 \circ \eta_1, \ldots, \phi_n \circ \eta_n).$$

We now set

$$s(r(\tilde{g})) := \tilde{g} - \Phi(\tilde{g}).$$

It is straightforward to verify that s is well-defined and is Lipschitz, provided that each s_i is. We thus obtain:

Corollary 9.154. *Given a finitely-generated free abelian group A and a hyperbolic group G, each central co-extension* (9.13) *admits a Lipschitz section $s : G \to \tilde{G}$ and a Lipschitz projection $\Phi : \tilde{G} \to A$ given by*

$$\Phi(\tilde{g}) = \tilde{g} - s(r(\tilde{g})).$$

Using this corollary, we define the map

$$h : G \times A \to \tilde{G}, \quad h(g, a) = s(g) + \iota(a)$$

and its inverse

$$h^{-1} : \tilde{G} \to G \times A, \quad \hat{h}(\tilde{g}) = (r(\tilde{g}), \Phi(\tilde{g})).$$

Since homomorphisms are 1-Lipschitz while the maps r and Φ are Lipschitz, we conclude that h is a bi-Lipschitz quasiisometry. □

Remark 9.155. The same proof goes through in the case of an arbitrary finitely-generated group G and a central co-extension (9.13) given by a *bounded 2-nd cohomology class*, cf. **[Ger92]**.

Example 9.156. Let $G = \mathbb{Z}^2$, $A = \mathbb{Z}$. Since $H^2(G, \mathbb{Z}) = H^2(T^2, \mathbb{Z}) \cong \mathbb{Z}$, the group G admits nontrivial central co-extensions with the kernel A, for instance, the integer Heisenberg group H_3. The group $\tilde{G}$ for such a co-extension is nilpotent but not virtually abelian. Hence, by Pansu's theorem (Theorem 14.26), $\tilde{G}$ is not quasiisometric to $G \times A = \mathbb{Z}^3$.

One can ask if Theorem 9.150 generalizes to other normal co-extensions of hyperbolic groups G. We note that Theorem 9.150 does not extend, say, to the case where A is a non-elementary hyperbolic group and the action $G \curvearrowright A$ is trivial. The reason is the *quasiisometric rigidity* for products of certain types of groups proven in [**KKL98**]. A special case of this theorem says that if $G_1, ..., G_n$ are non-elementary hyperbolic groups, then quasiisometries of the product $G = G_1 \times ... \times G_n$ quasipreserve the product structure:

Theorem 9.157. *Let $\pi_j : G \to G_j, j = 1, \ldots, n$ be natural projections. Then for each (L, A)–quasiisometry $f : G \to G$, there is $C = C(G, L, A) < \infty$, such that, up to composing with a permutation of quasiisometric factors G_k, the map f is within distance $\leqslant C$ from a product map $f_1 \times \ldots \times f_n$, where each $f_i : G_i \to G_i$ is a quasiisometry and C depends only on δ, n, L and A.*

9.20. Characterization of hyperbolicity using asymptotic cones

The goal of this section is to strengthen the relation between hyperbolicity of geodesic metric spaces and 0-hyperbolicity of their asymptotic cones.

PROPOSITION 9.158 (§2.A, **[Gro93]**). *Let* (X, dist) *be a geodesic metric space. Assume that either of the following two conditions holds:*

(a) *There exists a non-principal ultrafilter* ω *such that for all sequences* $\boldsymbol{e} = (e_n)_{n\in\mathbb{N}}$ *of base-points* $e_n \in X$ *and* $\boldsymbol{\lambda} = (\lambda_n)_{n\in\mathbb{N}}$ *of scaling constants with* $\omega\text{-}lim\, \lambda_n = 0$*, the asymptotic cone* $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *is a real tree.*

(b) *For every non-principal ultrafilter* ω *and every sequence* $\boldsymbol{e} = (e_n)_{n\in\mathbb{N}}$ *of base-points, the asymptotic cone* $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *is a real tree, where* $\boldsymbol{\lambda} = (n^{-1})$.

Then (X, dist) *is hyperbolic.*

The proof of Proposition 9.158 relies on the following lemma, whose proof follows closely the proof of the Morse Lemma (Theorem 9.40).

LEMMA 9.159. *Assume that a geodesic metric space* (X, dist) *satisfies either the property (a) or the property (b) in Proposition 9.158. Then there exists* $M > 0$ *such that for every geodesic triangle* $\Delta(x, y, z) \subset X$ *with* $\text{dist}(y, z) \geqslant 1$*, the two edges with the endpoint* x *are at Hausdorff distance at most* $M\text{dist}(y, z)$.

PROOF. Suppose to the contrary that there exist sequences of triples of points x_n, y_n, $z_n \in X$, such that $\text{dist}(y_n, z_n) \geqslant 1$ and

$$\text{dist}_{Haus}(x_n y_n, x_n z_n) = M_n \text{dist}(y_n, z_n),$$

such that $M_n \to \infty$. Let a_n be a point on $x_n y_n$ such that

$$\delta_n := \text{dist}(a_n, x_n z_n) = \text{dist}_{Haus}(x_n y_n, x_n z_n).$$

Since $\delta_n \geqslant M_n$, it follows that $\delta_n \to \infty$.

(a) Assume that the condition **(a)** holds. Consider the sequence of base-points $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ and the sequence of scaling constants $\boldsymbol{\lambda} = (1/\delta_n)_{n\in\mathbb{N}}$. In the asymptotic cone $\text{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\lambda})$, the limits of $x_n y_n$ and $x_n z_n$ are at Hausdorff distance 1.

The triangle inequalities imply that the limits

$$\omega\text{-lim}\, \frac{\text{dist}(y_n, a_n)}{\delta_n} \text{ and } \omega\text{-lim}\, \frac{\text{dist}(z_n, a_n)}{\delta_n}$$

are either both finite or both infinite. It follows that the limits of $x_n y_n$ and $x_n z_n$ are either two distinct geodesics joining the points $x_\omega = (x_n)$ and the point $y_\omega = (y_n) = z_\omega = (z_n)$, or two distinct asymptotic rays with common origin, or two distinct geodesics asymptotic on both sides. As in the proof of Theorem 9.40, all these cases are impossible in a real tree.

(b) Suppose that the condition **(b)** holds. Define an infinite subset

$$\mathcal{S} = \{\lfloor \delta_n \rfloor \,:\, n \in \mathbb{N}\} \subset \mathbb{N}.$$

By Exercise 7.17, there exists a nonprincipal ultrafilter ω on $\mathbb{N}$ such that $\omega(\mathcal{S}) = 1$. We define sequences (x'_m), (y'_m), (z'_m) and (a'_m) in X, as follows. For every $m \in \mathcal{S}$ we choose an $n \in \mathbb{N}$ with $\lfloor \delta_n \rfloor = m$ and set

$$(x'_m, y'_m, z'_m, a'_m) = (x_n, y_n, z_n, a_n).$$

For m not in $\mathcal{S}$ we make an arbitrary choice of the quadruple (x'_m, y'_m, z'_m, a'_m). Lastly, define the scaling sequence $\boldsymbol{\lambda} = (m^{-1})$.

We now repeat the arguments in the part (a) of the proof for the asymptotic cone $\text{Cone}_\omega(X, \boldsymbol{a}', \boldsymbol{\lambda})$ and the limit geodesics $\omega\text{-lim}\, x'_m y'_m$ and $\omega\text{-lim}\, x'_m z'_m$. $\square$

Proof of Proposition 9.158. Suppose that the geodesic space X is not hyperbolic. For every geodesic triangle Δ in X and a point $a \in \Delta$ we define the quantity $d_\Delta(a)$, which is the minimal distance from a to the union of the two opposite sides of Δ. Since X is assumed to be non-hyperbolic, for every $n \in \mathbb{N}$ there exists a geodesic triangle

$$\Delta_n = \Delta(x_n,\, y_n,\, z_n)$$

(with the sides $x_n y_n, y_n z_n, z_n x_n$) and points $a_n \in x_n y_n, b_n \in y_n z_n$, such that

$$d_n := d_{\Delta_n}(a_n) = \text{dist}(a_n, b_n) \geqslant n.$$

Here we choose $a_n \in x_n y_n$ to maximize the function d_{Δ_n}. We also pick a point $c_n \in x_n z_n$ which realizes the distance

$$\delta_n := \text{dist}(a_n, x_n z_n) \geqslant d_n.$$

Figure 9.14. Fat triangles.

(a) Suppose that the condition **(a)** is satisfied. We use the sequence of base-points $\boldsymbol{a} = (a_n)$ and scaling factors $\boldsymbol{\lambda} = (1/d_n)$ to define the asymptotic cone

$$\mathbf{K} = \text{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\lambda})\,.$$

We next analyze the ultralimit of the sequence of geodesic triangles Δ_n.

There are two cases to consider:

A) $\omega\text{-lim}\, \frac{\delta_n}{d_n} < +\infty$.

By Lemma 9.159, we have that

$$\text{dist}_{Haus}(a_n x_n, c_n x_n) \leqslant M \cdot \delta_n.$$

Therefore the limits of $a_n x_n$ and $c_n x_n$ are either two geodesic segments with a common endpoint or two asymptotic rays. The same is true of the pairs of segments $a_n y_n$, $b_n y_n$ and $b_n z_n, c_n z_n$, respectively. It follows that the limit $\omega\text{-lim}\, \Delta_n$ is a geodesic triangle Δ with vertices $x, y, z \in \mathbf{K} \cup \partial_\infty \mathbf{K}$. The point $a = \omega\text{-lim}\, a_n \in xy$ is such that $\text{dist}(a, xz \cup yz) \geqslant 1$, which implies that Δ is not a tripod. This contradicts the fact that $\mathbf{K}$ is a real tree.

B) $\omega\text{-lim}\, \frac{\delta_n}{d_n} = +\infty$.

This also implies that

$$\omega\text{-lim}\, \frac{\text{dist}(a_n, x_n)}{d_n} = +\infty \text{ and } \omega\text{-lim}\, \frac{\text{dist}(a_n, z_n)}{d_n} = +\infty.$$

By Lemma 9.159, we have

$$\text{dist}_{Haus}(a_n y_n, b_n y_n) \leqslant M \cdot d_n.$$

Thus, the respective limits of the sequences of segments $x_n y_n$ and $y_n z_n$ are either two rays with the common origin origin $y = \omega\text{-lim}\, y_n$ or two complete geodesics asymptotic in one direction. We denote them xy and yz, respectively, with $y \in$

$\mathbf{K} \cup \partial_\infty \mathbf{K}, x, z \in \partial_\infty \mathbf{K}$. The limit of x_n, z_n in this case is empty (it is "out of sight").

The choice of a_n implies that any point of $b_n z_n$ must be at a the distance at most d_n from $x_n y_n \cup x_n z_n$. Therefore, all points on the ray bz are at the distance at most 1 from xy. It follows that xy and yz are either asymptotic rays emanating from y or complete geodesics asymptotic in both directions and at the Hausdorff distance 1. We again obtain a contradiction with the fact that $\mathbf{K}$ is a real tree.

We conclude that the condition in **(a)** implies that X is δ-hyperbolic, for some $\delta > 0$.

Suppose the condition **(b)** holds. Define $\mathcal{S} = \{\lfloor d_n \rfloor \ : \ n \in \mathbb{N}\}$, and let ω be a non-principal ultrafilter on $\mathbb{N}$ such that $\omega(\mathcal{S}) = 1$ (see Exercise 7.17). We consider a sequence (Δ'_m) of geodesic triangles and a sequence (a'_m) of points on these triangles with the property that whenever $m \in \mathcal{S}$, $\Delta'_m = \Delta_n$ and $a'_m = a_n$, for some n such that $\lfloor d_n \rfloor = m$.

In the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{a}', (m^{-1}))$, with $\boldsymbol{a}' = (a'_m)$ we consider the limit of the sequence of triangles (Δ'_m). We then argue as in the case when the condition **(a)** holds and, similarly, obtain a contradiction with the fact that the cone is a real tree. It follows that the condition **(b)** also implies hyperbolicity of X. □

REMARK 9.160. An immediate consequence of Proposition 9.158 is an alternative proof of quasiisometric invariance of Rips-hyperbolicity among geodesic metric spaces: A quasiisometry between two spaces induces a bi-Lipschitz map between asymptotic cones, and a metric space bi-Lipschitz equivalent to a real tree is a real tree.

As a special case, consider Proposition 9.158 in the context of hyperbolic groups: A finitely-generated group G is hyperbolic if and only if every asymptotic cone of G is a real tree. A finitely-generated group G is called *lacunary-hyperbolic* if at least one asymptotic cone of G is a tree. Theory of such groups is developed in [**OOS09**], where many examples of non-hyperbolic lacunary hyperbolic groups are constructed. Thus, having one tree as an asymptotic cone is not enough to guarantee hyperbolicity of a finitely-generated group. On the other hand:

THEOREM 9.161 (M. Kapovich, B. Kleiner [**OOS09**]). *Suppose that G is a finitely–presented group. Then G is hyperbolic if and only if one asymptotic cone of G is a tree.*

PROOF. Below we present a proof of this theorem which we owe to Thomas Delzant. We will need the following "local-to-global" characterization of hyperbolic spaces, which is a variation on Gromov's "local-to-global" criterion established in [**Gro87**]:

THEOREM 9.162 (B. Bowditch, [**Bow91a**], Theorem 8.1.2). *For every δ there exists δ', such that for every m there exists R for which the following holds. If Y be an m-locally simply-connected R-locally δ-hyperbolic geodesic metric space, then Y is δ'-hyperbolic.*

Here, a space Y is *R-locally δ-hyperbolic* if every R-ball in Y with the path-metric induced from Y is δ-hyperbolic. Instead of defining m-locally simply-connected spaces, we note that every simply-connected simplicial complex equipped with the standard metric, satisfies this condition for every $m > 0$. We refer to [**Bow91a**, Section 8.1] for the precise definition. We will be applying this theorem in the case when $\delta = 1$, $m = 1$ and let δ' and R denote the resulting constants.

We now proceed with the proof suggested to us by Thomas Delzant. Suppose that G is a finitely-presented group, one of whose asymptotic cones is a real tree. Let X be a simply-connected simplicial complex on which G acts freely, simplicially and cocompactly. We equip X with the standard path-metric dist. Then (X, dist) is quasiisometric to G. Suppose that ω is an ultrafilter on $\mathbb{N}$, (λ_n) is a scaling sequence converging to zero, and X_ω is the asymptotic cone of X with respect to this sequence, such that X_ω is isometric to a tree. Consider the sequence of metric spaces $X_n = (X, \lambda_n \text{dist})$. Then, since X_ω is a tree, by taking a diagonal sequence, there exists a pair of sequences r_n, δ_n with

$$\omega\text{-}\lim r_n = \infty, \quad \omega\text{-}\lim \delta_n = 0,$$

such that for ω-all n, every r_n-ball in X_n is δ_n-hyperbolic. In particular, for for ω-all n, every R-ball in X_n is 1-hyperbolic. Therefore, by Theorem 9.162, the space X_n is δ'-hyperbolic for ω-all n. Since X_n is a rescaled copy of X, it follows that X (and, hence, G) is hyperbolic as well. $\square$

We now continue discussion of properties of trees which appear as asymptotic cones of hyperbolic spaces.

THEOREM 9.163. *Let X be a geodesic hyperbolic space which admits a geometric action of a group G. Then all the asymptotic cones of X are real trees where every point is a branch-point with valence equal the cardinality of $\partial_\infty X$.*

PROOF. STEP 1. By Theorem 5.35, the group G is finitely generated and hyperbolic and every Cayley graph Γ of G is quasiisometric to X. It follows that there exists a bi-Lipschitz bijection between the asymptotic cones

$$\Phi : \text{Cone}_\omega(G, \boldsymbol{e}, \boldsymbol{\lambda}) \to \text{Cone}_\omega(X, \boldsymbol{x}, \boldsymbol{\lambda}),$$

where x is a base-point in X, and $\boldsymbol{e}, \boldsymbol{x}$ denote the constant sequences equal to $e \in G$ (the neutral element in G), and respectively to $x \in X$. Moreover, $\Phi(\boldsymbol{e}_\omega) = \boldsymbol{x}_\omega$. The map Φ thus determines a bijection between the space of directions $\Sigma_{\boldsymbol{e}_\omega}$ in the cone of Γ and the space of directions $\Sigma_{\boldsymbol{x}_\omega}$ in the cone of X. It suffices, therefore, to compute the cardinality of $\Sigma_{\boldsymbol{e}_\omega}$. For simplicity, in what follows, we denote the asymptotic cone $\text{Cone}_\omega(G, \boldsymbol{e}, \boldsymbol{\lambda})$ by G_ω.

STEP 2. We now construct an injective map from $\partial_\infty G$ to the space of directions at $\boldsymbol{e}_\omega$ in the asymptotic cone G_ω. Each point $\xi \in \partial_\infty G$ determines a collection of rays $e\xi$ in G within distance $\leqslant 2\delta$ from each other. The ultralimits of all these rays determine the same geodesic ray in G_ω. Taking the direction of this ray at the origin, we obtain a map

$$Log : \partial_\infty G \to \Sigma_{\boldsymbol{e}_\omega}.$$

We need to verify injectivity of this map. To this end, consider two geodesic rays $\rho_i : [0, \infty) \to \Gamma$, $\rho_i(0) = 1 \in G$, asymptotic to distinct points $\xi_i \in \partial_\infty G$, $i = 1, 2$. The ultralimits ρ_i^ω of these geodesic rays are geodesic rays in G_ω emanating from the point $\boldsymbol{e}_\omega$. Proving that $Log(\xi_1) \neq Log(\xi_2)$ amounts to showing that for all $s > 0, t > 0$,

$$\text{dist}(\rho_1^\omega(s), \rho_2^\omega(t)) = s + t.$$

Since $\xi_1 \neq \xi_2$, for all positive values of s and t, the sequence of Gromov-products

$$\left(\rho_1\left(\frac{s}{\lambda_n}\right), \rho_2\left(\frac{t}{\lambda_n}\right)\right)_e$$

ω-converges to $(\xi_1, \xi_2)_e \in \mathbb{R}$. Therefore,

$$\operatorname{dist}(\rho_1^\omega(s), \rho_2^\omega(t)) = \omega\text{-lim}\, \lambda_n \operatorname{dist}\left(\rho_1\left(\frac{t}{\lambda_n}\right), \rho_2\left(\frac{s}{\lambda_n}\right)\right) =$$

$$\omega\text{-lim}\left[t + s - 2\lambda_n\left(\rho_1\left(\frac{t}{\lambda_n}\right), \rho_2\left(\frac{s}{\lambda_n}\right)\right)_e\right] = t + s.$$

Thus, $Log(\xi_1) \neq Log(\xi_2)$.

STEP 3. We argue that every direction of Γ_ω at $\boldsymbol{e}_\omega$ is determined by a sequence of geodesic rays emanating from e in Γ. The argument below was suggested to us by Panos Papasoglu.

Elements of $\Sigma_{\boldsymbol{e}_\omega}$ are represented by nondegenerate geodesic segments $\boldsymbol{e}_\omega g_\omega$, where $g_\omega \in G_\omega$ is represented by a sequence (g_n) in G with $|g_n| \asymp \lambda_n^{-1}$ as $n \to \infty$.

We will need:

LEMMA 9.164 (Geodesic segments are uniformly close to geodesic rays). *Let Γ be a δ-hyperbolic, in the sense of Rips, Cayley graph of a group G. Then there exists a constant M such that each geodesic segment $s \subset \Gamma$ is contained in the M-neighborhood of a geodesic ray $v\xi$ in Γ.*

PROOF. We will consider the case when the group G is infinite, otherwise, there is nothing to prove.

Without loss of generality, we may assume that δ is a natural number. Furthermore, taking into account the isometric G-action on Γ, we may assume that the geodesic segment s is represented by an edge-path in Γ starting at the vertex $1 \in G = V(\Gamma)$. Let X denote the generating set of G used to define the Cayley graph Γ. Then vertex-paths in Γ starting at 1, can be described as finite or semi-infinite words in the alphabet $X \cup X^{-1}$. By abusing the terminology, we will use the same notation for paths in Γ as for the corresponding words. We will refer to a word as *geodesic* if it represents a geodesic path in Γ. Consider the set T of words in this alphabet, which define geodesics of length $k = 6\delta$ in Γ. Then there exists $R = R(k)$ such that each finite geodesic word p of length $\geqslant R$ contains at least two *disjoint subwords* equal to w for some $w \in T$, i.e., p has the form

$$w_0 w w_1 w w_2$$

where w_i's are subwords of p. Given such partition of a finite geodesic path p, we define an infinite path q:

$$w_0 w w_1 w w_1 w w_1 w \ldots$$

alternating the subwords w and w_1 infinitely many times. As w has length k, the path q is k-local geodesic, since each length k subword u in q appears as a subword in p, and p is geodesic.

Consider now a finite geodesic word p of length $\geqslant R$ and break p as the product of subwords:

$$p = p_1 p_2,$$

where p_2 has length R. Then, as above, partition p_2 as the product $w_0ww_1ww_2$ and define an infinite word q using this partition. Lastly, take the infinite word

$$q' = p_1 q.$$

Since w_2 has length $\leqslant R$, the path p is contained in the R-neighborhood of q'. By the construction, q' is a k-local geodesic. Taking into account Theorem 9.45, we conclude that q' is an $(3, 4\delta)$-quasigeodesic ray in Γ. By the Extended Morse Lemma, q' is $D = D(3, 4\delta)$-Hausdorff close to a geodesic ray 1ξ in Γ. Therefore, for $M = R + D$, the original path p is contained in the M-neighborhood of 1ξ. □

We conclude (using this Lemma) that every direction of Γ_ω in $\boldsymbol{e}_\omega$ is the germ of a limit ray. We then have a surjective map from the set of sequences in $\partial_\infty G$ to $\Sigma_{[\boldsymbol{e}_\omega]}$:

$$\{(\xi_n)_{n\in\mathbb{N}} \ : \ \xi_n \in \partial_\infty\Gamma\} = (\partial_\infty\Gamma)^{\mathbb{N}} \to \Sigma_{[\boldsymbol{e}_\omega]}.$$

Steps 2 and 3 imply that for a non-elementary hyperbolic group, the cardinality of $\Sigma_{[\boldsymbol{e}_\omega]}$ is the same as of $\partial_\infty G$, i.e., continuum. If G is an elementary hyperbolic group, then its asymptotic cone is a line and, theorem holds in this case as well. □

A. Dyubina–Erschler and I. Polterovich ([**DP01**], [**DP98**]) proved a stronger result than Theorem 9.163:

THEOREM 9.165 ([**DP01**], [**DP98**]). *Let $\mathcal{A}$ be the $2^{\aleph_0}$–universal tree, as defined in Theorem 9.19. Then:*

(a) *Every asymptotic cone of a non-elementary hyperbolic group is isometric to $\mathcal{A}$.*
(b) *Every asymptotic cone of a complete, simply connected Riemannian manifold with strictly sectional curvature (i.e, curvature $\leqslant -\kappa < 0$), is isometric to $\mathcal{A}$.*

A consequence of Theorem 9.165 is that asymptotic cones of non-elementary hyperbolic groups and of complete, simply connected Riemannian manifolds of strictly negative sectional curvature cannot be distinguished from each other.

9.21. Size of loops

In this section we show that the characterization of hyperbolicity using asymptotic cones allows one to define hyperbolicity of a space in terms of *size of its loops.*Throughout this section, X denotes a geodesic metric space.

9.21.1. The minsize. One quantity that measures the size of geodesic triangles is the *minimal size* introduced in Definition 5.107 for topological triangles, which, of course, apply to geodesic triangles in X. This leads to the following definition:

DEFINITION 9.166. The *m*inimal size function (the *minsize function*),

$$\mathrm{minsize} = \mathrm{minsize}_X : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\},$$

$isdefinedby \mathrm{minsize}(\ell) = \sup\{\mathrm{minsize}(\Delta) \ : \ \Delta \text{ a geodesic triangle of perimeter } \leqslant \ell\}$.

Note that according to (9.1), for each δ–hyperbolic (in the sense of Rips) metric space X, the function minsize is bounded above by 2δ. We will see below that the "converse" is also true, i.e., when the function minsize is bounded, the space X is hyperbolic. Moreover, M. Gromov proved [**Gro87**, §6] that sublinear growth of minsize is enough to conclude that a space is hyperbolic. With the characterization of hyperbolicity using asymptotic cones, the proof of this result is straightforward:

PROPOSITION 9.167. *A geodesic metric space X is hyperbolic if and only if* $\text{minsize}(\ell) = o(\ell)$.

PROOF. As noted above, one implication immediately follows from Lemma 9.57. Conversely, assume that $\text{minsize}(\ell) = o(\ell)$. We begin by proving that in each asymptotic cone of X, every finite geodesic is a limit geodesic, in the sense of Definition 7.50. More precisely:

LEMMA 9.168. *Let $\gamma = a_\omega b_\omega$ be a geodesic segment in the asymptotic cone $X_\omega = \text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, where the points a_ω, b_ω are represented by the sequences $(a_i), (b_i)$ respectively. Then for every geodesic $a_i b_i \subset X$ connecting a_i to b_i,*

$$\omega\text{-}lim\, a_i b_i = \gamma.$$

PROOF. Let $c_\omega = (c_i)$ be a point on γ. Consider an arbitrary geodesic triangle $\Delta_i \subset X$ with vertices a_i, b_c, c_i and the perimeter ℓ_i. Since

$$2\ d(a_\omega, b_\omega) = \omega\text{-lim}\, \lambda_i \ell_i < \infty$$

and $\text{minsize}(\Delta_i) = o(\ell_i)$, we get

$$\omega\text{-lim}\, \lambda_i\, \text{minsize}(\Delta_i) = 0.$$

Taking the points x_i, y_i, z_i on the sides of Δ_i realizing the minsize of Δ_i, we conclude:

$$\omega\text{-lim}\, \lambda_i\, \text{diam}(x_i, y_i, z_i) = 0.$$

In particular, the sequences $(x_i), (y_i), (z_i)$ represent the same point $x_\omega \in X_\omega$. Then

$$\text{dist}(a_\omega, b_\omega) \leqslant \text{dist}(a_\omega, x_\omega) + \text{dist}(x_\omega, b_\omega) \leqslant$$
$$\text{dist}(a_\omega, x_\omega) + \text{dist}(x_\omega, b_\omega) + 2\text{dist}(x_\omega, c_\omega) = \text{dist}(a_\omega, c_\omega) + \text{dist}(c_\omega, b_\omega)\,.$$

The first and the last term in the above sequence of inequalities are equal, hence all inequalities become equalities, in particular $c_\omega = x_\omega$. Thus c_ω belongs to the ultralimit $\omega\text{-lim}\, a_i b_i$ and lemma follows. □

If one asymptotic cone $X_\omega = \text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is not a real tree, then it contains a geodesic triangle Δ_ω which is not a tripod. Without loss of generality we may assume that the geodesic triangle is a simple loop in X_ω. By the above lemma, the geodesic triangle Δ_ω is the ultralimit of a sequence of geodesic triangles (Δ_i), with perimeters of the order $O\left(\frac{1}{\lambda_i}\right)$. The fact that $\text{minsize}(\Delta_i) = o\left(\frac{1}{\lambda_i}\right)$ implies that the three edges of Δ have a common point, a contradiction. □

We note that Gromov in [**Gro87**, Proposition 6.6.F] proved a stronger version of Proposition 9.167:

THEOREM 9.169. *There exists a universal constant $\varepsilon_0 > 0$ such that if in a geodesic metric space X all geodesic triangles with length $\geqslant L_0$, for some L_0, have*

$$\text{minsize}(\Delta) \leqslant \varepsilon_0 \cdot \text{perimeter}(\Delta)\,,$$

then X is hyperbolic.

9.21.2. The constriction. Another way of measuring the size of loops in a space X is through their *constriction* function. We define the constriction function only for simple loops in X primarily for the notational convenience, the definition and the results generalize without difficulty if one considers non-simple loops.

We fix a constant $\lambda \in \left(0, \frac{1}{2}\right)$. For a Lipschitz loop $c : \mathbb{S}^1 \to X$ of length ℓ, we define the λ–*constriction of the loop* c as $\mathrm{constr}_\lambda(c)$, which is the infimum of $d(x, y)$, where the infimum is taken over all all points x, y separating $c(\mathbb{S}^1)$ into two arcs of length at least $\lambda\ell$. Thus, the higher constriction means less distortion of c in X. The λ–*constriction function*, $\mathrm{constr}_\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, of a metric space X is defined as

$$\mathrm{constr}_\lambda(\ell) = \sup\{\mathrm{constr}_\lambda(c) \ : \ c \text{ is a Lipschitz simple loop in } X \text{ of length } \leqslant \ell\} .$$

Note that when $\lambda \leqslant \mu$, $\mathrm{constr}_\lambda \leqslant \mathrm{constr}_\mu$, and $\mathrm{constr}_\lambda(\ell) \leqslant \ell$.

Proposition 9.170 ([**Dru01**], Proposition 3.5). *For geodesic metric spaces X the following are equivalent:*

(1) *X is δ–hyperbolic in the sense of Rips, for some $\delta > 0$;*

(2) *there exists $\lambda \in \left(0, \frac{1}{4}\right]$ such that* $\mathrm{constr}_\lambda(\ell) = o(\ell)$ *;*

(3) *for all $\lambda \in \left(0, \frac{1}{4}\right]$ and $\ell > 1$,*

$$\mathrm{constr}_\lambda(\ell) \leqslant 4\delta\left(\log_2(\ell + 12\delta) + 3\right) + 2 .$$

Remark 9.171. One cannot obtain a better order than $O(\log \ell)$ for the constriction function in hyperbolic spaces. This can be seen by considering, metric circles of length ℓ lying on a horosphere in $\mathbb{H}^3$.

Proof. Our main tool, as before, are asymptotic cones of X.

We begin by arguing that (2) implies (1). In what follows we define *limit triangles* in an asymptotic cone $X_\omega = \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, to be the triangles in X_ω whose edges are limit geodesics. Note that such triangles *a priori* need not be themselves limits of sequences of geodesic triangles in X: They are merely limits of sequences of geodesic hexagons.

First note that (2) implies that every limit triangle in every asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is a tripod. Indeed, if one assumes that one limit triangle is not a tripod, without loss of generality one can assume that this triangle forms a simple loop in X_ω. This triangle is the limit of a sequence of geodesic hexagons (H_i), with three edges of lengths of the order $O\left(\frac{1}{\lambda_i}\right)$, alternating with three edges of lengths of the order $o\left(\frac{1}{\lambda_i}\right)$. (We leave it to the reader to verify that such hexagons may be chosen to be simple.) Since $\mathrm{constr}_\lambda(H_i) = o\left(\frac{1}{\lambda_i}\right)$ we obtain that $\omega\text{-}\lim H_i$ is not simple, a contradiction.

It remains to show that every geodesic segment in every asymptotic cone of X is a limit geodesic. The proof is similar to that of Lemma 9.168.

Let $\gamma = a_\omega b_\omega$ be a geodesic in a cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, where $a_\omega = (a_i)$ and $b_\omega = (b_i)$. We let $c_\omega = (c_i)$ be an arbitrary point in γ. We already know that every limit geodesic triangle $\Delta(a_\omega, b_\omega, c_\omega) \subset X_\omega$ is a tripod. If c_ω does not coincide with the center of this tripod, then

$$\mathrm{dist}(a_\omega, c_\omega) + \mathrm{dist}(c_\omega, b_\omega) > \mathrm{dist}(a_\omega, b_\omega),$$

a contradiction. Thus, c_ω belongs to $\omega\text{-}\lim a_i b_i$ and, hence, $\gamma = \omega\text{-}\lim a_i b_i$.

We thus proved that every geodesic triangle in every asymptotic cone of X is a tripod, hence every asymptotic cone is a real tree. It follows that X is hyperbolic.

Clearly, (3) implies (2). We will prove that (1) implies (3). By monotonicity of the constriction function (as a function of λ), it suffices to prove (3) for $\lambda = \frac{1}{4}$. We denote $\mathrm{constr}_{\frac{1}{4}}(\mathfrak{c})$ simply by constr.

Consider an arbitrary closed Lipschitz curve $\mathfrak{c} : \mathbb{S}^1 \to X$ of length ℓ. We orient the circle and will use the notation α_{pq} to denote the oriented arc of the image of $\mathfrak{c}$ connecting p to q. Let x, y, z be three points on $\mathfrak{c}(\mathbb{S}^1)$ connected by the arcs $\alpha_{xy}, \alpha_{yz}, \alpha_{zx}$ in $\mathfrak{c}(\mathbb{S}^1)$, such that the first two arcs have length $\frac{\ell}{4}$. Let $t \in \alpha_{zx}$ be the point minimizing the distance to y in X. Clearly,

$$R := \mathrm{dist}(y, t) \geqslant \mathrm{constr}$$

and for each point $s \in \alpha_{zx}$, $d(s, y) \geqslant R$. The point t splits the arc α_{zx} into two subarcs α_{zt}, α_{tx}. Without loss of generality, we can assume that the length of α_{tx} is $\geqslant \frac{\ell}{4}$. Let $\alpha_{xx'}$ be the maximal subarc of α_{xy} disjoint from $B(y, r)$ (we allow $x = x'$). We set

$$r := \frac{d(x', t)}{2}.$$

Since α_{tx} has length $\geqslant \frac{\ell}{4}$, we obtain

$$2r \geqslant \mathrm{constr}\,.$$

FIGURE 9.15. Constriction.

The arc $\alpha_{tx'}$ connects the points t, x' of the metric sphere $S(y, R)$ outside of the open ball $B(y, R)$. Therefore, according to Lemma 9.65,

$$\ell \geqslant \ell(\alpha_{tx'}) \geqslant 2^{\frac{r-1}{2\delta} - 3} - 12\delta$$

and, thus,

$$\mathrm{constr} \leqslant 2r \leqslant 4\delta\left(\log_2(\ell + 12\delta) + 3\right) + 2.$$

The inequality in (3) follows. ☐

9.22. Filling invariants of hyperbolic spaces

Recall that for every μ-simply connected geodesic metric space X we defined (see §5.10) the *filling area function* (or, *isoperimetric function*) $Ar(\ell) = Ar_{\mu,X}(\ell)$, which computes the least upper bound on the areas of disks bounding loops of lengths $\leqslant \ell$ in X. We also defined the *filling radius function* $r_{\mu,X}(\ell)$ which computes the least upper bounds on radii of such disks. The goal of this section is to relate both invariants to hyperbolicity of the space X.

9.22.1. Filling area. Recall also that hyperbolicity of X implies linearity of $Ar(\ell)$, see Corollary 9.140. In this section we will prove the converse. Moreover, we will prove that there is a gap between the quadratic filling order and the linear isoperimetric order. Namely, as soon as the isoperimetric inequality is less than quadratic, it has to be linear and the space has to be hyperbolic:

THEOREM 9.172 (Subquadratic filling, §2.3, §6.8, [**Gro87**]). *If a coarsely simply-connected geodesic metric space X has the isoperimetric function $Ar(\ell) = o(\ell^2)$, then X is hyperbolic.*

Note that there is a second gap for the possible filling orders of groups.

REMARK 9.173 ([**Ol′91b**], [**Bat99**]). If a group G admits a finite presentation which has the Dehn function $Dehn(\ell) = o(\ell)$, then G is either free or finite.

Proofs of Theorem 9.172 can be found in [**Ol′91b**], [**Pap95b**], [**Bow95**] and [**Dru01**]. B. Bowditch makes use of only two properties of the area function in his proof: The *quadrangle* (or *Besikovitch*) inequality (see Proposition 5.109) and a certain *theta–property*. In fact, as we will see below, only the quadrangle inequality or its triangle counterpart, *the minsize inequality* (see Proposition5.110) are needed. Also, we will see that it suffices to have subquadratic isoperimetric function for geodesic triangles.

Proof of Theorem 9.172. Let X be a μ-simply-connected geodesic metric space with the isoperimetric function Ar_X and the minsize function $\mathrm{minsize}_X : \mathbb{R}_+ \to \mathbb{R}_+$, see Definition 9.166. According to Proposition 5.110, for every $\delta \geqslant \mu$,

$$[\mathrm{minsize}_X(\ell)]^2 \leqslant 2\sqrt{3}\mu^2 Ar_X(\ell)\,,$$

whence $Ar_X(\ell) = o(\ell^2)$ implies that $\mathrm{minsize}_X(\ell) = o(\ell)$. Using Proposition 9.167, we conclude that X is hyperbolic. □

The strongest known version of the converse to Corollary 9.140 is:

THEOREM 9.174 (Strong subquadratic filling theorem,see §2.3, §6.8 of [**Gro87**], and also [**Ol′91b**], [**Pap96**]). *Let X be a μ-simply connected geodesic metric space. If there exist sufficiently large N and L, and $\epsilon > 0$ sufficiently small, such that every loop $\mathfrak{c}$ in X with $N \leqslant \mathrm{A}r_\delta(\mathfrak{c}) \leqslant LN$ satisfies*

$$\mathrm{A}r_\delta(\mathfrak{c}) \leqslant \epsilon[\mathrm{length}(\mathfrak{c})]^2\,,$$

then the space X is hyperbolic.

It seems impossible to prove this theorem using asymptotic cones.

In Theorem 9.174 it suffices to consider only geodesic triangles Δ instead of all loops, and to replace the condition $N \leqslant Ar_\delta(\Delta) \leqslant LN$ by length $(\Delta) \geqslant N$. This follows immediately from Theorem 9.169 and the minsize inequality in Proposition 5.110.

M. Coornaert, T. Delzant and A. Papadopoulos have shown that if X is a complete simply connected Riemannian manifold which is *reasonable* (see [**CDP90**, Chapter 6, §1] for a definition of this notion; for instance if X admits a geometric group action, then X is reasonable) then the constant ϵ in the previous theorem only has to be smaller than $\frac{1}{16\pi}$, see [**CDP90**, Chapter 6, Theorem 2.1].

In terms of the multiplicative constant, a sharp inequality was proved by S. Wenger.

THEOREM 9.175 (S. Wenger [**Wen08**]). *Let X be a geodesic metric space. Assume that there exists $\varepsilon > 0$ and $\ell_0 > 0$ such that every Lipschitz loop $\mathfrak{c}$ of length* $\mathrm{length}(\mathfrak{c})$ *at least ℓ_0 in X bounds a Lipschitz disk $\mathfrak{d} : \mathbb{D}^2 \to X$ with*

$$Area(\mathfrak{d}) \leqslant \frac{1-\varepsilon}{4\pi}\,\mathrm{length}(\mathfrak{c})^2\,.$$

Then X is hyperbolic.

In the Euclidean space one has the classical isoperimetric inequality

$$Area(\mathfrak{d}) \leqslant \frac{1}{4\pi} \operatorname{length}(\mathfrak{c})^2 ,$$

with equality if and only if $\mathfrak{c}$ is a circle and $\mathfrak{d}$ a planar disk.

Note that the quantity $Area(\mathfrak{d})$ appearing in Theorem 9.175 is a generalization of the notion of the geometric area used in this book. If the Lipschitz map $\phi : \mathbb{D}^2 \to X$ is injective almost everywhere, then $Area(\phi)$ is the 2-dimensional Hausdorff measure of its image. In the case of a Lipschitz map to a Riemannian manifold, $Area(\phi)$ is the *area of a map* defined in §2.4. When the target is a general geodesic metric space, $Area(\phi)$ is obtained by suitably interpreting the *Jacobian* $J_x(\phi)$ in the integral formula

$$Area(\phi) = \int_{\mathbb{D}^2} |J_x\phi(x)|.$$

9.22.2. Filling radius. Another application of the results of §9.21 is a characterization of hyperbolic spaces in terms of their filling radii.

PROPOSITION 9.176 ([**Gro87**], §6, [**Dru01**], §3). *For a geodesic μ-simply connected metric space X the following statements are equivalent:*

(1) *X is hyperbolic;*

(2) *the filling radius of X has sublinear growth: $r_X(\ell) = o(\ell)$;*

(3) *the filling radius is X grows at most logarithmically: $r_X(\ell) = O(\log \ell)$.*

PROOF. In what follows, we let $\mathrm{Ar} = \mathrm{Ar}_\mu$ denote the μ-filling area function in the sense of §5.10, defined for loops in the space X.

We first prove that (1) $\Rightarrow$ (3). According to the linear isoperimetric inequality for hyperbolic spaces (see Corollary 9.140), there exists a constant K depending only on X such that

$$\mathrm{Ar}(\mathfrak{c}) \leqslant K\ell_X(\mathfrak{c}). \tag{9.19}$$

Here $\mathrm{Ar}(\mathfrak{c})$ is the μ-area of a least-area μ-disk $\mathfrak{d} : \mathcal{D}^{(0)} \to X$ bounding $\mathfrak{c}$. Recall also that the *combinatorial length* and *area* of a simplicial complex is the number of 1-simplices and 2-simplices respectively in this complex. Thus, for a loop $\mathfrak{c}$ as above, we have

$$\ell_X(\mathfrak{c}) \leqslant \mu \ \operatorname{length}(\mathcal{C}),$$

where $\mathcal{C}$ is the triangulation of the circle $\mathbb{S}^1$ so that vertices of any edge are mapped by $\mathfrak{c}$ to points within distance $\leqslant \mu$ in X.

Consider now a loop $\mathfrak{c} : \mathbb{S}^1 \to X$ of metric length ℓ and a least area μ-disk $\mathfrak{d} : \mathcal{D}^{(0)} \to X$ filling $\mathfrak{c}$; thus, $\mathrm{Ar}(\mathfrak{c}) \leqslant K\ell$.

Let $v \in \mathcal{D}^{(0)}$ be a vertex such that its image $a = \mathfrak{d}(v)$ is at maximal distance r from $\mathfrak{c}(\mathbb{S}^1)$. For every $1 \leqslant j \leqslant k$, with

$$k = \lfloor \frac{r}{\mu} \rfloor$$

we define a subcomplex $\mathcal{D}_j$ of $\mathcal{D}$: $\mathcal{D}_j$ is the maximal connected subcomplex in $\mathcal{D}$ containing v, so that every vertex in $\mathcal{D}_j$ could be connected to v by a *gallery* (in the sense of §1.7.1) of 2-dimensional simplices σ in $\mathcal{D}$ such that

$$\mathfrak{d}\left(\sigma^{(0)}\right) \subset \overline{B}(a, j\mu).$$

For instance, $\mathcal{D}_1$ contains the star of v in $\mathcal{D}$. Let Ar_j be the number of 2-simplices in $\mathcal{D}_j$.

For each $j \leqslant k-1$ the geometric realization $\mathcal{D}_j$ of the subcomplex $\mathcal{D}_j$ is homeomorphic to a 2-dimensional disk with several disks removed from the interior. (As usual, we will conflate a simplicial complex and its geometric realization.) Therefore, the boundary $\partial\mathcal{D}_j$ of $\mathcal{D}_j$ in $\mathbb{D}^2$ is a union of several disjoint topological circles, while all the edges of $\mathcal{D}_j$ are interior edges for $\mathcal{D}$. We denote by s_j the outermost circle in $\partial\mathcal{D}_j$, i.e., s_j bounds a triangulated disk $\mathcal{D}'_j \subset \mathcal{D}$, such that $\mathcal{D}_j \subset \mathcal{D}'_j$. Let $\mathrm{length}(\partial\mathcal{D}_j)$ and $\mathrm{length}(s_j)$ denote the number of edges of $\partial\mathcal{D}_j$ and of s_j respectively.

By the definition, every edge of $\mathcal{D}_j$ is an interior edge of $\mathcal{D}_{j+1}$ and belongs to a 2-simplex of $\mathcal{D}_{j+1}$. Note also that if σ is a 2-simplex in $\mathcal{D}$ and two edges of σ belong to $\mathcal{D}_j$, then σ belongs to $\mathcal{D}_j$ as well. Therefore,

$$\mathrm{Ar}_{j+1} \geqslant \mathrm{Ar}_j + \frac{1}{3}\mathrm{length}(\partial\mathcal{D}_j) \geqslant \mathrm{Ar}_j + \frac{1}{3}\mathrm{length}(s_j).$$

Since $\mathfrak{d}$ is a least area filling disk for $\mathfrak{c}$ it follows that each disk $\mathfrak{d}|_{\mathcal{D}'_j}$ is a least area disk bounding the loop $\mathfrak{d}|_{\int_j}$. In particular, by the isoperimetric inequality in X,

$$\mathrm{Ar}_j = Area(\mathcal{D}_j) \leqslant Area(\mathcal{D}'_j) \leqslant K\ell_X(\mathfrak{d}(s_j)) \leqslant K\mu\ length(s_j)$$

We have thus obtained that

$$\mathrm{Ar}_{j+1} \geqslant \left(1 + \frac{1}{3\mu K}\right)\mathrm{Ar}_j\,.$$

It follows that

$$K\ell \geqslant \mathrm{Ar}(\mathfrak{d}) \geqslant \left(1 + \frac{1}{3\mu K}\right)^k$$

whence,

$$r \leqslant \mu(k+1) \leqslant \mu\left(\frac{\ln\ell + \ln K}{\ln\left(1+\frac{1}{3\mu K}\right)} + 1\right).$$

Clearly (3) $\Rightarrow$ (2). It remains to prove the implication (2) $\Rightarrow$ (1).

We first show that (2) implies that in every asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, all geodesic triangles that are limits of geodesic triangles in X (i.e. $\boldsymbol{\Delta} = \omega\text{-}\lim\Delta_i$) are tripods. We assume that $\boldsymbol{\Delta}$ is not a point. Every geodesic triangle Δ_i can be regarded as a loop $\mathfrak{c}_i : \mathbb{S}^1 \to \Delta_i$, and can be filled by a μ-disk $\mathfrak{d}_i : \mathcal{D}^{(1)} \to X$ of the filling radius $r_i = r(\mathfrak{d}_i) = o\left(\mathrm{length}\left(\Delta_i\right)\right)$. In particular, $\omega\text{-}\lim_i \lambda_i r_i = 0$.

Let x_iy_i, y_iz_i and z_ix_i be the three geodesic edges of Δ_i, and let $\overline{x}_i, \overline{y}_i, \overline{z}_i$ be the three points on $\mathbb{S}^1$ corresponding to the three vertices x_i, y_i, z_i. Consider a path $\overline{\mathfrak{p}}_i$ in the 1-skeleton of $\mathcal{D}$ with endpoints $\overline{y}_i$ and $\overline{z}_i$ such that $\overline{\mathfrak{p}}_i$ together with the arc of $\mathbb{S}^1$ with endpoints $\overline{y}_i, \overline{z}_i$ encloses a maximal number of triangles with $\mathfrak{d}_i$–images in the r_i–neighborhood of y_iz_i. Every edge of $\overline{\mathfrak{p}}_i$ that is not in $\mathbb{S}^1$ is contained in a 2-simplex whose third vertex has $\mathfrak{d}_i$–image in the r_i–neighborhood of $y_ix_i \cup x_iz_i$. The edges in $\overline{\mathfrak{p}}_i$ that are in $\mathbb{S}^1$ are either between $\overline{x}_i, \overline{y}_i$ or between $\overline{x}_i, \overline{z}_i$.

Thus, $\overline{\mathfrak{p}}_i$ has $\mathfrak{d}_i$–image $\mathfrak{p}_i$ in the $(r_i+\mu)$–neighborhood of $y_ix_i \cup x_iz_i$. See Figure 9.16.

Consider an arbitrary vertex $\overline{u}$ on $\mathbb{S}^1$ between $\overline{y}_i, \overline{z}_i$ and its image $u \in y_iz_i$. We have that $\mathfrak{p}_i \subset \overline{\mathcal{N}}_{r_i+\mu}(y_iu) \cup \overline{\mathcal{N}}_{r_i+\mu}(uz_i)$, where y_iu and uz_i are sub-geodesics of y_iz_i.

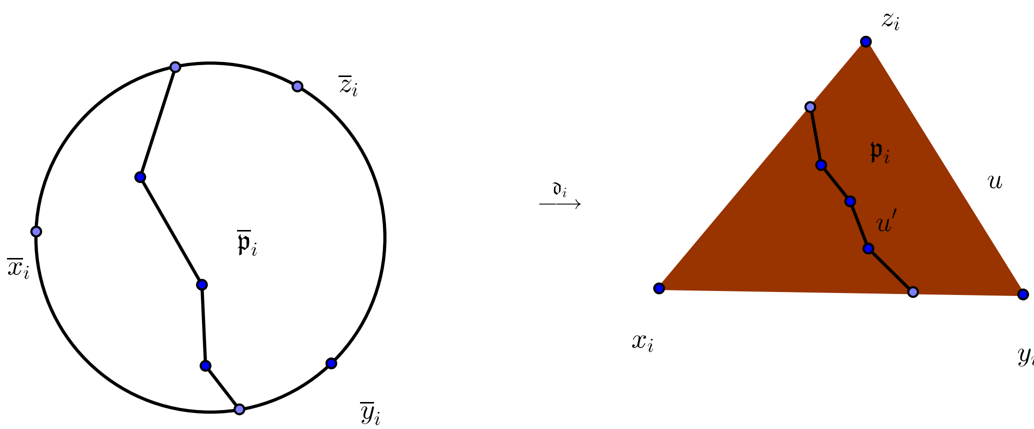


FIGURE 9.16. The path $\overline{\mathfrak{p}}_i$ and its image $\mathfrak{p}_i$.

By connectedness, there exists a point $u' \in \mathfrak{p}_i$ at distance at most $r_i + \mu$ from a point $u_1 \in y_i u$, and from a point $u_2 \in u z_i$. As the three points u_1, u, u_2 are aligned on a geodesic and $\operatorname{dist}(u_1, u_2) \leqslant 2(r_i + \mu)$ it follow that, say, $\operatorname{dist}(u_1, u) \leqslant r_i + \mu$, whence $\operatorname{dist}(u, u') \leqslant 3(r_i + \mu)$. Since the point $\overline{u}$ was arbitrary, we have thus proved that $y_i z_i$ is in $\overline{\mathcal{N}}_{3r_i+3\mu}(\mathfrak{p}_i)$, therefore, it is in $\overline{\mathcal{N}}_{4r_i+4\mu}\left(y_i x_i \cup x_i z_i\right)$. This implies that in $\boldsymbol{\Delta}$ one edge is contained in the union of the other two. The same argument done for each edge implies that $\boldsymbol{\Delta}$ is a tripod.

From this, one can deduce that every triangle in the cone is a tripod. In order to do this it suffices to show that every geodesic in the cone is a limit geodesic. Consider a geodesic in $\operatorname{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ with the endpoints $x_\omega = (x_i)$ and $y_\omega = (y_i)$ and an arbitrary point $z_\omega = (z_i)$ on this geodesic. Geodesic triangles $\Delta_i = \Delta(x_i, y_i, z_i)$ yield a tripod $\Delta_\omega = \Delta(x_\omega, y_\omega, z_\omega)$ in the asymptotic cone, but since,

$$\operatorname{dist}(x_\omega, z_\omega) + \operatorname{dist}(z_\omega, y_\omega) = \operatorname{dist}(x_\omega, y_\omega),$$

it follows that the tripod must be degenerate. Thus $z_\omega \in \omega\text{-lim}\, x_i y_i$. □

REMARK 9.177. 1. One can show that in Part (3) of the proposition, given a loop $\mathfrak{c} : \mathbb{S}^1 \to X$ of length ℓ, a filling disk $\mathfrak{d}$ minimizing the area has the filling radius $r(\mathfrak{d}) = O(\log \ell)$.

2. The logarithmic order in (3) cannot be improved, as shown by the example of the horizontal circle in the upper half-space model of $\mathbb{H}^3$.

3. In view of this proposition (as in the case of the filling area) there is a gap between the linear order of the filling radius and the logarithmical one.

Analogously to the filling area, for the radius too there is a stronger version of the implication *sublinear filling radius* $\Longrightarrow$ *hyperbolicity*, similar to Theorem 9.174.

PROPOSITION 9.178 (M. Gromov; P. Papasoglu [**Pap98**]). *Let Γ be a finitely presented group. If there exists $\ell_0 > 0$ such that*

$$r(\ell) \leqslant \frac{\ell}{73}, \ \forall \ell \geqslant \ell_0 \,,$$

then the group Γ is hyperbolic.

According to [**Pap98**], the best possible constant expected is not $\frac{1}{73}$, but $\frac{1}{8}$. Note that the proof of Proposition 9.178 cannot be extended from groups to metric spaces, because it relies on the bigon criterion for hyperbolicity [**Pap95c**], which only works for groups. There is probably a similar statement for general metric spaces, with a constant that can be made effective for complete simply connected Riemannian manifolds.

9.22.3. Orders of Dehn functions of non-hyperbolic groups and higher Dehn functions. As we saw earlier, there is a gap between linear and quadratic orders for Dehn functions of groups. It is natural to ask what happens to the growth orders. For each $k \in \mathbb{N}$ define the subset $A_k \subset \mathbb{R}$ consisting of numbers α such that n^α is the order of the k-th Dehn function of a group G of type $\mathbf{F}_{k+1}$. Since there are only countably many finitely-presented functions, each set A_k is countable. M. Sapir, J. Birget and E. Rips in [**SBR02**] gave a detailed description of the set A_1, which, in particular, implies that the intervals $(0, 1)$ and $(1, 2)$ are the only gaps in A_1:

THEOREM 9.179 (M. Sapir, J. Birget, E. Rips). *The closure of A_1 contains the half-line $[2, \infty)$. Moreover, let M be a not necessarily deterministic Turing machine with time function $T(n)$ for which $T(n)^4$ is superadditive. Then there exists a finitely presented group $G(M)$ with Dehn function equivalent to $T(n)^4$.*

In particular, this theorem shows that subpolynomial Dehn functions need not be of the type n^α; according to the following theorem, this can happen even for nilpotent groups:

THEOREM 9.180 (S. Wenger, [**Wen11**]). *There exists a nilpotent group G with the order of $Dehn(G)$ strictly larger n^2 and at most $n^2 \log(n)$.*

Dehn functions can be very fast-growing:

THEOREM 9.181 (W. Dison and T. Riley, [**DR13**]). *For each k, there exists a finitely presented group $Hydra_k$ (a "hydra group"), whose Dehn function is equivalent to the k-th Ackermann function,*

$$Hydra_k = \left\langle a_1, \ldots, a_k, p, t | t^{-1} a_i t = a_i a_{i+1}, i > 1, t^{-1} a_1 t = a_1, [p, a_i t] = 1, \forall i \right\rangle$$

Recall that Ackermann functions are defined by the recursive formula:

$$A_1(n) := 2n$$

$$A_{k+1}(n) := A_k^{(n)}(1),$$

where (n) means the n-fold composition.

Surprisingly, there are no gaps in the orders of higher Dehn functions:

THEOREM 9.182 (N. Brady, M. Forester, [**BF10**]). *For each $k \geqslant 2$, the closure of A_k contains the half-line $[1, \infty)$.*

We recall that one can define isoperimetric functions using homological fillings rather than the homotopical fillings (as in the definition of the Dehn functions), see (2.3).

THEOREM 9.183 (A. Abrams, N. Brady, P. Dani and R. Young, [**ABDY13**]). *There are finitely presented groups for which the (classical) Dehn function is not equivalent to the homological Dehn function.*

9.23. Asymptotic cones, actions on trees and isometric actions on hyperbolic spaces

Let G be a finitely-generated group with the generating set $g_1, .., g_m$; let X be a metric space. Given a homomorphism $\rho : G \to \mathrm{Isom}(X)$, we define the following function:

$$d_\rho(x) := \max_k d(\rho(g_k)(x), x) \tag{9.20}$$

and set

$$d_\rho := \inf_{x \in X} d_\rho(x).$$

The function $d_\rho(x)$ does not necessarily attain its infimum in X, hence, we choose $x_\rho \in X$ to be a point such that

$$d_\rho(x) - d_\rho \leqslant 1.$$

Such points x_ρ are called *min-max* points of ρ for obvious reasons. The set of min-max points could be unbounded, but, as we will see, this does not matter. Thus, high value of d_ρ means that *all* points of X move a lot by at least one of the generators of $\rho(G)$, while small value of d_ρ means that at least one point in X is moved only a bit by all the generators of G.

EXAMPLE 9.184. 1. Let $X = \mathbb{H}^n$, $G = \langle g \rangle$ be infinite cyclic group, where $\rho(g) \in \mathrm{Isom}(X)$ is a hyperbolic translation along a geodesic $L \subset X$ with the translation number $t > 1$, e.g., $\rho(g)(\mathbf{x}) = e^t \mathbf{x}$ in the upper half-space model. Then $d_\rho = t$ and we can take $x_\rho \in L$, since L is the set of point of minima for $d_\rho(x)$.

2. Suppose that $X = \mathbb{H}^n = \mathbf{U}^n$ and G are the same as above, but $\rho(g)$ is a parabolic isometry, e.g. $\rho(g)(\mathbf{x}) = \mathbf{x} + \mathbf{u}$, where $\mathbf{u} \in \mathbb{R}^{n-1}$ is a unit vector. Then d_ρ does not attain its infimum, $d_\rho = 0$ and we can take as x_ρ any point $\mathbf{x} \in \mathbf{U}^n$ with $x_n \geqslant 1$.

3. Suppose that X is the same, but G is no longer required to be cyclic. Assume that $\rho(G)$ fixes a unique point $x_o \in X$. Then $d_\rho = 0$ and the set of min-max points is contained in a metric ball centered at x_o. The radius of this ball could be estimated from above independently of G and ρ. (The latter is nontrivial.)

Suppose $\sigma \in \mathrm{Isom}(X)$ and we replace the original representation ρ with the conjugate representation

$$\rho' = \rho^\sigma : g \mapsto \sigma\rho(g)\sigma^{-1}, g \in G.$$

EXERCISE 9.185. Verify that $d_\rho = d_{\rho'}$ and that as $x_{\rho'}$ one can take $\sigma(x_\rho)$.

Thus, conjugating ρ by an isometry, does not change the geometry of the action, but moves min-max points in a predictable manner.

The set $Hom(G, \mathrm{Isom}(X))$ embeds in $(\mathrm{Isom}(X))^m$ since every ρ is determined by the m-tuple

$$(\rho(g_1), ..., \rho(g_m)).$$

As usual, we equip the group $\mathrm{Isom}(X)$ with the topology of uniform convergence on compacts and the set $Hom(G, \mathrm{Isom}(X))$ with the subspace topology.

EXERCISE 9.186. Show that the topology on $Hom(G, \mathrm{Isom}(X))$ is independent of the finite generating set. Hint: Embed $Hom(G, \mathrm{Isom}(X))$ in the product of countably many copies of $\mathrm{Isom}(X)$ (indexed by the elements of G) and relate the topology on $Hom(G, \mathrm{Isom}(X))$ to the product topology on the infinite product.

Suppose now that the metric space X is proper. Pick a base-point $o \in X$. Then the Arzela-Ascoli theorem implies that for every D the subset

$$Hom(G, \mathrm{Isom}(X))_{o,D} = \{\rho : G \to \mathrm{Isom}(X) | d_\rho(o) \leqslant D\}$$

is compact. We next consider the quotient

$$Rep(G, \mathrm{Isom}(X)) = Hom(G, \mathrm{Isom}(X)) / \mathrm{Isom}(X),$$

where $\mathrm{Isom}(X)$ acts on $Hom(G, \mathrm{Isom}(X))$ by conjugation $\rho \mapsto \rho^\sigma$. We equip $Rep(G, \mathrm{Isom}(X))$ with the quotient topology. In general, this topology is not Hausdorff:

EXAMPLE 9.187. Let $G = \langle g \rangle$ is infinite cyclic, $X = \mathbb{H}^n$. Show that the trivial representation $\rho_0 : G \to 1 \in \mathrm{Isom}(X)$ and the representation ρ_1, where $\rho_1(g)$ acts as a parabolic translation, project to points $[\rho_i]$ in $Rep(G, \mathrm{Isom}(X))$, such that every neighborhood of $[\rho_0]$ contains $[\rho_1]$. Hence, $Rep(G, \mathrm{Isom}(X))$ is not even T_1 in this example.

EXERCISE 9.188. Let X be a graph (not necessarily locally-finite) with the standard metric and consider the subset $Hom_f(G, \mathrm{Isom}(X))$ consisting of representations ρ which give rise to the free actions $G/Ker(\rho) \curvearrowright X$. Then

$$Rep_f(G, \mathrm{Isom}(X)) = Hom_f(G, \mathrm{Isom}(X)) / \mathrm{Isom}(X)$$

is Hausdorff.

We will be primarily interested in compactness rather than Hausdorff properties of $Rep(G, \mathrm{Isom}(X))$. Define

$$Hom_D(G, \mathrm{Isom}(X)) = \{\rho : G \to \mathrm{Isom}(X) | d_\rho \leqslant D\}.$$

Similarly, for a subgroup $H \subset \mathrm{Isom}(X)$, define

$$Hom_D(G, H) = Hom_D(G, \mathrm{Isom}(X)) \cap Hom(G, H).$$

LEMMA 9.189. *Suppose that $H \subset \mathrm{Isom}(X)$ is a closed subgroup whose action on X is cobounded. Then for every $D \in \mathbb{R}_+$, the quotient $Rep_D(G, H) = Hom_D(G, H)/H$ is compact.*

PROOF. Let $o \in X, R < \infty$ be such that the orbit of $\bar{B}(o, R)$ under the H-action is the entire space X. For every $\rho \in Hom(G, H)$ we pick $\sigma \in H$ such that *some* min-max point x_ρ of ρ satisfies:

$$\sigma(x_\rho) \in \bar{B}(o, R).$$

Then, using conjugation by such σ's, for each equivalence class $[\rho] \in Rep_D(G, H)$ we choose a representative ρ with $x_\rho \in \bar{B}(o, R)$. It follows that for every such ρ

$$\rho \in Hom(G, H) \cap Hom(G, \mathrm{Isom}(X))_{o,D'}, \quad D' = D + 2R.$$

This set is compact and, hence, its projection $Rep_D(G, H)$ is also compact. ☐

In view of this lemma, even if X is not proper, we say that a sequence of representations $\rho_i : G \to \mathrm{Isom}(X)$ *diverges* if

$$\lim_{i\to\infty} d_{\rho_i} = \infty.$$

DEFINITION 9.190. We say that an isometric action of a group on a real tree T is *unbounded* if the group does not fix a point in T.

PROPOSITION 9.191 (M. Bestvina; F. Paulin, []). *Suppose that (ρ_i) is a diverging sequence of representations $\rho_i : G \to H \subset \mathrm{Isom}(X)$, where X is a Rips–hyperbolic metric space. Then G admits an unbounded isometric action on a real tree.*

PROOF. Let $p_i = x_{\rho_i}$ be min-max points of ρ_i's. Take $\lambda_i := (d_{\rho_i})^{-1}$ and consider the corresponding asymptotic cone $\mathbf{X} = Cone_\omega(X, \mathbf{p}, \lambda)$ of the space X; here $\mathbf{p} = (p_i)$. According to Lemma 9.37, the metric space $\mathbf{X}$ in this asymptotic cone is a real tree. Furthermore, the sequence of group actions ρ_i converges to an isometric action $\rho_\omega : G \curvearrowright \mathbf{X}$, defined by:

$$\rho_\omega(g)(x_\omega) = (\rho_i(x_i)),$$

the key here is that *all generators $\rho_i(g_k)$ of $\rho_i(G)$ move the base-point $p_i \in \lambda_i X$ by* $\leqslant \lambda_i(d_{\rho_i} + 1)$. The ultralimit of the latter quantity is equal to 1. Furthermore, for ω-all i one of the generators, say $g = g_k$, satisfies

$$|d_{\rho_i} - d(\rho_i(g)(p_i), p_i)| \leqslant 1$$

in X. Thus, the element $\rho_\omega(g)$ will move the point $\mathbf{p} \in \mathbf{X}$ exactly by 1. Because p_i was a min-max point of ρ_i, it follows that

$$d_{\rho_\omega} = 1.$$

In particular, the isometric action $\rho_\omega : G \curvearrowright \mathbf{X}$ has no fixed point, i.e., is unbounded. □

One of the important applications of this proposition is:

THEOREM 9.192 (F. Paulin, [**Pau91a**]). *Suppose that G is a finitely-generated group with the property FA and H is a hyperbolic group. Then, up to conjugation in H, there are only finitely many homomorphisms $G \to H$.*

PROOF. Let X be a Cayley graph of H, then $H \subset \mathrm{Isom}(X)$, X is proper and Rips–hyperbolic. By the above proposition, if $Hom(G, H)/H$ is noncompact, then G has an unbounded isometric action on a real tree. This contradicts the assumption that G has the property FA. Suppose, therefore, that $Hom(G, H)/H$ is compact. If this quotient is infinite, pick a sequence $\rho_i \in Hom(G, H)$ of pairwise non-conjugate representations. Without loss of generality, by replacing ρ_i's by their conjugates, we can assume that min-max points p_i of ρ_i's are in $\overline{B}(e, 1)$. Therefore, after passing to a subsequence if necessary, the sequence of representations ρ_i converges. However, the action of H on itself is free, hence, for every generator g of G, the sequence $\rho_i(g)$ is eventually constant. It follows that the entire sequence (ρ_i) consists of only finitely many distinct representations. Contradiction. Thus, $Hom(G, H)/H$ is finite. □

This theorem is one of many results bounding number of homomorphisms from a group to a hyperbolic group. Having the Property FA is a very strong restriction on the group, thus, typically, one improves the Proposition 9.191 by making stronger assumptions on representations $G \to H$ and, accordingly, stronger conclusions about the action of G on the tree, for instance:

<u>Theorem</u> 9.193. *Suppose that G is a group, H is a hyperbolic group, X is a Cayley graph of H and $\rho_i : G \to H, i \in \mathbb{N}$ is a sequence of faithful representations as in Proposition 9.191. Then the action $G \curvearrowright T$ of G on a real tree as in Proposition 9.191, is* small*, i.e., stabilizer of every nontrivial geodesic segment is virtually cyclic.*

Given this theorem, whose proof can be found e.g. in [**Pau91b**], one then (typically) uses the *Rips Theory*, which converts small actions (satisfying some mild restrictions which will hold in the case of groups G which embed in hyperbolic groups) $G \curvearrowright T$, into graph-of groups decompositions of G with virtually cyclic edge groups. We refer the reader to [**BF95, RS94, Kap01**] for the details. As the result, one obtains:

<u>Theorem</u> 9.194 (E. Rips, Z. Sela, [**RS94**]). *Suppose that G does not split over a virtually cyclic subgroup. Then for every hyperbolic group H, $Hom_{inj}(G,H)/H$ is finite, where Hom_{inj} consists of injective homomorphisms. In particular, if G is itself hyperbolic, then $Out(G) = Aut(G)/Inn(G)$ is finite.*

Some interesting and important groups G, like surface groups, do split over virtually cyclic subgroups. In this case, one cannot in general expect $Hom_{inj}(G,H)/H$ to be finite. However, it turns out that the only reason for the lack of finiteness is the fact that one can precompose homomorphisms $G \to H$ with automorphisms of G itself:

<u>Theorem</u> 9.195 (E. Rips, Z. Sela, [**RS94**]). *Suppose that G is a 1-ended finitely-generated group. Then for every hyperbolic group H, the set*

$$Aut(G)\backslash Hom_{inj}(G,H)/H$$

is finite. Here $Aut(G)$ acts on $Hom(G,H)$ by precomposition.

9.24. Summary of equivalent definitions of hyperbolicity

Below we give a list of equivalent definitions of hyperbolicity for a (finitely-generated) group G (some of these definitions we saw earlier).

(1) Some/every Cayley graph of G is Rips-hyperbolic, i.e., has δ-thin triangles.
(2) G is Gromov-hyperbolic, i.e., when equipped with the word metric for some (every) finite generating set, the group G satisfies Gromov's inequality for the Gromov–product.
(3) Some/every Cayley graph of G has δ-thin bigons for some $\delta < \infty$, P. Papasoglu [**Pap95c**].
(4) G admits a Dehn–presentation.
(5) G is finitely presented and satisfies a linear isoperimetric inequality.
(6) G is finitely presented and satisfies subquadratic isoperimetric inequality.
(7) G is finitely presented and its isoperimetric function satisfies

$$IP(\ell) \le \frac{1-\epsilon}{4\pi}\ell^2$$

for some $\epsilon > 0$ and all sufficiently large ℓ.
(8) G is finitely presented and has sublinear filling radius.
(9) Every asymptotic cone of G is a real tree, M. Gromov [**Gro93**], see also Proposition 9.158.

(10) G is finitely presented and one asymptotic cone of G is a real tree, M. Kapovich, B. Kleiner [**KK09**], see also Theorem 9.161.
(11) The minsize function of one (every) Cayley graph of G is sublinear, see Proposition 9.167.
(12) For some $\lambda \in (0, \frac{1}{4}]$ the λ-constriction function constr_λ of G is sublinear, C. Drutu [**Dru01**], see also Proposition 9.170.
(13) Some/every Cayley graph of G has proper uniform divergence, P. Papasoglu [**Pap95c**].
(14) Some/every Cayley graph of G has exponential uniform divergence, P. Papasoglu [**Pap95c**].
(15) Some/every Cayley graph of G admits a thin bicombing, B. Bowditch and U. Hamenstädt, [**Ham07**, Proposition 3.5], see Theorem 9.12.
(16) G is finitely-presented and the canonical map between bounded and ordinary cohomology groups with coefficients in Banach $\mathbb{Z}G$-modules,

$$H_b^*(G, V) \to H^*(G, V)$$

is surjective, I. Mineyev [**Min01**].
(17) G is finitely-presented and

$$\ell_1 H_1(G, \mathbb{R}) = \overline{\ell_1 H_2}(G, \mathbb{R}) = 0,$$

D. Allcock and S. Gersten [**AG99**].
(18) Either G is virtually cyclic or G acts topologically on a compact perfect metrizable space Z of infinite cardinality, such that the induced action $G \curvearrowright \text{Trip}(Z)$ is properly discontinuous and cocompact, B. Bowditch [**Bow98b**], see also Theorem 9.134.

9.25. Further properties of hyperbolic groups

We conclude this chapter with a list of properties of hyperbolic groups not discussed earlier in the chapter:

1. Hyperbolic groups are ubiquitous:

<u>Theorem</u> 9.196 (See e.g. [**Del96**]). *Let G be a non-elementary δ-hyperbolic group. Then there exists N, such that for every collection $g_1, \ldots, g_k \in G$ of elements of length $\geq 1000\delta$, the following holds:*

i. The subgroup generated by the elements g_i^N and all their conjugates is free.

ii. Then the quotient group $G / \langle\langle g_1^n, \ldots, g_k^n \rangle\rangle$ is again non-elementary hyperbolic for all sufficiently large n. In particular, infinite hyperbolic groups are never simple.

Thus, by starting with, say, a nonabelian free group $F_n = G$, and adding to its presentation one relator of the form w^n at a time (where n's are large), one obtains non-elementary hyperbolic groups.

"Most" groups are hyperbolic:

<u>Theorem</u> 9.197 (A. Ol′shanskiĭ [**Ol′92**]). *Fix $k \in \mathbb{N}$, $k \geq 2$ and let $A = \{a^{\pm 1}, a^{\pm 2}, \ldots, a_k^{\pm 1}\}$ be an alphabet. Fix $i \in \mathbb{N}$ and let $(n_1, \ldots, n_i)$ be a sequence of natural numbers. Let $N = N(k, i, n_1, \ldots, n_i)$ be the number of group presentations*

$$\langle a_1, \ldots, a_k | r_1, \ldots, r_i \rangle$$

such that $r_1, \ldots, r_i$ are reduced words in the alphabet A such that the length of r_j is n_j, $j = 1, 2, \ldots, i$. If N_h is the number of presentations as above which define hyperbolic groups and if $n = \min\{n_1, \ldots, n_i\}$, then

$$\lim_{n\to\infty} \frac{N_h}{N} = 1,$$

and convergence is exponentially fast.

The model of randomness which appears in this theorem is, by no means, unique; below are two other models (among many others). We refer the reader to [**Gro03**], [**Ghy04**], [**Oll04b**], [**KS08**] for further discussion of random groups.

i. Fix the number $k \geqslant 2$ and consider the set $B(n)$ of presentations

$$\langle x_1, \ldots, x_k | R_1, \ldots, R_l \rangle ,$$

where the total length of the words $R_1, ..., R_l$ is $\leqslant n$. Then a class $\mathcal{C}$ of k-generated groups is said to consist of random groups if

$$\lim_{n\to\infty} \frac{\text{card } (B(n) \cap C)}{\text{card } B(n)} = 1.$$

ii. Here is another notion of randomness: Fix the number l of relators, assume that all relators have the same length n; this defines a class of presentations $S(k, l, n)$. Then require

$$\lim_{n\to\infty} \frac{\text{card } (S(k, l, n) \cap C)}{\text{card } S(k, l, n)} = 1.$$

See [**KS08**] for a comparison of various notions of randomness for groups. In all existing models of randomness, once random groups are infinite, they are hyperbolic with Menger curve as the ideal boundary, see e.g. [**DGP11**].

2. Hyperbolic groups have quotients with "exotic" properties:

THEOREM 9.198 (A. Ol′shanskiĭ, [**Ol′91c**]). *Every non-elementary torsion-free hyperbolic group admits a quotient which is an infinite torsion group, where every nontrivial element has the same order.*

THEOREM 9.199 (A. Ol′shanskiĭ, [**Ol′95**], T. Delzant [**Del96**]). *Every non-elementary hyperbolic group G is SQ-universal, i.e., every countable group embeds in a quotient of G.*

Theorems 9.200, 9.201, 9.202 below first appeared in Gromov's paper [**Gro87**]; other proofs could be found for instance in [**Aea91**], [**BH99**], [**ECH$^+$92**], [**ECH$^+$92**], [**GdlH90**].

3. Hyperbolic groups have finite type:

THEOREM 9.200. *Let G be δ-hyperbolic. Then there exists $D_0 = D_0(\delta)$ such that for all $D \geq D_0$ the Rips complex $\mathrm{Rips}_D(G)$ is contractible. In particular, G has type F_∞.*

4. Hyperbolic groups have controlled torsion:

THEOREM 9.201. *Let G be hyperbolic. Then G contains only finitely many conjugacy classes of finite subgroups.*

5. Hyperbolic groups have solvable algorithmic problems:

THEOREM 9.202. *Every δ-hyperbolic group has solvable word and conjugacy problems.*

Furthermore:

THEOREM 9.203 (I. Kapovich, [**Kap96**]). *The membership problem is decidable for quasiconvex subgroups of hyperbolic groups: Let G be hyperbolic and $H < G$ be a quasiconvex subgroup of a δ-hyperbolic group. Then the problem of membership in H is decidable.*

The isomorphism problem is decidable:

THEOREM 9.204 (Z. Sela, [**Sel95**]; F. Dahmani and V. Guirardel [**DG11**]). *There is an algorithm whose input is a pair P_1, P_2) of finite presentations of δ-hyperbolic groups G_1, G_2, has the output YES if G_1, G_2 are isomorphic and NO if they are not.*

Note that Sela proved this theorem only for torsion-free 1-ended hyperbolic groups. This result was extended to all hyperbolic groups by Dahmani and Guirardel.

6. Hyperbolic groups are hopfian:

THEOREM 9.205 (Z. Sela, [**Sel99**]). *For every hyperbolic group G and every epimorphism $\phi : G \to G$, $\mathrm{Ker}(\phi) = 1$.*

Note that every finitely generated residually finite group is hopfian, but the converse, in general, is false. An outstanding open problem is to determine if all hyperbolic groups are residually finite (it is widely expected that the answer is negative). Every finitely generated linear group is residually finite, but there are nonlinear hyperbolic groups, see [**Kap05**]. It is very likely that some (or even all) of the nonlinear hyperbolic groups described in [**Kap05**] are not residually finite.

7. Hyperbolic groups tend to be co-hopfian:

THEOREM 9.206 (Z. Sela, [**Sel97a**]). *For every 1-ended hyperbolic group G, every monomorphism $\phi : G \to G$ is surjective, i.e., such G is co-hopfian.*

8. All hyperbolic groups admit QI embeddings in some hyperbolic space $\mathbb{H}^n$:

THEOREM 9.207 (M. Bonk, O. Schramm [**BS00**]). *For every hyperbolic group G there exists n, such that G admits a quasiisometric embedding in $\mathbb{H}^n$.*

9.26. Relatively hyperbolic spaces and groups

Relatively hyperbolic groups were introduced by M. Gromov in the same paper [**Gro87**] as hyperbolic groups. While hyperbolic groups are modeled uniform lattices in negatively curved symmetric spaces, relatively hyperbolic groups are modeled on non-uniform lattices in negatively curved spaces and, more generally, fundamental groups of complete Riemannian manifolds of finite volume and curvature $\leqslant -a^2 < 0$. A good picture is that of truncated hyperbolic spaces defined in Chapter 10 (see Figure 10.1). These are metric spaces hyperbolic relative to the boundary horospheres. In general, one considers a geodesic metric space X and a collection $\mathcal{A}$ of subsets of it (called *peripheral subsets* when the relative hyperbolicity conditions are fulfilled).

The metric definition of relative hyperbolicity consists of three conditions, the main one being very similar to the condition of thin triangles for hyperbolic spaces.

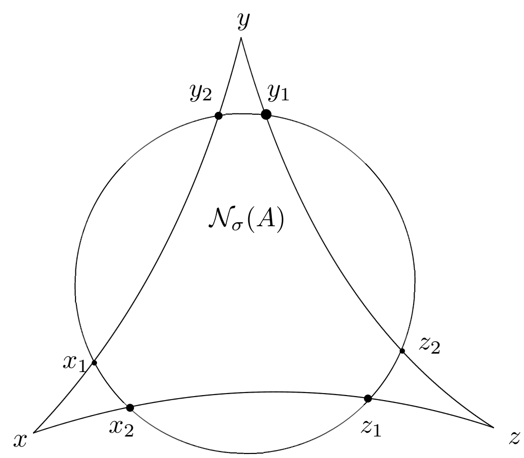


FIGURE 9.17. Second case of Definition 9.208.

DEFINITION 9.208. We say that X is $(*)$–*relatively hyperbolic with respect to* $\mathcal{A}$ if for every $C \geqslant 0$ there exist two constants σ and δ such for every triangle $T \subset X$ with $(1, C)$–quasi-geodesic edges, either there exists a point at distance at most σ from each of the sides of T, or there exists a subset $A \in \mathcal{A}$ such that its σ-neighborhood $\mathcal{N}_\sigma(A)$ intersects each of the sides of the triangle; moreover, for every vertex of the triangle, the two edges issuing from it enter $\mathcal{N}_\sigma(A)$ in two points at distance at most δ away from each other.

Clearly $(*)$–relative hyperbolicity is a rather weak condition. For instance every geodesic hyperbolic space is $(*)$–hyperbolic relative to every family of subsets covering it.

DEFINITION 9.209. The space X is *hyperbolic relative* to $\mathcal{A}$ if it is $(*)$–hyperbolic relative to $\mathcal{A}$, and moreover, the following properties are satisfied:

(α_1) For every $r > 0$, the r–neighborhoods of any two distinct subsets in $\mathcal{A}$ intersect in a set of diameter at most $D = D(r)$.

(α_2) Every geodesic of length ℓ with endpoints at distance at most $\frac{\ell}{3}$ from a set $A \in \mathcal{A}$, intersects the M-tubular neighborhood of A, with some uniform constant M.

DEFINITION 9.210. A finitely generated group G is *hyperbolic relative to* a finite set of subgroups $H_1, ..., H_n$ if, endowed with a word metric, G is hyperbolic in the sense of Definition 9.209 relative to the collection $\mathcal{A}$ of left cosets

$$\{gH_i : i \in \{1, 2, ..., n\}, g \in G\}.$$

It follows from the definition that all H_i's are finitely generated, since the three metric conditions imply that the peripheral subsets are quasi-convex. The groups H_i are called the *peripheral subgroups* of the relatively hyperbolic structure on G.

THEOREM 9.211 (C. Druţu, D. Osin, M. Sapir, [**DS05b**],[**Osi06**],[**Dru09**]). *Relative hyperbolicity in the sense of Definition 9.210 is equivalent to (strong) relative hyperbolicity as defined in* [**Gro87**].

Other characterizations of (strong) relative hyperbolicity can be found in the papers [**Bow12**], [**Far98**], [**Dah03b**], [**DS05b**], [**Osi06**]. Here and in what follows, by *relative hyperbolicity* we always mean *strong* relative hyperbolicity; we will always assume that *every* H_i *has infinite index in* G.

In the list of properties in Definition 9.209, one cannot drop the property (α_1), as shown by the examples of groups in [**OOS09**] and in [**BDM09**, §7.1].

Many properties similar to those of hyperbolic groups are proved in the relatively hyperbolic case, in particular a Morse lemma, a characterization in terms of asymptotic cones [**DS05b**], of relative linear filling [**Osi06**], and of action on the boundary as a convergence group [**Yam04**].

Hyperbolic groups are clearly relatively hyperbolic with the peripheral subgroup $\{1\}$. Other examples of relatively hyperbolic groups include:

(1) G is hyperbolic and each H_i is quasiconvex and *almost malnormal* in G (see [**Far98**]). Almost malnormality of a subgroup $H \leqslant G$ means that for every $g \in G \setminus H$,
$$|gHg^{-1} \cap H| < \infty.$$
(2) G is the fundamental group of a finite graph of groups with finite edge groups; then G is hyperbolic relative to the vertex groups, see [**Bow12**].
(3) Fundamental groups of complete finite volume manifolds of pinched negative curvature; the peripheral subgroups are the fundamental groups of their cusps ([**Bow12**], [**Far98**]).
(4) Fully residually free groups, also known as limit groups of Sela; they have as peripheral subgroups a finite list of maximal abelian non-cyclic subgroups [**Dah03a**].

Similarly to hyperbolic groups, relatively hyperbolic groups are used to construct examples of infinite finitely generated groups with exotic properties. Denis Osin used in [**Osi10**] direct limits of relatively hyperbolic groups to construct torsion-free two-generated groups with exactly two conjugacy classes (i.e., all elements $\neq 1$ are conjugate to each other).

Study of relatively hyperbolic groups is a very active and rapidly developing area of Geometric Group Theory. We refer the reader to [**Bow12, DG08, Dru09, BDM09, DS05b, DS05a, DS07, Ger09, Ger12, GP13, MR08, Osi06, Osi10, Yam04**] for further reading.

CHAPTER 10

Lattices in Lie groups

In §3.6.5 we defined lattices in general locally compact groups. In this chapter we consider lattices in Lie groups. While our main motivation comes from lattices in the Lie group $PO(n,1) \cong SO(n,1)$, the isometry group of the hyperbolic n-space, most of our discussion here is general. Lattices in Lie groups (as well as in "p-adic Lie groups") play prominent role in geometric group theory for several reasons:

1. Rigidity theorems for lattices (proven by Mostow and Margulis) played key role in the development of basic concepts and tools of geometric group theory.

2. Lattices act on homogeneous spaces, which provide nice *geometric models* for studying coarse geometry of lattices themselves.

3. Having a nice geometric model helps to prove QI rigidity theorems for lattices: Somehow, the rigid geometric nature of homogeneous (primarily symmetric) spaces, translates into QI rigidity of lattices.

While the exact nature of QI rigidity for lattices in general connected Lie groups is still unclear, QI rigidity of lattices in semisimple Lie groups is now well-understood, see Chapter 23. We refer the reader to the Gelander's survey **[Gel14]** for a detailed review of properties of lattices in Lie groups.

10.1. Semisimple Lie groups and their symmetric spaces

Consider a Lie group G with a compact subgroup $K < G$. In view of uniqueness (up to scaling) of the Haar measure μ on G, we can define μ as follows. Pick arbitrarily a positive definite bilinear form on the tangent space T_eG, where $e \in G$ is the identity element. Then, using the fact that G acts on itself smoothly and simply-transitively *via* the left multiplication, we spread this bilinear form from T_eG to the rest of the tangent bundle TG. The result is a left-invariant Riemannian metric $h = \langle \cdot, \cdot \rangle$ on G and, hence, a G-invariant volume form. This volume form yields, *via* integration, the measure μ. This basic construction has an important modification. We let the compact subgroup $K < G$ act on G *via* the right multiplication. Compactness of K allows us to average the metric h:

$$Av(h) = \frac{1}{mes(K)} \int_K R_k^*(h)dk.$$

Here the integration is with respect to the Haar measure on K and R_k is the right multiplication:

$$R_k(g) = gk.$$

The metric $Av(h)$ is then both left-invariant (with respect to the action of G) and right-invariant (with respect to the action of K). This *left-right* invariant Riemannian metric on G descends to a G-invariant Riemannian metric on the manifold $X = G/K$ and we obtain a homogeneous Riemannian manifold X. Conversely, as we explained in §3.6.5, one can lift the measure (defined via an invariant volume form) from X to a Haar measure on G.

The homomorphism

$$\rho : G \to \mathrm{Isom}(X)$$

defined by the isometric action of G on X, is not, in general, injective, as G might have a normal subgroup contained in K. For instance, the action of the group $G = SL(2,\mathbb{R})$ on the hyperbolic plane $\mathbb{H}^2 \cong G/K$, $K = SO(2)$, has nontrivial kernel, equal to the center $\{\pm I\}$ of G. Furthermore, the image of the group G in $\mathrm{Isom}(X)$ can have infinite index. Nevertheless, in view of transitivity of the action of G on X and compactness of K, both the kernel $\mathrm{Ker}(\rho)$ of ρ and its *cokernel*, the quotient $\mathrm{Isom}(X)/\rho(G)$, are compact.

Example 10.1. Consider the group $SU(2)$ with a *biinvariant Riemannian metric*, i.e., a metric invariant under both left and right multiplication. The group $SU(2) \times SU(2)$ maps to $\mathrm{Isom}(X)$, $X = SU(2)$, where the first factor acts em via the left multiplication, while the second factor acts *via* the right multiplication. The kernel of the homomorphism $SU(2) \times SU(2) \to \mathrm{Isom}(X)$ equals $\mathbb{Z}_2 = \langle(-I, -I)\rangle$, where $-I$ is the negative of the identity matrix in $SU(2)$. We leave it to the reader to verify that the Riemannian manifold X is isometric to a round 3-dimensional sphere (of some radius).

Some Lie groups, and their lattices, are better-behaved and more interesting than others. In §3.6.4 we defined the class of *semisimple Lie groups*. We now impose further conditions on the groups in this class:

Convention 10.2. In order to simplify the terminology, from now on, when referring to a Lie group G as *semisimple*, we will always assume that G is linear (i.e., admits a monomorphism $G \to GL(N,\mathbb{R})$ for some N), has finitely many components and does not contain nontrivial compact connected subgroups.

The latter assumption guarantees that the kernel of $\rho : G \to \mathrm{Isom}(X)$ is finite. Every semisimple group G contains unique, up to conjugation, maximal compact subgroup K. Moreover, the quotient homogeneous space $X = G/K$ (with the projection of any left-right invariant metric on G) is a nonpositively curved simply-connected complete Riemannian manifold. The G-invariant metric on X is essentially unique; for instance, if the group G is simple, then this metric is unique up to a constant factor. The manifolds X obtained this way are *symmetric spaces of noncompact type*. One interesting feature of such spaces is that $\rho(G)$ has finite index in $\mathrm{Isom}(X)$. In other words, the homomorphism ρ has both finite kernel and cokernel.

The *rank of the symmetric space* X, $\mathrm{rank}\,(X)$, is defined as the dimension of a *maximal flat* in the associated symmetric space $X = G/K$, i.e., the maximal r such that there exists an isometric embedding of the Euclidean r-space into X:

$$\mathbb{E}^r \to (X, \mathrm{dist}_X),$$

where dist_X is the Riemannian distance function on X. For instance, X has rank one if and only if X is negatively curved. The (real rank), $\mathrm{rank}\,(G)$, of the group G can be defined geometrically as the rank of the symmetric space $X = G/K$. (There is an alternative, algebraic, definition of the rank of G, which we will not give here.)

Definition 10.3. Suppose that X is a nonpositively curved simply-connected symmetric space. Then X admits a *de Rham decomposition*, i.e., an isometric splitting as the direct product

$$X = \prod_{i=1}^{n} X_i$$

of irreducible symmetric spaces X_i, i.e., spaces which themselves do not split as nontrivial products.

The de Rham splitting is preserved by the group $\mathrm{Isom}(X)$, except that some factors might be permuted by isometries. Furthermore, the splitting is unique (up to permuting the indices) and it reflects the algebraic decomposition of the Lie algebra of G as the direct sum of simple subalgebras

$$\mathfrak{g} = \oplus_{i=1}^{n} \mathfrak{g}_i.$$

Namely, after rearranging the indices, for each i, the Lie algebra $\mathfrak{g}_i$ is isomorphic to the Lie algebra of $\mathrm{Isom}(X_i)$.

10.2. Lattices

Suppose now that G is a semisimple Lie group, $K < G$ is a compact subgroup, $X = G/K$ is the associated symmetric space, as in the previous section. Every subgroup Γ of G also acts isometrically on X and we will frequently identify Γ with its image $\rho(\Gamma)$, in $\mathrm{Isom}(X)$ (since ρ has finite kernel, this abuse of terminology is *mostly harmless*). Compactness of K (and, hence, finiteness of its Haar measure) has three immediate, but important, consequences:

Lemma 10.4. *1. A subgroup $\Gamma < G$ is discrete if and only if Γ is discrete as a subgroup of $\mathrm{Isom}(X)$, if and only if Γ acts properly discontinuously on X.*

2. A discrete subgroup $\Gamma < G$ is a lattice if and only if the quotient $M = \Gamma \setminus X$ has finite volume.

3. A lattice $\Gamma < G$ is uniform (see §3.6.5) if and only if the quotient $\Gamma \setminus X$ is compact.

We have to warn the reader at this point that the quotient space M is, in general, not a manifold but an *orbifold*, since the group Γ can fail to act freely on X. If Γ acts freely on X, we can compute the volume of M via the projection of the Riemannian metric from X. Otherwise, one can define the volume of M, say, *via* fundamental sets as in §3.6.5. It is a nontrivial fact that each lattice in G is finitely-generated: This result is clear for uniform lattices in Lie groups with finitely many components, but is difficult in general. Furthermore, T. Gelander **[Gel11]** established the following relation between the minimal number of generators of lattices and volumes of their quotient spaces:

<u>Theorem</u> 10.5. $\mathrm{rank}\,(\Gamma) \leqslant C \cdot Vol(\Gamma \setminus X)$*, where C is a constant depending only on X.*

Here $\mathrm{rank}\,(\Gamma)$, the *rank* of Γ, is the minimum of cardinalities of its generating sets.

Since we are assuming that the Lie group G is linear, each lattice in G is virtually torsion free (according to Selberg's lemma, Theorem 4.111). For torsion-free lattices $\Gamma < G$, the projection $X \to M = \Gamma \setminus X$ is a covering map and the

Riemannian metric projects from X to a Riemannian metric on the smooth manifold M.

The quotients $M = \Gamma \setminus X$ of symmetric spaces X by discrete subgroups $\Gamma < G$ are called *locally-symmetric spaces*. Quotients defined by torsion-free rgoups Γ are Riemannian manifolds locally isometric to X.

EXERCISE 10.6. Recall that two subgroups Γ_1, Γ_2 of a group G are called *commensurable* if $\Gamma_1 \cap \Gamma_2$ has finite index in Γ_1, Γ_2. Show that if two subgroups Γ_1, Γ_2 in a locally compact group G are commensurable, then Γ_1 is a lattice if and only if Γ_2 is a lattice. Furthermore, show that the lattice Γ_1 is uniform if and only if Γ_2 is uniform.

Let

$$X \cong \prod_{i=1}^{n} X_i$$

be the de Rham decomposition of the symmetric space X. A lattice Γ in the isometry group $\mathrm{Isom}(X)$ is called *irreducible* if it is not commensurable to a lattice $\Gamma' < \mathrm{Isom}(X)$ of the product form:

$$\Gamma' = \prod_{i=1}^{n} \Gamma'_i,$$

$\Gamma'_i < \mathrm{Isom}(X_i), i = 1, \ldots, n$.

10.3. Examples of lattices

The most basic examples of lattices are given by the integer linear groups $SL(n, \mathbb{Z})$ and their finite index subgroups. Proving that $\Gamma = SL(n, \mathbb{Z})$ is a lattice in $G = SL(n, \mathbb{R})$ is not easy: While discreteness is clear, finiteness of volume of the quotient is not obvious. Proving finiteness requires constructing a certain subset S, called a *Siegel set* in G, which contains a fundamental domain of Γ. Then one verifies that S indeed has finite volume, from which it is immediate that $Vol(\Gamma \setminus G) < \infty$.

Arithmetic groups generalize this example and provide a rich and interesting source of lattices in all semisimple Lie groups.

DEFINITION 10.7. An *arithmetic subgroup* in a semisimple Lie group G is a subgroup of G commensurable to a subgroup of the form

$$\Gamma := \phi^{-1}(GL(N, \mathbb{Z})),$$

for a (continuous) homomorphism $\phi : G \to GL(N, \mathbb{R})$ with compact kernel.

As in the case of $SL(n, \mathbb{Z})$, it is clear that every arithmetic subgroup is discrete in G. It is a much deeper theorem that every arithmetic subgroup is a lattice in a Lie subgroup $H \leqslant G$, see e.g. [**Mar91, Rag72**]. We refer the reader to [**Bor63**] and [**Rag72**] for proofs of the following theorem:

THEOREM 10.8 (A. Borel). *Every semisimple Lie group G contains both uniform and nonuniform arithmetic lattices. Furthermore, G contains infinitely many commensurability classes of arithmetic lattices, both uniform and nonuniform.*

In other words, arithmetic lattices are ubiquitous. Arithmetic groups are also interesting, since they provide connections between various fields of mathematics (geometry, topology, analysis, ergodic theory) to number theory: Many number-theoretic results and conjectures can be stated (and proven!) in the form of properties of various lattices and their quotient spaces.

Fuchsian groups. We will refer to lattices in $PO(2,1) = \mathrm{Isom}(\mathbb{H}^2)$ as *Fuchsian groups*. Apart from the groups $SL(n,\mathbb{Z})$, these are the most-studied lattices, whose investigation goes back to the 2nd half of the 19th century. For instance, every finitely generated free group and the fundamental group of every closed connected surface of negative Euler characteristic is isomorphic a Fuchsian group. If S is a compact oriented surface of genus $p > 1$, and Π_p is its fundamental group, then the space of conjugacy classes of isomorphisms from Π_p to lattices in $PSL(2,\mathbb{R}) < PO(2,1)$ is a manifold of dimension $3p-3$.

EXAMPLE 10.9. Consider the group $G = PO(2,1)$ and a non-uniform lattice $\Gamma < G$. After passing to a finite-index subgroup in Γ, we may assume that Γ is torsion-free. Then the quotient $\mathbb{H}^2/\Gamma$ is a non-compact surface with fundamental group Γ. Therefore, Γ is a free group of finite rank.

EXERCISE 10.10. Show that the groups Γ in the above example cannot be cyclic.

Bianchi groups. We now describe a very concrete class of non-uniform arithmetic lattices in the isometry group of the hyperbolic 3-space, called *Bianchi groups*. Let $D \in \mathbb{Z}$ be a *square-free negative integer*, i.e., an integer which is not divisible by the square of a prime number. Consider the *imaginary quadratic field*

$$\mathbb{Q}(\sqrt{D}) = \{a + \sqrt{D}b : a, b \in \mathbb{Q}\}$$

in $\mathbb{C}$. Set

$$\omega := \sqrt{D}, \text{ if } D \equiv 2, 3, \mod 4$$

$$\omega := \frac{1+\sqrt{D}}{2}, \text{ if } D \equiv 1, \mod 4$$

Then the ring of integers of $\mathbb{Q}(\sqrt{D})$ is

$$O_D = \{a + \omega b : a, b \in \mathbb{Z}\}.$$

For instance, if $D = -1$, then O_D is the ring of *Gaussian integers*

$$\{a + ib : a, b \in \mathbb{Z}\}.$$

A *Bianchi group* is a subgroup of the form

$$SL(2, O_D) < SL(2, \mathbb{C})$$

for some D. Since the ring O_D is discrete in $\mathbb{C}$, it is immediate the every Bianchi subgroup is discrete in $SL(2,\mathbb{C})$. By abusing the terminology, one also refers to the group $PSL(2, O_D)$ as a Bianchi subgroup of $PSL(2,\mathbb{C})$.

Bianchi groups Γ are arithmetic lattices in $SL(2,\mathbb{C})$; in particular, the quotients $\mathbb{H}^3/\Gamma$ have finite volume. Furthermore, every nonuniform arithmetic lattice in $SL(2,\mathbb{C})$ is commensurable to a Bianchi group. We refer the reader to [**MR03**] for the detailed discussion of these and other facts about Bianchi groups.

10.4. Rigidity and Superrigidity

The example of Fuchsian groups shows that lattices in $PO(2,1)$ are highly flexible: They typically admit continuum of nonconjugate representations (as lattices) into $PO(2,1)$. The theory of lattices in general semisimple Lie groups took off in the late 1950s, when it was discovered that lattices in other Lie groups actually tend to be quite rigid. This development culminated in the fundamental rigidity theorems due to Mostow and Margulis which we describe below. Of course, in order to get rigidity, in addition to Fuchsian groups, one has to exclude their products in the products of $PO(2,1)$'s. This explains the *irreducibility* assumptions in rigidity theorems. In order to keep the statements simple, we first formulate the rigidity results in the context of *simple* Lie groups and, after that, for semisimple groups.

The proof of the next theorem can be found in Mostow's book **[Mos73]**

<u>Theorem</u> 10.11 (G. D. Mostow, Strong Rigidity Theorem). *1. Let G_1, G_2 be connected linear noncompact simple Lie group with trivial centers, not isomorphic to $PSL(2,\mathbb{R})$. Then for any two lattices $\Gamma_1 < G_1, \Gamma_2 < G_2$, every isomorphism*

$$\phi : \Gamma_1 \to \Gamma_2$$

extends to an isomorphism $G_1 \to G_2$. Geometrically speaking, the isomorphism ϕ is induced by a similarity $f : X_1 \to X_2$ of the associated symmetric spaces $X_i = G_i/K_i$. (The mapping f becomes an isometry after one replaces the metric on X_2 by its scalar multiple.)

2. Assume that the groups G_i are connected semisimple, without nontrivial normal compact subgroups. Assume also that both G_1 and G_2 are not isomorphic to $PSL(2,\mathbb{R})$. Then for any two irreducible lattices $\Gamma_1 < G_1, \Gamma_2 < G_2$, every isomorphism

$$\phi : \Gamma_1 \to \Gamma_2$$

extends to an isomorphism $G_1 \to G_2$.

Mostow originally proved his theorem only for lattices in $G = PO(n,1)$, the isometry group of the hyperbolic n-space, $n \geqslant 3$. In §22.3 we will give another proof of Mostow's theorem for $PO(n,1)$.

In the case when the spaces X_i have rank 1 (i.e., are negatively curved), Mostow's proof is along the same lines as for the real-hyperbolic space: He first constructs an equivariant quasiisometry between symmetric spaces, then extends this quasiisometry to the ideal boundaries, establishes that the extension is quasiconformal and then proves that this quasiconformal extension is, in fact, conformal. It is the last step where the assumption that X is not isometric to the hyperbolic plane is used. Mostow used ergodic theory arguments in the last step of his proof; we will be using the *zooming argument*, which seems to have its origin in Gromov's paper **[Gro81b]**.

In the case of symmetric spaces of rank $\geqslant 2$ Mostow's proof starts in the same fashion, but then he uses the theory of Tits buildings at infinity of X_i's instead of quasiconformal analysis.

Since Mostow's pioneering work, other, very different proofs of his rigidity theorem have emerged. For instance, for $X = \mathbb{H}^n$ there are very different proofs due to Gromov and Thurston **[BP92]** (based on bounded cohomology) and due to

Besson, Gourtois and Gallot [**BCG96, BCG98**]. The latter proof is differential-geometric in nature and avoids analyzing the boundary maps. In the complex-hyperbolic setting, Siu [**Siu80**] gave a proof using *Kähler geometry*; his proof is based on harmonic maps between compex–hyperbolic manifolds.

In his theorem Mostow assumes an isomorphism between two lattices. Margulis' Superrigidity Theorem below goes one step further: Margulis considers *arbitrary* homomorphisms from lattices $\Gamma < G$ into the group $GL(N, \mathbb{R})$, allowing, for instance, images to be nondiscrete. Of course, there is a price to be paid for this level of generality on the side of homomorphisms, one has to restrict the class of Lie groups G. Namely, for every n, there exist arithmetic lattices Γ (both uniform and non-unuform) in the groups $PO(n, 1)$ and $PU(n, 1)$, such that Γ has infinite abelianization, i.e., there exists an epimorphism $\Gamma \to \mathbb{Z}$ (see [**Mil76**], [**Kaz75**]).

<u>Theorem</u> 10.12 (G. Margulis, Archimedean Superrigidity Theorem). *1. Suppose that G is a simple connected (linear) Lie group and $\Gamma < G$ is an lattice. Assume, in addition, that G has rank at least 2. Then for every homomorphism*

$$\phi : \Gamma \to GL(N, \mathbb{R}),$$

either the image $\phi(\Gamma)$ is relatively compact, or there exists a finite index subgroup $\Gamma' < \Gamma$ such that the restriction $\phi|_{\Gamma'}$ extends to a homomorphism $G \to GL(N, \mathbb{R})$.

2. The same conclusion holds if the group G is semisimple, of rank $\geqslant 2$ and $\Gamma < G$ is an irreducible lattice.

Margulis also proved a *nonarchimedean superrigidity theorem*, which deals with representations of lattices Γ as above into the groups $GL(N, \mathbb{Q}_p)$, where the conclusion is exactly the same as before. Instead of trying to formulate the nonarchimedean superrigidity theorem, we will only state its special case:

<u>Theorem</u> 10.13. *Suppose that Γ is an irreducible lattice in a semisimple Lie group G of rank $\geqslant 2$. Then each action of Γ on a simplicial tree has a fixed point.*

The full nonarchimedean superrigidity theorem states existence of a fixed point for isometric actions of irreducible lattices on higher-dimensional generalizations of trees, which are called *Euclidean buildings*. As an application of these remarkable rigidity theorems, Margulis proved:

<u>Theorem</u> 10.14 (G. Margulis, Arithmeticity Theorem). *Every lattice Γ satisfying the hypothesis of Theorem 10.12 is arithmetic.*

We refer the reader to Margulis' book [**Mar91**] for the proofs. The Margulis Arithmeticity theorem was extended to lattices in the groups $\mathrm{Isom}(\mathbf{H}\mathbb{H}^n)$ $(n \geqslant 2)$ and $\mathrm{Isom}(\mathbf{O}\mathbb{H}^2)$ by K. Corlette [**Cor92**] and by M. Gromov and R. Schoen [**GS92**]. The combination of these arithmeticity theorems yields:

<u>Theorem</u> 10.15. *Suppose that $\Gamma < \mathrm{Isom}(X)$ is an irreducible lattice, where X is a nonpositively curved symmetric space not isometric (up to rescaling) to a real-hyperbolic space $\mathbb{H}^n$ and complex-hyperbolic space $\mathbf{C}\mathbb{H}^n$. Then Γ is arithmetic.*

It follows from the work of M. Gromov and I. Piatetsky-Shapiro [**GPS88**] that for each $n \geqslant 3$, the group $PO(n, 1) = \mathrm{Isom}(\mathbb{H}^n)$, contains infinitely many VI classes of both uniform and nonuniform nonarithmetic lattices. On the other hand, only finitely many VI classes of non-arithmetic lattices are known in $PU(2, 1)$ and $PU(3, 1)$ (the groups of biholomorphic isometries of complex-hyperbolic plane and complex-hyperbolic 3-space). No non-arithmetic lattices are currently known in the groups $PU(n, 1)$, $n \geqslant 4$.

10.5. Commensurators of lattices

Recall (see §3.2) that the commensurator of a subgroup Γ in a group G is the subgroup $Comm_G(\Gamma) < G$ consisting of elements $g \in G$ such that the groups $g\Gamma g^{-1}$ and Γ are commensurable, i.e.,

$$|\Gamma : g\Gamma g^{-1} \cap \Gamma| < \infty$$

and

$$|g\Gamma g^{-1} : g\Gamma g^{-1} \cap \Gamma| < \infty.$$

Below we consider commensurators in the situation when Γ is a lattice in a Lie group G.

EXERCISE 10.16. Let $\Gamma := SL(2, O_D) < G := SL(2, \mathbb{C})$ be a Bianchi group.

1. Show that $Comm_G(\Gamma) < SL(2, \mathbb{Q}(\omega))$. In particular, $Comm_G(\Gamma)$ is dense in G.

2. Show that the set of fixed points of parabolic elements in Γ (in the upper half-space model of $\mathbb{H}^3$) is

$$\mathbb{Q}(\omega) \cup \{\infty\}.$$

3. Show that $Comm_G(\Gamma) = SL(2, \mathbb{Q}(\omega))$.

G. Margulis proved (see [**Mar91**], Chapter IX, Theorem B and Lemma 2.7; see also [**Zim84**], Theorem 6.2.5) that a lattice in a semisimple real Lie group G is arithmetic if and only if its commensurator is dense in G.

10.6. Lattices in $PO(n, 1)$

We now turn to the case of lattices in the isometry group $PO(n, 1)$ of the hyperbolic n-space $\mathbb{H}^n$. This material discussed here will be used in Chapters 21, 22 in the proofs of QI rigidity theorems for lattices.

10.6.1. Zariski density. The next lemma is a basic results about discrete subgroups of $PO(n, 1)$.

LEMMA 10.17. *Suppose that α, β are hyperbolic and parabolic isometries of $\mathbb{H}^n$ respectively, which have a common fixed point ξ in the boundary sphere $\mathbb{S}^{n-1}$ of $\mathbb{H}^n$. Then the subgroup $\Gamma < PO(n, 1)$ generated by α and β is not discrete.*

PROOF. We will identify $\mathbb{S}^{n-1}$ with $\mathbb{R}^{n-1} \cup \{\infty\}$ so that ξ corresponds to the point ∞ and that the second fixed point of α corresponds to $0 \in R^{n-1}$. Then α is a similarity

$$\mathbf{x} \mapsto \lambda A\mathbf{x}, A \in O(n-1), \lambda \neq 0, |\lambda| \neq 1,$$

and β is a (skew) translation

$$\mathbf{x} \mapsto B\mathbf{x} + \mathbf{v}, B \in O(n-1), \mathbf{v} \neq 0.$$

Here and below $\mathbf{x}$'s are vectors in $\mathbb{R}^{n-1}$. Suppose first that $|\lambda| < 1$. Then consider the following sequence of conjugates of β in Γ:

$$\beta_k = \alpha^k \beta \alpha^{-k} : \mathbf{x} \mapsto C_k\mathbf{x} + \lambda^k A^k \mathbf{v},$$

where $C_k \in O(n-1)$ and, of course, $A^k \in O(n-1)$ as well. Since the group $O(n-1)$ is compact, the sequence C_k subconverges to some $C \in O(n-1)$. At the same time,

$$\lim_{k\to\infty} \lambda^k A^k \mathbf{v} = 0.$$

Therefore, a subsequential limit of the sequence of Moebius transformations (β_k) is the orthogonal transformation

$$\mathbf{x} \mapsto C\mathbf{x}.$$

It follows that Γ is a non-discrete subgroup of $Mob(\mathbb{S}^{n-1})$.

The case $|\lambda| > 1$ is similar: Instead of the sequence β_k above, one uses the sequence $\alpha^{-k}\beta\alpha^k$. □

COROLLARY 10.18. *Suppose that $\alpha_1, \alpha_2 \in PO(n,1)$ are hyperbolic elements which generate a discrete subgroup of $PO(n,1)$ and have a common fixed point ξ in $\mathbb{S}^{n-1}$. Then α_1, α_2 share both fixed point; in particular, the subgroup they generate has an invariant geodesic in $\mathbb{H}^n$, asymptotic to their common fixed points.*

PROOF. Suppose that the second fixed points of α_1, α_2 (the ones different from ξ) are distinct. The reader will verify that the commutator $\beta = [\alpha_1, \alpha_2]$ is a parabolic element of $PO(n,1)$. Since β clearly fixes ξ, we get a contradiction with Lemma 10.17. □

THEOREM 10.19. *Suppose that n is at least 2. Then:*

1. No lattice $\Gamma < PO(n,1)$ can have a proper invariant hyperbolic subspace $H \subset \mathbb{H}^n$.

2. No lattice $\Gamma < PO(n,1)$ can have a fixed point in the boundary sphere $\mathbb{S}^{n-1}$.

PROOF. 1. Suppose that such subspace H exists. We let $\pi : \mathbb{H}^n \to H$ be the nearest-point projection. Since the subspace H is Γ-invariant, the projection π is Γ-equivariant. The mapping π extends continuously to a Γ-equivariant mapping

$$\pi : Y := \mathbb{H}^n \cup (\mathbb{S}^{n-1} \setminus \partial_\infty H) \to H.$$

Proper discontinuity of the action of Γ on H implies proper discontinuity of the action of Γ on Y. Therefore, there exists a point $y \in Y \cap \mathbb{S}^{n-1}$ with trivial Γ-stabilizer and its neighborhood U in Y such that

$$\gamma U \cap U = \emptyset$$

for all $\gamma \in \Gamma \setminus \{1\}$. The neighborhood U contains metric balls $B(x,R) \subset U \cap \mathbb{H}^n$ of arbitrarily large radius R and, hence, arbitrarily large volume. Since the projection

$$\mathbb{H}^n \to M = \Gamma \setminus \mathbb{H}^n$$

is injective on the balls $B(x,R)$, we conclude that the space M has infinite volume. A contradiction.

2. The argument for fixed points at infinity is similar. Suppose that $\Gamma < PO(n,1)$ is a discrete subgroup fixing a point $\xi \in \mathbb{S}^{n-1}$. In view of Lemma 10.17, either all elements of Γ are parabolic and elliptic, or all its elements are hyperbolic and elliptic. We consider the former case and leave the latter to the reader as an exercise. Since Γ consists only of parabolic and elliptic elements, it preserves each horoball $B \subset \mathbb{H}^{n-1}$ centered at the point ξ. We now repeat the argument from Part 1 using the nearest-point projection to B instead of the nearest-point projection to an invariant hyperbolic subspace. □

Corollary 10.20. *If $\Gamma < PO(n,1)$ is a lattice, it cannot have a finite orbit in $\mathbb{S}^{n-1}$.*

Corollary 10.21. *If $\Gamma < PO(n,1)$ is a lattice, then Γ cannot contain nontrivial finite normal subgroups.*

Proof. Suppose that $\Phi \lhd \Gamma$ is a finite normal subgroup. According to Corollary 2.70, the fixed-point set F of Φ in $\mathbb{H}^n$ is nonempty. The set F is the intersection of fixed-point sets of the elements of Γ; the latter are hyperbolic subspaces of $\mathbb{H}^n$. Therefore, F itself is a hyperbolic subspace in $\mathbb{H}^n$. Since the subgroup Φ is normal in Γ, the set F has to be invariant under Γ. By Theorem 10.19a lattice in $PO(n,1)$ cannot have a proper invariant hyperbolic subspace; it follows that $F = \mathbb{H}^n$, i.e., the subgroup Φ is trivial. □

Note that every connected subgroup in $PO(n,1)$ either has index 2 (i.e., is the subgroup $PO_o(n,1)$ of orientation-preserving isometries of $\mathbb{H}^n$) or has a fixed point in $\mathbb{S}^{n-1}$, or has a proper invariant hyperbolic subspace in $\mathbb{H}^n$, see **[Gre62]**. Therefore, we conclude that a lattice in $PO(n,1)$ cannot be contained in a connected subgroup of $PO(n,1)$, other than in $PO_o(n,1)$.

The properties of lattices in $PO(n,1)$ established above, are elementary manifestations of a harder theorem, due to A. Borel **[Bor60]**:

Theorem 10.22 (Borel Density Theorem). *Suppose that G is an algebraic Lie group. Then every lattice $\Gamma < G$ is Zariski dense.*

Corollary 10.23. *Every lattice Γ in a semisimple algebraic group G has finite center.*

Proof. If Γ has infinite center, so does the Zariski closure $\overline{\Gamma} < G$. The Borel Density Theorem implies that G has infinite center. Since the center of an algebraic group is an algebraic subgroup, it follows that the center of G gas positive dimension. By passing to the Lie algebra $\mathfrak{g}$ of G, we conclude that the center of $\mathfrak{g}$ is also nontrivial. This, however, contradict the assumption that the group G is semisimple. □

10.6.2. Parabolic elements and noncompactness. Consider the upper half-space model of the hyperbolic space $\mathbb{H}^n$. Recall that (open) horoballs in $\mathbb{H}^n$ with center at the point $\infty \in \partial_\infty \mathbb{H}^n$ are Euclidean half-spaces of the form

$$B_t = \{(x_1, \ldots, x_n) : x_n > t\}, \quad t > 0.$$

Accordingly, horospheres centered at the point ∞ are boundaries of horoballs:

$$H_t = \{(x_1, \ldots, x_n) : x_n = t\}, \quad t > 0.$$

Define the projection $\Pi : \mathbb{R}^n_+ \to \mathbb{R}^{n-1}$,

$$\Pi : (x_1, \ldots, x_{n-1}, x_n) \mapsto (x_1, \ldots, x_{n-1}).$$

For each $\mathbf{x} \in \mathbb{R}^{n-1}$ we set $\mathbf{x}(t) := \Pi^{-1}(\mathbf{x}) \cap H_t$:

$$\mathbf{x}(t) = (x_1, \ldots, x_{n-1}, t),$$

where $\mathbf{x} = (x_1, \ldots, x_{n-1})$.

Lemma 10.24. *Suppose that $\Gamma < PO(n,1)$ is a discrete subgroup containing a parabolic element γ. Then Γ cannot be a uniform lattice in $PO(n,1)$.*

Proof. Suppose to the contrary, that Γ is a uniform lattice. Without loss of generality, by conjugating Γ by an element of $PO(n,1)$, we may assume that the unique fixed point of the parabolic element $\gamma \in \Gamma$ is the point $\infty \in \mathbb{S}^{n-1}$. Therefore, γ acts as a Euclidean isometry on $\mathbb{R}^n_+$, which, after making further conjugation, has the form:

$$\gamma : \mathbf{x} \mapsto A\mathbf{x} + \mathbf{v}, \quad A \in O(n-1), \mathbf{v} \in \mathbb{R}^{n-1} \setminus \{0\}, \quad A\mathbf{v} = \mathbf{v}.$$

The isometry γ preserves the Euclidean straight line $L \subset \mathbb{R}^{n-1}$ spanned by the vector $\mathbf{v}$. Furthermore, the restriction of γ to L is the translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{v}$. Then, for each $t > 0$ and $\mathbf{x} \in L$, by integrating the hyperbolic length element along the Euclidean line segment c_t connecting $\mathbf{x}(t)$ and $\gamma\mathbf{x}(t)$, we obtain:

$$\text{length}(c_t) = \frac{|\mathbf{v}|}{t}.$$

Therefore,

$$\lim_{t\to\infty} \text{dist}(\mathbf{x}(t), \gamma\mathbf{x}(t)) = 0.$$

Since the action $\Gamma \curvearrowright \mathbb{H}^n$ is cocompact, for each t exists $\alpha_t \in \Gamma$ such that

$$\text{dist}(\alpha_t(\mathbf{x}(t)), p) \le R,$$

where $p \in \mathbb{H}^n$ is a base-point and R is a constant. Then the conjugate element of Γ

$$\gamma_t = \alpha_t \gamma \alpha_t^{-1}$$

moves the point $\alpha_t(x(t)) \in B(p,R)$ by the distance not exceeding $\frac{|\mathbf{v}|}{t}$. In view of compactness of the ball $K = \bar{B}(p,R)$, there exists a sequence t_i diverging to infinity, such that the sequence

$$q_i = \alpha_t(\mathbf{x}(t_i))$$

converges to some $q \in \mathbb{H}^n$. Then

$$\lim_{i\to\infty} \text{dist}(q, \gamma_{t_i}(q)) = 0.$$

Since the elements $\gamma_{t_i} \in \Gamma$ are not elliptic, we obtain a contradiction with discreteness of the group Γ. ☐

10.6.3. Thick-thin decomposition. The idea of the *thick-thin decomposition* of locally symmetric spaces $M = \Gamma \setminus X$ is that such M splits naturally into a *thin part* M_{thin}, which has reasonably simple topological structure, and a *thick part* M_{thick} which, typically, has complicated topology, but whose geometry is bounded. In the case when Γ is a lattice, the thick part of M turns out to be compact.

For simplicity, we will state and use the tick-thin decomposition only for lattices $\Gamma < PO(n,1)$, even though its version also holds for general discrete subgroups of $PO(n,1)$ and other semisimple Lie groups.

Theorem 10.25 (Thick-thin decomposition). *Suppose that Γ is a nonuniform lattice in $PO(n,1)$. Then:*

1. There exists an (infinite) collection C of open horoballs $C := \{B_j, j \in J\}$, with pairwise disjoint closures, such that

$$\Omega := \mathbb{H}^n \setminus \bigcup_{j\in J} B_j$$

is Γ-invariant and $M_c := \Omega/\Gamma$ is compact.

2. Every parabolic element $\gamma \in \Gamma$ preserves (exactly) one of the horoballs B_j.

The proof of this theorem is based on a mild generalization of the Zassenhaus theorem due to Kazhdan and Margulis, see e.g. [**BP92**], [**Kap01**], [**Rat06**], [**Thu97**].

We note that the stabilizer Γ_j of each horoball B_j in this theorem cannot contain hyperbolic elements (since they do not preserve horoballs); therefore, Γ_j consists only of parabolic and elliptic elements. In view of compactness of M_c, the quotient $T_j := \Sigma_j/\Gamma_j$ of each horosphere$\Sigma_j \subset \mathbb{H}^n$ bounding B_j, is compact. On the other hand, since Γ_j preserves horospheres with the same center as Σ_j, we have

$$\bar{B}_j/\Gamma_j \cong T_j \times [0, \infty).$$

The quotient M_c is called the *thick part*, M_{thick}, of $M = \mathbb{H}^n/\Gamma$, while its (noncompact) complement $M \setminus M_c$ is called the *thin part* of M. If Γ is torsion-free, then it acts freely on $\mathbb{H}^n$ and M has a natural structure of a hyperbolic manifold of finite volume. If Γ is not torsion-free, then M is a *hyperbolic orbifold*. In view of the above observations, M is compact if and only if $M_{thin} = \emptyset$, equivalently, $C = \emptyset$,

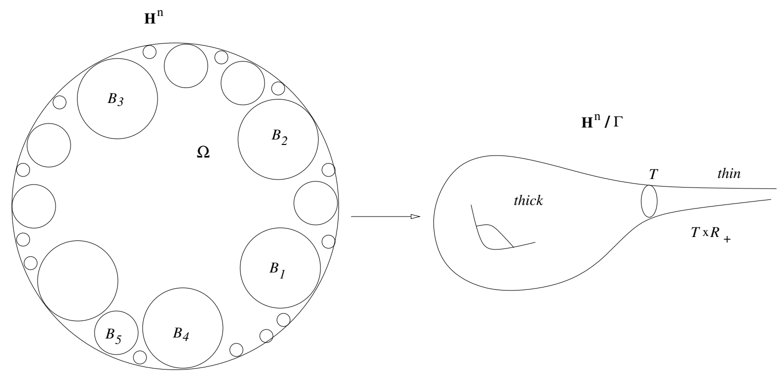


FIGURE 10.1. Truncated hyperbolic space and thick-thin decomposition.

The set Ω is called a *truncated hyperbolic space*. The boundary horospheres Σ_j of Ω are called *peripheral horospheres.*

LEMMA 10.26. *The truncated hyperbolic space* Ω *is contractible.*

PROOF. Since all closed horoballs $\bar{B}_j$ and all horospheres Σ_j are simply-connected, Seifert – van Kampen theorem implies that $\pi_1(\mathbb{H}^n)$ is isomorphic to the group $\pi_1(\Omega)$. Hence, Ω is simply-connected. Vanishing of all homology groups $H_k(\Omega)$, $k \geqslant 2$, follows from the Mayer –Vietoris sequence. Therefore, Hurewicz theorem implies that Ω is contractible. □

COROLLARY 10.27. *The group* Γ *is virtually torsion-free and has the type* $\mathbf{F}_\infty$.

PROOF. The group Γ acts properly discontinuously and cocompactly on the simply-connected space Ω. Therefore, Γ is finitely-generated. Since the group $PO(n, 1) \cong SO(n, 1)$ is linear and Γ is finitely-generated, the group Γ is virtually torsion-free (by Selberg's lemma). Let $\Gamma' < \Gamma$ be a finite index torsion-free subgroup. This subgroup acts smoothly, properly discontinuously, freely and cocompactly on the smooth manifold with boundary Ω. Therefore, the smooth quotient manifold Ω/Γ' admits a finite triangulation. Contractibility of Ω now implies that

Γ' has the type $\mathbf{F}$. Since the type $\mathbf{F}_\infty$ is a virtual isomorphism invariant (Corollary 6.59), the group Γ also has the type $\mathbf{F}_\infty$.

COROLLARY 10.28. *A lattice $\Gamma < PO(n,1)$ is uniform if and only if it does not contain parabolic elements. Moreover, if Γ is uniform, it contains a parabolic subgroup isomorphic to $\mathbb{Z}^{n-1}$.*

PROOF. Since Γ acts cocompactly on Ω, it follows that each subgroup $\Gamma_j < \Gamma$ acts cocompactly on the corresponding horosphere Σ_j, which is isometric to $\mathbb{R}^{n-1}$. Therefore, Γ_j is isomorphic to a uniform lattice in $\mathrm{Isom}(\mathbb{E}^{n-1})$. Bieberbach proved (see e.g. [**Rat06**, Theorem 7.5.2]) that each lattice in $\mathrm{Isom}(\mathbb{E}^{n-1})$ contains a finite-index subgroup isomorphic to $\mathbb{Z}^{n-1}$. All nontrivial elements of this subgroup of Γ_j have to be parabolic since they preserve the horosphere Σ_j. □

The next theorem is a sharpening of this corollary. We refer the reader to §9.13.4 for the definition of a conical limit point of a discrete group action on a hyperbolic space.

THEOREM 10.29. *If Γ is a lattice then every points $\xi \in \partial_\infty \mathbb{H}^n$ is either a conical limit point or a parabolic fixed point.*

PROOF. Let ρ be a geodesic ray in $\mathbb{H}^n$ asymptotic to ξ. This ray projects to a ray $\bar{\rho}$ in $M = \mathbb{H}^n/\Gamma$. Two things may occur:

Case 1: There exists $T \geqslant 0$ and a component $M_j = B_j/\Gamma_j$ of the thin part of M, such that for all $t \geqslant T$, $\bar{\rho}(t)$ belongs to M_j. Then the ray $\rho([T,\infty))$ is entirely contained in a Γ-translate B of the horoball B_j. However, if a horoball contains a geodesic ray, then this ray is asymptotic to the center of the horoball. It follows that the point ξ (to which ρ is asymptotic) is fixed by the subgroup $\Gamma_j < \Gamma$, which, as we know, contains parabolic elements.

Case 2: There exists a sequence $t_i \in \mathbb{R}_+$ diverging to ∞ such that for each i, $\bar{\rho}(t_i)$ belongs to the thick part M_c of M. Since M_c is compact, there exists a compact $C \subset \mathbb{H}^n$ and a sequence of elements $\gamma_i \in \Gamma$ such that $\rho(t_i) \in \gamma_i(C)$. Then, arguing as in the proof of Lemma 9.118, we conclude that ξ is a conical limit point of Γ. (In Lemma 9.118 we assumed that the action of Γ on a Gromov-hyperbolic space is cobounded, but, in fact, all what we needed was a geodesic ρ, a compact C and sequences t_i, γ_i as above.) □

REMARK 10.30. The above theorem holds for all symmetric spaces X of rank 1 and its converse holds as well (cf. [**Bis96**]): A discrete subgroup $\Gamma < \mathrm{Isom}(X)$ is a lattice if and only if every point of $\partial_\infty X$ is either a conical limit point of Γ or is fixed by a parabolic element of Γ.

10.7. Central coextensions

Recall that central coextensions

$$1 \to \mathbb{Z} \to \tilde{\Gamma} \to \Gamma \to 1$$

are classified by elements of the cohomology group $H^2(\Gamma, \mathbb{Z})$. In this section we describe some classes of lattices which admit nontrivial central coextensions.

For each subgroup $\Gamma' < \Gamma$ we have the *restriction homomorphism*:

$$H^2(\Gamma, \mathbb{Q}) \to H^2(\Gamma', \mathbb{Q})$$

defined by restricting cocycles to the subgroup Γ'. In general, this homomorphism may have large kernel (e.g., if we take $\Gamma' = \{1\}$). However, for finite-index subgroups $\Gamma' < \Gamma$ the behavior of cohomology classes is more predictible:

LEMMA 10.31. *Let $\Gamma' < \Gamma$ be a finite index subgroup. Then the restriction homomorphism $H^2(\Gamma, \mathbb{Q}) \to H^2(\Gamma', \mathbb{Q})$ is injective.*

A proof of this lemma can be found, for instance, in [**Bro82b**, Chapter III, Proposition 10.4], it is an application of the *transfer argument*, which allows one to push cochains from Γ' to Γ by *averaging them* (this is where finite index and rational coefficients are used).

In particular, if a central coextension of Γ is given by a cohomology class which has nonzero projection to $H^2(\Gamma, \mathbb{Q})$, then this central extension remains nontrivial over every finite index subgroups $\Gamma' < \Gamma$.

We next need a class of lattices with nontrivial second betti numbers, i.e., nonvanishing $H^2(\Gamma, \mathbb{Q})$.

THEOREM 10.32. *1. For every $n \geqslant 2$, the group $PO(n, 1)$ contains uniform lattices Γ with nonvanishing $H^2(\Gamma, \mathbb{Q})$.*

2. If Γ is a torsion-free uniform lattice in $PU(n, 1)$, then $H^2(\Gamma, \mathbb{Q}) \neq 0$.

3. Every torsion-free uniform lattice $\Gamma < SO_o(n, 2)$ has nonzero $H^2(\Gamma, \mathbb{Q})$. Here $SO_o(n, 2)$ is the identity component of the Lie group $SO(n, 2)$.

PROOF. 1. For every $n \geqslant 2$, the group $PO(n, 1)$ contains torsion-free uniform lattices Γ with nonvanishing $H^2(\Gamma, \mathbb{Q}) \cong H^2(\mathbb{H}^n/\Gamma\mathbb{Q})$, see [**MR81**].

2. A multiple of the Kähler form on the complex-hyperbolic space projects to the quotient manifold $M = \mathbf{C}\mathbb{H}^n/\Gamma$ and defines a nonzero element of $H^2(\Gamma, \mathbb{Q})$.

3. The same argument with the Kähler class applies in this case as well, see Toledo's appendix to [**Ger92**]. The nonzero cohomology class comes from the 1st Chern class in $H^2(M, \mathbb{Z})$, where M is the locally-symmetric space of Γ, $M = \Gamma \setminus SO(n, 2)/SO(2) \times SO(n)$. ☐

COROLLARY 10.33. *Every lattice Γ in Theorem 10.32 admits a central coextension*

$$1 \to \mathbb{Z} \to \tilde{\Gamma} \to \Gamma \to 1$$

which satisfies:

a. The coextension does not split over any finite index subgroup $\Gamma' < \Gamma$. In particular, $\tilde{\Gamma}$ is not virtually isomorphic to a product group $\Gamma' \times \mathbb{Z}$.

b. The group $\tilde{\Gamma}$ is quasiisometric to the product $\Gamma \times \mathbb{Z}$.

PROOF. (a) A multiple of a nonzero cohomology class $H^2(\Gamma, \mathbb{Q})$ is a nontrivial integral cohomology class. The latter defines a nontrivial central coextension

$$1 \to A = \mathbb{Z} \to \tilde{\Gamma} \to \Gamma \to 1. \tag{10.1}$$

As we noted above, this central coextension does not split over any finite index subgroup $\Gamma' < \Gamma$. In order to see that the group $\tilde{\Gamma}$ is not isomorphic to direct product $\Gamma \times \mathbb{Z}$, we note that Γ has trivial center (see Corollary 10.23). Therefore, any isomorphism

$$\tilde{\phi} : \tilde{\Gamma} \to \Gamma_1 \times A_1, \quad A_1 \cong \mathbb{Z}$$

would send A isomorphically to A_1. Therefore, $\tilde{\phi}$ would project to an isomorphism $\phi : \Gamma \to \Gamma_1$. This would imply that the central coextension (10.1), resulting in a contradiction. The same argument applies to finite index subgroups of Γ. Part (a) follows.

(b) The symmetric spaces $\mathbb{H}^n$ and $\mathbf{C}\mathbb{H}^n$ associated with the Lie groups $PO(n,1)$ and $PU(n,1)$ are negatively curved. Hence, the uniform lattices $\Gamma < PO(n,1), \Gamma < PU(n,1)$ are Gromov-hyperbolic. Theorem 9.150 implies that each a central coextension $\tilde{\Gamma}$ of such lattice is quasiisometric to the direct product $\Gamma \times \mathbb{Z}$.

The argument in the case of lattices in $SO(n,2)$ is less direct, since the associated symmetric space $X = SO(n,2)/SO(2) \times SO(n)$ is not Gromov-hyperbolic. However, for each uniform torsion-free lattice $\Gamma < SO(n,2)$, the notrivial class in Part 3 of Theorem 10.32 is *bounded*: It lies in the image of the natural homomorphism

$$H^2_b(\Gamma, \mathbb{Q}) \to H^2(\Gamma, \mathbb{Q}).$$

See Toledo's appendix to [**Ger92**]. Therefore, the central coextension $\tilde{\Gamma}$ defined by the class ω is quasiisometric to the product $\Gamma \times \mathbb{Z}$, see Remark 9.155. □

EXERCISE 10.34. Prove a generalization of Corollary 10.33 to central coextensions with the kernel $\mathbb{Z}^k$, $k \geqslant 2$.

CHAPTER 11

Solvable groups

This chapter covers basic properties of general solvable groups and some special classes of solvable groups: Abelian, nilpotent and polycyclic groups. These properties will be used in proofs of theorems about growth of groups. Much of this material is algebraic rather than geometric, we decided to keep it in the book for the sake of completeness. Solvable and polycyclic groups appear naturally in the framework of *poly-X-groups*, where X is a certain class of groups: A group G is said to be poly-X if it admits a finite sequence of normal subgroups:

$$\{1\} \lhd G_k \lhd \ldots \lhd G_2 \lhd G_1 \lhd G_0 = G,$$

such that each successive quotient G_i/G_{i+1} belongs to the class X. Solvable groups will be obtain by taking X to be the class of abelian groups, while polycyclic groups will use the class of cyclic groups (a further refinement of the definition uses X consisting of infinite cyclic groups, all isomorphic to each other, of course). As an aside, we note that there are other interesting classes of poly-X groups which we will not be discussing in the book, like *poly-free* groups, important examples of which are given by the *pure braid groups*.

Notation. For abelian groups G we will frequently use the notation mg or $m \cdot g$ for the m-fold sum

$$\underbrace{g + \ldots + g}_{m \text{ times}},$$

with $m \in \mathbb{N}$. This extends to $m \in \mathbb{Z}$ by declaring that

$$0 \cdot g = 0 \in G$$

and that

$$-(m \cdot g) = (-m) \cdot g.$$

11.1. Free abelian groups

Definition 11.1. A group G is called *free abelian* on a generating set S if it is isomorphic to the direct sum

$$\bigoplus_{s \in S} \mathbb{Z}.$$

The minimal cardinality of S is called *the rank of* G and denoted rank(G), the set S is called a *basis* of G.

Of course, if $|S| = n$, $G \cong \mathbb{Z}^n$. Given an abelian group G, we define its subgroup

$$2G = \{2x : x \in G\}.$$

Clearly, this subgroup is *characteristic* in G, i.e., is invariant under all automorphisms of G. Then, for the free abelian group $G = \oplus_{s\in S}\mathbb{Z}$, the quotient $G/2G$ is isomorphic to

$$\bigoplus_{s\in S} \mathbb{Z}_2,$$

which has natural structure of a vector space over $\mathbb{Z}_2$ with basis S. Since any two bases of a vector space have the same cardinality, it follows that two bases of a free abelian group have the same cardinality, equal to $\mathrm{rank}\,(G)$.

Exercise 11.2. Every free abelian group is torsion-free.

Below is a characterization of free abelian groups by a *universality property*:

Theorem 11.3. *Let G be an abelian group and X is a subset of G. The group G is free abelian with basis X if and only if it satisfies the following* universality property*: For every abelian group A, every map $f : X \to A$ extends to a unique homomorphism $f : G \to A$.*

Proof. Suppose that G is free abelian with the basis X. Every element $g \in G$ is uniquely represented as a sum

$$g = \sum_{x\in X} m_x \cdot x, m_x \in \mathbb{Z}$$

with only finitely many nonzero terms. Then, we extend f to G by

$$f(g) = \sum_{x\in X} m_x \cdot f(x).$$

It is clear that this extension is unique.

Conversely, assume that $(G_1, X_1), (G_2, X_2)$ satisfy the universality property and $f : X_1 \to X_2$ is a bijection. Then f and $f^{-1} = \bar{f} : X_2 \to X_1$ admit homomorphic extensions $F : G_1 \to G_2$, $\bar{F} : G_2 \to G_1$ respectively. The compositions $\bar{F} \circ F, F \circ \bar{F}$ are homomorphisms $\phi : G_1 \to G_1, \psi : G_2 \to G_2$, respectively. These homomorphisms extend the identity maps $X_2 \to X_2, X_1 \to X_1$. By the uniqueness part of the universality property, it follows that ϕ and ψ are the identity maps. Therefore, the homomorphism $F : G_1 \to G_2$ is an isomorphism. Applying this to $G_1 = G, X_1 = X$ and G_2 equal to the free abelian group with the basis $X_2 = X_1 = X$, we conclude that G is free abelian with the basis X. □

Corollary 11.4. *Let $0 \to A \to B \xrightarrow{r} C \to 0$ be a short exact sequence of abelian groups, where C is free abelian. Then this sequence splits and $B \cong A \oplus C$.*

Proof. Let $c_i, i \in I$, denote a basis of C. Then, since r is surjective, for every c_i there exists $b_i \in B$ such that $r(b_i) = c_i$. By the universal property of free abelian groups, the map $s : c_i \to b_i$ extends to a homomorphism $s : C \to B$ such that $r \circ s = Id$. □

Exercise 11.5. Show that a group G is free abelian with the basis S if and only if G admits the presentation

$$\langle S | [s, s'] = 1, \forall s, s' \in S \rangle .$$

The following theorem is the abelian analogue of the Nielsen–Schreier theorem (Theorem 4.42), although, we are unaware of a topological or geometric proof:

THEOREM 11.6. *1. Subgroups of free abelian groups are again free abelian.*
2. If $G < F$ is a subgroup of a free abelian group F, then $\operatorname{rank}(G) \leqslant \operatorname{rank}(F)$.

PROOF. Let X be a basis of a free abelian group $F = A_X$. For each subset Y of X let A_Y be the free group with the basis Y, thus A_Y embeds naturally as a free abelian subgroup A_Y in F. We fix a subgroup $G \leqslant F$ once and for all; for each $Y \subset X$ we let G_Y denote the intersection $G \cap A_Y$.

Define the set S consisting of triples (G_Y, B, ϕ), where Y ranges over the set of all subsets of X such that G_Y is free with a basis of cardinality at most the cardinality of X; the sets B are bases of such G_Y, and ϕ is an injective map $\phi : B \to X$.

The set S is nonempty, as we can take $Y = \emptyset$.

We define a partial order $\leqslant$ on S by:

$$(G_Y, B, \phi) \leqslant (G_Z, C, \psi) \iff Y \subset Z, B \subset C, \quad \phi = \psi|_B.$$

Suppose that L is a chain in the above order indexed by an ordered set M:

$$\{(G_{Y_m}, B_m, \phi_m), m \in M\}, (G_{Y_m}, B_m, \phi_m) \leqslant (G_{Y_n}, B_n, \phi_n) \iff m \leqslant n.$$

Then the union

$$\bigcup_{m \in M} G_{Y_m}$$

is again a subgroup in F and the set

$$C = \bigcup_{m \in M} B_m$$

is a basis in the above group. Furthermore, the maps ϕ_m determine an embedding $\psi : C \hookrightarrow X$. Thus,

$$(\bigcup_{m \in M} G_{Y_m}, C, \psi) \in S.$$

Therefore, by Zorn's Lemma, there exists a maximal element (G_Y, B, ϕ) of S. If $Y = X$ then $G_Y = G$ and we are done. Suppose that there exists $x \in X \setminus Y$. Set $Z := Y \cup \{x\}$. We will show that G_Z is still free abelian with a basis C containing B and ϕ extends to an embedding $\psi : Z \to X$. If $G_Z = G_Y$, we take $C = B$, $\psi = \phi$. Otherwise, assume that $G_Z/G_Y \neq 0$. The quotient A_Z/A_Y is isomorphic to $\mathbb{Z}$ and is generated by the image $\bar{x}$ of x. The image of G_Z in this quotient is isomorphic to G_Z/G_Y and is generated by some $n \cdot \bar{x}$, $n \in \mathbb{Z} \setminus 0$. Let $g \in G_Z$ be an element which maps to $n \cdot \bar{x}$. The mapping $G_Z/G_Y \to \langle g \rangle$ splits the sequence

$$0 \to G_Y \to G_Z \to G_Z/G_Y = \mathbb{Z} \to 0$$

and, hence,

$$G_Z \cong G_Y \oplus \langle g \rangle .$$

This means that $C := B \cup \{g\}$ is a basis of G_Z; we extend ϕ to C by $\psi(g) = x$. Thus, $(G_Z, C, \psi) \in S$. This contradicts maximality of (G_Y, B, ϕ).

We conclude that G is free abelian and its basis embeds in a basis of F. □

11.2. Classification of finitely generated abelian groups

THEOREM 11.7. *Every finitely generated abelian group A is isomorphic to a finite direct sum of cyclic groups.*

PROOF. The proof below is taken from **[Mil12]**. The proof is induction on the number of generators of A.

If A is 1-generated, the assertion is clear. Assume that the assertion holds for abelian groups with $\leqslant n-1$ generators and suppose that A is an abelian group generated by n elements. Consider all ordered generating sets $(a_1, ..., a_n)$ of A. Among such generating sets choose one, $S = (a_1, ..., a_n)$, such that the order of a_1 (denoted $|a_1|$) is the least possible. We claim that

$$A \cong \langle a_1 \rangle \oplus A' = \langle a_1 \rangle \oplus \langle a_2, ..., a_n \rangle .$$

(This claim will imply the assertion since, inductively, A' splits as a direct sum of cyclic groups.) Indeed, if A is not the direct sum as above, then we have a nontrivial relation

$$\sum_{i=1}^{n} r_i a_i = 0, r_i \in \mathbb{Z}, r_1 a_1 \neq 0. \tag{11.1}$$

Without loss of generality, $0 < r_1 < |a_1|$ and $r_i \geqslant 0, i = 1, ...n$ (otherwise, we replace a_i's with $-a_i$ whenever $r_i < 0$). Furthermore, let $d = gcd(r_1, ..., r_n)$ be the greatest common divisor of the numbers $r_i, i = 1, ..., n$. Set $q_i := \frac{r_i}{d}$.

LEMMA 11.8. *Suppose that $a_1, ..., a_n$ are generators of A and $q_1, ..., q_n \in \mathbb{Z}_+$ are such that $gcd(q_1, ..., q_n) = 1$. Then there exists a new generating set $b_1, ..., b_n$ of A such that*

$$b_1 = \sum_{i=1}^{n} q_i a_i.$$

PROOF. The proof of this lemma is a form of the Euclid's algorithm for computation of gcd. Note that $q := q_1 + ... + q_n \geqslant 1$. The proof of lemma is induction on q. If $q = 1$ then $b_1 \in \{a_1, ..., a_n\}$ and lemma follows. Suppose the assertion holds for all $q < m$, we will prove the claim for $q = m > 1$. After rearranging the indices, we can assume that $q_1 \geqslant q_2 > 0$.

Clearly, the set $\{a_1, a_1 + a_2, a_3, ..., a_n\}$ generates A. Furthermore,

$$gcd(q_1 - q_2, q_2, q_3, ..., q_n) = 1$$

and

$$q' := (q_1 - q_2) + q_2 + q_3 + ... + q_n < m$$

Thus, by the induction hypothesis, there exists a generating set $b'_1, ..., b'_n$ of A, where

$$b'_1 = (q_1 - q_2)a_1 + q_2(a_1 + a_2) + q_3 a_3 + ... + q_n a_n.$$

However, $b_1 = b'_1$. Lemma follows. ☐

In view of this lemma, we get a new generating set $b_1, ..., b_n$ of A such that

$$b_1 = \sum_{i=1}^{n} \frac{r_i}{d} a_i.$$

The equation (11.1) implies that $db_1 = 0$ and $d \leqslant r_1 < |a_1|$. Thus, the ordered generating set $(b_1, ..., b_n)$ of A has the property that $|b_1| < |a_1|$, contradicting our choice of S. Theorem follows. ☐

For a prime p, an abelian group A is called a *p-group* if every element $a \in A$ has the order which is a power of p. Clearly, each subgroup and each quotient of a p-group is again a p-group.

EXERCISE 11.9. A finite abelian group A is a p-group if and only if $|A| = p^\ell$ for some ℓ.

Given an abelian group A, we let $A(p)$ denote the subset of A consisting of elements whose order is a power of p. Since the sum of two elements of the orders p^k, p^m has the order p^n, where $n = \max(k, m)$, the subset $A(p)$ is a subgroup of A. A group T is said to be a *torsion group* if every element of T has finite order. For every abelian group G, the set $\mathrm{Tor}\,(G)$ of finite-order elements is a subgroup T of G, called the *torsion subgroup* $T \leqslant G$. This subgroup of G is characteristic.

EXERCISE 11.10. Every finitely generated abelian torsion group is finite.

THEOREM 11.11 (classification of abelian groups). *Suppose that A is a finitely generated abelian group. Then there exist an integer $r \geqslant 0$, and k-tuples of prime numbers $(p_1, \ldots, p_k)$ and natural numbers $(m_1, \ldots, m_k)$, for which*

$$A \simeq \mathbb{Z}^r \times \mathbb{Z}_{p_1^{m_1}} \times \cdots \times \mathbb{Z}_{p_k^{m_k}}\,. \tag{11.2}$$

Here $p_1 \leqslant p_2 \leqslant \ldots \leqslant p_k$, and whenever $p_i = p_{i+1}$, we have $m_i \geqslant m_{i+1}$. Furthermore, the number r, and the k-tuples $(p_1, \ldots, p_k)$ and $(m_1, \ldots, m_k)$ are uniquely determined by A.

PROOF. By Theorem 11.7, A is isomorphic to the direct product of finitely many cyclic groups

$$C_1 \times \ldots C_r \times C_{r+1} \times \ldots \times C_n,$$

where C_i is infinite cyclic for $i \leqslant r$ and finite cyclic for $i > r$.

EXERCISE 11.12. (Chinese remainder theorem) $\mathbb{Z}_s \times \mathbb{Z}_t \cong \mathbb{Z}_{st}$ if and only if the numbers s, t are coprime.

In view of this exercise, we can split every finite cyclic group C_i as a direct product of cyclic groups whose orders are prime powers. This proves existence of the decomposition (11.2).

We now consider the uniqueness part of the theorem. We first note that

$$\mathrm{Tor}\,(A) = C_{r+1} \times \ldots \times C_n,$$

which implies that

$$C_1 \times \ldots \times C_r \simeq \mathbb{Z}^r \simeq A/\mathrm{Tor}\,(A).$$

Since the subgroup $\mathrm{Tor}\,(A)$ is characteristic in A, it follows that the number r is uniquely determined by A.

Thus, in order to prove uniqueness of p_i's and m_i's it suffices to assume that A is finite. Since the primes p_i are the prime divisors of the order of A, the uniqueness question reduces to the case when $|A| = p^\ell$, i.e., when $A = A(p)$ is an abelian p-group. Suppose that A is an abelian p-group and

$$A \cong \mathbb{Z}_{p^{m_1}} \times \cdots \times \mathbb{Z}_{p^{m_k}}, \quad m_1 \geqslant \ldots \geqslant m_k.$$

Set $m = m_1$ and let $m_1 = m_2 = \ldots = m_d > m_{d+1}$. Clearly, the number p^m is the largest order of an element of A. The subgroup A_m of A generated by elements of this order is clearly characteristic and equals the d-fold direct product of copies of $\mathbb{Z}_{p^m}$,

$$\mathbb{Z}_{p^{m_1}} \times \cdots \times \mathbb{Z}_{p^{m_d}}$$

in the above factorization of A. Hence, the number m_k and the number d depend only on the group A. We then divide A by A_m and proceed by induction. □

EXERCISE 11.13. The number r equals the rank of a maximal free abelian subgroup of A.

We will refer to the number r as the *free rank* of the abelian group A, in order to distinguish it from the notion of rank in Definition 4.1. Theorem 11.7 implies that each finitely generated abelian group is isomorphic to a direct sum of finitely many cyclic groups C_i, which are unique up to an isomorphism.

DEFINITION 11.14. Generators of cyclic subgroups C_i such that

$$A = \oplus_{i=1}^{s} C_i$$

will be called *standard generators* of A. (These generators, of course, are not uniquely determined by A.)

Below are several immediate corollaries of Theorem 11.7.

COROLLARY 11.15. *Each finite abelian group A is isomorphic to the direct product of abelian p-groups:*

$$A \simeq A(p_1) \times \ldots A(p_k),$$

where $p_1, \ldots, p_k$ are the prime divisors of $|A|$.

COROLLARY 11.16. *Every finitely generated abelian group G is* polycyclic, *i.e., G possesses a finite descending series*

$$G = N_0 \geqslant N_1 \geqslant \ldots \geqslant N_n \geqslant N_{n+1} = \{1\}\,, \tag{11.3}$$

such that every quotient N_i/N_{i+1} is cyclic.

COROLLARY 11.17. *Every finitely generated abelian group A is isomorphic to the direct product $F \times \mathrm{Tor}\,(A)$, where F is a free abelian group.*

COROLLARY 11.18. *A finitely generated abelian group is free abelian if and only if it is torsion-free.*

EXERCISE 11.19. 1. Show that the torsion-free abelian group $\mathbb{Q}$ is not a free abelian group.

2. Show that the image of the free abelian group F in A is not a characteristic subgroup of A (unless $A \simeq F$ or $A = \mathrm{Tor}\,(A)$).

COROLLARY 11.20. *Let G be an abelian group generated by n elements. Then every subgroup H of G is finitely generated (with $\leqslant n$ generators).*

PROOF. Theorem 11.3 implies that there exists an epimorphism $\phi : \mathbb{Z}^n \to A$. Let $A := \phi^{-1}(H)$. Then, by Theorem 11.6, the subgroup A is free of rank $m \leqslant n$. Therefore, H is also m-generated. □

REMARK 11.21. Groups where every subgroup is finitely generated are called *noetherian*. We will see that all polycyclic groups has this property; we will discuss noetherian groups in more detail in §11.8.

Exercise 11.22. Construct an example of a finitely generated abelian group G and a subgroup $H \leqslant G$, such that there is no direct product decomposition $G = F \times \mathrm{Tor}\,(G)$ for which $H = (F \cap H) \times (\mathrm{Tor}\,(G) \cap H)$. Hint: Take $G = \mathbb{Z} \times \mathbb{Z}_2$ and H infinite cyclic.

Exercise 11.23. Let F be a free abelian group of rank n and $B = \{x_1, ..., x_n\}$ be a generating set of F. Then B is a basis of F. Conclude that n equals the minimal cardinality of all generating sets of F. Thus, the notion of *rank* for (finitely generated) free abelian groups agrees with the notion of rank introduced in the beginning of §4.1.

The classification of finitely generated abelian groups allows one find a simple geometric model for such groups:

Lemma 11.24. *Every finitely generated abelian group G of free rank n admits a geometric (in the sense of Definition 3.62) action on the Euclidean space $\mathbb{E}^n$, such that every element of G acts as a translation. In particular, G is quasiisometric to $\mathbb{E}^n$.*

Proof. Let $G = \mathbb{Z}^n \times \mathrm{Tor}\,(G)$. We let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ denote a basis of $\mathbb{Z}^n$, and let $\mathbb{R}^n$ be the vector space with the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. We equip $\mathbb{R}^n$ with the standard Euclidean metric where the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal and let $\mathbb{E}^n$ be the corresponding Euclidean n-space. Then every $g = \sum_{i=1}^n a_i \mathbf{e}_n \in \mathbb{Z}^n$ acts on $\mathbb{E}^n$ as the translation by the vector $\mathbf{a} = (a_1, \ldots, a_n)$. This action of $\mathbb{Z}^n$ extends to G by declaring that every $g \in \mathrm{Tor}\,(G)$ acts on $\mathbb{E}^n$ trivially. We leave it to the reader to check that this action is geometric and the quotient $\mathbb{E}^n/G$ is the n-torus T^n. □

11.3. Automorphisms of $\mathbb{Z}^n$

Theorem 11.25. *The group of automorphisms of $\mathbb{Z}^n$ is isomorphic to $GL(n, \mathbb{Z})$.*

Proof. Consider the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathbb{Z}^n$, where

$$\mathbf{e}_i = (\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, 1, \underbrace{0, \ldots, 0}_{n-i \text{ times}}).$$

Let $\phi : \mathbb{Z}^n \to \mathbb{Z}^n$ be an automorphism. Set

$$\phi(\mathbf{e}_i) = \sum_{j=1}^{n} m_{ij} \mathbf{e}_j. \tag{11.4}$$

We, thus, obtain a map $\mu : \phi \mapsto M_\phi = (m_{ij})$, where M_ϕ is a matrix with integer entries. We leave it to the reader to check that $\mu(\phi \circ \psi) = M_\phi M_\psi$. It follows that $\mu(\phi) \in GL(n, \mathbb{Z})$ for every $\phi \in Aut(\mathbb{Z}^n)$.

Given a matrix $M \in GL(n, \mathbb{Z})$, we define an endomorphism

$$\phi : \mathbb{Z}^n \to \mathbb{Z}^n,$$

using the equation (11.4). Since the map $\nu : M \mapsto \phi$ respects the composition, it follows that $\nu : GL(n, \mathbb{Z}) \to Aut(\mathbb{Z}^n)$ is a homomorphism and $\mu = \nu^{-1}$. □

Below we establish several properties of automorphisms of free abelian groups that are interesting by themselves and will also be useful in Chapter 12, in the proof of the Milnor–Wolf Theorem.

Lemma 11.26. *Let* $\mathbf{v} = (v_1, .., v_n) \in G = \mathbb{Z}^n$ *be a vector with* $\gcd(v_1, \ldots, v_n) = 1$*. Then* $H = G/\langle \mathbf{v} \rangle$ *is free abelian of rank* $n-1$*. Moreover, there exists a basis* $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \mathbf{v}\}$ *of* G *such that* $\{\mathbf{y}_1 + \langle \mathbf{v} \rangle, \ldots, \mathbf{y}_{n-1} + \langle \mathbf{v} \rangle\}$ *is a basis of* H*.*

Proof. First, let us show that the group H is free abelian; since this group is finitely generated, it suffices to verify that the quotient group is torsion-free. We will use the notation $x \mapsto \bar{x}$ for the quotient map $G \to H$.

Let $u \in G$ be such that $\bar{u} \in H$ has finite order k. Then $ku \in \langle \mathbf{v} \rangle$, i.e., $ku = m\mathbf{v}$ for some $m \in \mathbb{Z}$. Since $\gcd(v_1, \ldots, v_n) = 1$, it follows that $k|m$ and, hence, $u \in \langle \mathbf{v} \rangle$, $\bar{u} = \bar{1}$.

Thus, $H = \mathbb{Z}^n/\langle \mathbf{v} \rangle$ is torsion-free, and, hence, is free abelian of finite rank m. Next, the homomorphism $G \to H$ extends to a surjective linear map $\mathbb{R}^n \to \mathbb{R}^m$, whose kernel is the line spanned by v. Therefore, $m = n - 1$.

Let $\{\bar{x}_1, \ldots, \bar{x}_{n-1}\}$ be a basis on H. The map

$$\bar{x}_i \mapsto x_i, i = 1, \ldots, n-1,$$

extends to a group monomorphism $H \to G$; thus, the set $\{x_1, \ldots, x_{n-1}, v\}$ generates $\mathbb{Z}^n$. It follows that $\{x_1, \ldots, x_{n-1}, v\}$ is a basis of G. □

Lemma 11.27. *If a matrix* M *in* $GL(n, \mathbb{Z})$ *has all eigenvalues equal to* 1 *then there exists a finite ascending series of subgroups*

$$\{1\} = \Lambda_0 \leqslant \Lambda_1 \leqslant \cdots \leqslant \Lambda_{n-1} \leqslant \Lambda_n = \mathbb{Z}^n$$

such that $\Lambda_i \simeq \mathbb{Z}^i$*,* $\Lambda_{i+1}/\Lambda_i \simeq \mathbb{Z}$ *for all* $i \geqslant 0$*,* $M(\Lambda_i) = \Lambda_i$ *and* M *acts on* Λ_{i+1}/Λ_i *as the identity.*

Proof. Since M has eigenvalue 1, there exists a vector $v = (v_1, .., v_n) \in \mathbb{Z}^n$ such that $\gcd(v_1, .., v_n) = 1$ and $Mv = v$. Then M induces an automorphism of $H = \mathbb{Z}^n/\langle v \rangle \simeq \mathbb{Z}^{n-1}$ and the matrix $\bar{M}$ of this automorphism has only 1 as an eigenvalue. This follows immediately when writing the matrix of the automorphism M with respect to a basis $\{x_1, x_2, \ldots, x_{n-1}, v\}$ of $\mathbb{Z}^n$ as in Lemma 11.26 and looking at the characteristic polynomial. Now, lemma follows from induction on n. □

Lemma 11.28. *Let* $M \in GL(n, \mathbb{Z})$ *be a matrix such that each eigenvalue of* M *has the absolute value* 1*. Then all eigenvalues of* M *are roots of unity.*

Proof. We owe the following proof to Mark Sapir, it replaces an earlier, more complicated, argument. Recall that for each $n \times n$ matrix A with the eigenvalues $\mu_1, \ldots, \mu_n$ (here and below, we repeat the eigenvalues if necessary, according to their multiplicities) the characteristic polynomial $p_A(t)$ equals

$$\sum_{i=0}^{n} a_{n-i} t^i,$$

where, by Vieta's formulae,

$$a_i = \det(A)(-1)^n \sigma_i(\mu_1, \ldots, \mu_n),$$

and σ_i is the ith *elementary symmetric polynomial*:

$$\sigma_i(x_1, \ldots, x_n) = \sum_{1 \leq j_1 < \ldots < j_i \leq n} x_{j_1} \ldots x_{j_i}.$$

We now return to the integer square matrix M as in lemma and let $\lambda_1, \ldots, \lambda_n$ denote its eigenvalues. Consider the sequence of matrices $M^k, k \in \mathbb{N}$. The eigenvalues of M^k are $\lambda_1^k, \ldots, \lambda_n^k$, which, by the assumption, all have the absolute value 1. Therefore, the coefficients of the characteristic polynomials $p_k(t) := p_{M^k}(t)$ of M^k are uniformly bounded, independently on k. Since the matrices M^k belong to $GL(n, \mathbb{Z})$, there are only finitely many distinct characteristic polynomials of the matrices M^k. Hence, there exists an infinite sequence $k_1 < k_2 < k_3 < \ldots$, such that

$$p_{k_1}(t) = p_{k_2}(t) = p_{k_3}(t) = \ldots .$$

It follows that there are distinct members of this sequence, $q < r$, such that

$$\lambda_1^q = \lambda_1^r, \ldots, \lambda_n^q = \lambda_n^r.$$

Hence, for each $i = 1, \ldots, n$

$$\lambda_i^{r-q} = 1,$$

which means that each eigenvalue of M is a root of unity. □

LEMMA 11.29. *If a matrix M in $GL(n, \mathbb{Z})$ has one eigenvalue λ of absolute value at least 2 then there exists a vector $\mathbf{v} \in \mathbb{Z}^n$ such that the following map is injective:*

$$\begin{array}{rcc} \Phi : \bigoplus\limits_{n \in \mathbb{Z}_+} \mathbb{Z}_2 & \longrightarrow & \mathbb{Z}^n \\ \Phi : (s_n)_n & \mapsto & s_0 v + s_1 M\mathbf{v} + \ldots + s_n M^n \mathbf{v} + \ldots . \end{array} \tag{11.5}$$

PROOF. The matrix M defines an automorphism $\varphi : \mathbb{Z}^n \to \mathbb{Z}^n$, $\varphi(\mathbf{v}) = M\mathbf{v}$. The dual map φ^* has the matrix M^T in the dual canonical basis. Therefore, it also has the eigenvalue λ and, hence, there exists a linear form $f : \mathbb{C}^n \to \mathbb{C}$ such that $\varphi^*(f) = f \circ \varphi = \lambda f$.

Take $\mathbf{v} \in \mathbb{Z}^n \setminus \operatorname{Ker} f$. Assume that the map Φ is not injective. It follows that there exist some $(t_n)_n$, $t_n \in \{-1, 0, 1\}$, such that

$$t_0 \mathbf{v} + t_1 M\mathbf{v} + \ldots + t_n M^n \mathbf{v} + \ldots = 0.$$

Let N be the largest integer such that $t_N \neq 0$. Then

$$M^N \mathbf{v} = r_0 \mathbf{v} + r_1 M\mathbf{v} + \ldots + r_{N-1} M^{N-1} \mathbf{v}$$

where $r_i \in \{-1, 0, 1\}$. By applying f to the equality we obtain

$$\left(r_0 + r_1 \lambda + \cdots + r_{N-1} \lambda^{N-1}\right) f(\mathbf{v}) = \lambda^N f(\mathbf{v}),$$

whence

$$|\lambda|^N \leqslant \sum_{i=1}^{N-1} |\lambda|^i = \frac{|\lambda|^N - 1}{|\lambda| - 1} \leqslant |\lambda|^N - 1,$$

a contradiction. □

11.4. Nilpotent groups

Recall that $[x, y] = xyx^{-1}y^{-1}$ is the commutator of the elements x, y in a group G and that $x^g := gxg^{-1}$ is the g-conjugate of x in G. We begin the discussion of nilpotent groups with some useful commutator identities:

LEMMA 11.30. *Let $(G, \cdot)$ be a group and x, y, z elements in G. The following identities hold:*

(1) $[x, y]^{-1} = [y, x]$;

(2) $[x^{-1}, y] = [x^{-1}, [y, x]]\, [y, x]$;

(3) $[x, yz] = [x, y]\, [y, [x, z]]\, [x, z]$;

(4) $[xy, z] = [x, [y, z]]\, [y, z]\, [x, z]$.

(5) $[x, y]^g = [x^g, y^g]$.

PROOF. (1) and (2) are immediate, (4) follows from (3) and (1). It remains to prove (3). Since $[y, [x, z]]\, [x, z] = y[x, z]y^{-1}$ we have that

$$[x, y]\, [y, [x, z]]\, [x, z] = xyx^{-1}[x, z]y^{-1} = xyzx^{-1}z^{-1}y^{-1} = [x, yz]\, .$$

We leave the last identity as an exercise to the reader. □

NOTATION 11.31. For every $x_1, \ldots, x_n$ in a group G we denote by $[x_1, \ldots, x_n]$ the *n-fold left-commutator*

$$[[[x_1, x_2], \ldots, x_{n-1}], x_n].$$

We declare that 1-fold left commutator $[x]$ is simply x.

EXERCISE 11.32. $[x_1, \ldots, x_n]^g = [x_1^g, \ldots, x_n^g]$.

Recall that for subsets A, B in a group G, $[A, B]$ denotes the subgroup of G generated by all commutators $[a, b], a \in A, b \in B$. In what follows we also use:

NOTATION 11.33. Given n subgroups $H_1, H_2, \ldots, H_n$ in a group G we denote by $[H_1, \ldots, H_n]$ the subgroup $[\ldots[H_1, H_2], \ldots, H_n] \leqslant G$.

We define *the lower central series* of a group G,

$$C^1G \trianglerighteq C^2G \trianglerighteq \ldots \trianglerighteq C^nG \trianglerighteq \ldots,$$

inductively by:

$$C^1G = G\, ,\ C^{n+1}G = [C^nG, G]\, .$$

In particular, each C^kG is a *characteristic subgroup* of G. We will see later on (Proposition 11.62) that

$$[C^iG, C^kG] \leqslant C^{i+k}G.$$

Note that $C^2G = [G, G] = G'$ is the commutator subgroup, or the *derived subgroup*, of G.

EXERCISE 11.34. 1. The subgroup $C^kG \leqslant G$ is normal in G.
2. $C^{n+1}G = [G, C^nG]$.

DEFINITION 11.35. A group G is called *k-step nilpotent* if $C^{k+1}G = \{1\}$. The minimal k for which G is k-step nilpotent is called the (nilpotency) *class* of G.

EXAMPLES 11.36. (1) Every nontrivial abelian group is nilpotent of class 1.

(2) The group $\mathcal{U}_n(\mathbb{K})$ of upper triangular $n \times n$ matrices with 1 on the diagonal and entries in a ring $\mathbb{K}$, is nilpotent of class $n-1$ (see Exercise 11.38).

(3) The *Heisenberg group*

$$H_{2n+1}(\mathbb{K}) = \left\{ \begin{pmatrix} 1 & x_1 & x_2 & \ldots & \ldots & x_n & z \\ 0 & 1 & 0 & \ldots & \ldots & 0 & y_n \\ 0 & 0 & 1 & \ldots & \ldots & 0 & y_{n-1} \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & 1 & 0 & y_2 \\ 0 & 0 & \ldots & \ldots & 0 & 1 & y_1 \\ 0 & 0 & \ldots & \ldots & \ldots & 0 & 1 \end{pmatrix} ;\ x_1, \ldots, x_n, y_1, \ldots, y_n, z \in \mathbb{K} \right\}$$

is nilpotent of class 2.

Taking $\mathbb{K} = \mathbb{Z}$, we obtain the *integer Heisenberg group*

$$H_{2n+1}(\mathbb{Z}).$$

The group $H_{2n+1}(\mathbb{Z})$ is finitely generated; we can take as generators the elementary matrices $N_{ij} = I + E_{ij}$ with

$$(i,j) \in \{(1,2), \ldots, (1, n+1), (2,n), \ldots, (n+1, n)\}.$$

All the groups $H_{2n+1}(\mathbb{K})$ are nilpotent of class 2. Indeed $C^2 H_{2n+1}(\mathbb{K})$ is the subgroup $x_i = y_i = 0, i = 1, \ldots, n$.

(4) We will see later (Proposition 12.1) that semidirect products $\mathbb{Z}^n \rtimes_A \mathbb{Z}$ with the matrix $A \in \mathcal{U}_n(\mathbb{Z})$, are nilpotent.

EXERCISE 11.37. Which of the permutation groups S_n are nilpotent? Which of these groups are solvable?

EXERCISE 11.38. The goal of this exercise is to prove that the group $\mathcal{U}_n(\mathbb{K})$ is nilpotent of class $n-1$.

Let $\mathcal{U}_{n,k}(\mathbb{K})$ be the subset of $\mathcal{U}_n(\mathbb{K})$ formed by matrices (a_{ij}) such that $a_{ij} = \delta_{ij}$ for $j < i + k$. Note that $\mathcal{U}_{n,1}(\mathbb{K}) = \mathcal{U}_n(\mathbb{K})$.

(1) Prove that for every $k \geqslant 1$ the map

$$\begin{array}{rcl} \varphi_k : \mathcal{U}_{n,k}(\mathbb{K}) & \to & \left(\mathbb{K}^{n-k}, +\right) \\ A = (a_{i,j}) & \mapsto & (a_{1,k+1}, a_{2,k+2}, \ldots, a_{n-k,n}) \end{array}$$

is a homomorphism. Deduce that $(\mathcal{U}_{n,k}(\mathbb{K}))' \subset \mathcal{U}_{n,k+1}(\mathbb{K})$ and that $\mathcal{U}_{n,k+1}(\mathbb{K}) \lhd \mathcal{U}_{n,k}(\mathbb{K})$ for every $k \geqslant 1$.

(2) Let E_{ij} be the matrix with all entries 0 except the (i,j)–entry, which is equal to 1. Consider the triangular matrix $T_{ij}(a) = I + aE_{ij}$.

Deduce from (1), using induction, that $\mathcal{U}_{n,k}$ is generated by the set

$$\{T_{ij}(a) \mid j \geqslant i + k, a \in \mathbb{R}\}.$$

(3) Prove that for every three distinct numbers i, j, k in $\{1, 2, \ldots, n\}$

$$[T_{ij}(a), T_{jk}(b)] = T_{ik}(ab)\,,\ [T_{ij}(a), T_{ki}(b)] = T_{kj}(-ab)\,,$$

and that for all quadruples of distinct numbers i, j, k, ℓ,

$$[T_{ij}(a), T_{k\ell}(b)] = I\,.$$

(4) Prove that $C^k\mathcal{U}_n(\mathbb{K}) \leqslant \mathcal{U}_{n,k+1}(\mathbb{K})$ for every $k \geqslant 0$. Deduce that $\mathcal{U}_n(\mathbb{K})$ is nilpotent.

Remark. All the arguments above work also when all matrices have integer entries. In this case (2) implies that $\mathcal{U}_n(\mathbb{Z})$ is generated by $\{T_{ij}(1) \mid j \geqslant i+1\}$.

EXERCISE 11.39. The group $\mathcal{U}_n(\mathbb{K})$ is torsion-free provided that $\mathbb{K}$ has zero characteristic.

A combination of deep theorems by Mal'cev and Ado shows that each finitely generated torsion-free nilpotent group embeds in $\mathcal{U}_n$, for some n:

THEOREM 11.40 (A. I. Mal'cev [**Mal49b**]). *Every finitely generated torsion-free nilpotent group Γ of class k embeds as a uniform lattice in a simply-connected nilpotent Lie group N of class k. Furthermore, the group N and the embedding $\Gamma \to N$ are unique up to an isomorphism.*

THEOREM 11.41 (Ado–Engel theorem). *Every simply-connected nilpotent Lie group N embeds into $\mathcal{U}_n(\mathbb{R})$ for some n.*

REMARK 11.42. We are attributing this theorem to Ado and Engel, but, as usual, the history is more complicated. Ado (see Theorem 3.55) proved linearity of finite-dimensional real Lie algebras; in the case of nilpotent Lie algebras \, the faithful linear representation

$$r : \mathfrak{n} \to End(\mathbb{R}^n) = gl_n(\mathbb{R})$$

sends each element of $\mathfrak{n}$ to a nilpotent linear transformation, i.e., a linear endomorphism A such that $A^k = 0$ for some k. Much earlier, Engel sketched a proof, details of which were written by his student, Umlauf in his PhD thesis [**Uml91**], that any subalgebra of $gl_n(\mathbb{R})$ consisting entirely of nilpotent endomorphisms is conjugate to a subalgebra of the algebra of upper–traingular matrices with zeroes on the diagonal, we refer to [**FH94**, Theorem 9.9] for a modern proof, cf. Theorem 12.45. Thus, we can assume that $r(\mathfrak{n})$ is contained in the algebra u_n of such matrices. Now, if N is a simply-connected Lie group with the Lie algebra $\mathfrak{n}$, then, via exponentiation, r induces a representation $\rho : N \to \mathcal{U}_n(\mathbb{R})$, which has to be faithful, since the exponential map $\exp : u_n \to \mathcal{U}_n(\mathbb{R})$ is bijective. Note that this proof, in particular, implies that N is contractible, since its exponential map has to be a homeomorphism as well.

Since simply-connected nilpotent Lie groups are contractible, it follows that each finitely generated torsion-free nilpotent group Γ has type **F**, i.e., admits a finite $K(\Gamma, 1)$, namely, N/Γ. In particular, the cohomological dimension $cd(\Gamma)$ of Γ is at most $n = \dim(N)$. Since N/Γ is a closed orientable manifold, $H^n(\Gamma) \cong H^n(N/\Gamma) \cong \mathbb{Z}$. Therefore,

$$cd(\Gamma) = \dim(N).$$

COROLLARY 11.43. *Each finitely generated torsion-free nilpotent group is residually finite.*

We will see later on, Theorem 11.76, that all polycyclic groups are residually finite, which shows residual finiteness of all finitely generated nilpotent groups.

We now proceed with establishing some basic properties of lower central series and nilpotent groups.

LEMMA 11.44. *If S is a generating set of a group G (not necessarily nilpotent), then for every k the subgroup C^kG is generated by the k-fold left commutators in S and their inverses, together with $C^{k+1}G$.*

PROOF. We prove the assertion by induction on k. For $k = 1$ the statement is clear, since 1-fold commutators of elements of S are just elements of S. Assume that the assertion holds for some $k \geqslant 1$ and consider $C^{k+1}G$.

By definition, $C^{k+1}G$ is generated by all commutators $[c_k, g]$ with $c_k \in C^kG$ and $g \in G$. The induction hypothesis and normality of $C^{k+1}G$ in G imply that $c_k = \ell_1^{\pm 1} \cdots \ell_m^{\pm 1} x$, where $m \in \mathbb{N}$, ℓ_i are k-fold left commutators in S and $x \in C^{k+1}G$.

According to Lemma 11.30, (4),

$$[c_k, g] = [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1} x, g] = [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, [x, g]][x, g][\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, g] .$$

The first two factors are in $C^{k+2}G$, so it remains to deal with the third.

We write $g = s_1 \cdots s_r$, where $s_i \in S$, and we prove that $[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_r]$ is a product of $(k+1)$–fold left commutators in S and their inverses, and of elements in $C^{k+2}G$; our proof is another induction, this time on $m + r \geqslant 2$.

For the case $m + r = 2$ it suffices to note that $[\ell^{-1}, s] = [\ell^{-1}, [s, \ell]][s, \ell]$. The first factor is in $C^{k+2}G$, the second is the inverse of a $(k+1)$–fold left commutator.

Assume that the statement is true for $m + r = n \geqslant 2$. We now prove it for $m + r = n + 1$.

Suppose that $m \geqslant 2$. We apply Lemma 11.30, (4), and obtain that

$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \ldots s_r] = [\ell_1^{\pm 1} \cdots \ell_{m-1}^{\pm 1}, [\ell_m^{\pm 1}, g]]\, [\ell_m^{\pm 1}, s_1 \cdots s_r]\, [\ell_1^{\pm 1} \cdots \ell_{m-1}^{\pm 1}, s_1 \ldots s_r] .$$

The first factor is in $C^{k+2}G$, and for the second and the third the induction hypothesis applies.

Likewise, if $r \geqslant 2$ then we apply Part 3 of Lemma 11.30, and write

$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_r] =$$

$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_{r-1}]\, [s_1 \cdots s_{r-1}, [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_r]]\, [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_r] . \quad \square$$

COROLLARY 11.45. *If G is nilpotent, then C^nG is generated by k-fold left commutators in S and their inverses, where $k \geqslant n$. In particular, if G is finitely generated, so is each group C^nG.*

PROOF. Suppose that $C^{m+1}G = \{1\}$. Then C^mG is generated by the m-fold left commutators in S and their inverses. By applying the reverse induction in n, each C^nG is generated by the set of all k-fold left commutators of elements of S and their inverses, $k \geqslant n$. $\square$

Thus, if G is finitely generated, each quotient $C^iG/C^{i+1}G$ is a finitely generated abelian group and, hence, we define two important invariants of finitely generated nilpotent groups:

DEFINITION 11.46. Let G be a finitely generated nilpotent group of class k. Let m_i denote the free rank of the abelian group $C^iG/C^{i+1}G$; define the *Hirsch length* of G

$$h(G) = \sum_{i=1}^{k} m_i$$

and the *homogeneous dimension* of G,

$$d(G) = \sum_{i=1}^{k} i m_i.$$

In the next chapter we will give a geometric interpretation of the number $d(G)$. For now, we note that for torsion-free nilpotent groups G, the Hirsch length $h(G)$ equals the dimension of the simply-connected nilpotent group N into which G embeds as a uniform lattice, [**Mal49b**]; hence, $h(G)$ is the cohomological dimension of G in this case.

Definition 11.47. Given natural numbers k and m, the *k-step m-generated free nilpotent* group is the quotient $N_{m,k}$ of the free group of rank m, F_m, by the normal subgroup $C^{k+1}F_m$.

We will refer to the images of the free generators of F_m as *free nilpotent generators* of $N_{m,k}$.

Note that the free abelian group of rank m is the 1-step m-generated free nilpotent group.

A consequence of Proposition 4.22 is the following.

Proposition 11.48 (Universality property of free nilpotent groups). *For every k–step nilpotent group G equipped with a generating set $X = \{x_1, ..., x_m\}$, there exists an epimorphism $\psi : N_{m,k} \to G$ sending free nilpotent generators $s_1, ..., s_m$ of $N_{m,k}$ to the generators $x_1, ..., x_m$ respectively. In particular, every k–step m-generated nilpotent group is a quotient of $N_{m,k}$.*

Proof. Take a generating set X of a k–step nilpotent group G, such that X has cardinality m. The homomorphism $\phi : F(S) = F_m \to G$, sending $s_i \mapsto x_i, i = 1, \ldots, m$, defined in Proposition 4.22 contains $C^{k+1}F(S)$ in its kernel. Therefore, ϕ projects to an epimorphism $\psi : N_{m,k} \to G$ as required. □

So far, we were describing nilpotent groups "from the top–down", starting from the group G and then looking at the chain of decreasing subgroups. It is also useful to have a "bottom-up" description of nilpotent groups, which we present below.

Recall that the center of a group H is denoted $Z(H)$. Given a group G, consider the sequence of normal subgroups $Z_i(G) \lhd G$ defined inductively by:

- $Z_0(G) = \{1\}$.
- If $Z_i(G) \lhd G$ is defined and $\pi_i : G \to G/Z_i(G)$ is the quotient map, then
$$Z_{i+1}(G) = \pi_i^{-1}\left(Z(G/Z_i(G))\right).$$

Note that $Z_{i+1}(G)$ is normal in G, as the preimage of a normal subgroup of a quotient of G. In particular,

$$Z_{i+1}(G)/Z_i(G) \cong Z(G/Z_i(G)).$$

Proposition 11.49. *The group G is k-step nilpotent if and only if $Z_k(G) = G$.*

Proof. Assume that G is nilpotent of class k. We prove by induction on $i \geqslant 0$ that $C^{k+1-i}G \leqslant Z_i(G)$. For $i = 0$ we have equality. Assume that

$$C^{k+1-i}G \leqslant Z_i(G).$$

For every $g \in C^{k-i}G$ and every $x \in G$, $[g, x] \in C^{k+1-i}G \leqslant Z_i(G)$, whence $gZ_i(G)$ is in the center of $G/Z_i(G)$, i.e., $g \in Z_{i+1}(G)$. Thence, the inclusion follows by induction. For $i = k$ the inclusion becomes $C^1G = G \leqslant Z_k(G)$, hence, $Z_k(G) = G$.

Conversely, assume that there exists k such that $Z_k(G) = G$. We prove by induction on $j \geqslant 1$ that $C^jG \leqslant Z_{k+1-j}(G)$. For $j = 1$ the two are equal. Assume that the inclusion is true for j. The subgroup $C^{j+1}G$ is generated by commutators $[c, g]$ with $c \in C^jG$ and $g \in G$. Since $c \in C^jG \leqslant Z_{k+1-j}(G)$, by the definition of $Z_{k+1-j}(G)$, the element c commutes with g modulo $Z_{k-j}(G)$, equivalently $[c, g] \in Z_{k-j}(G)$. This implies that $[c, g] \in Z_{k-j}(G)$. It follows that $C^{j+1}G \leqslant Z_{k-j}(G)$.

For $j = k+1$ this gives $C^{k+1}G \leqslant Z_0(G) = \{1\}$, hence G is k-step nilpotent. □

Definition 11.50. The ascending series

$$Z_0(G) = \{1\} \lhd Z_1(G) \lhd \ldots \lhd Z_i(G) \lhd Z_{i+1}(G) \lhd \ldots$$

of normal subgroups of G is called the *upper central series* of G.

In view of Proposition 11.49, a group G is nilpotent if and only if its upper central series is finite, and its nilpotency class is the minimal k such that $Z_k(G) = G$.

Exercise 11.51. Any central coextension of a nilpotent group is again nilpotent.

Remark 11.52. Yet another equivalent definition a nilpotent group, is to require that the group admits a finite normal series

$$\{1\} = \Gamma_0 \lhd \ldots \Gamma_i \lhd \Gamma_{i+1} \lhd \ldots \Gamma_{n-1} \lhd \Gamma_n = G,$$

such that $\Gamma_{i+1}/\Gamma_i \leqslant Z(G/\Gamma_i)$, or, equivalently, $[G, \Gamma_{i+1}] \leqslant \Gamma_i$. In particular, the quotients Γ_{i+1}/Γ_i are abelian for each i. We will need only the fact that existence of such normal series implies that G is n-step nilpotent. Indeed, the condition $\Gamma_{i+1}/\Gamma_i \leqslant Z(G/\Gamma_i)$ implies that $\Gamma_i \leqslant Z_i(G)$ for every i. In particular, $G = Z_n(G)$. Now, the assertion follows from Proposition 11.49. We refer to [**Hal76**, Theorem 10.2.2] for further details.

The following example shows that the difference between lower and upper central series of groups can be quite substantial, in particular, $C^{k+1-i}G \leqslant Z_i(G)$ could be of infinite index:

Example 11.53. We start with the integer Heisenberg group H; it is 2-step nilpotent, $C^2H = H' = Z(H) \cong \mathbb{Z}$. Next, take $G = H \times \mathbb{Z}$. Then G is still 2-step nilpotent, but now $C^2G = C^2H \cong \mathbb{Z}$, while $Z(G) \cong \mathbb{Z}^2$.

Exercise 11.54. Construct an example of a 2-step nilpotent group G with torsion-free center, such that G/C^2G is not torsion-free.

The following useful lemma is a converse to Corollary 11.45:

Lemma 11.55. *Let S be a generating set of a group G. Suppose that all $N+1$-fold commutators $[s_1, \ldots, s_{N+1}]$ of elements of S are trivial. Then G is N-step nilpotent.*

Proof. Let G_n be the subgroup of Γ generated by the n-fold commutators $y_n = [s_1, \ldots, s_n]$ of generators $s_i \in S$ of the group G. For every generator x of G and every generator y_n of G_n we have:

$$[y_n, x] = y_n x y_n^{-1} x^{-1} \in G_{n+1} \leqslant G_n.$$

Since $y_n \in G_n$, it follows that $x y_n^{-1} x^{-1} \in G_n$ which implies that G_n is a normal subgroup of G.

We claim that for every n, G_{n-1}/G_n embeds (under the map induced by inclusion $G_{n-1} \hookrightarrow G$) in the center of G/G_n. To simplify the notation, we will regard G_{n-1}/G_n as a subgroup of G/G_n. The proof of this statement is the reverse induction on n.

The subgroup G_{N+1} is trivial, hence it is contained in the center of G. Suppose that the assertion holds for $n = k+1$, we will now prove it for $n = k$. To show that G_{k-1}/G_k is in the center of G/G_k it is enough to verify that for all elements $\bar{z}$ and

$\bar{w}$ of generating sets of G_{n-1}/G_n and G/G_n, respectively, the commutator $[\bar{z}, \bar{w}]$ is trivial.

The group G is generated by the set S, the group G_{n-1} is generated by the $n-1$-fold commutators y_{n-1} of elements $x \in S$. Thus, the groups G_{n-1}/G_n and G/G_n are generated by the projections $\bar{x}, \bar{y}_{n-1}$ of the elements x, y_{n-1}. By definition of G_n we have: $[y_{n-1}, x] \in G_n$, thus, dividing by G_n, we obtain $[\bar{y}_{n-1}, \bar{x}] = 1$. Thus, $G_{n-1}/G_n \leqslant Z(G/G_n)$ for every n and Lemma follows from Remark 11.52. □

LEMMA 11.56. (1) *Every subgroup of a nilpotent group is nilpotent.*
(2) *If G is nilpotent and $N \lhd G$ then G/N is nilpotent.*
(3) *The direct product of a family of nilpotent groups is again nilpotent.*

PROOF. (1) Let H be a subgroup in a nilpotent group G. Then $C^iH \leqslant C^iG$. Hence, if G is k-step nilpotent then $C^{k+1}H = \{1\}$.

(2) If $\pi : G \to G/N$ is the quotient map, $\pi(C^iG) = C^i(G/N)$.

(3) The assertion follows from the equality

$$C^j(\prod_{i\in I} G_i) = \prod_{i\in I} C^jG_i\ . \quad \square$$

THEOREM 11.57. *Every subgroup of a finitely generated nilpotent group is finitely generated, i.e., finitely generated nilpotent groups are* noetherian.

PROOF. We argue by induction on the class of nilpotency k. For $k = 1$ the group is abelian and the statement is already proven in Corollary 11.20. Assume that the assertion holds for k, let G be a nilpotent group of class $k+1$ and let $H \leqslant G$ be a subgroup. By the induction hypothesis $H_1 = H \cap C^2G$ and $H_2 = H/(H \cap C^2G)$ are both finitely generated. Thus, H fits in the short exact sequence

$$1 \to H_1 \to H \xrightarrow{\pi} H_2 \to 1,$$

where H_1, H_2 are finitely generated. Lemma 4.10 then shows that H is also finitely generated. □

Our next goal is to prove some structural results for nilpotent groups. We begin the "calculus of commutators."

LEMMA 11.58. *If A, B, C are normal subgroups in a group G, then the subgroup $[A, B, C] \leqslant G$ is generated by the commutators $[a, b, c]$ with $a \in A, b \in B, c \in C$.*

PROOF. By the definition, $[A, B, C]$ is generated by the commutators $[k, c]$ with $k \in [A, B]$ and $c \in C$. The element k is a product $t_1 \cdots t_n$, where each t_i is equal either to a commutator $[a, b]$ or to a commutator $[b, a]$, $a \in A, b \in B$.

We prove, by the induction on n, that $[k, c]$ is a product of finitely many commutators $[a, b, c]$ and their inverses. For $n = 1$ we only need to consider the case $[t^{-1}, c]$, where $t = [a, b]$. By Lemma 11.30, (2),

$$[t^{-1}, c] = [c, t]^{t^{-1}} = [c^{t^{-1}}, t] = [c', t] = [a, b, c']^{-1}\,.$$

In the second equality above we applied the identity $\phi([x, y]) = [\phi(x), \phi(y)]$ for the inner automorphism $\phi(x) = x^{t^{-1}}$.

Assume that the statement is true when k is a product of n commutators t_i and consider $k = k_1t$, where t is equal to either a commutator $[a, b]$ or a commutator $[b, a]$, and k_1 is a product of n such commutators. According to Lemma 11.30, (4),

$$[k_1t, c] = [t, c]^{k_1}[k_1, c]\,.$$

Both factors are products of finitely many commutators $[a, b, c]$ and their inverses, by the induction hypothesis and the fact that A, B, C are normal subgroups and, thus, are invariant under conjugation. $\square$

Exercise 11.59. Prove the same result for $[H_1, \dots, H_n]$, where all H_i are normal subgroups of G.

Lemma 11.60 (The Hall identity). *Given a group G and three arbitrary elements x, y, z in G, the following identity holds:*

$$\left[x^{-1}, y, z\right]^x \left[z^{-1}, x, y\right]^z \left[y^{-1}, z, x\right]^y = 1 \,. \tag{11.6}$$

Proof. The factor $\left[x^{-1}, y, z\right]^x$ equals $yxy^{-1}zyx^{-1}y^{-1}xz^{-1}x^{-1}$. The other two factors can be obtained by proper cyclic permutation and a direct calculation shows that all the terms cancel and the product is 1. $\square$

Corollary 11.61. *Assume that A, B, C are normal subgroups in G. Then*

$$[A, B, C] \leqslant [B, C, A][C, A, B] \,. \tag{11.7}$$

The next proposition shows that the lower central series of G is *graded* with respect to commutators:

Proposition 11.62. *Let $C^k G$ be the k-th group in the lower central series of a group G. Then for every $i, j \geqslant 1$*

$$\left[C^i G, C^j G\right] \leqslant C^{i+j} G \,. \tag{11.8}$$

Proof. We prove by induction on $i \geqslant 1$ that for every $j \geqslant 1$, the inclusion (11.8) holds.

For $i = 1$ this follows from the definition of $C^k G$. Assume that the statement is true for i. Consider $j \geqslant 1$ arbitrary.

$$[C^{i+1}G, C^j G] = [C^i G, G, C^j G] \leqslant [G, C^j G, C^i G][C^j G, C^i G, G] \leqslant$$
$$[C^{j+1}G, C^i G][C^{j+i}G, G] = [C^i G, C^{j+1}G][C^{j+i}G, G] \leqslant C^{j+i+1}G \,,$$

since $[C^i G, C^{j+1} G] \leqslant C^{j+i+1} G$ by the induction hypothesis. $\square$

We now prove that, as for abelian groups, all elements of finite order in a finitely generated nilpotent group form a finite subgroup. We will need the following lemma:

Lemma 11.63. *Let G be a nilpotent group of class k. For every $x \in G$ the subgroup H generated by x and $C^2 G$ is a normal subgroup, which is nilpotent of class $\leqslant k - 1$.*

Proof. By normality of $C^2 G$ in G, the subgroup H can be described as

$$H = \{x^m c \mid m \in \mathbb{Z},\, c \in C^2 G\}.$$

For every $g \in G$, and $h \in H$, $h = x^m c$, $ghg^{-1} = x^m [x^{-m}, g] gcg^{-1}$, and, since the last two factors are in $C^2 G$, the whole product is in H. Hence, H is normal in G.

We now prove that $C^2 H \leqslant C^3 G$, which will imply that H is of class $\leqslant k - 1$ and, thereby conclude the proof of lemma.

Let h, h' be two elements in H, $h = x^m c_1$, $h' = x^n c_2$ with $c_i \in C^2 G$. Then, according to Lemma 11.30, (3),

$$[h, h'] = [h, x^n c_2] = [h, x^n]\, [x^n, [h, c_2]]\, [h, c_2].$$

The last term is in $C^3 G$, hence the middle term is in $C^4 G$.

For $[h, x^n] = [x^m c_1, x^n]$ we apply Lemma 11.30, (4), and obtain

$$[h, h'] = [x^m, [c_1, x^n]][c_1, x^n].$$

Since the last term is in C^3G and the first in C^4G, lemma follows. □

THEOREM 11.64. *Let G be a nilpotent group. The set of all finite order elements forms a characteristic subgroup of G, called the torsion subgroup of G and denoted by* $\mathrm{Tor}\, G$.

PROOF. We argue by induction on the class of nilpotency k of G. For $k = 1$ the G group is abelian and the assertion is clear. Assume that the statement is true for all nilpotent groups of class $\leqslant k$, and consider a $(k+1)$–step nilpotent group G.

It suffices to prove that for two arbitrary elements a, b of finite order in G, the product ab is likewise of finite order. The subgroup $B = \langle b, C^2G \rangle$ is nilpotent of class $\leqslant k$, according to Lemma 11.63. By the induction hypothesis, the set of finite order elements of B is a subgroup $\mathrm{Tor}\, B \leqslant B$, which is necessarily characteristic in B. Since B is normal in G it follows that $\mathrm{Tor}\, B$ is normal in G.

Assume that a is of order m. Then

$$(ab)^m = aba^{-1}a^2ba^{-2}a^3b \cdots a^{-m+1}a^mba^{-m}\,,$$

and right-hand side is a product of conjugates of b, hence it is in $\mathrm{Tor}\, B$. We conclude that $(ab)^m$ is of finite order. □

PROPOSITION 11.65. *A finitely generated nilpotent torsion group is finite.*

PROOF. We again argue by induction on the nilpotency class n of the group G. For $n = 1$ we apply Exercise 11.10. Assume that the property holds for all nilpotent groups of class at most n and consider G, a finitely generated torsion group that is $(n + 1)$–step nilpotent. Then C^2G and G/C^2G are finite, by the induction hypothesis, whence G is finite as well. □

COROLLARY 11.66. *Let G be a finitely generated nilpotent group. Then the torsion subgroup* $\mathrm{Tor}\, G$ *is finite.*

COROLLARY 11.67. $h(\Gamma) = cd_{\mathbb{Q}}(\Gamma)$*, the cohomological dimension of* Γ *over* $\mathbb{Q}$.

PROOF. Let $K := \mathrm{Tor}\,(\Gamma)$, $\Lambda := \Gamma/\mathrm{Tor}\,(\Gamma)$ and M be a $\mathbb{Q}\Gamma$-module. We also let M^K denote the submodule of K-invariants in M. Since K is finite, we obtain $H^i(K, M) = 0$ for all $i > 0$. Then, the Lyndon-Hochschild-Serre spectral sequence for group cohomology, yields isomorphisms

$$H^i(\Lambda, M^K) \cong H^i(\Gamma, M), \quad i > 0,$$

cf. Theorem 2 in [**HS53**]. It follows that $cd_{\mathbb{Q}}(\Gamma) \leqslant h = h(\Lambda) = cd_{\mathbb{Q}}(\Lambda)$: Vanishing of cohomology of Λ in degrees $> h$ implies vanishing of cohomology of Γ in the same degrees. To see the converse, consider $M = \mathbb{Q}$, the trivial $\mathbb{Q}\Gamma$-module (and trivial $\mathbb{Q}\Lambda$-module); in this case, of course, $M = M^K$. Since

$$\mathbb{Q} \cong H^h(\Lambda, \mathbb{Q}),$$

it follows that

$$H^h(\Lambda, \mathbb{Q}) = H^h(\Gamma, \mathbb{Q}) \neq 0.$$

Therefore, $cd_{\mathbb{Q}}(\Gamma) = h = h(\Gamma)$. □

EXERCISE 11.68. Let D_∞ be the infinite dihedral group.

(1) Give an example of two elements a, b of finite order in D_∞ such that their product ab is of infinite order.

(2) Is D_∞ a nilpotent group ?

(3) Are any of the finite dihedral groups D_{2n} nilpotent?

LEMMA 11.69 (A. I. Mal'cev, **[Mal49a]**). *If G is a nilpotent group with torsion-free center, then:*

(a) Each quotient $Z_{i+1}(G)/Z_i(G)$ is torsion-free.

(b) G is torsion-free.

PROOF. (a) We argue by induction on the nilpotency class n of G. The assertion is clear for $n = 1$; assume it holds for all nilpotent groups of class $< n$. We first prove that the group $Z_{n-1}(G)/Z_n(G)$ is torsion-free.

We will show that for each nontrivial element $\bar{x} \in Z_2(G)/Z_1(G)$, there exists a homomorphism $\varphi \in \mathrm{Hom}(Z_2(G)/Z_1(G), Z_1(G))$ such that $\varphi(\bar{x}) \neq 1$. Since $Z_1(G)$ is torsion-free this would imply that $Z_2(G)/Z_1(G)$ is torsion-free as well. Let $x \in Z_2(G)$ be the element which projects to $\bar{x} \in Z_1(G)/Z_n(G)$. Thus $x \notin Z_1(G)$, therefore there exists an element $g \in G$ such that $[g, x] \in Z_1(G) - \{1\}$. Define the map $\tilde{\varphi} : Z_2(G) \to Z_1(G)$ by:

$$\tilde{\varphi}(y) := [y, g],$$

where $g \in G$ is an element above (such that $[g, x] \neq 1$). Clearly, $\tilde{\varphi}(x) \neq 1$; since $Z_1(G)$ is the center of G, the map $\tilde{\varphi}$ descends to a map $\varphi : Z_2(G)/Z_1(G) \to Z_1(G)$. It follows from Part 3 of Lemma 11.30 that $\tilde{\varphi}$ is a homomorphism. Hence, φ is a homomorphism as well. Since $Z_n(G)$ is torsion-free, it follows that $Z_2(G)/Z_1(G)$ is torsion-free too. Now, we replace G by the group $\bar{G} = G/Z_1(G)$.

Since $Z_2(G)/Z_1(G)$ is torsion-free, the group $\bar{G}$ has torsion-free center. Hence, by the induction hypothesis, $Z_{i+1}(\bar{G})/Z_i(\bar{G})$ is torsion-free for every i. However,

$$Z_{i+1}(\bar{G})/Z_i(\bar{G}) \cong Z_i(G)/Z_{i-1}(G)$$

for every $i \geqslant 1$. Thus, every group $Z_i(G)/Z_{i-1}(G)$ is torsion-free, proving (a).

(b) In view of (a), for each i, $m \neq 0$ and each $x \in Z_i(G) \setminus Z_{i+1}(G)$ we have: $x^m \notin Z_{i+1}(G)$. Thus $x^m \neq 1$. Therefore, G is torsion-free. □

COROLLARY 11.70. *If G is nilpotent then $\bar{G} := G/\mathrm{Tor}\, G$ is torsion-free.*

PROOF. Each element $\bar{x} \in \bar{G}$, is the image of $x = ty \in G$ under the quotient map $\pi : G \to \bar{G}$, where $t \in \mathrm{Tor}\, G$. Then $1 = (\bar{x})^k$ would imply that

$$1 = (\bar{x})^k = \pi(y^k),$$

$y^k \in \mathrm{Tor}\, G$ and, hence, $y \in \mathrm{Tor}\, G$. It follows that $\bar{x} = 1$. □

Note that this lemma *does not imply* that for torsion-free nilpotent groups the quotients $C^iG/C^{i+1}G$ are torsion-free (this is, in general, false).

11.5. Polycyclic groups

DEFINITION 11.71. A group G is *polycyclic* if it admits a subnormal descending series

$$G = N_0 \rhd N_1 \rhd \ldots \rhd N_n \rhd N_{n+1} = \{1\} \tag{11.9}$$

such that N_i/N_{i+1} is cyclic for all $i \geqslant 0$.

A series as in (11.9) is called a *cyclic series*, and its *length* is the number of non-trivial groups in this sequence, this number is $\leqslant n+1$ in (11.9). The *length* $\ell(G)$ of a polycyclic group is the least length of a cyclic series of G.

If, moreover, N_i/N_{i+1} is infinite cyclic for all $i \geqslant 0$, then the group G is called *poly-C_∞* and the series is called a *C_∞–series.*

We declare the trivial group to be poly-C_∞ as well.

REMARK 11.72. If G is poly-C_∞ then Corollary 4.25 implies that $N_i \simeq N_{i+1} \rtimes \mathbb{Z}$ for every $i \geqslant 0$; thus, the group G is obtained from $N_n \simeq \mathbb{Z}$ by *successive semi-direct products with $\mathbb{Z}$.*

For general polycyclic groups G the above is no longer true, for instance, G could be a finite group. However, the above property is *almost true* for G: *Every polycyclic group contains a normal subgroup of finite index which is poly-C_∞* (see Proposition 11.81).

PROPOSITION 11.73. (1) *A polycyclic group has the bounded generation property in the sense of Definition 4.16. More precisely, let G be a group with a cyclic series* (11.9) *of length n and let t_i be such that t_iN_{i+1} is a generator of N_i/N_{i+1}. Then every $g \in G$ can be written as $g = t_1^{k_1} \cdots t_n^{k_n}$, with $k_1, \ldots, k_n$ in $\mathbb{Z}$.*

(2) *A polycyclic torsion group is finite.*

(3) *Any subgroup of a polycyclic group is polycyclic, and, hence, finitely generated.*

(4) *If N is a normal subgroup in a polycyclic group G, then G/N is polycyclic.*

(5) *If $N \lhd G$ and both N and G/N are polycyclic then G is polycyclic.*

(6) *Properties (3) and (5) hold with 'polycyclic' replaced by 'poly-C_∞', but not (4).*

PROOF. Part (1) follows by an easy induction on n.

Part (2) follows immediately from (1).

(3). Let H be a subgroup in G. Given a cyclic series for G as above, the intersections $H \cap N_i$ define a cyclic series for H.

(4). The proof is by induction on the length $\ell(G) = n$. For $n = 1$, G is cyclic and any quotient of G is also cyclic.

Assume that the statement is true for all $k \leqslant n$, and consider a group G with $\ell(G) = n+1$. Let N_1 be the first term distinct from G in this cyclic series. By the induction hypothesis, $N_1/(N_1 \cap N) \simeq N_1N/N$ is polycyclic. The subgroup N_1N/N is normal in G/N and $(G/N)/(N_1N/N) \simeq G/N_1N$ is cyclic, as it is a quotient of G/N_1. It follows that G/N is polycyclic.

(5) Consider the cyclic series

$$G/N = Q_0 \geqslant Q_1 \geqslant \cdots \geqslant Q_n = \{\bar{1}\}$$

and

$$N = N_0 \geqslant N_1 \geqslant \cdots \geqslant N_k = \{1\}\,.$$

Given the quotient map $\pi : G \to G/N$ and $H_i := \pi^{-1}(Q_i)$, the following is a cyclic series for G:

$$G \geqslant H_1 \geqslant \ldots \geqslant H_n = N = N_0 \geqslant N_1 \geqslant \ldots \geqslant N_k = \{1\}\,.$$

(6) The proofs of properties (3) and (5) with 'polycyclic' replaced by 'poly-C_∞' are identical. A counter-example for (4) with 'polycyclic' replaced by 'poly-C_∞' is $G = \mathbb{Z}$, $N = 2\mathbb{Z}$. □

REMARKS 11.74. (1) If G is polycyclic then, in general, the subset $\operatorname{Tor} G \subset G$ of finite order elements in G is neither a subgroup nor is a finite set.

Consider for instance the infinite dihedral group D_∞. This group can be realized as the group of isometries of $\mathbb{R}$ generated by the symmetry $s : \mathbb{R} \to \mathbb{R}$, $s(x) = -x$ and the translation $t : \mathbb{R} \to \mathbb{R}$, $t(x) = x + 1$, and as noted before (see Section 3.3) $D_\infty = \langle t \rangle \rtimes \langle s \rangle$. Therefore D_∞ is polycyclic by Proposition 11.73, (5), but $\operatorname{Tor} D_\infty$ is the union of a left coset and the trivial subgroup:

$$\operatorname{Tor} G = s\,\langle t \rangle \cup \{1\}.$$

(2) Every polycyclic group is virtually torsion-free (see Proposition 11.81).

PROPOSITION 11.75. *Every finitely generated nilpotent group is polycyclic.*

PROOF. This may be proved using Proposition 11.73, Part (5), and an induction on the nilpotency class or directly, by constructing a series as in (11.9) as follows: Consider the finite descending series with terms C^kG. For every $k \geqslant 1$, $C^kG/C^{k+1}G$ is finitely generated abelian (see Corollary 11.45). According to the classification of finitely generated abelian groups, there exists a finite subnormal descending series

$$C^kG = A_0 \geqslant A_1 \geqslant \cdots \geqslant A_n \geqslant A_{n+1} = C^{k+1}G$$

such that every quotient A_i/A_{i+1} is cyclic. By inserting all these finite descending series into the one defined by C^kG's, we obtain a finite subnormal cyclic series for G. □

THEOREM 11.76 (K. A. Hirsch, [**Hir38**]). *All polycyclic groups are residually finite.*

PROOF. The theorem is an immediate corollary of the following lemma:

LEMMA 11.77. *Suppose that H is a finitely generated residually finite group and we have a cyclic extension of H, i.e., a group G which appears in a short exact sequence*

$$1 \to H \to G \xrightarrow{f} C \to 1,$$

where C is a cyclic group. Then G is also residually finite.

PROOF. Consider $g \in G \setminus 1$. If g does not belong to H, then $f(g) \neq 1$ and residual finiteness of C implies that there exists a homomorphism of C to a finite cyclic group which sends sends $f(g)$ to a nontrivial element. By composing the homomorphisms, we obtain a homomorphism of G to a finite group which sends g to a nontrivial element.

Suppose, therefore, that g is in H. Let $F \leqslant H$ be a finite index subgroup which does not contain g. If C is finite, then F has finite index in G and we are done. Assume, therefore, that C is infinite cyclic. Since H is finitely generated, there exists a finite index subgroup $A \leqslant F$ which is a characteristic subgroup of G. Consider the quotient groups $\bar{H} = H/A, \bar{G} = G/A$ and the element $\bar{g} \in \bar{G}$, the image of g in $\bar{G}$. As above, it suffices to find a finite index subgroup of $\bar{G}$ which does not contain $\bar{g}$.

The group $\bar{G}$ is virtually isomorphic to $\mathbb{Z}$. According to Lemma 4.105, the group $\bar{G}$ is residually finite. Since $\bar{g} \neq 1$, we obtain a finite index subgroup of $\bar{G}$ not containing $\bar{g}$. Hence, G is residually finite. $\square$

This also proves the theorem. $\square$

REMARK 11.78. There exist coextensions

$$1 \to \mathbb{Z}_2 \to G \to H \to 1$$

where H is finitely generated residually finite, while G is not residually finite, see [**Mil79**].

An edifying example of a polycyclic group is the following.

PROPOSITION 11.79. *Let $m, n \geqslant 1$ be two integers, and let $\varphi : \mathbb{Z}^n \to \mathrm{Aut}(\mathbb{Z}^m)$ be a homomorphism.*

The semidirect product $G = \mathbb{Z}^m \rtimes_\varphi \mathbb{Z}^n$ is a poly-C_∞ group.

PROOF. The quotient $G/\mathbb{Z}^m$ is isomorphic to $\mathbb{Z}^n$. Therefore by Proposition 11.73, (6), the group G is poly-C_∞. $\square$

EXERCISE 11.80. Let $\mathcal{T}_n(\mathbb{K})$ be the group of invertible upper-triangular $n \times n$ matrices with entries in a field $\mathbb{K}$.

(1) Prove that $\mathcal{T}_n(\mathbb{K})$ is a semi-direct product of its nilpotent subgroup $\mathcal{U}_n(\mathbb{K})$ introduced in Exercise 11.38, and the subgroup of diagonal matrices.
(2) Prove that, if $\mathbb{K}$ hsa zero characteristic, the subgroup of $\mathcal{T}_n(\mathbb{K})$ generated by $I + E_{12}$ and by the diagonal matrix with $(-1, 1, \ldots, 1)$ on the diagonal is isomorphic to the infinite dihedral group D_∞. Deduce that $\mathcal{T}_n(\mathbb{K})$ is not nilpotent.

PROPOSITION 11.81. *A polycyclic group G contains a normal subgroup of finite index which is poly-C_∞.*

PROOF. We argue by induction on the length $\ell(G) = n$. For $n = 1$ the group G is cyclic and the statement obviously true. Assume that the assertion is true for n and consider a polycyclic group G having a cyclic series (11.9).

The induction hypothesis implies that N_1 contains a normal subgroup S of finite index which is poly-C_∞. Lemma 3.10 shows that S has a finite index subgroup S_1 which is normal in G. Proposition 11.73, Part (6), implies that S_1 is poly-C_∞ as well.

If G/N_1 is finite then S_1 has finite index in G.

Assume that G/N_1 is infinite cyclic. Then the group $K = G/S_1$ contains the finite normal subgroup $F = N_1/S_1$ such that K/F is isomorphic to $\mathbb{Z}$. Corollary 4.25 implies that K is a semidirect product of F and an infinite cyclic subgroup $\langle x \rangle$. The conjugation by x defines an automorphism of F and since $\mathrm{Aut}(F)$ is finite, there exists r such that the conjugation by x^r is the identity on F. Hence $F \langle x^r \rangle$ is a finite index subgroup in K and it is a direct product of F and $\langle x^r \rangle$. We conclude that $\langle x^r \rangle$ is a finite index normal subgroup of K. We have that $\langle x^r \rangle = G_1/S_1$, where G_1 is a finite index normal subgroup in G, and G_1 is poly-C_∞ since S_1 is poly-C_∞. $\square$

Corollary 11.82. (a) *A poly-C_∞ group is torsion-free.*

(b) *A polycyclic group is virtually torsion-free.*

Proof. In view of Proposition 11.81, it suffices to prove (a). Consider a poly-C_∞ group G. We argue by induction on cyclic length $\ell(G) = n$. For $n = 1$, the group G is infinite cyclic and the statement obviously holds. Assume that the statement is true for all groups of cyclic length at most n and consider a group G with $\ell(G) = n + 1$ and the cyclic series (11.9). Let g be an element of finite order in G. Then its image in the infinite cyclic quotient G/N_1 is the identity, hence $g \in N_1$. The induction hypothesis implies that $g = 1$. □

Proposition 11.83. *Let G be a finitely generated nilpotent group. The following are equivalent:*

(1) *G is poly-C_∞;*
(2) *G is torsion-free;*
(3) *the center of G is torsion-free.*

Proof. Implication (1)⇒ (2) is Corollary 11.82, (a), while the implication (2)⇒ (3) is obvious. The implication (3)⇒ (1) follows from Lemma 11.69. □

Remark 11.84. 1. Lemma 11.69 also implies that for each torsion-free nilpotent group G, every quotient $Z_{i+1}(G)/Z_i(G)$ is also torsion-free.

2. In contrast, the lower central series of a (finitely generated) nilpotent torsion-free group may have abelian quotients $C^{i+1}G/C^iG$ with non-trivial torsion. Indeed, given an integer $p \geqslant 2$, consider the following subgroup G of the integer Heisenberg group $H_3(\mathbb{Z})$:

$$G = \left\{ \begin{pmatrix} 1 & k & n \\ 0 & 1 & pm \\ 0 & 0 & 1 \end{pmatrix} ;\ k, m, n \in \mathbb{Z} \right\} .$$

Since $H_3(\mathbb{Z})$ is poly-C_∞, so is G. On the other hand, the commutator subgroup in G is:

$$G' = \left\{ \begin{pmatrix} 1 & 0 & pn \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} ;\ n \in \mathbb{Z} \right\} .$$

The quotient G/G' is isomorphic to $\mathbb{Z}^2 \times \mathbb{Z}_p$.

Proposition 11.85. *Every polycyclic group is finitely presented.*

Proof. The proof is an easy induction on the minimal length of a cyclic series, combined with Proposition 4.31. □

One parameter measuring the complexity of the “poly-C_∞ part” of any polycyclic group is the *Hirsch number* (generalizing the Hirsch length for nilpotent groups), defined as follows:

Proposition 11.86. *The number of infinite factors in a cyclic series of a polycyclic group G is the same for all series. This number is called the* Hirsch number *(or* Hirsch length*) of G.*

Proof. The proof will follow from the following observation on cyclic series:

LEMMA 11.87. *Any refinement of a cyclic series is also cyclic. Moreover, the number of quotients isomorphic to $\mathbb{Z}$ is the same for both series.*

PROOF. Consider a cyclic series

$$H_0 = G \geqslant H_1 \geqslant \ldots \geqslant H_n = \{1\}.$$

A refinement of this series is composed of the following sub-series

$$H_i = R_k \geqslant R_{k+1} \geqslant \ldots \geqslant R_{k+m} = H_{i+1}.$$

Each quotient R_j/R_{j+1} embeds naturally as a subgroup in H_i/R_{j+1}, and the latter is a quotient of the cyclic group H_i/H_{i+1}; hence all quotients are cyclic. If H_i/H_{i+1} is finite then all quotients R_j/R_{j+1} are finite.

Assume now that $H_i/H_{i+1} \simeq \mathbb{Z}$. We prove by induction on $m \geqslant 1$ that exactly one among the quotients R_j/R_{j+1} is isomorphic to $\mathbb{Z}$, and the other quotients are finite. For $m = 1$ the statement is clear. Assume that it is true for m and consider the case of $m + 1$.

If H_i/R_{k+m} is finite then all R_j/R_{j+1} with $j \leqslant k + m - 1$ are finite. On the other, under this assumption, hand R_{k+m}/R_{k+m+1} cannot be finite, otherwise H_i/H_{i+1} would be finite.

Assume that $H_i/R_{k+m} \simeq \mathbb{Z}$. The induction hypothesis implies that exactly one quotient R_j/R_{j+1} with $j \leqslant k + m - 1$ is isomorphic to $\mathbb{Z}$ and the others are finite. The quotient R_{k+m}/R_{k+m+1} is a subgroup of $H_i/R_{k+m} \simeq \mathbb{Z}$ such that the quotient by this subgroup is also isomorphic to $\mathbb{Z}$. This can only happen when R_{k+m}/R_{k+m+1} is trivial. □

Proposition 11.86 now follows from Lemmas 11.87 and 3.6. □

EXERCISE 11.88. Show that for each finitely generated nilpotent group the Hirsch number equals the *Hirsch length* $h(G)$, defined earlier.

In view of this exercise, the Hirsch number for a polycyclic group G will be again denoted $h(G)$.

A natural question to ask is the following.

QUESTION 11.89. Since poly-C_∞ groups are constructed by successive semi-direct products with $\mathbb{Z}$, is there a way to detect during this construction whether the group is nilpotent or not?

The answer to this question will be given in Section 12.3 and it has some interesting relation to the growth of groups.

11.6. Solvable groups: Definition and basic properties

Recall that G' denotes the *derived subgroup* $[G, G]$ of the group G. Given a group G, we define its *iterated commutator subgroups* $G^{(k)}$ inductively by:

$$G^{(0)} = G, G^{(1)} = G', \ldots, G^{(k+1)} = \left(G^{(k)}\right)', \ldots$$

The descending series

$$G \trianglerighteq G' \trianglerighteq \ldots \trianglerighteq G^{(k)} \trianglerighteq G^{(k+1)} \trianglerighteq \ldots$$

is called the *derived series* of the group G.

Note that all subgroups $G^{(k)}$ are *characteristic* in G.

Definition 11.90. A group G is *solvable* if there exists k such that $G^{(k)} = \{1\}$. The minimal k such that $G^{(k)} = \{1\}$ is called the *derived length* of G and the group G itself is called *k-step solvable*. A solvable group of derived length at most two is called *metabelian*.

We will use the notation $\ell_{\mathrm{der}}(G)$ for the derived length

In particular, every solvable group G of derived length k satisfies the law:

$$\llbracket x_1, \ldots, x_{2^k} \rrbracket = 1, \forall x_1, \ldots x_{2^k} \in G. \tag{11.10}$$

Here and in what follows,

$$\llbracket x_1, \ldots, x_{2^k} \rrbracket := [\ \llbracket x_1, \ldots, x_{2^{k-1}} \rrbracket, \llbracket x_{2^{k-1}+1}, \ldots, x_{2^k} \rrbracket\ ] \tag{11.11}$$

and $\llbracket x_1, x_2 \rrbracket = [x_1, x_2]$.

Exercise 11.91. 1. Find the values $n \in \mathbb{N}$ for which the symmetric group S_n is solvable. Show that the groups S_3, S_4 are nilpotent.

2. Show that if a group G satisfies the law (11.10), then it is solvable of the derived length $\leqslant k$.

Proposition 11.92.
(1) *If N is a normal subgroup in G and both N and G/N are solvable, then G is solvable. If the derived lengths of G/N and N are at most d, d' respectively, then the derived length of G is at most $d + d'$. In other words, the derived length is subadditive:*

$$\ell_{\mathrm{der}}(G) \leqslant \ell_{\mathrm{der}}(N) + \ell_{\mathrm{der}}(G/N).$$

(2) *Every subgroup H of a solvable group G is solvable and*

$$\ell_{\mathrm{der}}(H) \leqslant \ell_{\mathrm{der}}(G).$$

(3) *If G is solvable and $N \lhd G$, then G/N is solvable and*

$$\ell_{\mathrm{der}}(G/N) \leqslant \ell_{\mathrm{der}}(G).$$

Note that the statement (1) is not true when 'solvable' is replaced by 'nilpotent', consider, for instance, the infinite dihedral group D_∞.

Proof. (1) We are assuming that G/N is solvable of derived length d and N is solvable of derived length d'. Since $(G/N)^{(d)} = \{\bar{1}\}$ it follows that $G^{(d)} \leqslant N$. Then, as $G^{(d+i)} \leqslant N^{(i)}$, we obtain $G^{(d+d')} = \{1\}$.

(2) Note that for every subgroup H of a group G, $H' \leqslant G'$. Thus, by induction,

$$H^{(i)} \leqslant G^{(i)}.$$

If G is solvable of derived length k then $G^{(k)} = \{1\}$; thus $H^{(k)} = \{1\}$ as well and, hence, H is also solvable.

(3) Consider the quotient map $\pi : G \to G/N$. It is immediate that $\pi\left(G^{(i)}\right) = (G/N)^{(i)}$, in particular if G is solvable then G/N is solvable. $\square$

For the next exercise, we will need the following definition: A finite sequence of vector subspaces

$$V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_k$$

in a vector space V is called a *flag* in V. If the number of the subspaces in such a sequence is maximal possible (equal $\dim(V)+1$), the flag is called *full* or *complete*. In other words, $\dim(V_i) = i$ for all members of this sequence.

EXERCISE 11.93. (1) Prove that the subgroup $\mathcal{T}_n(\mathbb{K})$ of upper-triangular matrices in $GL(n, \mathbb{K})$, where $\mathbb{K}$ is a field, is solvable. [Hint: you may use Exercise 11.38.]

(2) Use Part (1) to show that for a finite-dimensional vector space V, the subgroup G of $GL(V)$ consisting of elements g preserving a complete flag in V (i.e., $gV_i = V_i$, for every $g \in G$ and every i) is solvable.

(3) Let V be a $\mathbb{K}$-vector space of dimension n, and let

$$V_0 = 0 \subset V_1 \subset \cdots \subset V_{k-1} \subset V_k = V$$

be a flag, not necessarily complete. Let G be a subgroup of $GL(V)$ preserving this flag. For every $i \in \{1, 2, \ldots, k-1\}$ let ρ_i be the projection $G \to GL(V_{i+1}/V_i)$. Prove that if every $\rho_i(G)$ is solvable, then G is also solvable.

EXERCISE 11.94. 1. Let $\mathbb{F}_k$ denote the field with k elements. Use the 1-dimensional vector subspaces in $\mathbb{F}_k^2$ to construct a homomorphism $GL(2, \mathbb{F}_k) \to S_n$ for an appropriate n.

2. Prove that $GL(2, \mathbb{F}_2)$ and $GL(2, \mathbb{F}_3)$ are solvable.

THEOREM 11.95 (Direct limits of virtually solvable groups). *Suppose that we have a direct system $G_i, i \in I$, of virtually solvable groups satisfying the following:*

1. The derived length of each solvable subgroup in G_i is at most d for all i.

2. Each G_i contains a normal solvable subgroup H_i of index $\leqslant c$.

Then the direct limit G of this system is again virtually solvable and contains a normal solvable subgroup H of index $\leqslant c$ and derived length at most d.

PROOF. We start the proof with several simple observations. If a group M is virtually solvable and $N_1, N_2 \lhd M$ are normal solvable subgroups of finite index, then the subgroup $N \leqslant F$ generated by N_1, N_2 is again solvable and normal in M. Therefore, without loss of generality, we may assume that each H_i is maximal among all normal solvable subgroups of finite index in G_i. By the hypothesis, the derived length of each H_i does not exceed d. Clearly, we retain the property that $|G_i : H_i| \leqslant c$ for all i.

Suppose now that $m \in I$ is such that the index $n = |G_m : H_m| \leqslant c$ is maximal among all indices $|G_i : H_i|, i \in I$. Define $J = \{j \in I : m \leqslant j\}$. Then for each homomorphism $f_{mi} : G_m \to G_i$ of the direct system, we have:

(1) $H_{mi} := f_{mi}^{-1}(H_i)$ is a normal solvable subgroup of finite index in G_m. Hence, $H_{im} \leqslant H_m$.

(2)

$$n = |G_m : H_m| \leqslant |G_m : H_{im}| = |\langle H_i, f_{mi}(G_m)\rangle : H_i| \leqslant |G_i : H_i| \leqslant n.$$

In particular, $|G_i : H_i| = n$ and $f_{mi}(H_m) \leqslant H_i$. Applying this to any pair $i, j \in J$, $i \leqslant j$, we obtain:

$$f_{ij}(H_i) \leqslant H_j.$$

Hence, for $i, j \in J$, the restrictions of the homomorphisms f_{ij} to H_i define a direct system of solvable groups of derived length $\leqslant d$. Taking the direct limit $\lim_{j \in J} H_j$, we obtain a solvable subgroup $H \leqslant G$ (see Exercise 1.27). The reader will verify (using Exercises 1.28 and 11.91) that:

(1) The subgroup H is normal in G.
(2) $|G : H| = n$.
(3) The derived length of H is at most d.

This concludes the proof. □

11.7. Free solvable groups and Magnus embedding

As in the case of nilpotent groups, there exist universal objects in the class of solvable groups that we now describe.

Definition 11.96. Given two integers $k, m \geqslant 1$, the *free solvable group of derived length k with m generators* is the quotient of the free group F_m by the normal subgroup $F_m^{(k)}$.

When $k = 2$ we call the corresponding group *free metabelian group with m generators.*

Notation 11.97. In what follows we use the notation $S_{m,k}$ for the free solvable group of derived length k and with m generators. Note that $S_{m,1}$ is $\mathbb{Z}^m$.

Proposition 11.98 (Universal property of free solvable groups). *Every solvable group with m generators and of derived length k, is a quotient of $S_{m,k}$.*

Proof. Let G be a solvable group of derived length k and let X be a generating set of G of cardinality m. X of cardinality m. The map defined in Proposition 4.22 contains $F(X)^{(k)}$ in its kernel, therefore it defines an epimorphism from the free solvable group $S_{m,k}$ to G. □

Our next goal is to define the *Magnus embedding* of the free solvable group $S_{r,k+1}$ into the wreath product $\mathbb{Z}^r \wr S_{r,k}$. Since $\mathbb{Z}^r \wr S_{r,k}$ is a semidirect product, Remark 3.112, (2), implies that in order to define a homomorphism

$$S_{r,k+1} \to \mathbb{Z}^r \wr S_{r,k}$$

one has to specify a homomorphism $\pi : S_{r,k+1} \to S_{r,k}$ and a derivation

$$d \in Der(S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r).$$

Here we will use the following action of $S_{r,k+1}$ on $\bigoplus_{S_{r,k}} \mathbb{Z}^r$: We compose π with the action of $S_{r,k}$ on itself *via* left multiplication.

To simplify the notation, we let $F = F_r$ denote the free group on r generators $x_1, \ldots, x_r$. First, since $F/F^{(m)} = S_{r,m}$ for every m, and $F^{(k+1)} \leqslant F^{(k)}$, we have a natural quotient homomorphism

$$\pi : S_{r,k+1} \to S_{r,k}.$$

We now proceed to construct the derivation d. We will use definitions and results of Section 3.9.4. Note that $\bigoplus_{S_{r,k}} \mathbb{Z}^r$ is isomorphic (as a free abelian group) to

$$M_1 \oplus \ldots \oplus M_r$$

where for every i, $M_i = M = \mathbb{Z}S_{r,k}$, the group algebra of $S_{r,k}$. Since $S_{r,k}$ is the quotient of $F = F_r$, every derivation $\partial \in Der(\mathbb{Z}F, \mathbb{Z}F)$ projects to a derivation (denoted $\widehat{\partial}$) in $Der(\mathbb{Z}F, \mathbb{Z}S_{r,k})$. Thus, derivations $\partial_i \in Der(\mathbb{Z}F, \mathbb{Z}F)$ introduced in Section 3.9.4, projects to derivations $\widehat{\partial}_i \in Der(\mathbb{Z}F, M)$. Furthermore, every derivation $\widehat{\partial}_i \in Der(\mathbb{Z}F, M)$ extends to a derivation $d_i : \mathbb{Z}F \to \bigoplus_{S_{r,k}} \mathbb{Z}^r$ by

$$d_i : w \mapsto (0, \ldots, \widehat{\partial}_i(w), \ldots 0)$$

where we place $\widehat{\partial}_i(w)$ in the i-th slot. Since a sum of derivations is again a derivation, we obtain a derivation

$$d = (\widehat{\partial}_1, \ldots, \widehat{\partial}_r) = d_1 + \ldots + d_r \in Der(\mathbb{Z}F, \bigoplus_{S_{r,k}} \mathbb{Z}^r).$$

For simplicity, in what follows, we denote $F^{(k)}$ by N and, accordingly, $F^{(k+1)}$ by N'. Thus, $S_{r,k} = F/N$ and $S_{r,k+1} = F/N'$.

LEMMA 11.99. *The derivation d projects to a derivation*

$$\bar{d} \in Der(\mathbb{Z}S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r).$$

PROOF. Let us check that N' is in the kernel of d. Indeed, given a commutator $[x, y]$ with x, y in N, property (P_3) in Exercise 3.108 implies that (by computing in $\mathbb{Z}F$)

$$\partial_i[x, y] = (1 - xyx^{-1})\partial_i x + x(1 - yx^{-1}y^{-1})\partial_i y\,.$$

Since both $x, y \in N$ project to 1 in $S_{r,k}$, they act trivially on $M = \mathbb{Z}S_{r,k}$, it follows that

$$(1 - xyx^{-1}) \cdot \xi = 0 \text{ and } x(1 - yx^{-1}y^{-1}) \cdot \eta = 0, \quad \forall \xi, \eta \in M.$$

Hence, $d_i([x, y]) = 0$ for every i and, thus, $d([x, y]) = 0$. Therefore, $d(N') = 0$ since the group N' is generated by commutators $[x, y], x, y \in N$. For arbitrary $g \in F, h \in N'$, we have

$$d(gn) = d(g) + g \cdot d(n) = d(g).$$

Thus, the derivation d projects to a derivation $\bar{d} \in Der((\mathbb{Z}S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r)$,

$$\bar{d}(gN') = d(g). \quad \square$$

Thus, according to Remark 3.112, the pair (d, π) determines a homomorphism

$$\mathfrak{M} : S_{r,k+1} \to \mathbb{Z}^r \wr S_{r,k}\,.$$

THEOREM 11.100 (W. Magnus [**Mag39**]). *The homomorphism $\mathfrak{M}$ is injective; $\mathfrak{M}$ is called the* Magnus embedding*.*

We refer to [**Fox53**, Section (4.9)] for the proof of injectivity of $\mathfrak{M}$. Remarkably, the Magnus embedding also has nice geometric features. The following theorem was proven independently by A. Sale [**Sal12**] and S. Vassileva [**Vas12**]:

THEOREM 11.101 (A. Sale, S. Vassileva). *The Magnus embedding is a quasi-isometric embedding.*

Clearly, the Magnus embedding is a useful tool for studying free solvable groups by induction on the derived length.

11.8. Solvable versus polycyclic

Proposition 11.102. *Every polycyclic group G is solvable.*

Proof. This follows immediately by the induction argument on the cyclic length of G and Part (1) of Proposition 11.92. □

Definition 11.103. A group is said to be *noetherian*, or satisfies the *maximal condition* if for every increasing sequence of subgroups

$$H_1 \leqslant H_2 \leqslant \cdots \leqslant H_n \leqslant \cdots \tag{11.12}$$

there exists N such that $H_n = H_N$ for every $n \geqslant N$.

Proposition 11.104. *A group G is noetherian if and only if every subgroup of G is finitely generated.*

Proof. Assume that G is a Noetherian group, and let $H \leqslant G$ be a subgroup which is not finitely generated. Pick $h_1 = H \setminus \{1\}$ and let $H_1 = \langle h_1 \rangle$. Inductively, assume that

$$H_1 < H_2 < ... < H_n$$

is a strictly increasing sequence of finitely generated subgroups of H, pick $h_{n+1} \in H \setminus H_n$, and set $H_{n+1} = \langle H_n, h_{n+1} \rangle$. We thus have a strictly increasing infinite sequence of subgroups of G, contradicting the assumption that G is Noetherian.

Conversely, assume that all subgroups of G are finitely generated, and consider an increasing sequence of subgroups as in (11.12). Then $H = \bigcup_{n \geqslant 1} H_n$ is a subgroup, hence generated by a finite set S. There exists N such that $S \subseteq H_N$, hence $H_N = H = H_n$ for every $n \geqslant N$. □

Proposition 11.105. *A solvable group is polycyclic if and only if it is noetherian.*

Proof. The 'only if' part follows immediately from Parts (1) and (3) of Proposition 11.73. Let G be a noetherian solvable group. We prove by induction on the derived length k that G is polycyclic.

For $k = 1$ the group is abelian, and since, by hypothesis, G is finitely generated, it is polycyclic.

Assume that the statement is true for k and consider a solvable group G of derived length $k + 1$. The commutator subgroup $G' \leqslant G$ is also Noetherian and solvable of derived length k. Hence, by the induction hypothesis, G' is polycyclic. The abelianization $G_{ab} = G/G'$ is finitely generated (because G is, by hypothesis), hence it is polycyclic. It follows that G is polycyclic by Proposition 11.73 (5). □

By Proposition 11.102 every nilpotent group is solvable. A natural question to ask is to find a relationship between nilpotency class and derived length.

Proposition 11.106. (1) *For every group G and every $i \geqslant 0$,*

$$G^{(i)} \leqslant C^{2^i} G.$$

(2) *If G is a k-step nilpotent group then its derived length is at most*

$$[\log_2 k] + 1\,.$$

PROOF. (1) The statement is obviously true for $i = 0$. Assume that it is true for i. Then

$$G^{(i+1)} = \left[G^{(i)}, G^{(i)}\right] \leqslant \left[C^{2^i}G, C^{2^i}G\right] \leqslant C^{2^{i+1}}G.$$

In the last inclusion we applied Proposition 11.62.

(2) follows immediately from (1). □

REMARK 11.107. The derived length can be much smaller than the nilpotency class: the dihedral subgroup D_{2n} with $n = 2^k$ is k-step nilpotent and metabelian.

EXERCISE 11.108. If G_1 is noetherian and G_2 is virtually isomorphic to G_1, then G_2 is also noetherian.

REMARK 11.109. There are noetherian groups which are not virtually polycyclic, e.g. *Tarski monsters*, which are finitely generated groups G, such that every proper subgroup of G is cyclic, see [**Ol'91a**].

An instructive example of solvable group is the *lamplighter group*. This group is the wreath product $G = \mathbb{Z}_2 \wr \mathbb{Z}$ in the sense of Definition 3.31.

EXERCISE 11.110. Prove that if K, H are solvable groups then $K \wr H$ is solvable. In particular, the lamplighter group G is solvable (even metabelian).

In view of Lemma 4.11, since wreath products of finitely generated are finitely generated as well, the lamplighter group is finitely generated. On the other hand:

(1) *Not all subgroups in the lamplighter group G are finitely generated:* the subgroup $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$ of G is not finitely generated.

(2) *The lamplighter group G is not virtually torsion-free:* For any finite index subgroup $H \leqslant G$, $H \cap \bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$ has finite index in $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$; in particular this intersection is infinite and contains elements of order 2.

   Both (1) and (2) imply that the lamplighter group is *not polycyclic*.

(3) The commutator subgroup G' of the lamplighter group G coincides with the following subgroup of $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$:

$$C = \{f : \mathbb{Z} \to \mathbb{Z}_2 \mid \mathrm{Supp}(f) \text{ has even cardinality}\}\,, \tag{11.13}$$

   where $\mathrm{Supp}(f) = \{n \in \mathbb{Z} \mid f(n) = 1\}$.

   [*NB. The notation here is additive, the identity element is* 0.]

   In particular, G' *is not finitely generated* and the group G is metabelian (since G' abelian).

We prove (3). First of all, C is clearly a subgroup. Note also that

$$(f, m)^{-1} = (-\varphi(-m)f, -m)\,,$$

where φ is the action of $\mathbb{Z}$ on the space of functions $f : \mathbb{Z} \to \mathbb{Z}_2$ via *shift*: For $m \in \mathbb{Z}$,

$$\varphi(m) : f(x) \mapsto f(x + m).$$

If we think of functions f as biinfinite sequences, then $\varphi(m)$ acts on a sequence via shifting all the indices by m. A straightforward calculation gives

$$[(f,m),(g,n)] = (f - g - \varphi(n)f + \varphi(m)g\,,\,0)\,.$$

Now, observe that either $\mathrm{Supp}(f)$ and $\mathrm{Supp}(\varphi(n)f)$ are disjoint, in which case $\mathrm{Supp}(f-\varphi(n)f)$ has cardinality twice the cardinality of $\mathrm{Supp}\, f$, or they overlap on a set of cardinality k; in the latter case, $\mathrm{Supp}(f-\varphi(n)f)$ has cardinality twice the cardinality of $\mathrm{Supp}\, f$ minus $2k$. The same holds for $\mathrm{Supp}(-g+\varphi(m)g)$. Since C is a subgroup,

$$(f - g - \varphi(n)f + \varphi(m)g) = (f - \varphi(n)f - (g - \varphi(m)g)) \in C.$$

This shows that $G' \leqslant C$.

Consider the opposite inclusion. The subgroup C is generated by functions $f:\mathbb{Z}\to\mathbb{Z}_2\,, \mathrm{Supp}\, f = \{a,b\}$, where a,b are distinct integers; thus, it suffices to show that $(f,0)\in G'$. Let $\delta_a : f:\mathbb{Z}\to\mathbb{Z}_2\,,\, \mathrm{Supp}\,\delta_a = \{a\}$. Then

$$[(\delta_a,0),(0,b-a)] = (\delta_a - \varphi(b-a)\delta_a, 0) = (f,0)$$

which implies that $(f,0)\leqslant G'$.

We conclude this section by noting that, unlike polycyclic groups, solvable groups may not be finitely presented. An example of such group is the wreath product $\mathbb{Z}\wr\mathbb{Z}$ [**Bie79**]. We refer to the same paper for a survey on finitely presented solvable groups. Nevertheless, a solvable group may be finitely presented without being polycyclic; for instance the Baumslag–Solitar group

$$G = BS(1,p) = \left\langle a,b | aba^{-1} = b^p \right\rangle$$

is metabelian but not polycyclic (for $|p|\geqslant 2$). The derived subgroup G' of G is isomorphic to the additive group of p-adic rational numbers, i.e., rational numbers whose denominators are powers of p. In particular, G' is not finitely generated. Hence, in view of Proposition 11.73, G is not polycyclic.

Exercise 11.111. Show that the group $G = BS(1,p)$ is metabelian.

CHAPTER 12

Geometric aspects of solvable groups

In this chapter we discuss several geometric aspects of solvable groups:

- Distortion of subgroups in nilpotent groups.
- Growth of solvable groups: We will compute growth rates of nilpotent groups and prove Milnor–Wolf theorem that a solvable group has polynomial growth if and only if it is virtually nilpotent.
- Erschler's examples establishing failure of quasiisometry invariance of the class of (virtually) solvable groups.
- Theorems of Zassenhaus and Jordan dealing with discrete subgroups of Lie groups. Jordan's theorem shows that finite subgroups of $GL(n,\mathbb{R})$ contain abelian subgroups of *uniformly bounded index*.

Many of these results will play important role in the proof of Gromov's theorem on groups of polynomial growth.

12.1. Wolf's Theorem for semidirect products $\mathbb{Z}^n \rtimes \mathbb{Z}$

In this section we explain how to prove Conjecture 5.84 in the case of semidirect products $\mathbb{Z}^n \rtimes \mathbb{Z}$. This easy example helps to understand the general case of polycyclic groups and the general Wolf's Theorem.

Note that the semidirect product is defined by a homomorphism $\varphi : \mathbb{Z} \to Aut(\mathbb{Z}^n) = GL(n,\mathbb{Z})$, and the latter is determined by $\theta = \varphi(1)$, which is represented by a matrix $M \in GL(n,\mathbb{Z})$. Therefore the same semidirect product is also denoted $\mathbb{Z}^n \rtimes_\theta \mathbb{Z} = \mathbb{Z}^n \rtimes_M \mathbb{Z}$.

Proposition 12.1. *A semidirect product $G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is*

(1) *either virtually nilpotent (when M has all eigenvalues of absolute value 1);*

(2) *or of exponential growth (when M has at least one eigenvalue of absolute value $\neq 1$).*

Remarks 12.2. (1) The group $G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is nilpotent if M has all eigenvalues equal to 1 (see Case (1) of the proof of the proposition).

(2) The same is not in general true if M has all eigenvalues of absolute value 1. The group $G = \mathbb{Z} \rtimes_M \mathbb{Z}$ with $M = (-1)$ is a counter-example: It admits a quotient which is the infinite dihedral group and the latter is not nilpotent. In this example, the group $G = \mathbb{Z} \rtimes_M \mathbb{Z}$ is polycyclic, virtually nilpotent but not nilpotent. In particular, the statement (1) in Proposition 12.1 cannot be improved to '$G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is nilpotent'.

Proof. Note that $\mathbb{Z}^n \rtimes_{\theta^N} \mathbb{Z}$ is a subgroup of finite index in $G = \mathbb{Z}^n \rtimes_\theta \mathbb{Z}$ (corresponding to the replacement of the second factor $\mathbb{Z}$ by $N\mathbb{Z}$). Thus, we may replace M by some power of M, and replace G with a finite-index subgroup. We will retain the notation G and M for the finite-index subgroup and the power of M. Then, the matrix $M \in GL(n, \mathbb{Z})$ will have no nontrivial roots of unity as eigenvalues. In view of Lemma 11.28, this means that for every eigenvalue $\lambda \neq 1$ of M, $|\lambda| \neq 1$.

We have two cases to consider.

(1) We prove the statement by induction on n. For $n = 0$ there is nothing to prove; we assume, therefore, that the statement holds for $n-1$. The matrix M has only eigenvalues equal 1. Lemma 11.27 then implies that there exists a finite series

$$\{1\} = H_n \leqslant H_{n-1} \leqslant \ldots \leqslant H_1 \leqslant A = H_0 = \mathbb{Z}^n$$

such that $H_i \simeq \mathbb{Z}^{n-i}$, each quotient H_i/H_{i+1} is cyclic, the automorphism θ preserves each H_i and induces the identity automorphism on H_i/H_{i+1}. Thus, θ acts via the identity on H_{n-1}. In particular, the subgroup H_{n-1} is central in G; the automorphism θ projects to an automorphism $\bar{\theta} : \bar{A} \to \bar{A}$, $\bar{A} = A/H_{n-1}$. The automorphism $\bar{\theta}$ preserves the central series

$$\{1\} = \bar{H}_{n-1} \leqslant \ldots \leqslant \bar{H}_1 \cong \mathbb{Z}^{n-1},$$

(where $\bar{H}_i = H_i/H_{n-1}$) and induces trivial automorphism of each quotient

$$\bar{H}_i/\bar{H}_{i+1} \cong H_i/H_{i+1}.$$

By the induction hypothesis, the group

$$\bar{G} = \bar{A} \rtimes_{\bar{\theta}} \mathbb{Z} \cong G/H_{n-1},$$

is nilpotent. Since central coextensions of nilpotent groups are again nilpotent (Exercise 11.51), we conclude that the group G is nilpotent as well.

(2) Assume that M has an eigenvalue with absolute value strictly greater than 1. After replacing θ with its power θ^N if necessary, we may assume that the matrix M has an eigenvalue with absolute value at least 2.

Lemma 11.29 applied to M implies that there exists an element $v \in \mathbb{Z}^n$ such that distinct elements $s = (s_k) \in \bigoplus_{k \geqslant 0} \mathbb{Z}_2$ define distinct vectors

$$s_0 v + s_1 M v + \ldots + s_n M^k v + \ldots$$

in $\mathbb{Z}^n$. With the multiplicative notation for the binary operation in G, the above vectors correspond to distinct elements

$$g_s = v^{s_0} (tvt^{-1})^{s_1} \cdots (t^k v t^{-k})^{s_k} \cdots \in G.$$

Now, consider the set Σ_K of sequences $s = (s_k)$ for which $s_k = 0\,, \forall k \geqslant K+1$. Then the map

$$\Sigma_K \to G, \quad s \mapsto g_s$$

is injective and its image consists of 2^{K+1} distinct elements g_s. Assume that the generating set of G contains the elements t and v. With respect to this generating set, the word-length $|g_s|$ is at most $3K+1$ for every $s \in \Sigma_K$. Thus, for every K we obtain 2^{K+1} distinct elements of G of length at most $3K+1$, whence G has exponential growth. □

Remark 12.3. What remains to be proven is that the two cases in Proposition 12.1 are mutually exclusive, i.e., that a nilpotent group cannot have exponential growth. We shall prove in Section 12.2 that nilpotent (hence virtually nilpotent) groups have in fact polynomial growth. In the next section we compute growth for the integer Heisenberg group $H_3(\mathbb{Z})$.

In order to analyze growth functions of solvable groups, we first have to discuss *distortion* (see §5.9) of subgroups of solvable groups. This will be done in sections 12.1.2 and 12.1.3.

12.1.1. Geometry of $H_3(\mathbb{Z})$. In this section we discuss in detail the geometric concepts introduced so far in the case of the *integer Heisenberg group*

$$G = H_3(\mathbb{Z}) = \left\{ U_{klm} = \begin{pmatrix} 1 & k & m \\ 0 & 1 & l \\ 0 & 0 & 1 \end{pmatrix} \; ; \; k,l,m \in \mathbb{Z} \right\},$$

When convenient, we will also use the notation $U_{k,l,m}$ instead of U_{klm}.

Exercise 12.4. (1) Show that the elements $x = I + U_{100}$, $y = I + U_{010}$, $z = I + U_{001}$ generate G and satisfy the relation

$$[x, y] = z.$$

(2) Prove that $U_{klm} = x^k y^l z^{m+kl}$ for every $k, l, m \in \mathbb{Z}$. This in particular shows that every element of G can be written as $x^k y^l z^m$ with $k, l, m \in \mathbb{Z}$, and that this decomposition is unique for every element (since it is entirely determined by its matrix entries).
(3) Prove that $[x^k, y^l] = z^{kl}$.

We let $S = \{x, y\}$ denote this generating set and let $|g|$ denote the distance $\mathrm{dist}_S(1, g)$, $g \in G$.

Exercise 12.5. Use Part 3 of Exercise 12.4 to show that

$$|x^k y^l z^m| \leqslant |k| + |l| + 4\sqrt{|m| + 1}\,. \tag{12.1}$$

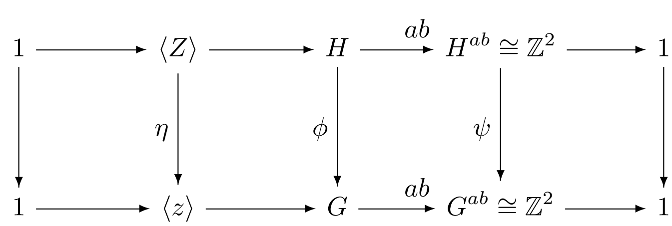
Lemma 12.6. *The group G has the presentation*

$$\langle X, Y, Z | [X, Y] = Z, [X, Z] = [Y, Z] = 1 \rangle\,.$$

Proof. The group $H = \langle X, Y, Z | [X, Y] = Z, [X, Z] = [Y, Z] = 1 \rangle$ has a homomorphism ϕ to G defined via

$$\phi(X) = x, \quad \phi(Y) = y, \quad \phi(Z) = z.$$

We will verify that ϕ is an isomorphism. To this end, consider the commutative diagram



where the homomorphisms $ab : H \to H^{ab}$ and $ab : G \to G^{ab}$ are the abelianization homomorphisms. We leave it to the reader to check that η and ψ are isomorphism and to conclude from this that ϕ is an isomorphism as well. □

By abusing the notation, we will continue to use the letters x, y for the images of the generators $x, y \in S$ under the abelianization homomorphism $G \to G^{ab}$. We will identify the Cayley graph of G^{ab} (with respect to the generating set $\{x, y\}$) with the coordinate grid in the plane $\mathbb{R}^2$. We will use the coordinates X, Y in the plane so that the generators x, y correspond to the vectors $(1, 0), (0, 1)$ respectively.

Each word $w = w(x, y)$, representing the identity element of G^{ab}, defines a piecwise-linear oriented loop L_w in the plane, with edges of the unit length, where every edge is parallel to one of the coordinate axes; the loop L_w starts and ends at the origin. This loop, treated as 1-cycle in $\mathbb{R}^2$, bounds a 2-chain D in $\mathbb{R}^2$ and we define the *signed area* $a(w) = a(L_w)$ as the integral

$$\int_D dXdY.$$

This integral, is, of course, independent of the choice of D. For instance,

$$a(x^p y^q x^{-p} y^{-q}) = pq.$$

EXERCISE 12.7. Show that

$$4\sqrt{a(w)} \leqslant \text{length}(L_w).$$

In the following lemma we describe a procedure of converting a word $w(x, y, z)$ into a *normal form* $x^k y^l z^m$; this is the simplest case of the similar process used in §12.1.3 for the proof of Proposition 12.20. Note that the redundant generating set $\{x, y, z\}$ for the group G, will be called a *closed lcs generating set* in §12.1.3.

LEMMA 12.8. *If* $w = w(x, y)$ *represents the element* $z^m \in G$, *then*

$$m = a(w).$$

PROOF. We will convert w to its normal form z^m by inductively moving all the letters $x^{\pm 1}$ to the left and the letters $z^{\pm 1}$ to the right. The induction is on the number of *inversions* in the word w, i.e. occurrences of letters $y^{\pm 1}$ to the left of the letters $x^{\pm 1}$.

If w has the form

$$y^q x^p u(x, y),$$

we convert it to the word

$$x^p y^q z^{-pq} u(x, y)$$

and then to

$$x^p y^q u(x, y) z^{-pq}$$

using the fact that z is a central element of G. If $L_w = \partial D$ and w' denotes $x^p y^q u(x, y)$, then $L_{w'}$ is bounded by the sum of two chains:

$$D + Q,$$

where Q is represented by the oriented rectangle bounding the loot L_c,

$$c = x^p y^q x^{-p} y^{-q}.$$

In particular,

$$pq = a(Q)$$

and, hence,

$$a(w) = a(w') - pq.$$

The word w' has less inversions than w and, hence, inductively, we obtain:

$$w' =_G z^{a'}, a' = a(w').$$

Since

$$z^m =_G w =_G w' z^{-pq}$$

we obtain

$$m = a' - pq = a(w).$$

The case when the word w has the form

$$x^p y^q x^r v(x, y)$$

is similar (we commute the elements y^q, x^r instead) and is left to the reader. $\square$

Since $\text{length}(w) \geqslant 4\sqrt{|m|}$, we conclude that

$$4\sqrt{|m|} \leqslant |z^m| \leqslant 4\sqrt{|m| + 1},$$

and, hence,

$$4\sqrt{|m|} \leqslant |z^m| \leqslant 8\sqrt{|m|},$$

In other words, the central subgroup $\langle z \rangle$ in G has quadratic distortion.

We now consider words $w = w(x, y)$ representing arbitrary elements g of G. Suppose that g projects to $x^p y^q \in G^{ab}$. Then the word

$$w' = w y^{-q} x^{-p}$$

represents the identity element of G^{ab} and, hence (by the lemma),

$$w' =_G z^a, \quad a = a(w').$$

It follows that

$$g = x^p y^q z^a.$$

We obtain:

Corollary 12.9. *If $w(x, y)$ represents $x^k y^l z^m \in G$, then:*
1. w represents the product $x^k y^l \in G^{ab}$ and, more importantly,
2. $m = a(w')$, where $w' = w y^{-l} x^{-k}$.
3.

$$|w| \geqslant \max(|k| + |l|, \sqrt{|m|}) \geqslant \frac{1}{2}(|k| + |l| + \sqrt{|m|}).$$

Corollary 12.10. *1. For each $n \in \mathbb{N}$ the map*

$$\varphi_n : x \mapsto x^n, \quad y \mapsto y^n$$

defines an endomorphism of G, such that

$$\varphi_n(U_{k,l,m}) = U_{nk,nl,n^2 m} \tag{12.2}$$

and $|G : \varphi_n(G)| = n^4$.

2. The endomorphism φ_n is expanding for each $n > 8$, see §5.7 for the definition.

PROOF. 1. We will check the equation (12.2) and leave the rest to the reader. The endomorphism φ_n descends to the endomorphism of the abelianization

$$\bar{\varphi}_n : \langle x\rangle \oplus \langle y\rangle \mapsto \langle x^n\rangle \oplus \langle y^n\rangle ,$$

which implies that φ_n sends $U_{k,l,m}$ to $U_{nk,nl,p}$ for some p. It remains to compute p. The endomorphism $\bar{\varphi}_n$ extends to the dilation $\mathbf{v} \mapsto n\mathbf{v}$ of $\mathbb{R}^2$, which scales all length by n and all the areas by n^2. Therefore, φ_n sends each word $w = w(x, y)$ to a word of the signed area

$$n^2 a(w).$$

Now, the claim follows from Corollary 12.9.

2. To verify the expansion property, note that

$$|U_{k,l,m}| \leqslant |k| + |l| + 4\sqrt{|m| + 1}$$

while

$$|\varphi_n(U_{k,l,m})| \geqslant (|nk| + |nl| + n\sqrt{|m|})/2.$$

Hence, for each $n > 8$, there exists $c > 1$ such that for all $g \in G \setminus \{1\}$,

$$|\varphi_n(g)| > c|g|,$$

which means that φ_n is expanding. □

We next compute the growth function of the group G:

LEMMA 12.11. $\mathfrak{G}_G(n) \asymp n^4$.

PROOF. We first note that the box

$$B_n := \{U_{klm} : -n \leqslant k, l, \leqslant n, -n^2 + 1 \leqslant m \leqslant n^2 - 1\}$$

contains at least $4n^2(n^2 - 1)$ elements and each $U_{klm} \in B_n$ satisfies

$$|U_{klm}| \leqslant 2n + 4\sqrt{n^2} = 6n.$$

Thus, for all $n > 2$, the ball $B(1, 6n) \subset G$ contains at least $3n^4$ elements and, hence the growth function of G satisfies

$$n^4 \preceq \mathfrak{G}_G(n).$$

We next estimate the growth of G from above. The image of the ball $B(1, n) \subset G$ under the abelianization homomorphism $f : G \to G^{ab}$ equals the ball $B(1, n) \subset \langle x\rangle \oplus \langle y\rangle$. The latter has $4n^2 + 1$ elements. Sicne each $U_{klm} \in B(1, n)$ satisfies

$$n \geqslant \sqrt{|m|},$$

it follows that

$$f^{-1}(x^k y^l) \cap B(1, n)$$

contains at most $2n^2 + 1$ elements. Thus, the ball $B(1, n) \subset G$ has cardinality at most

$$(4n^2 + 1)(2n^2 + 1)$$

and $\mathfrak{G}_G(n) \precsim n^4$. □

12.1.2. Distortion of subgroups of solvable groups.

LEMMA 12.12. *Let $G = \mathbb{Z}^m \rtimes_M \mathbb{Z}$, where $M \in GL(m, \mathbb{Z})$.*
If M has an eigenvalue with absolute value different from 1 then

$$\Delta_G^{\mathbb{Z}^m}(n) \asymp e^n \,. \tag{12.3}$$

PROOF. Note that (12.3) is equivalent to the existence of constants $b \geqslant a > 1$ and $c_i > 0$, $i = 1, 2$, such that for every $n \in \mathbb{N}$,

$$c_1 a^n \leqslant \Delta_G^{\mathbb{Z}^m}(n) \leqslant c_2 b^n \,. \tag{12.4}$$

Lower bound. There exists N such that M^N has an eigenvalue with absolute value at least 2. According to Proposition 5.96, we may replace in our arguments the group G by the finite index subgroup $\mathbb{Z}^m \rtimes (N\mathbb{Z})$. Thus, without loss of generality, we may assume that M has an eigenvalue with absolute value at least 2.

Lemma 11.29 implies that there exists a vector $v \in \mathbb{Z}^m$ such that the map

$$\begin{array}{ccc} \mathbb{Z}_2^{k+1} & \to & \mathbb{Z}^m \\ s = (s_n) & \mapsto & s_0 v + s_1 M v + \ldots + s_k M^k v \end{array}$$

is injective. If we denote by t the generator of the factor $\mathbb{Z}$ and we use the multiplicative notation for the operation in the group G, then the element

$$w_s = s_0 v + s_1 M v + \ldots + s_k M^k v \in \mathbb{Z}^m$$

can be rewritten as

$$w_s = v^{s_0} (t v t^{-1})^{s_1} \cdots (t^k v t^{-k})^{s_k} \,.$$

Thus we obtain 2^{k+1} elements of $\mathbb{Z}^m$ of the form w_s, and if we assume that t and v are in the generating set defining the metric, the length of all these elements is at most $3k + 1$.

In the subgroup $\mathbb{Z}^m$ we consider the generating set $X = \{e_i \mid 1 \leqslant i \leqslant m\}$, where e_i is the i-th element in the canonical basis. Then for every $w \in \mathbb{Z}^m$, $|w|_X = |w_1| + \cdots + |w_m|$, i.e., $|w|_X = \|w\|_1$, where $\| \, \|_1$ denotes the ℓ_1–norm on $\mathbb{R}^m$.

Define the number

$$r = \max\{\|w_s\|_1 : s = (s_n) \in \mathbb{Z}_2^{k+1}\} \,.$$

The ball in $(\mathbb{Z}^m, \| \, \|_1)$ with center 0 and radius r contains all the products w_s, i.e., 2^{k+1} elements, whence $r^m \succeq 2^{k+1}$, and $r \succeq a_1^k$, where $a_1 = 2^{\frac{1}{m}}$.

We have thus obtained that $\Delta_G^{\mathbb{Z}^m}(3k+1) \succeq a_1^k$, whence $\Delta_G^{\mathbb{Z}^m}(n) \succeq a^n$, where $a = a_1^{\frac{1}{3}}$.

Upper bound. Consider the generating set $X = \{e_i \mid 1 \leqslant i \leqslant m\}$ in $\mathbb{Z}^m$ and the generating set $S = X \cup \{t\}$ in G. Let w be an element of $\mathbb{Z}^m$ such that $|w|_S \leqslant n$. It follows that

$$w = t^{k_0} v_1 t^{k_1} v_2 \cdots t^{k_{\ell-1}} v_\ell t^{k_\ell} \,, \tag{12.5}$$

where $k_j \in \mathbb{Z}$, k_0 and k_ℓ possibly equal to 0 but all the other exponents of t are non-zero, $v_j \in \mathbb{Z}^m$, and

$$\sum_{j=0}^{\ell} |k_j| + \sum_{j=1}^{\ell} \|v_j\|_1 \leqslant n \,.$$

We may rewrite (12.5) as

(12.6)
$$w = \left(t^{k_0} v_1 t^{-k_0}\right)\left(t^{k_0+k_1} v_2 t^{-k_0-k_1}\right)\cdots\left(t^{k_0+\ldots+k_{\ell-1}} v_\ell t^{-k_0-\ldots-k_{\ell-1}}\right) t^{k_0+\ldots+k_{\ell-1}+k_\ell}\,.$$

The uniqueness of the decomposition of every element in G as wt^q with $w \in \mathbb{Z}^m$ and $q \in \mathbb{Z}$, implies that $k_0 + .. + k_{\ell-1} + k_\ell = 0$. With this correction, the decomposition in (12.6), rewritten with the additive notation and using the fact that $t^k v t^{-k} = M^k v$ for every $v \in \mathbb{Z}^m$, is as follows

$$w = M^{k_0} v_1 + M^{k_0+k_1} v_2 + \cdots + M^{k_0+\ldots+k_{\ell-1}} v_\ell\,.$$

Let α_+ be the maximum among absolute values of the eigenvalues of M, α_- be the maximum of absolute values of eigenvalue of M^{-1}; set $\alpha = \max(\alpha_+, \alpha_-)$.

In $GL(m, \mathbb{C})$ the matrix M can be written as $PDUP^{-1}$, where D is diagonal, U is upper triangular with entries 1 on the diagonal and $DU = UD$ (the multiplicative Jordan decomposition of M).

Then $M^k = PD^kU^kP^{-1}$, and $\|M^k\| \leqslant \lambda\|D^k\|\|U^k\| \leqslant \lambda'\alpha^{|k|}k^m \leqslant \mu\beta^{|k|}$, for an arbitrary $\beta > \alpha$ and all sufficiently large values of k. Therefore,

$$\|w\|_1 \preceq \|M^{k_0}\|\,\|v_1\|_1 + \|M^{k_0+k_1}\|\,\|v_2\|_1 + \cdots + \|M^{k_0+..+k_{\ell-1}}\|\,\|v_\ell\|_1 \preceq$$

$$\beta^{|k_0|}\,\|v_1\|_1 + \beta^{|k_0|+|k_1|}\,\|v_2\|_1 + \cdots + \beta^{|k_0|+..+|k_{\ell-1}|}\,\|v_\ell\|_1 \preceq \beta^n\, n \preceq \beta^{2n}\,.$$

We thus conclude that $\Delta_G^{\mathbb{Z}^m}(n) \preceq \beta^{2n}$. □

Example 12.13. Let $G := \left\langle a, b : aba^{-1} = b^p \right\rangle$, $p \geqslant 2$. Then the subgroup $H = \langle b \rangle$ is exponentially distorted in G.

Proof. To establish the lower exponential bound note that:

$$g_n := a^n b a^{-n} = b^{p^n},$$

hence $d_G(1, g_n) = 2n + 1$, $d_H(1, g_n) = p^n$, hence

$$\Delta_G^H(R) \geqslant p^{[(R-1)/2]}.$$

We will leave the upper exponential bound as an exercise. □

Exercise 12.14. Consider the group

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \;;\; a = 2^n\,,\, b = \frac{m}{2^k}\,,\, n, m, k \in \mathbb{Z} \right\}.$$

Note that G has a finite generating set consisting of matrices $d = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

(1) Prove that the group G has exponential growth.

(2) Prove that the cyclic subgroup generated by u has exponential distortion.

12.1.3. Distortion of subgroups in nilpotent groups. The goal of this section is to estimate the distortion function Δ_G^H of subgroups $H \leqslant G$ of nilpotent groups. These estimates will be used in the proof of the Bass–Guivarc'h Theorem (Theorem 12.26).

LEMMA 12.15. *Let G be a finitely generated nilpotent group of class k and let C^kG be the last non-trivial element in its lower central series. If S is a finite set of generators for G and g is an arbitrary element in C^kG then there exists a constant $\lambda = \lambda(S, g)$ such that*

$$|g^n|_S \leqslant \lambda n^{\frac{1}{k}} \text{ for every } n \in \mathbb{N}.$$

PROOF. We argue by induction on k. The statement is clearly true for $k = 1$. Assume that it is true for k and consider G, a $(k+1)$–step nilpotent group.

Note that $C^{k+1}G$ is central in G, in particular it is abelian. The subgroup $C^{k+1}G$ also has a finite set of generators of the form $[s, c]$, with $s \in S$ and $c \in C^kG$ (e.g., we can take as the generators of $C^{k+1}G$ the inverses of $(k+1)$–fold left commutators of generators of G, see Lemma 11.44). Since $C^{k+1}G$ is abelian, it suffices to prove the statement of lemma for g equal to one of these generators $[s, c]$.

The formulae (3) and (4) in Lemma 11.30, imply that for every $x, x' \in G$ and $y, y' \in C^kG$ we have

$$[x, yy'] = [x, y][x, y'] \text{ and } [xx', y] = [x, y][x', y]\,. \tag{12.7}$$

Here we used the fact that $C^{k+1}G$ is central to deduce that $[y, [x, y']] = 1$ and $[x, [x', y]] = 1$, and to swap $[x, y]$ and $[x', y]$.

In particular

$$[x, y^a] = [x, y]^a \text{ and } [x^b, y] = [x, y]^b\,. \tag{12.8}$$

Given n we let q denote the smallest integer such that $q > n^{\frac{1}{k+1}}$. Note that our goal is to show that $|[s, c]^n|_S$ is bounded by $\lambda\, q$ for a suitable choice of λ.

There exist two positive integers a, b such that $n = aq^k + b$ and $0 \leqslant b < q^k$; moreover, $n < q^{k+1}$ implies that $a < q$. The formulas in (12.8) then imply that

$$[s, c]^n = \left[s^a, c^{q^k}\right]\left[s, c^b\right]\,.$$

The induction hypothesis applied to the group $G/C^{k+1}G$ (where the finite generating set of the quotient is the image of S), and to the element $c\,C^{k+1}G$, implies that $c^{q^k} = k_1 z_1$ and $c^b = k_2 z_2$, where $|k_i|_S \leqslant \mu q$, for a constant $\mu = \mu(S, c)$, and $z_i \in C^{k+1}G$, for $i = 1, 2$.

The formulas (12.7) imply that for every $x \in G$, $[x, k_i z_i] = [x, k_i]$. Therefore

$$[s, c]^n = [s^a, k_1]\,[s, k_2]\,,$$

whence $[s, c]^n$ has S-length at most

$$2(a + \mu q) + 2(1 + \mu q) \leqslant 4(1 + \mu)q \leqslant 8(1 + \mu)n^{\frac{1}{k+1}}.$$

Thus, we can take $\lambda = 8(1 + \mu)$. □

COROLLARY 12.16. *Let G be a finitely generated nilpotent group of class k and let $H := C^kG$ be the last non-trivial element in its lower central series. Then:*

(1) *The restriction of the distance function from G to H satisfies $\mathrm{dist}_G(1, g) \preceq \mathrm{dist}_H(1, g)^{\frac{1}{k}}$, $g \in H$.*

(2) *If H is infinite then its distortion function satisfies $\Delta_G^H(n) \succeq n^k$.*

PROOF. The group $H = C^kG$ is abelian, hence isomorphic to $\mathbb{Z}^m \times F$ for some $m \in \mathbb{N}$ and a finite abelian group F. Let $\{t_1, \ldots, t_m\}$ be a basis for $\mathbb{Z}^m$ and $\tau_1, \ldots, \tau_q$ respective generators of the cyclic factors of F. We consider the word metric in H corresponding to the generating set $\{t_1, \ldots, t_m, \tau_1, \ldots, \tau_q\}$. Take the shortest word w in this generating set representing g,

$$g = t_1^{\alpha_1} \cdots t_m^{\alpha_m} \tau_1^{\beta_1} \cdots \tau_q^{\beta_q} .$$

Then

$$\mathrm{dist}_H(1, g) = \sum_{i=1}^{m} |\alpha_i| + \sum_{j=1}^{q} |\beta_j| . \tag{12.9}$$

Let D denote the diameter of the finite group F with respect to dist_S and let

$$\lambda := \max_i \lambda(S, t_i),$$

where $\lambda(S, t_i)$ is as in Lemma 12.15. Then:

$$|g|_S \leqslant \sum_{i=1}^{m} |t_i^{\alpha_i}|_S + |\tau_1^{\beta_1} \cdots \tau_q^{\beta_q}|_S \leqslant \lambda \sum_{i=1}^{m} |\alpha_i|^{\frac{1}{k}} + D . \tag{12.10}$$

Now, the statement (1) follows from (12.9) and (12.10). The statement (2) is an immediate consequence of (1). □

Our next goal is to prove the inequalities opposite to those in Corollary 12.16.

LEMMA 12.17. *If X is a finite generating set for a nilpotent group G, then there exists a finite generating set $\widehat{X}$ containing X and such that:*

1. For every $x, y \in \widehat{X}$, $[x, y] \in \widehat{X}$.

2. Whenever $x \in C^iG$ projects to an element of the order d in $C^iG/C^{i+1}G$, we have $x^d \in \widehat{X}$.

PROOF. We define inductively certain finite subsets of G. Let $T_1 = X$. Given T_i, we define T_{i+1} to consist of all commutators of the elements of T_i together with elements of the form x^d, where $x \in C^mG$ projects to an element of the order d in $C^mG/C^{m+1}G$ $(m \geq 1)$. We next claim that for every m,

$$T_m \subseteq C^mG.$$

Indeed, $T_2 \subseteq C^2G$. Assume inductively that $T_i \subseteq C^iG$. For the commutators $[a, b]$ with $a, b \in T_i \subset C^iG$, clearly,

$$[a, b] \in [C^iG, G] = C^{i+1}G.$$

Similarly, if x^d projects to 1 in $C^iG/C^{i+1}G$, then, $x^d \in C^{i+1}G$. Thus, $T_m \subseteq C^mG$ for each m. We then take

$$\widehat{X} := \bigcup_{i \geqslant 1} T_i.$$

□

Definition 12.18. Let G be a finitely generated nilpotent group. We call a finite generating set S of G an *lcs–generating set* (where lcs stands for the 'lower central series') if for every $i \geqslant 1$, the subset $F^i(S) := S \cap C^iG$ generates C^iG. For such a generating set we denote by $F_i(S)$ the complement $F^i(S) \setminus F^{i+1}(S)$. We say that an lcs–generating set T of G is *closed* if $T = \widehat{T}$, where $\widehat{T}$ is defined as in the proof of Lemma 12.17.

Note that for any generating set X, the set $\widehat{X}$ is a closed lcs-generating set, according to Lemma 11.44. Observe also that the projection of an lcs-generating set to each quotient $G/C^{i+1}G$ is again an lcs-generating set.

Definition 12.19. If G is a finitely generated nilpotent group and S is an *lcs–generating set* of G, then for any word w in $S \cup S^{-1}$ we define its length $|w|_S$ as usual and its *m–length* $|w|_m$ as the number of occurrences of letters from $F^m(S) \cup (F^m(S))^{-1}$ in the word w.

The *n-lcs–length of* a word w is the finite sequence $(|w|_0, |w|_1, \dots, |w|_n)$. An element g in G is said to have *lcs–length at most* $(r_0, r_1, .., r_n)$,

$$lcs_S(g) \leqslant (r_0, r_1, \dots, r_n),$$

if g can be expressed as a word in $S \cup S^{-1}$ of the n-lcs–length $(m_0, m_1, \dots, m_n)$, with $m_i \leqslant r_i$ for all $i, 0 \leqslant i \leqslant n$.

We now are ready to prove the following:

Proposition 12.20. *Let G be a finitely generated nilpotent group of class $N+1$ and let $H := C^kG$, $k \leqslant N$. Then:*

(1) *For $g \in H$, the distance function satisfies*

$$\mathrm{dist}_H(1, g) \asymp \mathrm{dist}_G(1, g)^k \,.$$

(2) *The distortion function $\Delta_G^H(n)$ of H in G satisfies*

$$\Delta_G^H(n) \asymp n^k.$$

The statement (2) is an immediate consequence of (1). We now prove (1). The relation $\mathrm{dist}_G(1, g)^k \preceq \mathrm{dist}_H(1, g)$ is proven in Corollary 12.16, (1). In what follows we prove $\mathrm{dist}_H(1, g) \preceq \mathrm{dist}_G(1, g)^k$. The main step in the proof is the following lemma.

Lemma 12.21. *Let G be a finitely generated torsion-free nilpotent group of class $N+1$ with an lcs generating set S. Then there exists a sequence of closed lcs–generating sets $S^{(k)}$ of subgroups C^kG, $k = 1, \dots, N$, such that the following holds:*

For every pair of numbers $\lambda \geqslant 1$, $r \geqslant 1$ and every element $g \in C^kG$ with $lcs_S(g) \leqslant (\lambda r, \lambda r^2, \dots, \lambda r^k)$, we have

$$lcs_{S^{(k)}}(g) \leqslant \lambda_k r^k$$

where λ_k depends only on S and on λ.

Proof. We construct the generating sets $S^{(n)}, n = 1, \dots, k$ by induction on n. The construction is clear for $n = 1$ since we can simply take $S^{(1)} = \widehat{S}$: The new lcs generating set of G is closed and the word length (as well as the lcs–length) increases only by a constant factor depending only on S, see Exercise 4.83 or Theorem 5.35.

We will describe only the induction step $1 \to 2$ since the general induction $n \to n+1$ is identical (replacing $G = C^1G$ with C^nG). We will also verify the inequality

$$lcs_{S^{(2)}}(g) \leqslant (\lambda_2 r^2, \lambda_2 r^3, \ldots, \lambda_2 r^k).$$

Step 1: Construction of the generating sets. Our goal is to define (given a generating set $S^{(1)}$ and a number $\lambda \geqslant 1$) a new lcs generating set T for G and a constant λ_2, such that whenever $g \in C^2G$ satisfies

$$lcs_{S^{(1)}}(g) \leqslant (\lambda r, \lambda r^2, ..., \lambda r^k),$$

we also have

$$lcs_T(g) \leqslant (0, \lambda_2 r^2, \ldots, \lambda_2 r^k).$$

We then will take $S^{(2)} := F^2(T)$.

We first modify the generating set $S^{(1)}$ by replacing it with another closed lcs generating set T such that

$$F_1(T) = \{t_1, \ldots, t_m\}$$

projects to a *standard* generating set of the abelianization $G_{ab} = G/G' = G/C^2G$ (see Definition 11.14). At the same time, whenever $t = t_j \in F_1(T)$ projects to an element of the order $d = d_j$ in $G^{ab} = G/C^2G$, we replace t^{-1} with $t^{d-1}t^d$, $t^d \in F^2(T)$. This increases lengths of all words by a uniformly bounded factor c and we obtain

$$lcs_T(g) = lcs_T(w) \leqslant (c\lambda r, c\lambda r^2, \ldots, c\lambda r^k).$$

Here $w = w_0$ is a word in T representing g.

Let $\ell = \ell_1$ denote $|w_0|_1$, i.e., the number of times the letters $t_p^{\pm 1}$, $t_p \in F_1(T)$, appear in w_0. We next construct, by induction on $j \leqslant \ell$, a sequence of words $(w_j)_{0\leqslant j\leqslant \ell}$ all representing g, such that for every j, $w_j = v_{j-1}t_{q_j}^{\pm 1}u_j$ for some $t = t_{q_j} \in F_1(T)$, and:

(a) $t^{\pm 1}$ occurs at most $\ell - j$ times in u_j.

(b) For every $i \geqslant 1$,

$$|w_j|_i \leqslant |w_{j-1}|_i + |w_{j-1}|_{i-1}$$

where, we set $|x|_0 := 0$ for every word x.

(c) $v_j = v_{j-1}t_{q_j}^{\pm 1}$ with v_0 being the empty word.

(d) In the words v_j, letters $t_p^{\pm 1}$ always appear in the increasing order with respect to p, i.e., $t_p^{\pm 1}$ is always to the left of $t_q^{\pm 1}$, whenever $p < q$.

Namely, w_1 is obtained from w_0 by considering the left-most occurrence of $t_1^{\pm 1}, t_1 \in F_1(T)$, and moving $t_1^{\pm 1}$ to the front of the word *via* the relations

$$xt_1^{\pm 1} = t_1^{\pm 1}x\left[x^{-1}, t_1^{\mp 1}\right], \quad x \in T.$$

We then find the second left-most appearance of $t_1^{\pm 1}$ and again move it all the way to the left. We then proceed inductively. Whenever we see the product $t_1t_1^{-1}$ or $t_1^{-1}t_1$, we, of course, cancel these generators. If t_1 projects to a generator of the order $d = d_1$ in G^{ab}, and, at some point of the process, we see a d-fold product

$$\underbrace{t_1 \ldots t_1}_{d \text{ times}},$$

we replace this product with the generator $t_1^d \in F_2(T)$. Once we are done with the generator t_1, we move the letters $t_2^{\pm 1}$ as far left as possible, but not beyond the power t_1 constructed earlier. Thus, inductively, w_{j+1} is obtained from w_j by considering the left-most occurrence of $t^{\pm 1}_{q_{j+1}}$ in u_j and moving it to the left in the similar fashion, where r_j is a weakly increasing sequence:

$$r_1 \leqslant r_2 \leqslant r_3 \ldots$$

Note that for each $t = t_{r_j}$, the commutator $\left[x^{-1}, t^{\mp 1}\right]$ is in the set $F^2(T)$ and the number of occurrences of $t^{\pm 1}$ does not increase in this process. Then properties (a)—(d) are immediate.

In the end of the induction process, we convert w to a word w_ℓ which has the form

$$t_1^{\epsilon_1} \ldots t_m^{\epsilon_m} w',$$

where w' is a word in the alphabet $F^2(T) \cup (F^2(T))^{-1}$ and, whenever t_j projects to an element of the finite order d_j in G^{ab}, we have

$$0 \leqslant \epsilon_j < d_j.$$

Since the set $F_1(T)$ projects to a standard generating set of G_{ab} and $g \in C^2G$, it follows that

$$t_1^{\epsilon_1} \ldots t_m^{\epsilon_m} = 1$$

in G; thus, the element $g \in G$ is represented by the word w'. This concludes the construction of the generating set

$$S^{(2)} = F^2(T)$$

. In the process, we also transform each word w representing $g \in C^2G$ to a new word w' in the new generating set.

Step 2: Estimating the length of the word w'.

Using (b) we obtain that for every $i = 1, \ldots, k$, and $j = 1, \ldots, \ell$,

$$|w_j|_i \leqslant \sum_{s=0}^{a} \begin{pmatrix} j \\ s \end{pmatrix} |w_0|_{i-s} \tag{12.11}$$

where $a = \min(i-1, j)$. The induction step follows from the formula

$$\begin{pmatrix} j \\ s \end{pmatrix} + \begin{pmatrix} j \\ s-1 \end{pmatrix} = \begin{pmatrix} j+1 \\ s \end{pmatrix}.$$

Since $|w_0|_{i-s} \leqslant c\lambda r^{i-s}$ and

$$\begin{pmatrix} j \\ s \end{pmatrix} \leqslant j^s \leqslant \ell^s \leqslant \lambda r^s,$$

we obtain:

$$|w_j|_i \leqslant \sum_{s=0}^{a} \ell^s c\lambda r^{i-s} \leqslant \sum_{s=0}^{a} c\lambda r^s \lambda r^{i-s} \leqslant i\lambda^i r^i.$$

In particular,

$$|w'|_i \leqslant |w_\ell|_i \leqslant ci\lambda^i r^i \leqslant \lambda_2 r^i,$$

where $\lambda_2 = ck\lambda^k$.

As we observed before, the construction of the generating sets $S^{(i)}$ is similar, we note that (for $g \in C^{i+1}G$) the inequality

$$lcs_{S^{(i)}}(g) \leqslant (\lambda_i r^i, \lambda_i r^{i+1}, \ldots, \lambda_i r^k)$$

implies

$$lcs_{S^{(i+1)}}(g) \leqslant (\lambda_{i+1} r^{i+1}, \lambda_{i+1} r^{i+2}, \ldots, \lambda_{i+1} r^k).$$

In particular, for $g \in C^k G$, we have

$$lcs_{S^{(k)}}(g) \leqslant \lambda_k r^k,$$

where λ_k is independent of r and g. □

Now, we can conclude the proof of Proposition 12.20, Part (2). We start with an lcs–generating set S of G. Applying Lemma 12.21, we obtain a new lcs generating set $T := S^{(k)}$ of $H = C^k G$. It suffices to prove the inequality

$$|g|_T \leqslant \mu |g|_S^k$$

for some constant μ independent of $g \in H$. Let w be a shortest word in the alphabet $S \cup S^{-1}$ representing an element $g \in H$, let r denote the length of w. Since $S^{(1)} = \widehat{S}$,

$$lcs_S(w) = (r, 0, \ldots, 0) \leqslant (r, r^2, \ldots, r^k).$$

According to Lemma 12.21, g is represented by a word w' in the alphabet $T \cup T^{-1}$ such that

$$|w'|_T \leqslant \mu r^k,$$

where $\mu = \lambda_k$ for $\lambda = 1$ and, thus, μ depends only on S. Therefore,

$$|g|_T \leqslant |w'|_T \leqslant \mu r^k = \mu |g|_S^k \ .$$

□

Another interesting consequence of Lemma 12.21 is a control on the exponents of the bounded generation property for nilpotent groups.

Proposition 12.22 (Controlled bounded generation for nilpotent groups). *Let G be a finitely generated nilpotent group of class N. For each i we let*

$$S_i = \{t_{i1}, \ldots, t_{iq_i}\} \subset C^i G$$

be a subset projecting bijectively to a standard generating set of the abelian group $C^i G / C^{i+1} G$. Define

$$S = \bigcup_{i=1}^{k} S_i$$

(thus, S is an lcs–generating set of G). Then every element $g \in G$ can be written as a product

$$g = \prod_{i=1}^{k} t_{i1}^{m_{i1}} \cdots t_{iq_i}^{m_{iq_i}} \ ,$$

so that:

1. $m_{ij} \in \mathbb{Z}$ and $0 \leqslant m_{ij} < d_{ij}$, if $d_{ij} < \infty$ is the order of the projection of t_{ij} to $C^i G / C^{i+1} G$.

2. $|m_{i1}| + \ldots + |m_{iq_i}| \leqslant C |g|_S^i$, for every $i \in \{1, \ldots, k\}$, , where C is a constant depending only on G and on S.

Proof. We argue by induction on the class N. For $N = 1$ the group is abelian and the statement is obvious. Assume that the statement is true for the class $N-1$ and let G be a nilpotent group of class $N \geqslant 2$. Let S_i and S be as in the statement of the proposition, and let g be an arbitrary element in G. The induction hypothesis implies that $g = pc$, where $c \in C^N G$ and

$$p = \prod_{i=1}^{k-1} t_{i1}^{m_{i1}} \cdots t_{iq_i}^{m_{iq_i}} ,$$

where $m_{ij} \in \mathbb{Z}$ are such that $0 \leqslant m_{ij} < d_{ij}$ (if the order d_{ij} is finite) and

$$|m_{i1}| + \ldots + |m_{iq_i}| \leqslant C|g|_S^i , \qquad (12.12)$$

for every $i \in \{1, .., k\}$, where C is a constant depending only on G and S.

Then, by the inequalities (12.12), the element $c = p^{-1}g$ in $C^N G$ has lcs–length with respect to S at most $(\lambda r, \lambda r^2, \ldots, \lambda r^N)$, where $r = |g|_S$ and $\lambda = C+1$. Lemma 12.21 then implies that there exists a new generating set T of $C^N G$, (determined by S), such that $|c|_T \leqslant \mu r^N$, where μ only depends on T. By the construction of this generating set, it is a standard generating set of the abelian group $C^N G$. Then

$$c = t_{k1}^{m_{k1}} \cdots {}_{kq_k}^{m_{kq_k}} ,$$

where

$$|m_{N1}| + \ldots + |m_{Nq_N}| \leqslant \eta |c|_S \leqslant \eta \mu r^N ,$$

where η depends only on S. Now, the assertion follows by combining product decompositions of p and c. □

Proposition 12.20 generalizes to all subgroups:

Proposition 12.23. *Let G be a finitely generated nilpotent group of class N. Then, for every subgroup H in a G, $\Delta_G^H(n) \preceq n^\gamma$, where $\gamma = 2^{N-1}$.*

Proof. We will first prove a general lemma about distortion of subgroups.

Suppose that we have a commutative diagram of finitely generated groups where the horizontal sequences are short exact and the vertical arrows are injective:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & H' & \longrightarrow & H & \xrightarrow{\psi} & H'' & \longrightarrow & 1 \\ & & \uparrow{\scriptstyle f'} & & \uparrow{\scriptstyle f} & & \uparrow{\scriptstyle f''} & & \\ 1 & \longrightarrow & G' & \longrightarrow & G & \xrightarrow{\phi} & G'' & \longrightarrow & 1 \end{array}$$

We will, henceforth, identify H', H, H'' with subgroups of G', G, G'' respectively. We let X, Y denote finite symmetric generating sets of G, H respectively, such that $f(Y) \subset X$, $X' = X \cap G', X'' = \phi(X)$, $Y'' = \psi(Y)$, $Y' = Y \cap H'$ are generating sets of the respective groups.

Lemma 12.24.

$$\Delta_G^H(m) \leqslant 2\Delta_{G'}^{H'} \circ \Delta_G^{G'}(2\Delta_{G''}^{H''}(m)).$$

Proof. We now take $h \in H$ such that $|f(h)|_X = m$. Then

$$|\phi(h)|_{X''} \leqslant n.$$

The distortion bound on the inclusion $H'' \hookrightarrow G''$ yields existence of a word w'' in the alphabet Y'' whose length is

$$|w''|_{Y''} \leqslant \Delta_{G''}^{H''}(m),$$

such that w'' represents the projection $\psi(h)$. We lift w'' to a word $\tilde{w}''$ in Y and let $\tilde{h}'' \in H$ be the element represented by $\tilde{w}''$. The product

$$h' := h \cdot (\tilde{h}'')^{-1},$$

belongs to the subgroup H'. We would like to bound the norm $|h'|_{Y'}$ from above using the distortion function

$$\Delta_{G'}^{H'}.$$

To this end, consider the word $v'' = f''(w'')$ (in the alphabet X'') and its lift $\tilde{v}''$ (in the alphabet X); the latter word represents the element $f(\tilde{h}'') \in G$. Clearly, v'' has length $\leqslant \Delta_{G''}^{H''}(m)$. We have

$$f'(h') = f(h) \cdot f(\tilde{h}'')^{-1},$$

which implies that

$$|f'(h')|_X \leqslant |f(h)|_X + |\tilde{v}''| \leqslant m + \Delta_{G''}^{H''}(m).$$

Therefore,

$$|f'(h')|_{X'} \leqslant \Delta_G^{G'}(m + \Delta_{G''}^{H''}(m)),$$

and, hence,

$$|h'|_{Y'} \leqslant \Delta_{G'}^{H'}\left(\Delta_G^{G'}(m + \Delta_{G''}^{H''}(m))\right).$$

By putting the inequalities together, we obtain

$$|h|_Y \leqslant |w''|_{Y''} + |h'|_{Y'} \leqslant \Delta_{G''}^{H''}(m) + \Delta_{G'}^{H'}\left(\Delta_G^{G'}(m + \Delta_{G''}^{H''}(m))\right).$$

Since each distortion function Δ_A^B satisfies

$$\Delta_A^B(k) \geqslant k,$$

we obtain:

$$|h|_Y \leqslant 2\Delta_{G'}^{H'}\left(\Delta_G^{G'}(2\Delta_{G''}^{H''}(m))\right),$$

as required. □

With this lemma in mind, we will now prove the proposition by induction on the nilpotency class N of G. The statement is obviously true for $N = 1$ since subgroups of abelian groups are undistorted. Assume that proposition holds for N and consider a subgroup H in a group G of nilpotency class $N+1$. Set $G' = C^2G$, $H' := H \cap G'$ and let $G'' = G/G' = G^{ab}$, $H'' = H/H'$, the projection of H into G^{ab}. We have (for suitable finite generating sets):

$$\Delta_{G''}^{H''}(m) = m,$$

(since subgroups of abelian groups are undistorted),

$$\Delta^{H'}G'(s) \leqslant C_1 \cdot s^{2^{N-2}}$$

(by the induction hypothesis), and

$$\Delta^{G'}G(t) \leqslant C_2 \cdot t^2$$

(by Proposition 12.20). Now, the inequality

$$\Delta^{H'} G'(m) \leqslant 2C_1 \left(C_2(2m)^2\right)^{2^{N-2}} = 2C_1(4C_2)^{2^{N-2}} m^{2^{N-1}}. \quad \square$$

In fact, a stronger statement holds:

THEOREM 12.25. *For every infinite subgroup H in a finitely generated nilpotent group G there exists a rational positive number α such that*

$$\Delta_G^H(n) \asymp n^{\alpha} .$$

Theorem 12.25 was originally proven by M. Gromov in **[Gro93]** (see also **[Var99]**); later on, an explicit computation of the possible exponents α was established by D. Osin in **[Osi01]**. More precisely, given an element of infinite order h in a nilpotent group G, its *weight* in G, $\nu_G(h)$, is the defined as the maximal i such that $\langle h\rangle \cap C^iG \neq \{1\}$. The exponent α in Theorem 12.25 is the maximum of the fractions

$$\frac{\nu_G(h)}{\nu_H(h)}$$

over all elements $h \in H$ of infinite order.

12.2. Polynomial growth of nilpotent groups

Let G be a finitely generated nilpotent group and $d = d(G)$ is the *homogeneous dimension* of G, see Definition 11.46.

THEOREM 12.26 (Bass–Guivarc'h Theorem). *Then the growth function of G satisfies*

$$\mathfrak{G}_G(n) \asymp n^d . \tag{12.13}$$

PROOF. In the proof below, m_i is the free rank of $C^iG/C^{i+1}G$; λ_i's are constants depending only on the generating set of the group G. We will use the notation $B_G(1,r)$ to denote the r-ball in the group G centered at $1 \in G$, with the word metric given by a suitable finite generating set of G.

We argue by induction on the class k of nilpotence of G. For $k = 1, d = m_1$ the group G is abelian of free rank d and the statement is obvious.

Assume that the statement holds for $k-1$ and consider G of class $k \geqslant 2$; let $H = C^kG$ be the last non-trivial subgroup in the lower central series of G. Let $d_1 := d - k\,m_k$ be the homogeneous degree of H. If H is finite then $m_k = 0$, we apply the induction hypothesis for G/H; since G and G/H have equivalent growth functions the result follows.

We now assume that H is infinite, i.e., $m_k \geqslant 1$. Fix a finite generating set S of G and use its projection as the generating set of G/H.

Lower bound. By our choice of generating sets, the ball $B_G(1,r)$ maps onto the ball $B_{G/H}(1,r)$ under the projection $G \to G/H$. The induction hypothesis applied to G/H implies that there exists λ_1 for which

$$N = card(B_{G/H}(1,n)) \geqslant \lambda_1 n^{d_1}.$$

Let $\{g_1, \ldots, g_N\} \subset B_G(1,n)$ denote the preimage of $B_{G/H}(1,n)$. Since the abelian group H has growth function t^{m_k}, and, according to Part (1) of Corollary 12.16, (for some μ independent of n)

$$B_G(1,n) \cap H \supset B_H(1, \mu n^k),$$

we conclude that

$$card(B_G(1,n)\cap H)\geqslant \lambda_2 n^{km_k},$$

for some λ_2 independent of n.

Therefore, the ball $B_G(1,2n)$ contains the set

$$\bigcup_{i=1}^{N} g_i(B_G(1,n)\cap H)$$

of cardinality at least

$$N\lambda_2 n^{km_k}\geqslant \lambda_1\lambda_2 n^{d_1+km_k}=\lambda_3 n^d=\lambda_3 2^{-d}(2n)^d.$$

Thus, for even $t=2n$,

$$\mathfrak{G}_G(t)\geqslant \lambda_4 t^d.$$

The case of odd t is left as an exercise to the reader.

PICTURE

Upper bound. The proof is analogous to the lower bound. Recall that the image of $B_G(1,n)$ in G/H is the ball $B_{G/H}(1,n)$. By the induction hypothesis there exist at most $\lambda_5 n^{d_1}$ elements

$$\bar{g}_1,\ldots,\bar{g}_N\in B_{G/H}(1,n),$$

which are projections of elements $g_i\in B_G(1,n)$, $i=1,\ldots,N$. For each element

$$g=g_ih\in g_iH\in B_G(1,n)$$

we have

$$|h|_S\leqslant |g|_S+|g_i|_S\leqslant 2n.$$

By Proposition 12.20 there are at most $\lambda_6 n^{km_k}$ elements $h\in H$ satisfying this inequality. It then follows that there are at most $\lambda_5\lambda_6 n^{d_1+km_k}=\lambda_7 n^d$ distinct elements $g\in B_G(1,n)$. ☐

12.3. Wolf's Theorem

In this section we classify virtually polycyclic groups according to their growth.

Notation 12.27. If G is a group, a semidirect product $G\rtimes_\Phi \mathbb{Z}$ is defined by a homomorphism $\Phi:\mathbb{Z}\to \mathrm{Aut}\,(G)$. The latter homomorphism is entirely determined by $\Phi(1)=\varphi$. Following the notation in Section 12.1, we set

$$S=G\rtimes_\varphi \mathbb{Z}:=G\rtimes_\Phi \mathbb{Z}$$

Proposition 12.28. *Let G be a finitely generated nilpotent group and let $\varphi\in Aut(G)$. Then the polycyclic group $P=G\rtimes_\varphi\mathbb{Z}$ is*

(1) *either virtually nilpotent;*
(2) *or has exponential growth.*

Remark 12.29. The statement (1) in Proposition 12.28 cannot be improved to 'P is nilpotent', see Remark 12.2, Part (2).

Proof. We note that replacing φ with its power will replace P with its finite index subgroup, and, hence, will not not affect virtual nilpotence of P and its growth rate. The automorphism φ preserves the lower central series of G; let θ_i denote the restriction of φ to C^iG. Then θ_i projects to an automorphism φ_i of the finitely generated abelian group $B_i:=C^iG/C^{i+1}G$. Therefore φ_i induces an automorphism ψ_i of $\mathrm{Tor}\,B_i$ and an automorphism $\overline{\varphi}_i$ of $B_i/\mathrm{Tor}\,B_i\simeq\mathbb{Z}^{m_i}$. Each automorphism $\overline{\varphi}_i$ is determined by a matrix M_i in $GL(m_i, \mathbb{Z})$. Analogously to the proof of Proposition 12.1, we have two case to consider:

(1) All matrices M_i only have eigenvalues of the absolute value 1; hence, by Lemma 11.28, all eigenvalues are roots of unity. Then there exists N such that the matrices of the automorphism φ_i^N have only eigenvalues equal to 1 and the induced automorphisms of finite abelian groups

$$\psi_i : \operatorname{Tor} B_i \to \operatorname{Tor} B_i$$

are all equal to the identity. Hence, without loss of generality we may therefore assume that the matrices M_i of all φ_i's have all eigenvalues equal to 1 and all the ψ_i are the identity automorphisms.

Lemma 11.27 applied to each $\overline{\varphi}_i$ and to each $\psi_i = \mathrm{id}_{\operatorname{Tor} B_i}$, imply that the lower central series of G is a sub-series of a cyclic series

$$\{1\} = H_n \leqslant H_{n-1} \leqslant \ldots \leqslant H_1 \leqslant H_0 = G,$$

where each H_i/H_{i+1} is cyclic, φ preserves each H_i and induces the identity map on H_i/H_{i+1}. We denote by t the generator of the semidirect factor $\mathbb{Z}$ in the decomposition $P = G \rtimes \mathbb{Z}$. Then, by the definition of the semidirect product, for every $g \in G$, $tgt^{-1} = \varphi(g)$. The fact that φ acts as the identity on each H_i/H_{i+1} implies that $t^k(hH_{i+1})t^{-k} = hH_{i+1}$ for every h in H_i; equivalently

$$[t^k, h] \in H_{i+1} \tag{12.14}$$

for every such h.

Since G contains the kernel $C^2P = [P, P]$ of the ableanization map $G \to \mathbb{Z}$, it follows that $C^2P \leqslant G$. We claim that for every $i \geqslant 0$, $[P, H_i] \subseteq H_{i+1}$. Indeed, consider an arbitrary commutator $[h, s]$, $h \in H_i, s \in P$. Since s has the form $s = gt^k$, with $g \in G$ and $k \in \mathbb{Z}$, we obtain:

$$[h, s] = [h, gt^k] = [h, g][g, [h, t^k]][h, t^k] \ ,$$

in view of the commutator identity (3) in Lemma 11.30.

According to (12.14), $[h, t^k] \in H_{i+1}$. Also, since the lower central series of G is a subseries of (H_i), there exists $r \geqslant 1$ such that $C^rG \geqslant H_i \geqslant H_{i+1} \geqslant C^{r+1}G$. Then, $h \in H_i \leqslant C^rG$ and $[h, g] \in C^{r+1}G \leqslant H_{i+1}$. Likewise, as $[h, t^k] \in H_{i+1} \leqslant C^rG$, the commutator $[g, [h, t^k]] \in C^{r+1}G \leqslant H_{i+1}$. By putting it all together, we conclude that $[h, s] \in H_{i+1}$ and, hence, $[P, H_i] \subseteq H_{i+1}$.

The easy induction now shows that $C^{i+2}P \leqslant H_i$ for every $i \geqslant 1$; in particular, $C^{n+2}P \leqslant H_n = \{1\}$. Therefore, P is virtually nilpotent.

(2) Assume that at least one matrix M_i has an eigenvalue with absolute value strictly greater than 1, in particular, $m_i \geqslant 2$. The group P contains the subgroup

$$P_i := C^iG \rtimes_{\theta_i} \mathbb{Z}.$$

Furthermore, the subgroup $C^{i+1}G$ is normal in P_i and

$$P_i/C^{i+1}G \cong B_i \rtimes_{\varphi_i} \mathbb{Z}$$

with a finitely-generated abelian group B_i. Lastly,

$$(B_i \rtimes_{\varphi_i} \mathbb{Z})/\operatorname{Tor} B_i \cong \mathbb{Z}^{m_i} \rtimes_{M_i} \mathbb{Z}.$$

According to Proposition 12.1, the group $\mathbb{Z}^{m_i} \rtimes_{M_i} \mathbb{Z}$ has exponential growth. Therefore, in view of Proposition 5.76, parts (a) and (c), the groups $B_i \rtimes_{\varphi_i} \mathbb{Z}$, $P_i/C^{i+1}G$,

P_i, and, hence, P, all have exponential growth. Thus, in the case (2), S has exponential growth. □

Proposition 12.28 combined with Proposition 3.11 on subgroups of finite index in finitely generated groups will be used to prove Wolf's Theorem, [**Wol68**]:

THEOREM 12.30 (Wolf's Theorem). *A polycyclic group is either virtually nilpotent or has exponential growth.*

PROOF. According to Proposition 11.81, it suffices to prove the statement for poly-C_∞ groups. Let G be a poly-C_∞ group, and consider a finite subnormal descending series

$$G = N_0 \geqslant N_1 \geqslant \ldots \geqslant N_n \geqslant N_{n+1} = \{1\}$$

such that $N_i/N_{i+1} \simeq \mathbb{Z}$ for every $i \geqslant 0$. We argue by induction on n. For $n = 0$ the group G is infinite cyclic and the statement is obvious. Assume that the assertion of theorem holds for n and consider the case of $n+1$. By the induction hypothesis, the subgroup $N_1 \leqslant G$ is either virtually nilpotent or has exponential growth. In the second case the group G itself has exponential growth.

Assume that N_1 is virtually nilpotent. Corollary 4.25 implies that G decomposes as a semidirect product $N_1 \rtimes_\theta \mathbb{Z}$, corresponding to a homomorphism $\Psi : \mathbb{Z} \to \mathrm{Aut}\,(N_1)$, $\theta = \Psi(1)$.

By hypothesis, N_1 contains a nilpotent subgroup H of finite index. According to Proposition 3.11, Part (2), we may moreover assume that H is characteristic in N_1. In particular H is invariant under the automorphisms ψ. We retain the notation θ for the restriction $\theta\big|_H$. Therefore, $H \rtimes_\theta \mathbb{Z}$ is a normal subgroup of G. Moreover, $H \rtimes_\theta \mathbb{Z}$ has finite index in G, since $G/(H \rtimes_\theta \mathbb{Z})$ is the quotient of the finite group N_1/H.

By Proposition 12.28, $H \rtimes_\theta \mathbb{Z}$ is either virtually nilpotent or of exponential growth. Therefore, the same alternative holds for $N_1 \rtimes_\theta \mathbb{Z} = G$. □

12.4. Milnor's theorem

THEOREM 12.31 (J. Milnor, [**Mil68a**]). *A finitely generated solvable group is either polycyclic or has exponential growth.*

We begin the proof by establishing a property of conjugates implied by sub-exponential growth:

LEMMA 12.32. *If a finitely generated group G has sub-exponential growth then for all $\beta_1, \ldots, \beta_m, g \in G$, the set of conjugates*

$$\{g^k \beta_i g^{-k} \mid k \in \mathbb{Z}, i = 1, \ldots, m\}$$

generates a finitely generated subgroup $N \leqslant G$.

PROOF. It suffices to prove lemma for $m = 1$, since N is generated by the subgroups

$$N_i = \left\langle g^k \beta_i g^{-k} \mid k \in \mathbb{Z} \right\rangle, \quad i = 1, \ldots, m.$$

NOTATION 12.33. We set $\alpha := \beta_1$ and let α_k denote $g^k \alpha g^{-k}$ for $k \in \mathbb{Z}$. In the proof we will be identifying $\mathbb{Z}_2$ with the subset $\{0, 1\}$ of $\mathbb{Z}$.

The goal is to prove that finitely many elements in the set $\{\alpha_k \mid k \in \mathbb{Z}\}$ generate the subgroup N.

Consider the map

$$\mu = \mu_m : \prod_{i=0}^{m} \mathbb{Z}_2 \to G$$
$$\mu : (s_i) \mapsto g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} .$$

EXERCISE 12.34. Verify that

$$g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} = \alpha_1^{s_0}\alpha_2^{s_1}\cdots\alpha_{m+1}^{s_m} g^{m+1}.$$

If for every $m \in \mathbb{N}$ the map μ is injective then for each sequence (s_i) we have 2^{m+1} products as above, and if $g, g\alpha$ are in the set of generators of G, all these products are in $B_G(1, m+1)$. This contradicts the hypothesis that G has subexponential growth. It follows that there exists some m and two distinct sequences $(s_i), (t_i)$ in $\prod_{i=0}^{m} \mathbb{Z}_2$ such that

$$g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} = g\alpha^{t_0} g\alpha^{t_1} \cdots g\alpha^{t_m} . \tag{12.15}$$

Assume that m is minimal with this property. This, in particular, implies that $s_0 \neq t_0$ and $s_m \neq t_m$. In view of the exercise, the equality (12.15) becomes

$$\alpha_1^{s_0}\alpha_2^{s_1}\cdots\alpha_{m+1}^{s_m} = \alpha_1^{t_0}\alpha_2^{t_1}\cdots\alpha_{m+1}^{t_m} .$$

Since $s_m \neq t_m$ and $s_m, t_m \in \{0, 1\}$, it follows that $s_m - t_m = \pm 1$. Then

$$\alpha_{m+1}^{\pm 1} = \alpha_m^{-s_{m-1}} \cdots \alpha_2^{-s_1}\alpha_1^{t_0 - s_0}\alpha_2^{t_1}\cdots\alpha_m^{t_{m-1}} . \tag{12.16}$$

If in (12.16) we conjugate by g, we obtain that

$$\alpha_{m+2}^{\pm 1} = \alpha_{m+1}^{-s_{m-1}} \cdots \alpha_3^{-s_1}\alpha_2^{t_0 - s_0}\alpha_3^{t_1}\cdots\alpha_{m+1}^{t_{m-1}} .$$

This and (12.16) imply that α_{m+2} is a product of powers of $\alpha_1, \dots, \alpha_m$. Then, by induction, every α_n with $n \in \mathbb{N}$ is a product of powers of $\alpha_1, \dots, \alpha_m$, and the same is true for α_n with $n \in \mathbb{Z}$ by replacing g with g^{-1}. Therefore, every generator α_n of N belongs to the subgroup of N generated by the elements $\alpha_1, \dots, \alpha_m$ and the elements $\alpha_1, \dots, \alpha_m$ generate N. □

EXERCISE 12.35. Use Lemma 12.32 to prove that the finitely generated group H described in Example 4.8 has exponential growth.

We now are ready to prove Theorem 12.31; our proof by induction on the derived length d of G. For $d = 1$ the group G is finitely generated abelian and the statement is immediate. Assume that the alternative holds for finitely generated solvable groups of derived length $\leqslant d$ and consider G of derived length $d+1$. Then $H = G/G^{(d)}$ is finitely generated solvable of derived length d. By the induction hypothesis, either H has exponential growth or H is polycyclic. If H has exponential growth then G has exponential growth too (see statement (c) in Proposition 5.76).

Assume therefore that H is polycyclic. In particular, H is finitely presented by Proposition 11.85. Theorem 12.31 will follow from:

LEMMA 12.36. *Consider a short exact sequence*

$$1 \to A \to G \xrightarrow{\pi} H \to 1 , \quad \textit{with } A \textit{ abelian and } G \textit{ finitely generated.} \tag{12.17}$$

If H is polycyclic then G is either polycyclic or has exponential growth.

PROOF. We assume that G has sub-exponential growth and will prove that G is polycyclic. The group G is polycyclic iff A is finitely generated. Since H is polycyclic, it has the bounded generation property (see Proposition 11.73); hence, there exist finitely many elements $h_1, \ldots, h_q$ in H such that every element $h \in H$ can be written as

$$h = h_1^{m_1} h_2^{m_2} \cdots h_q^{m_q}\,, \text{ with } m_1, m_2, \ldots, m_q \in \mathbb{Z}\,.$$

Choose $g_i \in G$ such that $\pi(g_i) = h_i$ for every $i \in \{1, 2, \ldots, q\}$. Then every element $g \in G$ can be written as

$$g = g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q} a\,, \text{ with } m_1, m_2, \ldots, m_q \in \mathbb{Z} \text{ and } a \in A\,. \tag{12.18}$$

Since H is finitely presented, by Lemma 4.30 there exist finitely many elements $a_1, \ldots, a_k$ in A such that every element in A is a product of G-conjugates of $a_1, \ldots, a_k$. According to (12.18), all the conjugates of a_j are of the form

$$g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q} a_j \left(g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q}\right)^{-1}\,. \tag{12.19}$$

By Lemma 12.32, the subgroup A_q generated by the conjugates $g_q^m a_j g_q^{-m}$ with $m \in \mathbb{Z}$ and $j \in \{1, \ldots, k\}$ is finitely generated. Let S_q be its finite generating set. Then the conjugates $g_{q-1}^n g_q^m a_j g_q^{-m} g_{q-1}^{-n}$ with $m, n \in \mathbb{Z}$ and $j \in \{1, \ldots, k\}$ are in the subgroup A_{q-1} of A generated by $g_{q-1}^n s g_{q-1}^{-n}$ with $n \in \mathbb{Z}$ and $s \in S_q$. Again Lemma 12.32 implies that the subgroup A_{q-1} is finitely generated. Continuing inductively, we conclude that the group A generated by all the conjugates in (12.19), is finitely generated. Hence, G is polycyclic. $\square$

This also concludes the proof of Milnor's theorem, Theorem 12.31. $\square$

By combining theorems of Milnor and Wolf we obtain:

THEOREM 12.37. *Every finitely generated solvable group either is virtually nilpotent or it has exponential growth.*

This was strengthened by J. Rosenblatt as follows:

THEOREM 12.38 (J. Rosenblatt, [**Ros74**]). *Every finitely generated solvable group either is virtually nilpotent or it contains a free non-abelian subsemigroup.*

Another application of Lemma 12.32 is the following proposition which will be used in the proof of Gromov's theorem on groups of polynomial growth:

PROPOSITION 12.39. *Suppose that G is a finitely generated group of sub-exponential growth, which fits in a short exact sequence*

$$1 \to K \to G \stackrel{\pi}{\to} \mathbb{Z} \to 1.$$

Then K is finitely generated. Moreover, if $\mathfrak{G}_G(R) \preceq R^d$ then $\mathfrak{G}_K(R) \preceq R^{d-1}$.

PROOF. Let $\gamma \in G$ be an element which projects to the generator 1 of $\mathbb{Z}$. Let $\{f_1, \ldots, f_k\}$ denote a set of generators of G. Then for each i there exists $s_i \in \mathbb{Z}$ such for $g_i := f_i \gamma^{s_i}$, $i = 1, \ldots, k$, we have

$$\pi(g_i) = 0 \in \mathbb{Z}.$$

Clearly, the set $\{g_1, \ldots, g_k, \gamma\}$ generates G. Without loss of generality we may assume that each generator g_i is nontrivial. Define

$$S := \{\gamma_{m,i} := \gamma^m g_i \gamma^{-m}\ ;\ m \in \mathbb{Z}, i = 1, \ldots, k\}.$$

We claim that the (infinite) set S generates K. Indeed, clearly, $S \subset K$. Every $g \in K$ can be written as a word $w = w(g_1, \ldots, g_k, \gamma)$ in the letters $g_1^{\pm 1}, \ldots, g_k^{\pm 1}, \gamma^{\pm 1}$. We then move all entries of powers of γ in the word w to the end of w by using the identities

$$\gamma^m g_i = \gamma_{m,i} \gamma^m.$$

As the result, we obtain a word $w' = u\gamma^a$ in the alphabet $S \cup S^{-1} \cup \{\gamma, \gamma^{-1}\}$, where u contains only the letters in $S \cup S^{-1}$ and $a \in \mathbb{Z}$. Since g projects to $0 \in \mathbb{Z}$, $a = 0$. Claim follows.

Lemma 12.32 implies that there exists $M(i)$ such that the subgroup K is generated by the finite set

$$\{\gamma_{l,i} \; ; \; |l| \leqslant M(i), i = 1, \ldots, k\}.$$

This proves the first assertion of the Proposition.

Now let us prove the second assertion which estimates the growth function of K. Take a finite generating set Y of the subgroup K and set $X := Y \cup \{\gamma\}$, where γ is as above. Then X is a generating set of G. Given $n \in \mathbb{N}$ let $N := \mathfrak{G}_Y(n)$, where $\mathfrak{G}_Y$ is the growth function of K with respect to the generating set Y. Thus, there exists a subset

$$H := \{h_1, \ldots, h_N\} \subset K$$

where $|h_i|_Y \leqslant n$ and $h_i \neq h_j$ for all $i \neq j$. We obtain a set T of $(2n+1) \cdot N$ pairwise distinct elements

$$h_i \gamma^j, \quad -n \leqslant j \leqslant n, \quad i = 1, \ldots, N.$$

It is clear that $|h_i \gamma^j|_X \leqslant 2n$ for each $h_j \gamma^j \in T$. Therefore,

$$n\mathfrak{G}_Y(n) \leqslant (2n+1)\mathfrak{G}_Y(n) = (2n+1)N \leqslant \mathfrak{G}_X(2n) \leqslant C(2n)^d = 2^d C \cdot n^d,$$

for some constant C depending only on X. It follows that

$$\mathfrak{G}_Y(n) \leqslant 2^d C \cdot n^{d-1} \preceq n^{d-1}. \quad \square$$

12.5. Failure of QI rigidity for solvable groups

THEOREM 12.40 (A. Dyubina (Erschler), [**Dyu00**]). *The class of finitely generated (virtually) solvable groups is not QI rigid: There are groups which are quasi-isometric to solvable groups, but not virtually isomorphic to such.*

PROOF. The groups that will be used in the proof are wreath products $G_A := A \wr C$ of finitely generated groups. Given (finite) generating sets $a_i, i \in I, c_j, j \in J$ of A and C, respectively, we will use the finite set of generators of G_A introduced in Lemma 4.11. In what follows we use multiplicative notation when dealing with wreath products.

Suppose now that A, B are finite groups of the same order, where A is abelian, say, cyclic, and B is a non-solvable group. For instance, we can take A to be the group $\mathbb{Z}_{60}$ and B is the alternating group A_5 (which is a simple nonabelian group of order 60). We declare each nontrivial element of these groups to be a generator. Let C be a finitely generated infinite abelian group, say, $\mathbb{Z}$, and consider the wreath products $G_A := A \wr C, G_B := B \wr C$. Let $\varphi : A \to B$ be a bijection sending 1 to 1. Extend this bijection to a map

$$\Phi : G_A \to G_B, \quad \Phi(f, c) = (\varphi \circ f, c).$$

LEMMA 12.41. Φ *extends to an isometry of Cayley graphs.*

Proof. First, the inverse map Φ^{-1} is given by $\Phi(f,c) = (\varphi^{-1} \circ f, c)$. We now check that Φ preserves edges of the Cayley graphs. The group G_A has two types of generators: $(1, c_j)$ and $(\delta_a, 1)$, where $c_j \in X$, a finite generating set of C and $a \in A$ are all nontrivial elements of A. The same holds for the group G_B.

1. Consider the edges connecting $(f(x), c)$ to $(f(xc_j^{-1}), cc_j)$ in $Cayley(G_A)$. Applying Φ to the vertices of such edges we obtain

$$(\varphi \circ f(x), c), \quad (\varphi \circ f(xc_j^{-1}), cc_j).$$

Clearly, they are again within unit distance in $Cayley(G_B)$, since they differ by $(1, c_j)$.

2. Consider the edges connecting $(f(x), c), (f(x)\delta_a(x), c)$ in $Cayley(G_A)$. Applying Φ to the vertices we obtain

$$(\varphi \circ f(x), c), \quad (\varphi \circ f(x)\delta_b(x), c),$$

where $b = \varphi(a)$. Again, we obtain vertices which differ by $(\delta_b, 1)$, so they are within unit distance in $Cayley(G_B)$ as well. □

Lemma 12.42. *The group G_B is not virtually solvable.*

Proof. Let $\psi : G_B \to F$ be a homomorphism to a finite group. Then the kernel of the restriction $\psi|_{\oplus_C B}$ is also solvable. The restriction of ψ to the subgroup $B_c < \oplus_C B$ consisting of maps $f : C \to B$ supported at $\{c\}$, is determined by a homomorphism $\psi_c : B \to F$. There are only finitely many of such homomorphisms, while C is an infinite group. Thus, we find $c_1 \neq c_2 \in C$ such that

$$\psi_{c_1} = \psi_{c_2} = \eta.$$

The kernel of the restriction $\psi|_{B_{c_1} \oplus B_{c_2}}$ consists of pairs

$$(b_1, b_2) \in B_{c_1} \oplus B_{c_2} = B \times B, \quad \eta(b_1) = \eta(b_2)^{-1}$$

and contains the subgroup

$$\{(b, b^{-1}), b \in B\}.$$

However, this subgroup is isomorphic to B (*via* projection to the first factor in the product $B \times B$). Thus, kernel of ψ contains a subgroup isomorphic to B and, hence is not solvable. □

Combination of these two lemmas implies the theorem. □

12.6. Virtually nilpotent subgroups of $GL(n)$

In this section we collect various properties about virtually nilpotent subgroups of $GL(n, \mathbb{K})$ for arbitrary fields $\mathbb{K}$, the main focus of which is to show that under certain conditions, a subgroup of $GL(n, \mathbb{K})$ is nilpotent or virtually nilpotent. These results will be used in the proof of the Tits Alternative in the next chapter. In what follows, V will denote a finite-dimensional vector space over a field $\mathbb{K}$; we will also use the notation W for a finite-dimensional vector space over the algebraic closure $\bar{\mathbb{K}}$ of K. We let $End(V)$ denote the algebra of (linear) endomorphisms of V and $GL(V)$ the group of invertible endomorphisms of V.

The main technical tool that we will use repeatedly is the following theorem originally proved by Burnside.

If $A \subset End(V)$ is a subsemigroup, then A is said to *act irreducibly* on V if V contains no proper subspaces $V' \subset V$ such that $aV' \subset V'$ for all $a \in A$. A proof of the following theorem can be found, for instance, in [**Lan02**, Chapter XVII, §3, Corollary 3.3]:

THEOREM 12.43 (Burnside's theorem). *Suppose that V is a finite dimensional vector space over an algebraically closed field $\mathbb{K}$. If $A \subset End(W)$ is a $\mathbb{K}$-subalgebra which acts irreducibly on W, then $A = End(W)$. In particular, if $G \leqslant End(W)$ is a subsemigroup acting irreducibly, then G spans $End(W)$ as a vector space.*

Another useful (and elementary) fact is:

LEMMA 12.44 (See e.g. []). *If V is a finite-dimensional vector space over a field $\mathbb{K}$, then*

$$\tau : End(V) \times End(V) \to \mathbb{K}, \tau(A, B) = tr(AB)$$

is a nondegenerate bilinear form on $End(V)$, regarded as a vector space over $\mathbb{K}$.

Below are two useful corollaries of Burnside's theorem. Recall that an endomorphism $h \in End(V)$ is *nilpotent* if $h^k = 0$ for some $k > 0$. Equivalently, in some basis, h can be written as an upper triangular matrix with zeroes on the diagonal. Automorphisms of V of the form $I+h$, h is nilpotent, are called *unipotent*. Here and in what follows, I is the identity map $V \to V$. Similarly, an automorphism g of V is called *quasiunipotent* if all eigenvalues of g are roots of unity in $\bar{\mathbb{K}}$. Equivalently, g is quasiunipotent if g^k is unipotent for some $k > 0$.

THEOREM 12.45 (Kolchin's theorem). *Suppose that $\mathbb{K} = \bar{\mathbb{K}}$ and $\Gamma < GL(V)$ consists only of unipotent elements. Then Γ is conjugate to a subgroup of the group of invertible upper-triangular matrices $\mathcal{T}_n(\mathbb{K})$.*

PROOF. The proof is by induction on the dimension n of V. The claim is clear for $n = 1$, hence, we assume that $n > 1$. The statement of the theorem amounts to the claim that Γ preserves a full flag

$$0 \subset V_1 \subset \ldots \subset V_{n-1} \subset V,$$

where $i = \dim(V_i)$ for each i. Indeed, given such a flag, we will inductively pick basis elements $\mathbf{e}_i \in V_i$ such that $\{e_1, \ldots, e_i\}$ is a basis in V_i. With respect to this basis, the subgroup of $GL(V)$ preserving the flag will be contained in $\mathcal{T}_n(\mathbb{K})$.

Suppose first that the action of Γ on V is reducible: Γ preserves a proper subspace $V' \subset V$. Then we obtain two induced actions of Γ on V' (by the restriction) and on $V'' = V/V'$ (by the projection). Since both actions preserve full flags in V', V'' (by the induction hypothesis), by combining these flags we obtain a full Γ-invariant flag in V.

Therefore, we will assume that the action of Γ on V is irreducible. For each $g \in \Gamma$ the endomorphism $g' = g - I$ is nilpotent, hence, has zero trace. Therefore, for all $x \in G$, we have

$$tr(g'x) = tr(gx - x) = tr(I) - tr(I) = 0.$$

Since, by Burnside's theorem, Γ spans $End(V)$, we conclude that for each $x \in End(V)$ and each $g \in \gamma$,

$$tr(g'x) = 0.$$

Using the fact that τ is a nondegenerate pairing on $End(V)$, we conclude that $g' = 0$ for all $g \in \Gamma$, i.e., $\Gamma = \{1\}$. □

The following theorem is a minor variation on Kolchin's theorem. A subgroup $\Gamma < GL(V)$ is quasiunipotent if every element of Γ is quasiunipotent.

Proposition 12.46. *Suppose that Γ is quasiunipotent and, moreover, there exists an upper bound on orders of all eigenvalues of elements $g \in \Gamma$. Then Γ contains a finite index subgroup conjugate into the group of upper triangular matrices $\mathcal{T}_n(\mathbb{K})$.*

Proof. The proof follows closely the proof of Kolchin's theorem. As in Kolchin's theorem, the proof is by induction on the dimension of V and it suffices to consider the case of subgroups acting irreducibly on V. For each $g \in G$ define a linear map

$$T_g : End(V) \to \mathbb{K}, \quad T_g(x) = tr(gx).$$

Since τ is a nondegenerate pairing on $End(V)$, for $g_1 \neq g_2 \in G$, we get $T_{g_1} \neq T_{g_2}$. As we assumed that the orders of eigenvalues of elements of G are uniformly bounded, the set of traces of the elements of G is finite as well. Therefore, for each $g \in G$, the set

$$\{T_g(x) : x \in \Gamma\}$$

is a certain finite set C, independent of g. By Burnside's theorem, Γ spans the algebra $End(V)$, which implies that for each $g \in \Gamma$, the map T_g is determined by its restriction to Γ. Thus, the set

$$\{T_g : End(V) \to \mathbb{K} | g \in \Gamma\}$$

is finite. We, therefore, conclude that the group Γ is finite. □

Suppose that $\Gamma < \mathcal{T}_n(\mathbb{K})$ is quasiunipotent with an upper bound on orders of eigenvalues. Then there exists $k > 0$ such that g^k is unipotent for each $g \in \Gamma$. Therefore, Γ contains a finite index subgroup Γ_1 contained in $\mathcal{U}_n(\mathbb{K})$. Since (see Example 11.36) the group $\mathcal{U}_n(\mathbb{K})$ is nilpotent, we obtain:

Corollary 12.47. *Suppose that Γ is quasiunipotent and, moreover, there exists an upper bound on orders of all eigenvalues of elements $g \in \Gamma$. Then Γ is virtually nilpotent.*

Restriction of scalars. Let field $\mathbb{F}$ be a finite extension of $\mathbb{E}$; in other words, $\mathbb{F}$ is a k-dimensional vector space over $\mathbb{E}$, where $k < \infty$. Thus, we obtain an isomorphism of abelian groups

$$\mathbb{F} \to \mathbb{E}^k.$$

Accordingly, we obtain a monomorphism

$$GL(n, \mathbb{E}) \hookrightarrow GL(nk, \mathbb{F}).$$

This construction, embedding $GL(n, \mathbb{E})$ into $G \hookrightarrow GL(nk, \mathbb{F})$ is called the *restriction of scalars.*

Restricting to finitely generated subgroups, allows one to eliminate the upper bound assumption on orders of eigenvalues:

Proposition 12.48. *Each finitely generated quasiunipotent subgroup $\Gamma < GL(V)$ contains a finite-index subgroup Γ' conjugate to a subgroup of $\mathcal{U}_n(\mathbb{K})$. Furthermore, Γ contains a nilpotent subgroup $\Gamma' < \Gamma$ of index $\leqslant q = q(V)$, which is independent of Γ.*

Proof. Since Γ is finitely-generated, it suffices to consider the case when the field $\mathbb{K}$ is finitely generated. Let $\mathbb{P} \subset \mathbb{K}$ be the prime field. Since $\mathbb{K}$ is finitely generated, we have inclusions

$$\mathbb{P} \subset \mathbb{P}(T) \subset \mathbb{K}$$

where $\mathbb{P} \subset \mathbb{P}(T)$ is a purely transcendental extension with the finite basis T and $\mathbb{P}(T) \subset \mathbb{K}$ is a finite algebraic extension. By applying the *restriction of scalars* procedure, we re-embed the subgroup $\Gamma < GL(n, \mathbb{K})$ into the group $GL(nd, \mathbb{P}(T))$, where $d = |\mathbb{K} : \mathbb{P}(T)|$. We leave it to the reader to verify that the image of the new embedding is still quasiunipotent. We claim that the set of orders of eigenvalues of elements $g \in \Gamma < GL(nd, \mathbb{P}(T))$ is finite.

Case 1: $\mathbb{K}$ has zero characteristic, i.e., $\mathbb{P} \cong \mathbb{Q}$. Let $\chi_g(x)$ denote the characteristic polynomial of $g \in \Gamma < GL(nd, \mathbb{P}(T))$; its roots are roots of unity. Since coefficients of $\chi_g(x)$ are symmetric polynomials of the roots of unity, it follows that all the coefficients of χ_g are algebraic integers. However, the extension $\mathbb{Q} \subset \mathbb{Q}(T)$ is purely transcendental, therefore, all coefficients of χ_g actually belong to $\mathbb{Z}$. As in the proof of Lemma 11.28, the set of coefficients of $\chi_g, g \in \Gamma$, is bounded, therefore, it is finite. This means that the orders of eigenvalues of the elements of Γ are uniformly bounded.

Case 2: $\mathbb{K}$ has characteristic p, i.e., $\mathbb{P} = \mathbb{Z}_p$ for some p. The argument is even simpler than the one in the case of zero characteristic. The coefficients of the characteristic polynomials of $g \in \Gamma < GL(nd, \mathbb{P}(T))$ all belong to $\mathbb{P}$. However, the field $\mathbb{P}$ is finite, therefore, the set of eigenvalues of the elements $g \in \Gamma$ is finite.

Now, the first claim of the proposition follows from Proposition 12.46. To verify the second claim we notice that the bound that we obtained on the orders of eigenvalues of $g \in \Gamma$ depends only on V. Then, examination of the proof of Proposition 12.46 shows that index in that theorem depends only on V and on the upper bound on the orders of the eigenvalues. □

12.7. Discreteness and nilpotence in Lie groups

The goal of this section is to prove theorems of Zassenhaus and Jordan. These theorems deal, respectively, with discrete and finite subgroups Γ of Lie groups G (with finitely many components). The Theorem of Zassenhaus shows that, appropriately defined, "small elements" of Γ generate a nilpotent subgroup of G. Jordan's theorem establishes that finite subgroups of G are "almost abelian": Every finite group Γ contains an abelian subgroup, whose index in Γ is uniformly bounded. Historically, Jordan's theorem was proven first and then, Zassenhaus proved his theorem. We will prove things in the reverse order and we will be using Zassenhaus' results in order to prove Jordan's theorem.

12.7.1. Some useful linear algebra. We begin by discussing some basic linear algebra which will be used in the proof of Jordan's theorem.

Suppose that V is a real Euclidean vector space (e.g., a Hilbert space, but we do not insist on completeness of the norm) with the inner product denoted $\mathbf{x} \cdot \mathbf{y}$ and the norm denoted $\|x\|$. We will endow the complexification $V^{\mathbb{C}}$ of V with the Euclidean structure

$$(\mathbf{x} + i\mathbf{y}) \cdot (\mathbf{u} + i\mathbf{v}) = \mathbf{x} \cdot \mathbf{u} + \mathbf{y} \cdot \mathbf{v}.$$

Recall that the *operator norm* of a bounded linear transformation $A \in End(V)$ is

$$\|A\| := \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

Then A extends naturally to a complex-linear transformation of $V^{\mathbb{C}}$ whose operator norm is $\leqslant \sqrt{2}\|A\|$. In what follows we will use the notation $\nu(A) = \|A - I\|$, the distance from A to the identity in $End(V)$.

LEMMA 12.49. *Suppose that $A \in O(V)$ and $\nu(A) < \sqrt{2}$. Then A is a* rotation *with rotation angles $< \pi/2$, i.e., for every nonzero vector $\mathbf{x} \in V$,*

$$A\mathbf{x} \cdot \mathbf{x} > 0.$$

PROOF. By the assumption,

$$\|A\mathbf{x} - \mathbf{x}\| < \sqrt{2}$$

for all unit vectors $\mathbf{x} \in V$. Denoting by y the difference vector $Ax - x$, we obtain:

$$2 > \mathbf{y} \cdot \mathbf{y} = (A\mathbf{x} - \mathbf{x}) \cdot (A\mathbf{x} - \mathbf{x}) = 2 - 2(A\mathbf{x} \cdot \mathbf{x}).$$

Hence, $A\mathbf{x} \cdot \mathbf{x} > 0$. □

COROLLARY 12.50. *The same conclusion holds for the complexification of A.*

PROOF. Let $\mathbf{v} = \mathbf{x} + i\mathbf{y} \in V^{\mathbb{C}}$. Then

$$A\mathbf{v} \cdot \mathbf{v} = (A\mathbf{x} + iA\mathbf{y}) \cdot (\mathbf{x} + i\mathbf{y}) = A\mathbf{x} \cdot \mathbf{x} + A\mathbf{y} \cdot \mathbf{y} > 0$$

by the above lemma. □

LEMMA 12.51. *Suppose that $A, B \in O(V)$ and $\nu(B) < \sqrt{2}$. Then*

$$[A, BAB^{-1}] = 1 \iff [A, B] = 1.$$

PROOF. Since one implication is clear, we assume that $[A, BAB^{-1}] = 1$. Let λ_j's be the (complex) eigenvalues of A. Then the complexification $V^{\mathbb{C}}$ splits as an A-invariant orthogonal sum

$$\oplus_j V^{\lambda_j},$$

where on each $V_j = V^{\lambda_j} = V_A^{\lambda_j}$ the orthogonal transformation A acts *via* multiplication by λ_j. Here we assume that for $j \neq k$, $\lambda_j \neq \lambda_k$. We refer to this orthogonal decomposition of $V^{\mathbb{C}}$ as $\mathcal{F}_A$. Then, clearly,

$$B(\mathcal{F}_A) = \mathcal{F}_{BAB^{-1}}$$

for any two orthogonal transformations $A, B \in O(V)$. Since A commutes with BAB^{-1}, A has to preserve the decomposition $\mathcal{F}_{BAB^{-1}}$ and, moreover, has to send each $W_j := V^{\lambda_j}_{BAB^{-1}} = B(V^{\lambda_j})$ to itself. What are the eigenvalues for this action of A on W_j? They are λ_k's for which V^{λ_k} has nontrivial intersection with W_j. However, if $\lambda_j \neq \lambda_k$ then V^{λ_j} is orthogonal to V^{λ_k} and, hence, by Corollary 12.50, B cannot send a (nonzero) vector from one space to the other. Therefore, in this case, $W_j \cap V_k = 0$. This leaves us with only one choice of the eigenvalue for the restriction $A|W_j$, namely λ_j. (Since the restriction has to have some eigenvalues!) Thus, $W_j \subset V_j$. However, B sends V_j to W_j injectively, so $W_j = V_j$ and we conclude that $B(V_j) = V_j$. Since A acts on V_j *via* multiplication by λ_j, it follows that $B\big|_{V_j}$ commutes with $A\big|_{V_j}$. This holds for all j, hence, $[A, B] = 1$. □

12.7.2. Zassenhaus neighborhoods. We now define 'smallness' in a Lie group: "Small" elements will be those which belong to a Zassenhaus neighborhood defined below.

DEFINITION 12.52. Let G be a topological group. A *Zassenhaus neighborhood* in G is an (open) neighborhood of the identity in G, denoted U or U_G, which satisfies the following:

1. The commutator map sends $U \times U$ to U.

2. There exists a continuous function $\sigma : U \to \mathbb{R}$ such that $0 = \sigma(1)$ is the minimal value of σ and

$$\sigma([A,B]) < \min(\sigma(A), \sigma(B))$$

for all $A \neq 1, B \neq 1$ in U.

We will refer to σ as *a Zassenhaus function*.

Note that if $H < G$ is a topological subgroup and U_G is a Zassenhaus neighborhood of G then $U_H := U_G \cap H$ is a Zassenhaus neighborhood of H.

We will see that every Lie group has Zassenhaus neighborhoods. We start with some examples. Let V be a Hilbert space (the reader can think of finite-dimensional V since this is the only case that we will need). We let $End(V)$ denote the semigroup of bounded linear operators in V and let $GL(V) \subset End(V)$ be the general linear group of V, i.e., the group of invertible operators A such that both A and A^{-1} are bounded. We again equip $End(V)$ with the operator norm.

LEMMA 12.53. *Let $G = O(V)$ be the orthogonal group of V. Then the set U given by the inequality $\nu(A) < 1/4$ is a Zassenhaus neighborhood in G.*

PROOF. We will use the function $\sigma = \nu$ as the Zassenhaus function. We will show that for all $A, B \in U \setminus 1$,

$$\nu([A,B]) < \min(\nu(A), \nu(B)),$$

which will also imply that $[\cdot,\cdot] : U \times U \to U$.

First, observe that the multiplication by orthogonal transformations in G preserves the operator norm on $End(V)$. Applying this twice to operators A, B such that $\nu(A) \leqslant \nu(B)$, we obtain:

$$\begin{aligned}\nu([A,B]) = \|AB - BA\| = \|(A-B)(A-I) - (A-I)(A-B)\| \leqslant \\ \|(A-B)(A-I)\| + \|(A-I)(A-B)\| \leqslant \\ 2\nu(A)\|A-B\| = 2\nu(A)(\nu(A) + \nu(B)).\end{aligned}$$

Since $\nu(A) \leqslant \nu(B) < 1/4$, we obtain

$$\nu([A,B]) < 2\nu(A)\left(\frac{1}{4} + \frac{1}{4}\right) = \nu(A). \quad \square$$

LEMMA 12.54. *Let $G = GL(V)$ be the general linear group of V, i.e., group of invertible operators A such that both A and A^{-1} are bounded. We set $\sigma(A) := \max(\nu(A), \nu(A^{-1}))$ for $A \in G$. Then the set U given by the inequality $\sigma(A) < 1/8$ is a Zassenhaus neighborhood in G with the Zassenhaus function σ.*

PROOF. Our proof follows the same lines as in the orthogonal case. We will show that

$$\|ABA^{-1}B^{-1} - I\| < \min(\sigma(A), \sigma(B)).$$

The inequality

$$\|(ABA^{-1}B^{-1})^{-1} - I\| < \min(\sigma(A), \sigma(B))$$

will follow by interchanging A and B. We again assume that $\sigma(A) \leqslant \sigma(B)$. Observe that $\|CD\| \leqslant \|C\| \cdot \|D\|$ for all $C, D \in End(V)$. Applying this twice, we get:

$$\|ABA^{-1}B^{-1} - I\| \leqslant \|B^{-1}\|\|ABA^{-1} - B\| \leqslant \|A^{-1}\|\|B^{-1}\|\|AB - BA\|.$$

If $\sigma(C) < c$ then $\|C^{-1}\| < 1 + c$ for every $C \in G$. Thus,

$$\|ABA^{-1}B^{-1} - I\| < (1+c)^2\|AB - BA\|$$

provided that $\sigma(A) < c, \sigma(B) < c$. The rest of the proof is the same as in the orthogonal case:

$$\|AB - BA\| \leqslant 2\sigma(A)(\sigma(A) + \sigma(B)) \leqslant 4\sigma(B)\sigma(A).$$

Putting it all together:

$$\|ABA^{-1}B^{-1} - I\| < 4c(1+c)^2\sigma(A).$$

Since for $c = 1/8$, $4c(1+c)^2 = \frac{1}{2}\left(\frac{9}{8}\right)^2 < 1$, we conclude that

$$\|ABA^{-1}B^{-1} - I\| < \sigma(A).$$

Thus,

$$\sigma([A, B]) < \min(\sigma(A), \sigma(B))$$

for all $A, B \in U$. □

REMARK 12.55. The above proofs, at first glance, look like a trickery. What is really happening in the proof? Consider $G = GL(n, \mathbb{R})$. The idea behind the proof is that the commutator map has zero 1-st derivative at the point $(1, 1) \in G \times G$ (which one can easily see by using the Taylor expansion $A^{-1} = I - a + a^2 ...$ for a matrix of the form $A = I + a$ where a has small norm). Thus, by the basic calculus, $[A, B]$ will be “closer” to $I \in G$ than $A = I + a$ and $B = I + b$ if a, b are sufficiently small. The above proofs provide explicit estimates for this argument.

We will say that a topological group G *admits a basis of Zassenhaus neighborhoods* if $1 \in G$ admits a basis of topology consisting of Zassenhaus neighborhoods.

COROLLARY 12.56. *Suppose that G is a linear Lie group. Then $1 \in G$ admits a basis of Zassenhaus neighborhoods.*

PROOF. First, suppose that $G = GL(V)$. Then the sets $U_t = \sigma^{-1}(t), t \in (0, \frac{1}{8})$ are Zassenhaus neighborhoods and their intersection is $1 \in G$. If G is a Lie group which admits a continuous closed embedding $\phi : G \to GL(V)$, the sets $\phi^{-1}(U_t)$ will serve as a Zassenhaus basis. □

Note that being a subgroup of $GL(n, \mathbb{R})$ is not really necessary for this corollary since the conclusion is *local* at the identity in G: If a topological group G_1 admits a basis of Zassenhaus neighborhoods and G_2 is a locally compact group which *locally embeds* (see §3.6.1) in G_1, then G_2 also admits a basis of Zassenhaus neighborhoods. In view of Ado's theorem (Theorem 3.55), every Lie group locally embeds in $GL(V)$ for some finite-dimensional real vector space V.

COROLLARY 12.57. *Every Lie group admits a basis of Zassenhaus neighborhoods.*

Why are Zassenhaus neighborhoods useful? We assume now that G is a locally compact group which admits a basis of Zassenhaus neighborhoods and fix a Zassenhaus neighborhood $\Omega \subset G$ whose closure is compact and is contained in another Zassenhaus neighborhood $U \subset G$. Define inductively subsets $\Omega^{(i)}$ as $\Omega^{(i+1)} = [\Omega, \Omega^{(i)}]$, $\Omega^{(0)} := \Omega$. Since Ω is Zassenhaus,

$$\Omega^{(i+1)} \subset \Omega^{(i)}$$

for all i.

LEMMA 12.58. $E := \bigcap_i \overline{\Omega^{(i)}} = \{1\}$.

Proof. Clearly, E is compact and $[\Omega, E] = E$. Suppose that $E \neq \{1\}$ and take $f \in E$ with maximal $\sigma(f) > 0$, where σ is the function in the definition of a Zassenhaus neighborhood. Then, $f = [g, h]$ for some $, g \in \Omega, h \in E$ and, hence,

$$\sigma(f) < \min(\sigma(g), \sigma(h)) \leqslant \sigma(h).$$

Contradiction. □

Theorem 12.59 (The Zassenhaus theorem). *Suppose that G is a locally compact group which admits a relatively compact Zassenhaus neighborhood Ω. Assume that $\Gamma < G$ is a discrete subgroup generated by the subset $X := \Gamma \cap \Omega$. Then Γ is nilpotent. In particular, this property holds for all Lie groups.*

Proof. In view of Lemma 11.44, it suffices to show that there exists n such that all n-fold iterated commutators of elements of X are trivial. By the definition of $\Omega^{(i)}$, all i-fold iterated commutators of X are contained in $\Omega^{(i)}$. Since Γ is discrete and Ω is relatively compact, we can have only finitely many distinct nontrivial iterated commutators of elements of X. Since

$$\bigcap_i \overline{\Omega^{(i)}} = \{1\},$$

there exists n such that $\Omega^{(n)}$ is disjoint from all these nontrivial commutators. Thus, all n-fold iterated commutators of the elements of X are trivial. Hence, by Lemma 11.55, the group Γ is nilpotent. □

In section 12.7.3 we will see how the Zassenhaus Theorem can be strengthened in the context of *finite subgroups* of Lie groups (Jordan's theorem).

Below is an application of the Zassenhaus Theorem to orthogonal groups. We equip a finite-dimensional real vector space V with a Euclidean structure and let $O(V)$ denote the group of orthogonal transformations. Recall that for $A \in O(V)$, $\nu(A)$ is the operator norm $\|A - I\|$.

Let U denote a Zassenhaus neighborhood of the identity in $O(V)$ such that

$$U \subset \{A : \nu(A) < \sqrt{2}\}.$$

For instance, in view of Lemma 12.53, we can take U given by the inequality $\nu(A) < 1/4$.

Lemma 12.60. *If G is a nilpotent subgroup of $O(V)$ generated by some elements $A_j \in U$, then G is abelian.*

Proof. Consider the lower central series of G:

$$G_1 = G, \ldots, G_i = [G, G_{i-1}], i = 1, \ldots, n,$$

where $G_{n+1} = 1$ and $G_n \neq 1$. We need to show that $n = 1$. Assume not and consider an $(n+1)$-fold iterated commutator of the generators A_i of G; this iterated commutator has the form:

$$[[B, A], A] \in G_{n+1} = 1$$

where $A = A_j$ and $B \in G_{n-1}$ is an $n-1$-fold iterated commutator of the generators of G. Thus, A commutes with $BAB^{-1}A^{-1}$. Since A commutes with A^{-1}, we then conclude that A commutes with BAB^{-1}. By the definition of a Zassenhaus neighborhood, if generators A_i of G are in U, then all their iterated commutators are also in U and, hence, B satisfies the inequality $\|B - I\| < \sqrt{2}$.

Thus, Lemma 12.51 implies that A commutes with B and, hence, every n-fold iterated commutator of generators in G is trivial. Thus, $G_n = 1$ by Lemma 11.44. Contradiction. ☐

12.7.3. Jordan's theorem. Notice that even the group $SO(2)$ contains finite subgroups of arbitrarily high order, but these subgroups, of course, are all abelian. Considering the group $O(2)$ we find some non-abelian subgroups of arbitrarily high order, but they all contain abelian subgroups of index 2. Jordan's theorem below shows that a similar statement holds for finite subgroups of other Lie groups as well:

THEOREM 12.61 (Jordan's theorem). *Let L be a Lie group with finitely many connected components. Then there exists a number $q = q(L)$ such that each finite subgroup G in L contains an abelian subgroup of index $\leqslant q$.*

PROOF. If L is not connected, we replace L with L_0, the identity component of L and G with $G_0 := G \cap L_0$. Since $|G : G_0| \leqslant |L : L_0|$, it suffices to prove theorem only for subgroups of connected Lie groups. Thus, we assume in what follows that L is connected.

1. We first consider the most interesting case, when the Lie group L is $K = O(V)$, the orthogonal group of a finite-dimensional Euclidean vector space V.

Let Ω denote a relatively compact Zassenhaus neighborhood of $O(V)$ given by the inequality

$$\{A : \nu(A) < 1/4\}.$$

The finite subgroup $G \subset K$ is clearly discrete, therefore the subgroup $H := \langle G \cap \Omega \rangle$ is nilpotent by the Zassenhaus Theorem. By Lemma 12.60, every nilpotent subgroup generated by elements of Ω is abelian. It, therefore, follows that H is abelian.

It remains to estimate the index $|G : H|$. Let $U \subset \Omega$ be a neighborhood of 1 in $K = O(V)$ such that $U \cdot U^{-1} \subset \Omega$ (i.e., products of pairs of elements xy^{-1}, $x, y \in U$, belong to Ω). Let q denote $Vol(K)/Vol(U)$, where Vol is the volume induced by a bi-invariant Riemannian metric on K.

LEMMA 12.62. $|G : H| \leqslant q$.

PROOF. Let $x_1, \ldots, x_{q+1} \in G$ be distinct coset representatives for G/H. Then

$$\sum_{i=1}^{q+1} Vol(x_i U) = (q+1)Vol(U) > Vol(K).$$

Hence there are $i \neq j$ such that $x_i U \cap x_j U \neq \emptyset$. It follows that $x_j^{-1} x_i \in UU^{-1} \subset \Omega$ and, hence, $x_j^{-1} x_i \in H$. Contradiction. ☐

This proves Jordan's theorem for subgroups of $O(V)$.

2. Consider now the case when either G or L has trivial center. Consider the adjoint representation $\rho : L \to GL(V)$, where V is the Lie algebra of L. This representation need not be faithful, but the kernel C of this representation is contained in the center of L, see Lemma 3.50. Hence, the kernel C of the homomorphism

$$\rho : G \to \bar{G} \leqslant GL(V)$$

is contained in the center of G. Under our assumptions, C is the trivial group and, hence, $G \cong \bar{G}$.

Next, we construct a G-invariant Euclidean metric on V: Start with an arbitrary positive-definite quadratic form μ_0 on V and then set

$$\mu := \sum_{g \in G} g^*(\mu_0).$$

It is clear that the quadratic form μ is again positive definite and invariant under $\bar{G}$. With respect to this quadratic form, $\rho(G) \leqslant O(V)$. Thus, by Part 1, there exists an abelian subgroup $A := H < G$ of index $\leqslant q(O(V))$, where q depends only on L.

3. We now consider the general case where we can no longer use elementary arguments. First, by Cartan–Iwasawa–Mal'cev theorem (Theorem 3.35), every connected locally compact contains unique, up to conjugation, maximal compact subgroup. We find such subgroup K in L. By Cartan's theorem (Theorem 3.53), every closed subgroup of a Lie group is again a Lie group. Hence, K is also a Lie group. Since $G < L$ is finite, it is compact, and, thus, is conjugate to a subgroup of K. Next, every compact Lie group is linear, Theorem 3.54. Thus, we can assume that K is contained in $GL(V)$ for some finite-dimensional vector space V. Now, we proceed as in the Part 2. This proves Jordan's theorem. □

REMARK 12.63. The above proof does not provide an explicit bound on $q(L)$. Boris Weisfeiler [**Wei84**] proved for $n > 63$, $q(GL(n, \mathbb{C})) \leqslant (n+2)!$, which is nearly optimal since $GL(n, \mathbb{C})$ contains the permutation group S_n which has the order $n!$. Weisfeiler did the work in 1984 shortly before he, tragically, disappeared in Chile in 1985. (On August 21 of 2012 a Chilean judge ordered the arrest of eight retired police and military officers in connection with the kidnapping and disappearance of Weisfeiler.)

12.8. Virtually solvable subgroups of $GL(n, \mathbb{C})$

In this section we prove an analogue of Jordan's theorem for virtually solvable subgroups of matrix groups. This material will needed only for the proof of the Tits' Alternative for infinitely generated subgroups of $GL(n, \mathbb{C})$; the reader not interested in infinitely generated groups can skip it.

Let G be a subgroup of $GL(V)$, where $V \cong \mathbb{C}^n$. We will think of V as a G-module. In particular we will talk about G-submodules and quotient modules: The former are G-invariant subspaces W of V, the latter are quotients V/W, where W is a G-submodule. The G-module V is *reducible* if there exists a proper G-submodule $W \subset V$. We say that G is *upper-triangular* (or the G-module V is *upper-triangular*) if it is conjugate to a subgroup of the group B of upper-triangular matrices in $GL(V)$. In other words, there exists a *complete flag*

$$0 \subset V_1 \subset ... \subset V_n = V$$

of G-submodules in V, where $\dim(V_i) = i$ for each i. Of course, reducibility makes sense only for modules of dimension > 1; however, by abusing the terminology, we will regard modules of dimension $\leqslant 1$ as reducible by default.

The group B (and its conjugates in $GL(V)$) is called the *Borel subgroup* of $GL(V)$.

Lemma 12.64. *Suppose that G is an abstract group so that every G-module $V \cong \mathbb{C}^k$ with $2 \leqslant k \leqslant n$ is reducible. Then every n-dimensional G-module V is upper-triangular.*

Proof. Since $G \curvearrowright V$ is reducible, there exists a proper submodule $W \subset V$. Thus $\dim(W) < n$ and $\dim(V/W) < n$. Now, the assertion follows by induction on the dimension. □

Let V be a finite-dimensional vector space V over a field $\mathbb{K}$ and let $P(V)$ denote the corresponding projective space.

Lemma 12.65. *Let $G < GL(V)$ be upper-triangular. Then the fixed-point set $\mathrm{Fix}(G)$ of the action of G on the projective space $P(V)$ is nonempty and consists of a disjoint union of projective subspaces $P(V_\ell), \ell = 1, ..., k$, so that the subspaces $V_i \subset V$ are* linearly independent*, i.e.:*

$$\mathrm{Span}(\{V_1, ..., V_k\}) = \bigoplus_{\ell=1}^{k} V_\ell.$$

In particular, $k \leqslant \dim(V)$.

Proof. For $g \in GL(V)$ we let $a_{ij}(g)$ denote the i, j matrix coefficient of g. Then, since G is upper-triangular, the maps $\chi_i : g \to a_{ii}(g)$ are homomorphisms $\chi : G \to \mathbb{C}^*$, called *characters* of G. The (multiplicative) group of characters of G is denoted $X(G)$. We let $J \subset \{1, ..., n\}$ be the set of all indices j such that

$$a_{ij}(g) = a_{ji}(g) = 0, \forall g \in G, \forall i \neq j.$$

We then break the set J into disjoint subsets $J_1, ..., J_m$ which are preimages of points $\chi \in X(G)$ under the map

$$j \in J \mapsto \chi_j \in X(G).$$

Set $V_\ell := \mathrm{Span}(\{e_i, i \in J_\ell\})$, where $e_1, ..., e_n$ form the standard basis in V. It is clear that G fixes each $P(V_\ell)$ pointwise since each $g \in G$ acts on V_ℓ *via* the scalar multiplication by $\chi_\ell(g)$. We leave it to the reader to check that

$$\bigcup_{\ell=1}^{m} P(V_\ell)$$

is the entire fixed-point set $\mathrm{Fix}(G)$. □

In what follows, the topology on subgroups of $GL(V)$ is always the Zariski topology, in particular, connectedness always means Zariski–connectedness.

Theorem 12.66 (A. Borel). *Let G be a connected solvable Lie group. Then every G-module V (where V is a finite-dimensional complex vector space) is upper-triangular.*

Proof. In view of Lemma 12.64, it suffices to prove that every such module V is reducible. The proof is an induction on the derived length d of G.

We first recall a few facts about eigenvalues of the elements of $GL(V)$. Let $Z_{GL(V)}$ denote the center of $GL(V)$, i.e. the group of matrices of the form $\mu \cdot I, \mu \in \mathbb{C}^*$, where I is the unit matrix.

Let $g \in GL(V)\backslash Z_{GL(V)}$. Then g has linearly independent eigenspaces $E_{\lambda_j}(g), j = 1, ..., k$, labeled by the corresponding eigenvalues $\lambda_j, 1 \leqslant j \leqslant k$, where $2 \leqslant k \leqslant n$. We let $\mathcal{E}(g)$ denote the set of (unlabeled) eigenspaces

$$\{E_{\lambda_j}(g), j = 1, ..., k\}.$$

Let B_g denote the abelian subgroup of $GL(V)$ generated by g and the center $Z_{GL(V)}$. Then for every $g' \in B_g$, $\mathcal{E}(g') = \mathcal{E}(g)$ (with the new eigenvalues, of course). Therefore, if $N(B_g)$ denotes the normalizer of B_g in G, then $N(B_g)$ preserves the set $\mathcal{E}(g)$, however, elements of $N(B_g)$ can permute the elements of $\mathcal{E}(g)$. (Note that $N(B_g)$ is, in general, larger than $N(\langle g \rangle)$, the normalizer of $\langle g \rangle$ in G.) Since $\mathcal{E}(g)$ has cardinality $\leqslant n$, there is a subgroup $N^o = N^o(B_g) < N(B_g)$ of index $\neq n!$ that fixes the set $\mathcal{E}(g)$ elementwise, i.e., every $h \in N^o$ will preserve each $E_\lambda(g)$, where $\lambda \in Sp(g)$, the spectrum of g. Of course, h need not act trivially on $E_\lambda(g)$. Since $g \notin Z_G$, this means that there exists a proper N^o-invariant subspace $E_\lambda(g) \subset V$.

We next prove several lemmas needed for the proof of Borel's theorem.

Lemma 12.67. *Let A be an abelian subgroup of $GL(V)$. Then the A-module V is reducible.*

Proof. If $A \leqslant Z_{GL(V)}$, there is nothing to prove. Assume, therefore, that A contains an element $g \notin Z_{GL(V)}$. Since $A \leqslant N(B_g)$, it follows that A preserves the collection of subspaces $\mathcal{E}(g)$. Since A is abelian, it cannot permute these subspaces. Therefore, A preserves the proper subspace $E_{\lambda_1}(g) \subset V$ and hence $A \curvearrowright V$ is reducible. □

Lemma 12.68. *Suppose that $G < GL(V)$ is a connected metabelian group, so that $G' = [G, G] \leqslant Z_{GL(V)}$. Then the G-module V is reducible.*

Proof. The proof is analogous to the proof of the previous lemma. If G is contained in the center of $GL(V)$, there is nothing to prove. Pick, therefore some $g \in G \setminus Z_{GL(V)}$. Since the image of G in $PGL(V)$ is abelian, the group G is contained in $N(B_g)$. Since G is connected, it cannot permute the elements of $\mathcal{E}(g)$. Hence G preserves each $E_{\lambda_i}(g)$. Since every subspace $E_{\lambda_i}(g)$ is proper, it follows that the G-module V is reducible. □

Similarly, we have:

Lemma 12.69. *Let $G < GL(V)$ be a metabelian group whose projection to $PGL(V)$ is abelian. Then G contains a reducible subgroup of index $\leqslant n!$.*

Proof. We argue as in the proof of the previous lemma, except G may permute the elements of $\mathcal{E}(g)$. However, it will contain an index $\leqslant n!$ subgroup which preserves each $E_{\lambda_j}(g)$ and the assertion follows. □

We can now prove Theorem 12.66. Lemma 12.67 proves the theorem for abelian groups, i.e., solvable groups of derived length 1. Suppose the assertion holds for all connected groups of derived length $< d$ and let $G < GL(V)$ be a connected solvable group of derived length d. Then $G' = [G, G]$ has derived length $< d$. Thus, by the induction hypothesis, G' is upper-triangular. By Lemma 12.65, $\mathrm{Fix}(G') \subset P(V)$ is a nonempty disjoint union of independent projective subspaces $P(V_i), i = 1, ..., \ell$. Since G' is normal in G, $\mathrm{Fix}(G')$ is invariant under G. Since G is connected, it cannot interchange the components $P(V_i)$ of $\mathrm{Fix}(G)$. Therefore, it has to preserve each $P(V_i)$. If one of the $P(V_i)$'s is a proper projective subspace in $P(V)$, then V_i is G-invariant and hence the G-module V is reducible. Therefore, we assume that $\ell = 1$ and $V_1 = V$, i.e., G' acts trivially on $P(V)$. This means that $G' < Z_{GL(V)}$ is abelian and hence G is 2-step nilpotent. Now, the assertion follows from Lemma 12.68. This concludes the proof of Theorem 12.66. □

The following is a converse to Theorem 12.66:

PROPOSITION 12.70. *For $V = \mathbb{C}^n$, the Borel subgroup $B < GL(V)$ is solvable of derived length n. Thus, a connected subgroup of $GL(V)$ is solvable if and only if it is conjugate to a subgroup of B, i.e., Borel subgroups are the maximal solvable connected subgroups of $GL(V)$. In particular, the derived length of every connected subgroup of $GL_n(\mathbb{C})$ is at most n.*

PROOF. The proof is induction on n. The assertion is clear for $n = 1$ as $GL_1(\mathbb{C}) \cong \mathbb{C}^*$ is abelian. Suppose it holds for $n' = n - 1$, we will prove it for n. Let $B^{(i)} := [B^{(i-1)}, B^{(i-1)}], B^{(0)} = B$ be the derived series of B.

Let $0 = V_0 \subset V_1 \subset ... \subset V_n$ be the complete flag invariant under B. Set $W := V/V_1$, let B_W be the image of B in $GL(W)$. The kernel K of the homomorphism $B \to B_W$ is isomorphic to $\mathbb{C}^*$. The group B_W preserves the complete flag

$$0 = W_0 := V_1/V_1 \subset W_1 := V_2/V_1 \subset ... \subset W = V/V_1.$$

Therefore, by the induction assumption it has derived length $n-1$. Thus $B^{(n)} := [B^{(n-1)}, B^{(n-1)}] \subset K \cong \mathbb{C}^*$. Since $\mathbb{C}^*$ is abelian $[B^{(n)}, B^{(n)}] = 0$, i.e., B has derived length n. □

REMARK 12.71. Theorem 12.66 is false for non-connected solvable subgroups of $GL(V)$. Take $n = 2$, let A be the group of diagonal matrices in $SL(2, \mathbb{C})$ and let

$$s = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then s normalizes A and $s^2 \in A$. We let G be the group generated by A and s which is isomorphic to the semidirect product of A and $\mathbb{Z}_2$. In particular, G is solvable of derived length 2. On the other hand, it is clear that the G-module $\mathbb{C}^2$ is irreducible.

THEOREM 12.72. *There exist functions $\nu(n), \delta(n)$ such that every virtually solvable subgroup $\Gamma \leqslant GL(V)$ contains a solvable subgroup Λ of index $\leqslant \nu(n)$ and derived length $\leqslant \delta(n)$.*

PROOF. Let d denote the derived length of a finite index solvable subgroup of Γ. Let $\overline{\Gamma}$ denote the Zariski closure of Γ in $GL(V)$. Then $\overline{\Gamma}$ has only finitely many (Zariski) connected components (see Theorem 3.80).

LEMMA 12.73. *The group $\overline{\Gamma}$ is contains a finite index subgroup which is a solvable group of derived length d.*

PROOF. We will use k-fold iterated commutators

$$[\![g_1, \ldots, g_{2^k}]\!]$$

defined in the equation (11.11). Let $\Gamma^o < \Gamma$ denote a solvable subgroup of derived length d and finite index m in Γ; thus

$$\Gamma = \gamma_1\Gamma^o \cup ... \cup \gamma_m\Gamma^o.$$

The group Γ^o satisfies polynomial equations of the form $(g_1, ..., g_{2^d}) = 1$. Therefore, Γ satisfies the polynomial equations in the variables g_j:

$$\gamma_i [\![g_1, \ldots, g_{2^d}]\!] = 1, i = 1, ..., m.$$

Hence, the Zariski closure $\overline{\Gamma}$ of Γ satisfies the same set of polynomial equations. It follows that $\overline{\Gamma}$ contains a subgroup of index m which is solvable of derived length d. □

Let G be the (Zariski) connected component of the identity of $\overline{\Gamma}$, which implies that $G \lhd \overline{\Gamma}$.

Lemma 12.74. *The group G is solvable of derived length $\leqslant n$.*

Proof. Let $H \lhd G$ be the maximal solvable subgroup of derived length d of finite index. Thus as above, H is given by imposing polynomial equations of the form $[\![g_1, \ldots, g_{2^d}]\!] = 1$ on tuples of the elements of G, i.e., H is Zariski closed. Since H has finite index in G, it is also open. Since G is connected, it follows that $G = H$, i.e., G is solvable and has derived length $\leqslant n$ by Proposition 12.70. □

It is clear that $\Gamma \cap G$ is a finite index subgroup of Γ whose index is at most $|\overline{\Gamma} : G|$. Unfortunately, the index $|\overline{\Gamma} : G|$ could be arbitrarily large. We will see, however, that we can enlarge G to a (possibly disconnected) subgroup $\widehat{G} \leqslant \overline{\Gamma}$ which is still solvable but has a uniform upper bound on $|\overline{\Gamma} : \widehat{G}|$ and a uniform bound on the derived length.

We will get a bound on the index and the derived length by the dimension induction. The base case where $n = 1$ is clear, so we assume that for each $n' < n$ and each virtually solvable subgroup $\Gamma' \leqslant GL_{n'}(\mathbb{C})$ there exists a solvable group $\widehat{G}'$

$$G' \leqslant \widehat{G}' \leqslant \overline{\Gamma}'$$

as required, with a uniform bound $\nu(n')$ on the index $|\overline{\Gamma}' : \widehat{G}'|$ and so that the derived length of $\widehat{G}'$ is at most $\delta(n') \leqslant \delta(n-1)$.

Let $\mathcal{V} := \{V_1, \ldots, V_\ell\}$ denote the maximal collection of (independent) subspaces in V so that G fixes each $P(V_i)$ pointwise (see Theorem 12.66 and Lemma 12.65). In particular, $\ell \leqslant n$. Since G is normal in $\overline{\Gamma}$, the collection $\mathcal{V}$ is invariant under $\overline{\Gamma}$. Let $K \leqslant \overline{\Gamma}$ denote the kernel of the action of $\overline{\Gamma}$ on the set $\mathcal{V}$. Clearly, $G \leqslant K$ and $|\overline{\Gamma} : K| \leqslant \ell! \leqslant n!$. We will, therefore, study the pair $G \leqslant K$.

Remark 12.75. Note that we just proved that every virtually solvable subgroup $\Gamma \leqslant GL(n, \mathbb{C})$ contains a reducible subgroup of index $\leqslant n!c(n)$, where $c(n) := q(PGL(n, \mathbb{C}))$ is the function from Jordan's Theorem 12.61. Indeed, if $\ell > 1$, the subgroup $K \cap \Gamma$ (of index $\neq n!$) preserves a proper subspace V_1. If $\ell = 1$, then G is contained in $Z_{GL(V)}$ and hence Γ projects to a finite subgroup $\Phi < PGL(V)$. After replacing Φ with an abelian subgroup A of index $\neq q(PGL(V))$ (see Jordan's Theorem 12.61), we obtain a metabelian group $\tilde{A} < \Gamma$ whose center is contained in $Z_{GL(V)}$. Now the assertion follows from Lemma 12.69.

The group K preserves each V_i and, by construction, the group G acts trivially on each $P(V_i)$. Therefore, the image Q_i of K/G in $PGL(V_i)$ is finite. (The finite group K/G need not act faithfully on $P(V_i)$.) By Jordan's Theorem 12.61, the group Q_i contains an abelian subgroup of index $\leqslant c(\dim(V_i)) \leqslant c(n)$. Hence, K contains a subgroup $N \lhd K$ of index at most

$$\prod_{i=1}^{\ell} c(\dim(V_i)) \leqslant c(n)^n$$

which acts as an abelian group on

$$\prod_{i=1}^{\ell} P(V_i).$$

We again note that $G \leqslant N$. The image of the restriction homomorphism $\phi : N \to GL(U)$,

$$U := V_1 \oplus \ldots \oplus V_\ell$$

is therefore a *metabelian* group M.

We also have the homomorphism $\psi : N \to GL(W)$, $W = V/U$ with the image N_W. This group contains the connected solvable subgroup $G_W := \psi(G)$ of finite index. To identify the intersection $\mathrm{Ker}(\phi) \cap \mathrm{Ker}(\psi)$ we observe that $V = U \oplus W$ and the group N acts by matrices of the block-triangular form:

$$\begin{bmatrix} x & y \\ 0 & z \end{bmatrix}$$

where $x \in M$, $z \in N_W$. Then the kernel of the homomorphism $\phi \times \psi : N \to M \times N_W$ consists of matrices of the upper-triangular form

$$\begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix}.$$

Thus by Proposition 12.70, $L = \mathrm{Ker}(\phi \times \psi)$ is solvable of derived length $\leqslant n$.

By the induction hypothesis, there exists a solvable group $\widehat{G_W}$ of derived length $\leqslant \delta(n-1)$, so that

$$G_W \leqslant \widehat{G_W} \leqslant N_W$$

and $|N_W : \widehat{G_W}| \leqslant \nu(n-1)$. Therefore, for $\widehat{G} := (\phi \times \psi)^{-1}(M \times \widehat{G_W})$, we obtain a commutative diagram

$$\begin{array}{ccccccc} 1 \to & L & \longrightarrow & N & \stackrel{\phi\times\psi}{\longrightarrow} & M \times N_W & \longrightarrow \\ & || & & \uparrow \iota' & & \uparrow \iota & \\ 1 \to & L & \longrightarrow & \widehat{G} & \stackrel{\phi\times\psi}{\longrightarrow} & M \times \widehat{G_W} & \longrightarrow \end{array}$$

where ι is the inclusion of index $i \leqslant \nu(n-1)$ subgroup and, hence, ι' is also the inclusion of index i subgroup. Furthermore, L is solvable of derived length $\leqslant n$, $M \times \widehat{G_W}$ is solvable of derived length$\leqslant \max(2, \delta(n-1))$. Putting it all together, we get

$$|\overline{\Gamma} : \widehat{G}| \leqslant \nu(n) := \nu(n-1)n!(c(n))^n,$$

where $\widehat{G}$ is solvable of derived length $\leqslant \delta(n) := \max(2, \delta(n-1)) + n$. Intersecting $\widehat{G}$ with Γ we obtain $\Lambda < \Gamma$ of index at most $\nu(n)$ and derived length $\leqslant \delta(n)$. Theorem 12.72 follows. ☐

Exercise 12.76. Prove that Theorem 12.72 also holds if we replace Zariski closure with the closure with respect to the standard topology.

CHAPTER 13

The Tits Alternative

In this chapter we will prove

THEOREM 13.1 (The Tits Alternative, [**Tit72**]). *Let L be a Lie group with finitely many connected components and let $\Gamma < L$ be a finitely generated subgroup. Then either Γ is virtually solvable or Γ contains a free nonabelian subgroup.*

The main corollary of the Tits Alternative that we will use (in the proof of Gromov's theorem on groups of polynomial growth) is:

COROLLARY 13.2. *Suppose that Γ is a finitely generated subgroup of $GL(n, \mathbb{R})$. Then Γ is either virtually nilpotent or has exponential growth. In particular, Γ has either polynomial or exponential growth.*

PROOF. By the Tits Alternative, either Γ contains a nonabelian free subgroup (and hence Γ has exponential growth) or Γ is virtually solvable. For virtually solvable groups the assertion follows from Theorem 12.37. □

Note that this corollary requires only a weaker version of the Tits Alternative, namely, existence of free subsemigroups, whose proof is sligthly easier than that of the full Tits Alternative.

In view of Corollary 13.2, Milnor's Conjecture (Conjecture 5.84), has affirmative answer for finitely generated subgroups of Lie groups L with finite $\pi_0(L)$. Since the Lie group L in Theorem 13.1 has only finitely many connected components, the intersection Γ_0 of Γ with the identity component $L_0 \leqslant L$ has finite index in Γ. Therefore, it suffices to consider the case of connected Lie groups G. Consider the adjoint representation $Ad : G \to GL(\mathfrak{g})$, where $\mathfrak{g}$ is the Lie algebra of G. The kernel of this homomorphism is a central subgroup of Γ (see Lemma 3.50). Since central coextension of a virtually solvable group is again virtually solvable and Γ contains a free nonabelian subgroup if and only if $Ad(\Gamma)$ does, we reduce the problem to the case of subgroups of $GL(n, \mathbb{R})$. As it turns out, even if we are interested in matrix groups with real coefficients, one still has to consider matrix groups over other fields, and we are lead to proving

THEOREM 13.3 (J. Tits). *The Tits alternative holds for finitely generated subgroups $\Gamma < GL(n, \mathbb{K})$, where $\mathbb{K}$ is an arbitrary field: Either Γ is virtually solvable or contains a free nonabelian subgroup.*

Since Γ is finitely-generated, we can assume that $\mathbb{K}$ is finitely generated as well: Otherwise, we replace it with the subfield generated (over the prime field of $\mathbb{K}$) by the matrix entries of the generators of Γ. Moreover, it clearly suffices to consider the case of infinite fields $\mathbb{K}$, which we will assume from now on.

REMARK 13.4. In §13.6, we prove an analogue of Theorem 13.3 in the case of subgroups $\Gamma < GL(n, \mathbb{F})$ which are not required to be finitely generated, but the field $\mathbb{F}$ has zero characteristic.

13.1. Outline of the proof

In this section we give an outline of the proof of the Tits alternative, breaking it into a sequence of theorems, which will be proven later on in this chapter. In what follows, V will denote a finite-dimensional vector space over a field $\mathbb{K}$ whose algebraic closure will be denoted $\bar{\mathbb{K}}$. We let $End(V)$ denote the algebra of (linear) endomorphisms of V and $GL(V)$ the group of invertible endomorphisms of V.

We present a step-by-step outline of the proof of the Tits alternative for finitely generated matrix groups, i.e., finitely-generated subgroups $\Gamma < GL(n, \mathbb{K})$, where $\mathbb{K}$'s is an arbitrary field. The proof is by induction on n. The statement is clear for $n = 1$, hence, we assume that it holds for all $m < n$.

Step 1: Reduction to the case of virtually absolutely irreducible subgroups.

Suppose that Γ contains a finite index subgroup Γ_1 which preserves a proper subspace of $\bar{\mathbb{K}}^n$, i.e., the action of Γ on $\bar{\mathbb{K}}^n$ is not irreducible. We will identify the proper subspace as above with $\bar{\mathbb{K}}^m$, $1 \le m < n$.

Then Γ_1 projects to a subgroup $\Gamma_2 < GL(n-m, \bar{\mathbb{K}})$ which, by the induction assumption, is either virtually solvable or contains a nonabelian free subgroup. In the latter case, Γ contains a nonabelian free subgroup as well. In the former case, we consider the kernel $\Gamma_3 \lhd \Gamma_1$ of the homomorphism $\Gamma_1 \to \Gamma_2$. The subgroup Γ_3 is contained in the semidirect product

$$N \rtimes GL(m, \bar{\mathbb{K}})$$

where N is the nilpotent subgroup of $GL(n, \bar{\mathbb{K}})$ consisting of the block-triangular matrices of the form

$$\begin{bmatrix} I_{n-m} & \star \\ 0 & I_m \end{bmatrix}.$$

This subgroup is the kernel of the restriction homomorphism r from the stabilizer of $\mathbb{K}^m$ in $GL(n, \mathbb{K})$ to the subspace $\bar{\mathbb{K}}^m$. If $r(\Gamma_3)$ contains a free nonabelian subgroup, then so does Γ. Otherwise, by the induction assumption, $r(\Gamma_3)$ is virtually solvable. Since $\Gamma_4 = \Gamma_3 \cap N$ is solvable, we obtain two short exact sequences

$$1 \to \Gamma_4 \to \Gamma_3 \to \Gamma_3/\Gamma_4 \to 1,$$

$$1 \to \Gamma_3 \to \Gamma_1 \to \Gamma_2 \to 1.$$

The projection $\Gamma_5 \cong \Gamma_3/\Gamma_4$ of Γ_3 to $GL(m, \bar{\mathbb{K}})$ is virtually solvably by the induction hypothesis; virtual solvability of Γ_4 then implies virtual solvability of Γ_3. Since Γ_2 is virtually solvable, so is Γ_1, see Proposition 11.92.

Therefore, it suffices to consider subgroups Γ of $GL(n, \mathbb{K})$ for various fields $\mathbb{K}$ such that each finite-index subgroup of Γ acts absolutely irreducibly on $\mathbb{K}^n$.

Step 2: Reduction to the case of subgroups of $SL(n, \mathbb{K})$.

Given a finitely generated subgroup $\Gamma < GL(n, \mathbb{K})$ we define the homomorphism

$$\phi : GL(n, \mathbb{K}) \to SL(n, \mathbb{K}), \quad \phi(g) = g \det(g)^{-1}.$$

The kernel of this homomorphism is abelian, contained in the center of $GL(n, \mathbb{K})$. Therefore, Γ is virtually solvable if and only if $\phi(\Gamma)$ is; similarly, Γ contains a free nonabelian subgroup if and only if $\phi(\Gamma)$ does. We leave it as an excercise to the reader to prove:

EXERCISE 13.5. Every finite index subgroup of Γ acts absolutely irreducibly on $\mathbb{K}^n$ if and only if the same is true for $\phi(\Gamma)$.

Therefore, it suffices to analyze the case of finitely generated subgroups of $SL(n, \mathbb{K})$ for various fields $\mathbb{K}$.

Step 3: Reduction to the case of subgroups which contain *non-distal elements*.

Recall that an element $a \in \mathbb{K}$ of a normed field $\mathbb{K}$ is a *unit* if a has unit norm. An element $g \in GL(n, \mathbb{K})$ is called *distal* if all the eigenvalues of g are units.

Consider a finitely generated virtually absolutely irreducible subgroup $\Gamma < SL(n, \mathbb{F})$, where $\mathbb{F}$ is a finitely generated field. If every eigenvalue of each element of Γ is a root of unity, then Γ is virtually nilpotent by Proposition 12.48. Suppose, therefore, that λ is an eigenvalue of some $\gamma \in \Gamma$, which is not a root of unity. Let $\mathbb{E}$ denote the extension of $\mathbb{F}$, which is the splitting field of the characteristic polynomial of γ. By Theorem 1.111, there exists an embedding $\mathbb{E} \hookrightarrow (\mathbb{K}, |\cdot|)$ into a normed local field, such that the image of λ in $\mathbb{K}$ is not a unit. Hence, the image of γ under the embedding $\Gamma \hookrightarrow SL(n, \mathbb{K})$, is non-distal.

This reduces the problem to the case of local fields $\mathbb{K}$ and finitely generated subgroups $\Gamma < SL(n, \mathbb{K})$, acting absolutely irreducibly on $\mathbb{K}^n$, such that at least one element $\gamma \in \Gamma$ is non-distal (cf. Exercise 3.44). The claim is that such Γ contains a free nonabelian subgroup.

Step 4: Finding very proximal elements. Suppose that $\Gamma < SL(n, \mathbb{K})$ is a subgroup satisfying the conclusion of Step 3 and $\gamma \in \Gamma$ is a non-distal element. Not all the norms of the eigenvalues of γ are the same, since, their product equals 1.

We let $\lambda_1, \ldots, \lambda_n \in \bar{\mathbb{K}}$ denote the eigenvalues of γ, ordered so that

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_k| > |\lambda_{k+1}| \geqslant \ldots \geqslant |\lambda_n|$$

We now consider the action of $SL(n, \mathbb{K})$ on the exterior product

$$W = \Lambda^k V, \quad V = \mathbb{K}^n.$$

According to Lemma 3.43, the action of G on $W = \Lambda^k V$ is again absolutely irreducible.

By the construction, the action of γ on W is *proximal*: The dominant eigenvalue of γ is λ_1^k, which has multiplicity one (see §1.15 for the definition of dominance). If one is interested only in finding free subsemigroups of Γ, proximal elements suffice for constructing semi ping-pong pairs in Γ, see §13.3. However, finding free subgroups requires more work:

THEOREM 13.6. Γ *contains an element* β *whose action on* W *is* very proximal β*, i.e., both* β *and* β^{-1} *are proximal elements of* $GL(W)$*.*

Our arguments in this part of the proof are inspired by the metric geometry arguments in the papers by Breuillard and Gelander, [**BG03, BG08b**], where they proved an *effective version* of the Tits alternative.

Step 5: Finding ping-pong partners. Suppose that $\Gamma < SL(n, \mathbb{K}) = SL(V)$ is absolutely irreducible and Γ contains an element γ whose action on $W = \Lambda^k V$ is very proximal.

Theorem 13.7. *The subgroup $\Gamma < GL(W)$ contains a pair of ping-pong partners α, β in the sense of Definition 4.67.*

Now that $\Gamma < GL(W)$ contains a pair of ping-pong partners α, β, there exists $m > 0$ such that the subgroup of Γ generated by α^m, β^m is free of rank 2, see Lemma 4.68. This step will conclude the proof of Theorem 13.3.

Clearly, the proof of Theorem 13.3 is now reduced to Theorems 13.6 and 13.7.

13.2. Separating sets

Definition 13.8. A subset $F \subset PGL(V)$ is called *m-separating* if for every choice of points $p_1, \ldots, p_m \in P = P(V)$ and hyperplanes $H_1, \ldots, H_m \subset P$, there exists $f \in F$ such that

$$f^{\pm 1}(p_i) \notin H_j, \forall i, j = 1, \ldots, m.$$

Proposition 13.9. *Let $\Gamma < GL(V)$ be a subgroup acting irreducibly on V with the property that its Zariski closure is Zariski–irreducible. For every m, Γ contains a finite m-separating subset F.*

Proof. Let $\overline{\Gamma}$ be the Zariski closure of Γ in $GL(V)$. Let $P^\vee$ denote the space of hyperplanes in P (i.e. the projective space of the dual of V). For each $g \in \overline{\Gamma}$ let $M_g \subset P^m \times (P^\vee)^m$ denote the collection of $2m$-tuples

$$(p_1, \ldots, p_m, H_1, \ldots, H_m)$$

such that

$$g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j$$

for some $i, j = 1, \ldots, m$.

Lemma 13.10. *If Γ is as in Proposition 13.9 then*

$$\bigcap_{g \in \Gamma} M_g = \emptyset\,.$$

Proof. Suppose to the contrary that the intersection is nonempty. Then there exists a $2m$-tuple $(p_1, \ldots, p_m, H_1, \ldots, H_m)$ such that for every $g \in \Gamma$,

$$\exists\, i, j \text{ such that } g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j\,. \tag{13.1}$$

The set of elements $g \in GL(V)$ such that (13.1) holds for the given $2m$-tuple is Zariski–closed, and G is the Zariski closure of Γ, hence all $g \in G$ also satisfy (13.1).

Let $G^{\pm}_{p_i, H_j} \subset \overline{\Gamma}$ denote the subset set consisting of those $g \in \overline{\Gamma}$ for which

$$g^{\pm 1}(p_i) \in H_j.$$

Clearly, these subsets are Zariski–closed and cover the group $\overline{\Gamma}$. Since $\overline{\Gamma}$ Zariski–irreducible, it follows that one of these sets, say $G^{+}_{p_i, H_j}$, is the entire of $\overline{\Gamma}$. Therefore, for every $g \in G$, $g(p_i) \in H_j$. Thus, projectivization of the vector subspace L spanned by the $\overline{\Gamma}$–orbit (of lines) $\overline{\Gamma} \cdot p_i$ is contained in H_j. The subspace L is proper and G-invariant. This contradicts the hypothesis that $\overline{\Gamma}$ acts irreducibly on V. $\square$

We now finish the proof of Proposition 13.9. Let M_g^c denote the complement of M_g in $P^m \times (P^\vee)^m$. This set is Zariski open. By Lemma 13.10, the sets M_g^c

($g \in \Gamma$) cover the space $P^m \times (P^\vee)^m$. By Lemma 3.68, there exists a finite subset $F \subset \Gamma$ so that

$$\bigcap_{f\in F} M_f = \bigcap_{g\in\Gamma} M_g = \emptyset.$$

In other words,

$$\bigcup_{f\in F} M_f^c = P^m \times (P^\vee)^m.$$

This subset F is the one whose existence is claimed by Proposition 13.9. □

13.3. Proof of existence of free subsemigroup

In this section we prove a weaker form of the Tits Alternative. The reader who is only interested in the proof of Gromov's theorem on groups of polynomial growth can read this proof instead of the one given in §13.4, since it is technically much simpler. We refer the reader to §1.15 for the definitions of subspaces $E_\alpha, A_\alpha \subset P(V)$ associated with linear transformations $\alpha \in GL(V)$ and to Definition 4.67 for the notion of *ping-partners*.

THEOREM 13.11. *Let $\Gamma < GL(n, \mathbb{K})$, $n > 1$, be a subgroup which acts virtually irreducibly on $V = \mathbb{K}^n$ and contains a proximal element α. Then Γ contains a free subsemigroup of rank 2.*

PROOF. Let $\overline{\Gamma}$ denote the Zariski closure of Γ in $GL(n,\mathbb{K})$. If $\overline{\Gamma}$ is Zariski–reducible, we replace Γ with the finite-index subgroup $\Gamma_0 < \Gamma$, the intersection of Γ with the identity component of $\overline{\Gamma}$, cf. Proposition 3.86. Thus, we will assume that $\overline{\Gamma}$ is Zariski–irreducible. We claim that Γ contains an element γ such that the elements $\alpha, \beta = \gamma\alpha\gamma^{-1}$ of the group Γ are *ping-partners* (Definition 4.67). Indeed, since Γ contains a 2-separating subset F (see Proposition 13.9), there exists $\gamma \in F$ such that

$$\{\gamma(A_\alpha), \gamma^{-1}(A_\alpha)\} \cap E_\alpha = \emptyset. \tag{13.2}$$

Being a conjugate of α, the element $\beta = \gamma\alpha\gamma^{-1}$ is also proximal and

$$A_\beta = \gamma(A_\alpha), \quad E_\beta = \gamma(E_\alpha).$$

Then (13.2) implies that

$$A_\alpha \notin E_\beta, \quad A_\beta \notin E_\alpha,$$

which means that α, β are ping-partners. Therefore, by Lemma 4.68, there exists $m > 0$ such that the subsemigroup of Γ generated by α^m, β^m is free of rank two. □

13.4. Existence of very proximal elements: Proof of Theorem 13.6

In what follows, V is an n-dimensional vector space over an (infinite) local field $\mathbb{K}$, $n = \dim(V) > 1$. We fix a basis $e_1, \ldots, e_n$ in V. Then the norm $|\cdot|$ on $\mathbb{K}$ determines the Euclidean norms $\|\cdot\|$ on V.

NOTATION 13.12. We let $P(V)$ denote the projective space of V. When there is no possibility of confusion we do not mention the vector space anymore, and simply denote the projective space by P.

In this section we show how to find very proximal elements in a subgroup $\Gamma < GL(V)$, assuming that Γ contains a proximal element g. We will find such very proximal elements in the form

$$f' g^i f g^{-i},$$

where f, f' belong to some finite 4-separating subset of Γ. While this appears to be a linear algebra problem, we will use geometric and topological arguments instead. To this end, in the following section we will present some *geometric conditions* for proximality, based on the contraction principle. Below, for each nonzero vector $v \in V$ we let $[v] \in P(V)$ denote its projection to the projective space; we will use the metric on $P(V)$ and metric balls $B([v], R) \subset P(V)$, see §1.14.

13.4.1. Proximality criteria.

PROPOSITION 13.13 (Proximality criterion-I). *Suppose that $g \in GL(V)$ and u is an eigenvector of g such that for some $R > 0$ the restriction of g to $B([u], R)$ is $\frac{1}{2}$-Lipschitz. Then g is proximal and $[u] = A_g$.*

PROOF. Suppose that g is not proximal and let $v \in V$ be an eigenvector of g linearly independent of u, such that $|\lambda_v| \geqslant |\lambda_u|$, where λ_u, λ_v are the eigenvalues for the eigenvectors u, v respectively. The linear transformation g preserves the 2-dimensional subspace $W = \text{Span}(u, v) \subset V$. The inequality $|\lambda_v| \geqslant |\lambda_u|$ implies that the sequence $g^i, i \in \mathbb{N}$ is either bounded in $GL(W)$ or its projection to $PGL(W)$ converges to $[v]$ uniformly on compacts in $P(W) \setminus \{[u]\}$, cf. Theorem 1.129. However, since the restriction of g to $B([u], R)$ is $\frac{1}{2}$-Lipschitz, $g(B([u], R)) \subset B([u], R)$ and $B([u], R) \neq \{[u]\}$ since the field $\mathbb{K}$ is infinite and local. Contradiction. □

PROPOSITION 13.14 (Proximality criterion-II). *Suppose that $g_i \in GL(V)$ is a sequence such that for some $a \in P$ and $R > 0$, the sequence $g_i\big|_{\bar{B}(a,R)}$ converges uniformly to a. Then for all but finitely many values of i the elements g_i are proximal. Moreover,*

$$\lim_{i \to \infty} A_{g_i} = a.$$

PROOF. Since the sequence g_i converges to a on the ball $B(a, R)$, it follows that the sequence of projective transformations g_i converges to the quasiconstant map with the image a on the projective space $P(V)$, see Theorem 1.129 and Exercise 1.127. It follows from Lemma 1.130 that

$$\lim_{i \to \infty} Lip(g_i\big|_{B(a,R)}) = 0.$$

Pick a positive number $\epsilon < R$. After discarding finitely many members of the sequence g_i, we can assume that the restrictions of g_i to $\bar{B}(a, R)$ are $\frac{1}{2}$-Lipschitz and that

$$g_i(\bar{B}(a, R)) \subset B(a, \epsilon).$$

Therefore, for each $g = g_i$ and $m \geq 1$, we have

$$\text{diam}(g^m(\bar{B}(a, R))) \leqslant 2^{-m}.$$

Since the normed field $\mathbb{K}$ is complete, the sequence of iterates g^m converges on $\bar{B}(a, R)$ to a point $[u]$ contained in $B(a, \epsilon)$, which, therefore, has to be fixed by g. By Proposition 13.13, u is an eigenvector of g with dominant eigenvalue. Hence, g is proximal with $A_g \in B(a, \epsilon)$. This also shows that

$$\lim_{i \to \infty} A_{g_i} = a. \quad \square$$

Proposition 13.15 (Proximality criterion-III). *Suppose that $g_i \in GL(V)$ is a sequence such that for some $c \in P$ and $R > 0$, the sequence $g_i\big|_{\bar{B}(c,R)}$ converges uniformly to a point $a \in P(V)$. Assume also that the sequence (g_i) converges to some quasiprojective transformation $\hat{g} \in End(V)$, whose kernel $\mathrm{Ker}_{\hat{g}} \subset P(V)$ does not contain the point a. Then all but finitely many members of the sequence g_i are proximal and*

$$\lim_{i\to\infty} A_{g_i} = a.$$

Proof. The proof is similar to that of Proposition 13.14. The sequence (g_i) converges to $\hat{g}$ uniformly on compacts in $P(V) \setminus \mathrm{Ker}_{\hat{g}}$. By Exercise 1.127, $\hat{g}$ is a quasiconstant, since it maps the entire ball $B(c,R)$ to the point a. Since $\mathrm{Ker}_{\hat{g}}$ does not contain a, the sequence of restrictions of g_i's to a small ball $B(a,\epsilon)$ converges uniformly to a. Now, the claim follows from Proposition 13.14. □

13.4.2. Constructing very proximal elements. We now prove one of the two main results of this section:

Theorem 13.16. *Let $F \subset SL(V)$ be a 4-separating subset and let $g \in SL(V)$ be a proximal element. Then there exist $f, f' \in F$ such that for infinitely many i's, the elements*

$$g_i = f' g^i f g^{-i}$$

are very proximal.

Proof. We break the proof into two propositions, first ensuring proximality of g_i's and the second verifying proximality of g_i^{-1}.

Proposition 13.17. *Let $g \in GL(V)$ be a proximal element and $F \subset GL(V)$ be a 2-separating subset. Then for each infinite subset $I \subset \mathbb{N}$ there exist $f, f' \in F$ and an infinite subset $J \subset I$, such that the products $g_i = f' g^i f g^{-i}, i \in J$, satisfy:*

1. Each g_i is proximal.

2.

$$\lim_{i\to\infty, i\in J} A_{g_i} = f'(A_g).$$

Proof. Since g is proximal, the sequence $(g^i)_{i\in\mathbb{N}}$ converges to a quasiconstant map

$$\hat{g} : P \setminus E_g \to A_g.$$

We first consider the sequence $g^{-i}, i \in I$. By the convergence property (Theorem 1.129), this sequence subconverges to some non-invertible quasiprojective transformation $\check{g} \in End(P(V))$. We let $J \subset I$ denote the subset such that

$$\lim_{i\to\infty, i\in J} g^{-i} = \check{g}.$$

We pick a small ball $\bar{B}(x,\epsilon)$ disjoint from $\mathrm{Ker}(\check{g})$, its image under $\check{g}$ is contained in a small ball $\bar{B}(\check{g}(x), L\epsilon) \subset \mathrm{Im}(\check{g})$, where L is the Lipschitz constant of $\check{g}$. Since the set F is 2-separating, there exists $f \in F$ such that

$$f(\check{g}(x)) \notin E_g.$$

By using a sufficiently small ϵ, we can assume that $f(\bar{B}(\check{g}(x), L\epsilon))$ is disjoint from E_g as well. We let

$$E_{\hat{g}\circ f\circ\check{g}} = \mathrm{Ker}_{\hat{g}\circ f\circ\check{g}}$$

denote the indeterminacy set of the quasiconstant map $\hat{g} \circ f \circ \check{g}$, whose image is A_g. To be more precise, this set is the hyperplane in P such that the suitable subsequence of

$$g^i \circ f \circ \circ g^{-i}$$

converges to the constant A_g away from this hyperplane. (The limit is indeed a quasiconstant map since its restriction to the ball $B(x, \epsilon)$ is a constant map, see Exercise 1.127.) Using again the fact that F is a 2-separating subset, we find $f' \in F$ such that

$$f'(A_g) \notin E_{\hat{g} \circ f \circ \check{g}}.$$

Thus, the composition

$$g_i := f' \circ g^i \circ f \circ \circ g^{-i}$$

converges to the constant $f'(A_g)$ uniformly on compacts away from the hyperplane

$$E_{\hat{g} \circ f \circ \check{g}},$$

which is disjoint from $f'(A_g)$. Therefore, according to Proposition 13.15, for all but finitely many $i \in J$, the composition g_i is proximal and

$$\lim_{i \to \infty, i \in J} A_{g_i} = f'(A_g). \quad \square$$

Our goal, of course, is to find *very proximal elements*, not just proximal ones. We will see now that this can be achieved by using compositions of the same type as in Proposition 13.17, but with a slightly more careful choice of the separating elements f, f'.

PROPOSITION 13.18. *Let $g \in GL(V)$ be a proximal element and $F \subset GL(V)$ be a 4-separating subset. Then for each infinite subset $I \subset \mathbb{N}$ there exist $f, f' \in F$ and an infinite subset $J \subset I$, such that the transformations $g_i = f g^i f g^{-i}, i \in J$, satisfy:*
1. g_i, g_i^{-1} are both proximal.
2.

$$\lim_{i \to \infty, i \in J} A_{g_i} = A_g, \quad \lim_{i \to \infty, i \in J} A_{g_i^{-1}} = A_g.$$

PROOF. We follow the proof of Proposition 13.17, except that we now impose more restrictions on the elements f, f' (using the 4-separation property).

1. We take $f \in F$ such that, in addition to the earlier requirement, we have:

$$f^{-1}(\check{g}(x)) \notin E_g.$$

2. Similarly, we take f' which, additionally, satisfies

$$f'(x) \notin E_{\hat{g} f \check{g}}.$$

Then, taking into account the fact that

$$g_i^{-1} = g^i \circ f^{-1} \circ g^{-i} \circ (f')^{-1},$$

and arguing as in the proof of Proposition 13.17, we obtain that the sequence g_i^{-1} converges on a small ball around $f'^{-1}(x)$ to the constant A_g. Hence, we conclude that (for infinitely many values of i), not only g_i is proximal, but also g_i^{-1} is proximal and

$$\lim_{i \to \infty} A_{g_i^{-1}} = A_g. \quad \square$$

This also concludes the proof of Theorem 13.16. $\square$

13.5. Finding ping-pong partners: Proof of Theorem 13.7

Let $F \subset GL(V)$ be a 2-separating subset and suppose that $g \in GL(V)$ is a very proximal element.

LEMMA 13.19. *There exists $f \in F$ such that the elements $g, h = fgf^{-1}$ are a ping-pong partners.*

PROOF. For each $f \in GL(V)$, setting $h = fgf^{-1}$, we get:

$$A_h = f(A_g), \quad E_h = f(E_g).$$

In order to ensure that g, h are ping-pong partners, we need to find f such that

$$A_h \notin E_g, \quad \text{equivalently} \quad f(A_g) \notin E_g,$$

and

$$A_g \notin E_h, \quad \text{equivalently} \quad f^{-1}(A_g) \notin E_g.$$

Existence of such f is immediate from the assumption that the set F is 2-separating. □

THEOREM 13.20. *Let $\Gamma \leqslant GL(V)$ be a finitely generated subgroup, acting virtually irreducibly on V and containing a proximal element. Then Γ contains a free non-abelian subgroup.*

PROOF. It remains to summarize what we already have done. After passing to a finite index subgroup in Γ, we can assume that the Zariski closure of Γ is a Zariski-irreducible subgroup of $GL(V)$. Then, by 13.9, Γ contains a 4-separating subset $F \subset \Gamma$. Using this subset, in §13.4.2, we converted a proximal element of Γ into a very proximal element and then (Lemma 13.19) into a pair of ping-pong partners $\alpha, \beta \in \Gamma$. Lastly, by Lemma 4.68, for some $m > 0$, the group generated by α^m, β^m is free of rank 2. □

This also concludes the proof of the Tits Alternative (Theorem 13.3) for finitely generated subgroups of $GL(n, \mathbb{K})$, where $\mathbb{K}$ is an arbitrary field, as well as the Tits Alternative for subgroups of Lie groups (Theorem 13.1).

As a consequence of the proof of the Tits Alternative we obtain:

THEOREM 13.21. *Let Γ be a finitely generated group that does not contain a free non-abelian subgroup. Then:*

1. *If Γ is a subgroup of a real algebraic group G then the Zariski closure $\overline{\Gamma}$ of Γ in G is virtually solvable.*
2. *If Γ is a subgroup of a Lie group G with finitely many connected components, then the closure $\widehat{\Gamma}$ of Γ in the Lie group G (with respect to the standard topology) is virtually solvable.*

Furthermore, $\overline{\Gamma}$ and $\widehat{\Gamma}$ contain a solvable subgroup S of derived length at most $\delta = \delta(G)$ and the index at most $\nu = \nu(G)$.

PROOF. We will prove the first statement as the proof of the second statement is completely analogous (cf. Exercise 12.76). First of all, after replacing G with its finite index subgroup, we may assume that the group G is irreducible (with respect to the Zariski topology). According to the Tits Alternative, the group Γ is virtually solvable. Furthermore, the adjoint representation $\rho : G \to GL(\mathfrak{g})$ has abelian kernel. As in the proof of Theorem 12.72, we conclude that the group $G_1 := \overline{\rho(\Gamma)}$ is virtually solvable. According to the same theorem, γ_1 contains a solvable subgroup S_1 of index $\leq \nu(n)$ and derived length $\leq \delta(n)$, where n is the dimension of G. The preimage $\rho^{-1}(G_1) \leq G$ contains the algebraic closure $\overline{\Gamma}$ and $|\rho^{-1}(G_1) : \rho^{-1}(S_1)| \leq \nu(n)$, while $\rho^{-1}(S_1)$ has derived length $\leq \delta(n)+1$. The same estimates hold for the intersections $S := \rho^{-1}(S_1) \cap \overline{\Gamma}$. Theorem follows. ☐

13.6. The Tits Alternative without finite generation assumption

THEOREM 13.22 (The Tits Alternative without finite generation assumption). *Let $\mathbb{K}$ be a field of zero characteristic and Γ be a subgroup of $GL(n, \mathbb{K})$. Then either Γ is virtually solvable or Γ contains a free nonabelian subgroup.*

PROOF. We will need the following elementary lemma:

LEMMA 13.23. *Every countable field $\mathbb{L}$ of zero characteristic embeds into $\mathbb{C}$.*

PROOF. Since $\mathbb{L}$ has characteristic zero, its prime subfield $\mathbb{P}$ is isomorphic to $\mathbb{Q}$. Then $\mathbb{L}$ is an extension of the form

$$\mathbb{P} \subset \mathbb{E} \subset \mathbb{L},$$

where $\mathbb{P} \subset \mathbb{E}$ is a purely transcendental extension and $\mathbb{E} \subset \mathbb{L}$ is an algebraic extension (see [**Hun80**, Chapter VI.1]). Since $\mathbb{L}$ is countable, $\mathbb{E}/\mathbb{P}$ has countable dimension and, therefore,

$$\mathbb{E} = \mathbb{P}(u_1, \ldots, u_m)$$

or

$$\mathbb{E} = \mathbb{P}(u_1, \ldots, u_m, \ldots).$$

Sending variables u_j to algebraically independent transcendental numbers $z_j \in \mathbb{C}$, we then obtain an embedding $\mathbb{E} \hookrightarrow \mathbb{C}$. Since $\mathbb{C}$ is algebraically closed, the algebraic closure of $\mathbb{E}$ embeds in $\mathbb{C}$. Therefore, $\mathbb{F}$ also embeds in $\mathbb{C}$. ☐

We now return to the proof of the theorem. The group Γ is the direct limit of the direct system of its finitely generated subgroups Γ_i. Let $\mathbb{K}_i \subset \mathbb{K}$ denote the subfield generated by the matrix entries of the generators of Γ_i. Then $\Gamma_i \leqslant GL(n, \mathbb{K}_i)$. Since $\mathbb{K}$ (and, hence, every $\mathbb{K}_i$) has zero characteristic, the field $\mathbb{K}_i$ embeds in $\mathbb{C}$ (see Lemma 13.23).

If one of the groups Γ_i contains a free nonabelian subgroup, then so does Γ. Assume, therefore, that this does not happen. Then, in view of the Tits Alternative (for finitely generated matrix groups), each Γ_i is virtually solvable. For $\nu = \nu(GL(n, \mathbb{C}))$ and $\delta = \delta(GL(n, \mathbb{C}))$, for every i there exists a subgroup $\Lambda_i \leqslant \Gamma_i$ of index $\leqslant \delta$, such that Λ_i has derived length $\leqslant \delta$ (see Theorem 12.72). In view of Theorem 11.95, the group Γ is also virtually solvable. ☐

COROLLARY 13.24. *$SU(2)$ contains a subgroup isomorphic to F_2.*

PROOF. The group $SU(2)$ is connected, therefore, it has no proper finite index subgroups. The group $SU(2)$ cannot be solvable, for instance, because it contains A_5, the alternating group on 5 symbols, which is a nonabelian finite simple group. Alternatively, if $SU(2) < SL(2, \mathbb{C})$ were solvable, it would preserve a proper subspace in $\mathbb{C}^2$ according to Theorem 12.66, which is not the case. Now, the claim follows from Theorem 13.22.

13.7. Groups satisfying the Tits Alternative

One says that a group G *satisfies the Tits' Alternative* if it is either virtually solvable or contains a free nonabelian subgroups.

Classes of groups (besides those covered by Theorems 13.1 and 13.22) satisfying the Tits' Alternative are:

(1) Subgroups of Gromov hyperbolic groups (see [**Gro87**, §8.2.F], [**GdlH90**, Chapter 8].

(2) Subgroups of the mapping class group, see [**Iva92**].

(3) Subgroups of $Out(F_n)$, see [**BFH00, BFH05, BFH04**].

(4) Fundamental groups of 3-dimensional manifolds.

(5) Fundamental groups of compact manifolds of nonpositive curvature, see [**Bal95**].

(6) Groups acting isometrically properly discontinuously and cocompactly on two-dimensional CAT(0) complexes [**BS99, BB95**].

(7) Subgroups of cube complex groups (Sageev-Wise, [**SW05**]): If G acts properly on a finite-dimensional cube complex and has a bound on order of finite subgroups, then each subgroup of G either contains F_2 or is virtually abelian.

(8) Certain classes of $CAT(0)$ groupos, see [**Xie06**], [**AM15**].

CHAPTER 14

Gromov's Theorem

The main objective of this chapter is to prove the converse of Bass–Guivarc'h Theorem 12.26. We refer the reader to §5.7 for the definition of the growth function $\mathfrak{G}$ for finitely generated groups.

THEOREM 14.1 (M. Gromov, [**Gro81a**]). *If* Γ *is a finitely generated group of polynomial growth then* Γ *is virtually nilpotent.*

We will actually prove a slightly stronger version (Theorem 14.3 below) of Theorem 14.1, which is due to van der Dries and Wilkie [**dDW84**] (our proof mainly follows [**dDW84**]).

DEFINITION 14.2. A finitely generated group Γ has *weakly polynomial growth* of degree $\leqslant a$ if there exists a sequence of positive numbers R_n diverging to infinity and a pair of numbers C and a, for which

$$\mathfrak{G}(R_n) \leqslant CR_n^a, \forall n \in \mathbb{N}.$$

THEOREM 14.3. *If* Γ *has weakly polynomial growth then it is virtually nilpotent.*

Gromov's proof of polynomial growth theorem relies heavily upon the work of Gleason, Yamabe, Montgomery and Zippin on topological transformation groups. Therefore in the following section we review some of the results in the theory of topological transformation groups.

14.1. Topological transformation groups

Throughout this section we will consider only *metrizable* topological spaces X. We will topologize the group of homeomorphisms $\mathrm{Homeo}(X)$ *via* the compact-open topology and, thus, obtain a continuous action $\mathrm{Homeo}(X) \times X \to X$.

The following definition (Property A of topological groups introduced by Montgomery and Zippin) should not be confused with the Property A in geometric group theory, introduced by G. Yu in [**Yu00**].

DEFINITION 14.4. [Property A, section 6.2 of [**MZ74**]] Suppose that H is a separable, locally compact topological group. Then H is said to satisfy Property A if for each neighborhood V of the identity $e \in H$ there exists a compact subgroup $K \subset H$ such that $K \subset V$ and H/K (equipped with the quotient topology) is a Lie group.

In other words, the group H can be *approximated by* the Lie groups H/K. Here is an example to keep in mind. Let H be the additive group $\mathbb{Q}_p$ of p-adic numbers. The sets

$$H_{i,p} := \{x \in \mathbb{Q}_p : |x|_p \leqslant p^{-i}\}, i \in \mathbb{N},$$

are open and form a basis of topology at $0 \in \mathbb{Q}_p$. For instance, for $i = 0$, $H_{0,p}$ is the group of p-adic integers O_p. Now, the fact that the p-adic norm $|x|_p$ is nonarchimedean implies that $H_{i,p}$ is a subgroup of H. Furthermore, this subgroup is closed and, therefore, compact, see Lemma 1.102. The quotient $H/H_{i,p}$ has discrete quotient topology since $H_{i,p}$ is open in H. Hence, $G_{i,p} = H/H_{i,p}$ is a Lie group. In particular, H has Property A.

THEOREM 14.5 (D. Montgomery and L. Zippin, [**MZ74**], Chapter IV). *Each separable locally compact group H contains an open subgroup $\hat{H} \leqslant H$ such that $\hat{H}$ satisfies Property A.*

The following theorem is proven in [**MZ74**], section 6.3, Corollary on page 243.

THEOREM 14.6 (D. Montgomery and L. Zippin). *Suppose that X is a topological space which is connected, locally connected, finite-dimensional and locally compact. Suppose that H is a separable locally compact group satisfying Property A, $H \times X \to X$ is a topological action which is effective and transitive. Then H is a Lie group.*

We are mainly interested in the following corollary for metric spaces.

THEOREM 14.7. *Let X be a metric space which is proper, connected, locally connected and finite-dimensional. Let H be a closed subgroup in* Homeo(X) *with the compact-open topology, such that $H \curvearrowright X$ is transitive. If there exists $L \in \mathbb{R}$ such that each $h \in H$ is L–Lipschitz, then the group H is a Lie group with finitely many connected components.*

PROOF. It is clear that $H \times X \to X$ is a continuous effective action. It follows from the Arzela-Ascoli theorem that H is locally compact.

LEMMA 14.8. *1. The group H is separable.*
2. For each open subgroup $U \subset H$, the quotient H/U is countable.

PROOF. 1. Pick a point $x \in X$. Given $r \in \mathbb{R}_+$, consider the subset

$$H_r = \{h \in H : \text{dist}(x, h(x)) \leqslant r\}.$$

By the Arzela–Ascoli theorem, each H_r is a compact set. Therefore

$$H = \bigcup_{r \in \mathbb{N}} H_r$$

is a countable union of compact subsets. Thus, it suffices to prove separability of each H_r. Recall that $\bar{B}(x, R)$ denotes the closed R-ball in X centered at the point x. For each $R \in \mathbb{R}_+$ define the map

$$\phi_R : H \to C_L(\bar{B}(x, R), X)$$

given by the restriction $h \mapsto h|\bar{B}(x, R)$. Here $C_L(\bar{B}(x, R), X)$ is the space of L-Lipschitz maps from $\bar{B}(x, R)$ to X. Observe that $C_L(\bar{B}(x, R), X)$ is metrizable *via*

$$\text{dist}(f, g) = \max_{y \in \bar{B}(x,R)} \text{dist}(f(y), g(y)).$$

Thus, the image of H_r in each $C_L(\bar{B}(x, R), X)$ is a compact metrizable space. We claim now that each $\phi_R(H_r)$ is separable. Indeed, for each $i \in \mathbb{N}$ take $\mathcal{E}_i \subset \phi_R(H_r)$ to be a $\frac{1}{i}$-net. The union

$$\bigcup_{i \in \mathbb{N}} \mathcal{E}_i$$

is a dense countable subset of $\phi_R(H_r)$. On the other hand, the group H (as a topological space) is homeomorphic to the *inverse limit*

$$\varprojlim_{R\in\mathbb{N}} \phi_R(H),$$

see Section 1.5.

Let $E_i \subset \phi_i(H_r)$ be a dense countable subset. For each element $h_i \in E_i$ consider a sequence $(g_j) = \tilde{h}_i$ in the above inverse limit such that $g_i = h_i$. Let $\tilde{h}_i \in H$ be the element corresponding to this sequence (g_j). It is clear now that

$$\bigcup_{i\in\mathbb{N}} \{\tilde{h}_i \in H \ ; \ h_i \in E_i\}$$

is a dense countable subset of H_r.

2. Let $I \subset H$ be a dense countable set. The subsets

$$hU, h \in H$$

are open subsets of H such that $hU = gU$ or $hU \cap gU = \emptyset$ for all $g, h \in H$. The countable subset I intersects every $hU, h \in H$. Therefore, the above collection of open subsets of H consists of countably many elements. □

Thus, we now know that the topological group H is locally compact and separable. Therefore, by Theorem 14.5, H contains an open subgroup $\hat{H} \leqslant H$ satisfying Property A.

LEMMA 14.9. *For every $x \in X$, the orbit $Y := \hat{H}x \subset X$ is open in X.*

PROOF. If Y is not open then it has empty interior (since $\hat{H}$ acts transitively on Y). Since $\hat{H} \subset H$ is closed, the Arzela–Ascoli theorem implies that Y is closed as well. Since $\hat{H}$ is open in H, by Lemma 14.8 the coset $S := H/\hat{H}$ is countable. Choose representatives g_i of S, $i \in I$, where I is countable. Then

$$\bigcup_{i\in I} g_i Y = X.$$

Therefore, the space X is a countable union of closed subsets with empty interior. However, by Baire's theorem, each first category subset in the locally compact metric space X has empty interior. Contradiction. □

We now can conclude the proof of Theorem 14.7. Let $Z \subset Y$ be the connected component of x in $Y := \hat{H}x$ as above. The stabilizer $F \subset \hat{H}$ of Z is both closed and open in $\hat{H}$. Therefore, F again has the Property A and the assumptions of Theorem 14.6 are satisfied by the action $F \curvearrowright Z$. It follows that F is a Lie group. Since $F \subset H$ is an open subgroup, the group H is a Lie group as well. Let K be the stabilizer of x in H. The subgroup K is a compact Lie group and, therefore, has only finitely many connected components. Since the action $H \curvearrowright X$ is transitive, X is homeomorphic to H/K, see Lemma 3.37. Connectedness of X now implies that H has only finitely many connected components. □

14.2. Regular Growth Theorem

We now proceed to construct, for a group Γ of weakly polynomial growth, a representation $\rho : \Gamma \to \mathrm{Isom}(X)$, where X is a homogeneous metric space as in Theorem 14.7.

The first naive attempt would be to take X to be a Cayley graph $\mathrm{Cayley}(\Gamma, S)$ of Γ. But in that case $\mathrm{Isom}(X)$ does not act transitively on X. If we replace the Cayley graph with its set of vertices, then we achieve homogeneity but loose connectedness. The ingenious idea of Gromov is to take X to be a limit of rescaled Cayley graphs $(\mathrm{Cayley}(\Gamma, S), \lambda_n \mathrm{dist})$, where λ_n is a sequence of positive numbers converging to 0. Gromov originally used Gromov-Hausdorff convergence to define the limit; we will take X to be an asymptotic cone of $\mathrm{Cayley}(\Gamma, S)$ instead; equivalently X is an asymptotic cone of Γ with the word metric. Such an asymptotic cone inherits both the homogeneity from Γ (see Proposition 7.71) and the property of being geodesic from $\mathrm{Cayley}(\Gamma, S)$ (see (2) in Proposition 7.67). In particular it is connected and locally connected. The asymptotic cone X is also complete, by Proposition 7.69. These properties and the Hopf-Rinow Theorem 1.60 imply that in order to prove properness of X it suffices to prove local compactness.

To sum up, if we wish to apply Theorem 14.7 to an asymptotic cone, it remains to use the hypothesis of polynomial growth to find an asymptotic cone that is locally compact and finite dimensional. In what follows we explain how to choose a scaling sequence $\boldsymbol{\lambda}$ such that $X_\omega = \mathrm{Cone}_\omega(\Gamma, \mathbf{1}, \boldsymbol{\lambda})$ has both properties.

Remark 14.10. We note that once it is proven that $\mathrm{Isom}(X_\omega)$ is a Lie group, one still has to address the issue that the homomorphisms $\Gamma \to \mathrm{Isom}(X_\omega)$ arising as ultralimits of sequences of isometric actions of Γ on its Cayley graph (with rescaled metric), may have finite images.

A metric space X is called *p-doubling* if each R-ball in X can be covered by p balls of radius $R/2$. One way to show that a metric space X is doubling is to estimate the *packing number* of R-balls in X. The *packing number* $p(\bar{B})$ of a ball $\bar{B} = \bar{B}(x, R) \subset X$ is the supremum of cardinalities of $R/2$-separated subsets $\mathcal{N}$ of $\bar{B}$. If $\mathcal{N}$ is a maximal subset as above, then

$$\forall x \in \bar{B} \ \exists y \in \mathcal{N} \text{ such that } d(x, y) \leqslant R/2.$$

(This condition is slightly stronger than the one of being an $R/2$-net.) In other words, the collection of closed balls $\{\bar{B}(x, R/2) : x \in \mathcal{N}\}$ is a covering of B. Thus, there exist a covering of B by $p(B)$ balls of radius $R/2$. If $p(\bar{B}(x, R)) \leqslant p$ for every x and R, then X *has packing number* $\leqslant p$; such X is necessarily p-doubling. The reader should compare this (trivial) statement with the statement of the Regular Growth Theorem below.

Exercise 14.11. Show that doubling implies polynomial growth for uniformly discrete spaces.

Note that, being scale-invariant and invariant under ultralimits, the doubling property passes to asymptotic cones. The following lemma, although logically unnecessary for the proof of Gromov's theorem, motivates its arguments.

Lemma 14.12. *If X is p-doubling then the Hausdorff dimension of X is at most $\log_2(p)$.*

PROOF. Consider a metric ball $\bar{B} = \bar{B}(o, R) \subset X$. We first cover $\bar{B}$ by balls $\bar{B}(x_i, R/2)$, $i = 1, \ldots, p$. We then cover each of the new balls by balls of radius $R/4$ and proceed inductively. On the n-th step of induction we have a covering of $\bar{B}$ by p^n balls of radius $2^{-n}R$. The corresponding finite sum of in the definition of the α-Hausdorff measure of $\bar{B}$ (see (1.12)) then equals

$$\sum_{i=1}^{p^n} 2^{-n\alpha} R^\alpha = R^\alpha \left(\frac{p}{2^\alpha}\right)^n .$$

This quantity is converges to 0 as $n \to \infty$ provided that $p < 2^\alpha$, i.e., $\alpha > \log_2(p)$. Thus, $\mu_\alpha(B) = 0$ for every metric ball in X. Representing X as a countable union of concentric metric balls, we conclude that $\mu_\alpha(X) = 0$ for every $\alpha > \log_2(p)$. □

Thus, every asymptotic cone of a doubling metric space has finite Hausdorff and, hence, finite covering, dimension, see Theorem 1.89.

Although there are spaces of polynomial growth which are not doubling, the Regular Growth Theorem below shows that groups of polynomial growth exhibit doubling-like behavior, which suffices for proving that suitable asymptotic cones are finite-dimensional.

Our discussion below follows the paper of L. Van den Dries and A. Wilkie, [**dDW84**], Gromov's original statement and proof of the Regular Growth Theorem were different (although, some key arguments were quite similar).

THEOREM 14.13 (Regular growth theorem). *Let Γ be a finitely generated group. Assume that there exists a sequence (R_n) such that $R = (R_n)^\omega$ is an infinitely large number in the ultrapower $\mathbb{R}_+^\omega$ and that the growth function satisfies:*

$$\mathfrak{G}_\Gamma(R_n) = |B_\Gamma(1, R_n)| \leqslant CR_n^a \text{, for } \omega\text{-all} n \in \mathbb{N} , \tag{14.1}$$

where $C > 0$ and $a \in \mathbb{N}$ are constants independent of n. Let $\epsilon > 0$.

Then there exists $\eta \in [\log R\, , \, R] \subset \mathbb{R}_+^\omega$ such that the ball $B\left(1, \frac{\eta}{4}\right)$ in the ultrapower Γ^ω endowed with the nonstandard metric defined in (7.3) satisfies the following:

For every $i \in \mathbb{N}$, $i \geqslant 4$, all the $\frac{\eta}{i}$-separated subsets in the ball $\bar{B}\left(1, \frac{\eta}{4}\right)$ have cardinality at most $i^{a+\epsilon}$.

In particular, taking $i = 8$, we see that every $\frac{\eta}{4}$-ball in Γ^ω has packing number $\leqslant 8^{a+\epsilon}$ (with respect to the nonstandard metric).

We refer the reader to Definition 7.32 and Exercise 7.33 for the discussion of infinitely large numbers. The difference between the assertion of this theorem and the statement that Γ^ω has finite packing number is that *we are not estimating packing numbers of all metric balls, but only of metric balls of certain radii.*

PROOF. Suppose to the contrary that for every $\eta \in [\log R\, , \, R] \subset \mathbb{R}_+^\omega$ there exists $i \in \mathbb{N}$, $i \geqslant 4$, such that the ball $B\left(1, \frac{\eta}{4}\right)$ contains more than $i^{a+\epsilon}$ points that are $\frac{\eta}{i}$-separated.

Then we define the function

$\iota : [\log R, R] \to \mathbb{N}^\omega$, $\iota(\eta)$ is the smallest $i \in \mathbb{N}$ for which the above holds.

The range of ι is $\mathbb{N}$, which is identified with $\widehat{\mathbb{N}} \subset \mathbb{N}^\omega$.

It is easy to check that ι is an internal map defined by the sequence of maps:

$\iota_n : [\log R_n\, , \, R_n] \to \mathbb{N}$, $\iota_n(r) =$ the minimal $i \in \mathbb{N}$, $i \geqslant 4$, such that $B_\Gamma\left(1\, , \, \frac{r}{4}\right)$ contains more than $i^{a+\epsilon}$ points that are $\frac{r}{i}$-separated.

The image of ι is therefore internal, and contained in $\widehat{\mathbb{N}} \subset \mathbb{N}^\omega$. According to Lemma 7.36, the image of ι has to be finite. Thus, there exists $K \in \mathbb{N}$ such that

$$\iota(\eta) \in [4, K], \quad \forall \eta \in [\log R, R].$$

This implies that for every $\eta \in [\log R, R]$ there exists $i = \iota(\eta) \in \{4, \ldots, K\}$ such that the ball $B\left(1, \frac{\eta}{2}\right)$ contains more than $i^{a+\epsilon}$ disjoint balls of radii $\frac{\eta}{2i}$.

In particular, taking $\eta = R$, we see that there exists $i_1 = \iota(R) \in \{4, \ldots, K\}$ such that the ball $B\left(1, \frac{R}{4}\right) \subset \Gamma^\omega$ contains at least $i_1^{a+\epsilon}$ disjoint balls

$$B\left(x_1(1), \frac{R}{2i_1}\right), B\left(x_2(1), \frac{R}{2i_1}\right), \ldots, B\left(x_{t_1}(1), \frac{R}{2i_1}\right) \text{ with } t_1 \geqslant i_1^{a+\epsilon}.$$

Since Γ^ω is a group which acts on itself isometrically and transitively, all the balls in this list are isometric to $B\left(1, \frac{R}{2i_1}\right)$.

Exercise 14.14. For every natural number k and every infinitely large number R,

$$k \log(R) < R.$$

Thus, $\frac{R}{i_1} \in [\log R, R]$; hence there exists $i_2 = \iota\left(\frac{R}{i_1}\right)$ such that the ball $B\left(1, \frac{R}{4i_1}\right)$ contains at least $i_2^{a+\epsilon}$ disjoint balls of radii $\frac{R}{2i_1 i_2}$.

It follows that $B\left(1, \frac{R}{2}\right)$ contains a family of disjoint balls

$$B\left(x_1(2), \frac{R}{2i_1i_2}\right), B\left(x_2(2), \frac{R}{2i_1i_2}\right), \ldots, B\left(x_{t_2}(2), \frac{R}{2i_1i_2}\right) \text{ with } t_2 \geqslant i_1^{a+\epsilon} i_2^{a+\epsilon}.$$

We continue *via* the nonstandard induction. Consider $u \in \mathbb{N}^\omega$ such that $B\left(1, \frac{R}{2}\right)$ contains a family of disjoint balls

$$B\left(x_1(u), \frac{R}{2i_1i_2\cdots i_u}\right), B\left(x_2(u), \frac{R}{2i_1i_2\cdots i_u}\right), \ldots, B\left(x_{t_u}(u), \frac{R}{2i_1i_2\cdots i_u}\right),$$

with $t_u \geqslant (i_1 i_2 \cdots i_u)^{a+\epsilon}$.

We construct the next generation of points

$$x_1(u+1), \ldots, x_{t_{u+1}}(u+1)$$

by considering, within each ball

$$B\left(x_j(u), \frac{R}{2i_1i_2\cdots i_u}\right)$$

the centers of $i_{u+1}^{a+\epsilon}$ disjoint balls of radii

$$\frac{R}{2i_1i_2\cdots i_u i_{u+1}}, \quad 1 \leqslant j \leqslant t_{u+1}.$$

Here and below

$$i_{u+1} = \iota\left(\frac{R}{i_1 i_2 \cdots i_u}\right),$$

where the product $i_1 \cdots i_{u+1}$ is defined *via* the nonstandard induction as in the end of Section 7.3.

Thus, we obtain injective maps sending $[1, t_{u+1}] \subset \mathbb{N}^\omega$ to $B(1, R/2)$, $j \mapsto x_j(u+1)$.

This induction process continues as long as $R/(i_1 \cdots i_u) \geqslant \log R$. Recall that $i_j \geqslant 2$, hence

$$\frac{R}{i_1 \cdots i_u} \leqslant 2^{-u} R.$$

Therefore, if $u > \log R - \log\log R$ then

$$\frac{R}{i_1 \cdots i_u} < \log R.$$

Thus, there exists $u \in \mathbb{N}^\omega$ such that

$$\frac{R}{i_1 i_2 \cdots i_{u+1}} < \log R \leqslant \frac{R}{i_1 i_2 \cdots i_u} \leqslant \frac{KR}{i_1 i_2 \cdots i_{u+1}} \Leftrightarrow$$

$$\frac{R}{\log R} < i_1 i_2 \cdots i_{u+1} \leqslant \frac{KR}{\log R}\,. \tag{14.2}$$

Let's "count" the "number" (nonstandard of course!) of *centers*, points $x_i(k)$, we constructed between the first step of the induction and the u-th step of the induction:

We get $i_1^{a+\epsilon} i_2^{a+\epsilon} \cdots i_{u+1}^{a+\epsilon}$ points, i.e., we obtain a bijection from the interval

$$[1, i_1^{a+\epsilon} i_2^{a+\epsilon} \cdots i_{u+1}^{a+\epsilon}] \subset \mathbb{N}^\omega$$

to the set of centers $x_i(k)$. On the other hand, from (14.2) we get:

$$\left(\frac{R}{\log R}\right)^{a+\epsilon} \leqslant (i_1 i_2 \cdots i_{u+1})^{a+\epsilon}\,.$$

What does this inequality actually mean? Recall that R and u are represented by sequences of real and natural numbers R_n, u_n, respectively. The above inequality thus implies that for ω–all $n \in \mathbb{N}$, we have:

$$\left(\frac{R_n}{\log R_n}\right)^{a+\epsilon} \leqslant |B(1, R_n)| \leqslant C R_n^a\,.$$

Accordingly,

$$R_n^\epsilon \leqslant C(\log(R_n))^{a+\epsilon},$$

for ω–all $n \in \mathbb{N}$. This contradicts the assumption that R is infinitely large, cf. Exercise 14.14. □

14.3. Consequences of the Regular Growth Theorem

PROPOSITION 14.15. *Let Γ be a finitely generated group for which there exists an infinitely large number $R = (R_n)^\omega$ in the ultrapower $\mathbb{R}_+^\omega$ such that the growth function satisfies (14.1). Fix real numbers a and $\epsilon > 0$ as in Theorem 14.13 and let $\eta = (\eta_n)$ be a sequence provided by the conclusion of Regular Growth Theorem 14.13; let $\lambda = (\lambda_n)$ with $\lambda_n = \frac{1}{\eta_n}$.*

Then the asymptotic cone $X_\omega = \mathrm{Cone}_\omega(\Gamma; 1, \lambda)$ is

(a) *locally compact;*

(b) *has Hausdorff dimension at most $a + \epsilon$. In particular, in view of Theorem 1.89, X_ω has finite covering dimension.*

Proof. (a) Since X_ω is homogeneous, it suffices to prove that the closed ball $C = \bar{B}\left(1, \frac{1}{4}\right) \subset X_\omega$ is compact. Since C is complete, it suffices to show that it is *totally bounded*, i.e., for every $\delta > 0$ there exists a finite cover of C by δ-balls (see **[Nag85]**).

Let dist denote the word metric on Γ. By Theorem 14.13, the closed ball $\bar{B}(1, \frac{1}{4}) \subset (\Gamma, \lambda_n dist)$ satisfies the property that for every integer $i \geqslant 4$, every $\frac{1}{i}$-separated subset $E \subset \bar{B}(1, \frac{1}{4})$ contains at most $i^{a+\epsilon}$ points. The same assertion clearly holds for the ultralimit X_ω. Therefore, we pick some $i \in \mathbb{N}$ such that $\frac{1}{i} < \delta$ and choose (by Zorn's lemma) a maximal $\frac{1}{i}$-separated subset $E \subset C$. Then, by maximality (see Lemma 1.69),

$$C \subset \bigcup_{x \in E} \bar{B}(x, \frac{1}{i}) \subset \bigcup_{x \in E} \bar{B}(x, \delta).$$

We, thus, have a finite cover of C by δ-balls and, therefore, C is compact.

(b) We first verify that the Hausdorff dimension of the ball $\bar{B}(1, 1/4)$ is at most $a + \epsilon$. Pick $\alpha > a + \epsilon$. For each i consider a maximal $\frac{1}{i}$-separated subset $x_{1\omega}, x_{2\omega}, \ldots, x_{t\omega}$ in $B(1, 1/4)$, with $t \leqslant i^{a+\epsilon}$.

Then $\bar{B}(1, 1/4)$ is covered by the balls

$$\bar{B}(x_{j\omega}, 1/i), j = 1, \ldots, t\,.$$

We get:

$$\sum_{j=1}^{t} (1/i)^\alpha \leqslant i^{a+\epsilon}/i^\alpha = i^{a+\epsilon-\alpha}.$$

Since $\alpha > a + \epsilon$, $\lim_{i\to\infty} i^{a+\epsilon-\alpha} = 0$. Hence $\mu_\alpha(B(1, 1/4)) = 0$.

Thus by homogeneity of X_ω, $\dim_{Haus}(\bar{B}(x, 1/4)) \leqslant a + \epsilon$ for each $x \in X_\omega$.

By (a) and Theorem 1.60 X_ω is proper, hence it is covered by countably many balls $\bar{B}(x_n, 1/4)$, $n \in \mathbb{N}$. For every $\alpha > a + \epsilon$, additivity of μ_α implies that

$$\mu_\alpha(X_\omega) \leqslant \sum_{n=1}^{\infty} \mu_\alpha\left(\bar{B}(x_n, 1/4)\right) = 0.$$

Therefore $\dim_{Haus}(X_\omega) \leqslant a + \epsilon$. □

14.4. Weakly polynomial growth

Here we prove several elementary properties of groups of weakly polynomial growth (Definition 14.2) that will be used in the next section.

Lemma 14.16. *If Γ has weakly polynomial growth then for every normal subgroup $N \lhd \Gamma$, the quotient Γ/N also has weakly polynomial growth.*

Proof. We equip $Q = \Gamma/N$ with the generating set which is the image of the finite generating set of Γ. Then $B_Q(1, R)$ is the image of $B_\Gamma(1, R)$. Hence,

$$|B_Q(1, R)| \leqslant |B_\Gamma(1, R)|$$

and, therefore, Q also has weakly polynomial growth. □

Lemma 14.17. *If Γ has exponential growth then it cannot have weakly polynomial growth.*

Proof. Since Γ has exponential growth,

$$\lim_{r\to\infty} \frac{1}{r} \log(\mathfrak{G}(r)) > 0.$$

Suppose that Γ has weakly polynomial growth. This means that growth function of Γ satisfies

$$\mathfrak{G}(R_n) = |B_\Gamma(1, R_n)| \leqslant CR_n^a$$

for a certain sequence (R_n) diverging to infinity and constants C and a. Hence,

$$\frac{1}{R_n} \log(\mathfrak{G}(R_n)) \leqslant \frac{\log(C)}{R_n} + \frac{a}{R_n} \log(R_n).$$

However,

$$\lim_{R\to\infty} \left(\frac{\log(C)}{R} + \frac{a}{R} \log(R) \right) = 0.$$

Contradiction. □

Lemma 14.18. *Let Γ be a finitely generated subgroup of a Lie group G with finitely many components. If Γ has weakly polynomial growth then Γ is virtually nilpotent.*

Proof. According to Tits' alternative, either Γ contains a free nonabelian subgroup or is virtually solvable. In the former case, Γ cannot have weakly polynomial growth (see Lemma 14.17). Thus Γ is virtually solvable. Similarly, by applying Theorem 12.37, since Γ has weakly polynomial growth, Γ has to be is virtually nilpotent. □

14.5. Displacement function

In this section we discuss certain metric properties of action of a finitely generated group Γ on itself by left translations. These properties will be used to prove Gromov's theorem. We fix a finite generating set S of Γ, the Cayley graph $\mathrm{Cayley}(\Gamma, S)$ and the corresponding word metric on Γ.

We define certain *displacement functions* Δ for the action $\Gamma \curvearrowright \Gamma$ by left multiplication. For every $\gamma \in \Gamma$, $x \in \mathrm{Cayley}(\Gamma, S)$ and $r \geqslant 0$ we define the function measuring the maximal displacement by γ on the ball $\bar{B}(x, r) \subset \mathrm{Cayley}(\Gamma, S)$:

$$\Delta(\gamma, x, r) = \max\{\mathrm{dist}(y, \gamma y) \ : \ y \in \bar{B}(x, r)\}\,.$$

When $x = 1$ we use the notation $\Delta(\gamma, r)$ for the displacement function.

For a subset of $F \subset \Gamma$, define

$$\Delta(F, x, r) = \sup_{\gamma\in F} \Delta(\gamma, x, r)\,.$$

Likewise, we write $\Delta(F, r)$ when $x = 1$.

Clearly, for every $g \in \Gamma$,

$$\Delta(F, g, r) = \Delta(g^{-1}Fg, r)\,.$$

Lemma 14.19. *Fix $r > 0$ and a finite subset F in Γ. Then the function $\mathrm{Cayley}(\Gamma, S) \to \mathbb{R}$, $x \mapsto \Delta(F, x, r)$ is 2–Lipschitz.*

Proof. Let x, y be two points in $\mathrm{Cayley}(\Gamma, S)$. Let p be an arbitrary point in $B(x,r) \subset \mathrm{Cayley}(\Gamma, S)$. A geodesic in $\mathrm{Cayley}(\Gamma, S)$ connecting p to y has length at most $r + \mathrm{dist}(x,y)$, hence it contains a point $q \in B(y,r)$ with $\mathrm{dist}(p,q) \leqslant \mathrm{dist}(x,y)$. For an arbitrary $\gamma \in F$,

$$\mathrm{dist}(p, \gamma p) \leqslant \mathrm{dist}(q, \gamma q) + 2\mathrm{dist}(x,y) \leqslant \Delta(F, y, r) + 2\mathrm{dist}(x,y).$$

It follows that $\Delta(F,x,r) \leqslant \Delta(F,y,r) + 2\mathrm{dist}(x,y)$. The inequality $\Delta(F,y,r) \leqslant \Delta(F,x,r) + 2\mathrm{dist}(x,y)$ is proved similarly. □

Lemma 14.20. *Suppose that $\Delta(S,r)$ is bounded as a function of r. Then Γ is virtually abelian.*

Proof. Suppose that $\mathrm{dist}(sx, x) \leqslant C$ for all $x \in \Gamma$ and $s \in S$. Then

$$\mathrm{dist}(x^{-1}sx, 1) \leqslant C,$$

and, therefore, the conjugacy class of s in Γ has cardinality $\leqslant \mathfrak{G}_\Gamma(C) = N$. We claim that the centralizer $Z_\Gamma(s)$ of s in Γ has finite index in Γ: Indeed, if $x_0, \ldots, x_N \in \Gamma$ then there are i, k, $0 \leqslant i \neq k \leqslant N$, such that

$$x_i^{-1} s x_i = x_k^{-1} s x_k \Rightarrow [x_k x_i^{-1}, s] = 1 \Rightarrow x_k x_i^{-1} \in Z_\Gamma(s).$$

Thus, the intersection

$$A := \bigcap_{s \in S} Z_\Gamma(s)$$

has finite index in Γ. Therefore, A is an abelian subgroup of finite index in Γ. □

Note that there are virtually abelian groups Γ with unbounded displacement functions $\Delta(S,r)$, the simplest example is $\mathbb{Z}_2 \star \mathbb{Z}_2$.

Exercise 14.21. Show that the displacement function $\Delta(S,r)$ of Γ is bounded as a function of r if and only if Γ contains a finite normal subgroup K such that Γ/K is free abelian.

14.6. Proof of Gromov's theorem

In this section we prove Theorem 14.3 and, hence, Theorem 14.1 as well. Let Γ be a group satisfying the assumptions of Theorem 14.3 and $a, \epsilon, R \in \mathbb{R}^*, \eta \in \mathbb{R}^*$ be the quantities appearing in Theorem 14.3. In what follows we fix a finite generating set S of Γ and the corresponding Cayley graph $\mathrm{Cayley}(\Gamma, S)$.

Suppose that Γ has weakly polynomial growth with respect to a sequence (R_n) diverging to infinity. Take the diverging sequence (η_n) given by the Regular Growth Theorem applied to the group Γ. Set $\lambda = (\lambda_n)$ with $\lambda_n = \frac{1}{\eta_n}$. Construct the asymptotic cone $X_\omega = \mathrm{Cone}_\omega(\Gamma; 1, \lambda)$ of the Cayley graph of Γ *via* rescaling by the sequence λ_n and considering the constant sequence e of base-points equal to the identity in Γ. By Proposition 14.15, the metric space X_ω is connected, locally connected, finite-dimensional and proper.

According to Proposition 7.71, we have a homomorphism

$$\alpha : \Gamma_e^\omega \to L := \mathrm{Isom}(X_\omega)$$

such that $\alpha(\Gamma_e^\omega)$ acts on X_ω transitively. We also get a homomorphism

$$\rho : \Gamma \to L, \rho = \iota \circ \alpha,$$

where $\iota : \Gamma \hookrightarrow \Gamma_e^\omega$ is the diagonal embedding $\iota(\gamma) = (\gamma)^\omega$. Since the isometric action $L \curvearrowright X_\omega$ is effective and transitive, according to Theorem 14.7, the group L is a Lie group with finitely many components.

REMARK 14.22. Observe that, in view of local compactness of X_ω, the point-stabilizer L_y for $y \in X_\omega$ is a compact subgroup in L. Therefore X_ω (homeomorphic to L/L_x, see Lemma 3.37) can be given an L-invariant Riemannian metric ds^2. Hence, since X_ω is connected, by using the exponential map with respect to ds^2 we see that if $g \in L$ fixes an open ball in X_ω pointwise, then $g = id$.

The subgroup $\rho(\Gamma) \leqslant L$ has weakly polynomial growth because Γ has weakly polynomial growth (see Lemma 14.16). By Lemma 14.18, $\rho(\Gamma)$ is virtually nilpotent.

The main problem is that ρ may have large kernel. Indeed, if Γ is abelian then the homomorphism ρ is actually trivial. An induction argument on the degree d of weakly polynomial growth allows to get around this problem and prove Gromov's Theorem. In the induction step, we shall use ρ to construct an epimorphism $\Gamma \to \mathbb{Z}$, and then apply Proposition 12.39.

If $d = 0$, then $\mathfrak{G}_\Gamma(R_n)$ is bounded. Since the growth function is monotonic, it follows that Γ is finite and there is nothing to prove.

Suppose that each group Γ of weakly polynomial growth of degree $\leqslant d - 1$ is virtually nilpotent. Let Γ be a group of weakly polynomial growth of degree $\leqslant d$, i.e.,

$$\mathfrak{G}_\Gamma(R_n) \leqslant C_\Gamma R_n^d,$$

for some sequence (R_n) diverging to infinity. There are two cases to consider:

(a) The image of the homomorphism ρ above is infinite. Then there exists a finite index subgroup $\Gamma_1 \leqslant \Gamma$ such that $\rho(\Gamma_1)$ is a torsion-free infinite nilpotent group. The latter has infinite abelianization, hence, we get an epimorphism $\phi : \Gamma_1 \to \mathbb{Z}$. If $K = \mathrm{Ker}(\phi)$ is not finitely generated, then Γ_1 has exponential growth (see Proposition 12.39), which is a contradiction. Therefore, K is finitely generated. Repeating the arguments in the proof of Proposition 12.39 verbatim we see that K has weakly polynomial growth of degree $\leqslant d - 1$. It follows that, by the induction hypothesis, K is a virtually nilpotent group. Therefore, Γ_1 is solvable. Applying Lemma 14.17, we conclude that Γ_1 (and, hence, Γ) is virtually nilpotent.

(b) $\rho(\Gamma)$ is finite. First we note that we can reduce to the case when $\rho(\Gamma) = \{1\}$. Indeed, consider the subgroup of finite index $\Gamma' := \mathrm{Ker}(\rho) \leqslant \Gamma$. For every $\gamma \in \Gamma'$, we have that

$$\mathrm{dist}_\Gamma(x_n, \gamma x_n) = o(\eta_n), \tag{14.3}$$

for every sequence $(x_n) \in \Gamma^{\mathbb{N}}$ with $\mathrm{dist}_\Gamma(1, x_n) = O(\eta_n)$. Since the inclusion map $\Gamma' \hookrightarrow \Gamma$ is a quasiisometry, the estimate (14.3) holds for sequences (x_n) in Γ' and $\mathrm{dist}_{\Gamma'}$. Thus, Γ' acts trivially on its own asymptotic cone $\mathrm{Cone}_\omega(\Gamma'; 1, \lambda)$, and it clearly suffices to prove that Γ' is virtually nilpotent.

Hence, from now on we assume that $\rho(\Gamma) = \{1\}$. The next exercise clarifies the metric significance of this condition.

EXERCISE 14.23. Let Δ denote the displacement function for the action of Γ on itself *via* left multiplication introduced in Section 14.5. Show that the condition

$\operatorname{Ker}\rho = \Gamma$ is equivalent to the property that

$$\omega\text{-lim}\,\frac{\Delta(S, R\eta_n)}{\eta_n} = 0\,, \quad \text{for every } R > 0\,. \tag{14.4}$$

In other words, all generators of Γ act on Γ with *sublinear* (with respect to $(\eta_n)^\omega$) displacement.

Let $q = q(L)$ denote the constant given by Jordan's theorem applied to the group L. Consider the intersection Γ' of all the subgroups in Γ of index at most q, and let S' denote a finite set generating Γ'. We keep the notation S for a finite generating set of the group Γ.

If the function $\Delta(S', r)$ were bounded then Γ' (and, hence, Γ) would be virtually abelian (Lemma 14.20), which would conclude the proof. Thus, we assume that $\Delta(S', r)$ diverges to infinity as $r \to \infty$.

Lemma 14.24. *For every $\epsilon \in (0, 1]$ there exists a sequence (x_n) in Γ such that*

$$\omega\text{-}lim\,\frac{\Delta(x_n^{-1}S'x_n, \eta_n)}{\eta_n} = \epsilon. \tag{14.5}$$

Proof. By (14.4), for ω–all $n \in \mathbb{N}$ we have $\Delta(S', \eta_n) \leqslant \epsilon\eta_n/2$. Thus, there exists a subset $I \subset \mathbb{N}$ of ω-measure 1 such that for all $n \in I$, there is $p_n \in \Gamma$ such that $\Delta(S', p_n, \eta_n) \leqslant \eta_n/2$. Fix $n \in I$. Since the function $\Delta(S', r)$ diverges to infinity, there exists $q_n \in \Gamma$ such that

$$\Delta(S', q_n, \eta_n) \geqslant \max_{s\in S'} \operatorname{dist}(q_n, sq_n) > 2\eta_n.$$

The Cayley graph $\mathrm{Cayley}(\Gamma, S)$ is connected and the function $\mathrm{Cayley}(\Gamma, S) \to \mathbb{R}$, $p \mapsto \Delta(S', p, \eta_n)$ is continuous by Lemma 14.19. Hence, for ω-all n, there exists $y_n \in \mathrm{Cayley}(\Gamma, S)$ such that

$$\Delta(S', y_n, \eta_n) = \epsilon\eta_n.$$

The point y_n is not necessarily in the vertex set of the Cayley graph $\mathrm{Cayley}(\Gamma, S)$. Pick a point $x_n \in \Gamma$ within the distance $\frac{1}{2}$ from y_n. Again by Lemma 14.19

$$|\Delta(S', x_n, \eta_n) - \epsilon\eta_n| \leqslant 1.$$

It follows that $|\Delta(x_n^{-1}S'x_n, \eta_n) - \epsilon\eta_n| \leqslant 1$ and, therefore,

$$\omega\text{-lim}\,\frac{\Delta(x_n^{-1}S'x_n, \eta_n)}{\eta_n} = \epsilon. \quad \square$$

For every $0 < \epsilon \leqslant 1$ we consider a sequence (x_n) as in Lemma 14.24 and define the homomorphism

$$\rho_\epsilon : \Gamma \to \Gamma^\omega,\ \rho_\epsilon(g) = \left(x_n^{-1} g x_n\right)^\omega \in \Gamma^\omega.$$

Since $\Delta(x_n^{-1}S'x_n, \eta_n) = O(\epsilon\eta_n)$, the image of ρ_ϵ is contained in L. Clearly, the image of ρ_ϵ is nontrivial. If for some $\epsilon > 0$, $\rho_\epsilon(\Gamma)$ is infinite, then we are done as in (a). Hence we assume that $\rho_\epsilon(\Gamma)$ is finite for all $\epsilon \in (0, 1]$.

Next, we reduce the problem to the case when all groups $\rho_\epsilon(\Gamma)$ are finite abelian. For each ϵ consider the preimage Γ_ϵ in Γ of the abelian subgroup in $\rho_\epsilon(\Gamma)$ which is given by Jordan's theorem applied to L. The index of Γ_ϵ in Γ is at most q. This implies that the group Γ' described before Lemma 14.24 is contained in Γ_ϵ for every $\epsilon > 0$. It follows that $\rho_\epsilon(\Gamma')$ is finite abelian for every $\epsilon > 0$. Since L is a Lie group, there exists U_δ, a neighborhood of $1 \in L$, which contains no nontrivial finite subgroups. Without loss of generality, we may assume that U_δ has the form

$$U_\delta = \{u \in L : \Delta(u, e_\omega, 1) < \delta\},$$

for some $\delta > 0$. Thus, for each natural number M and $i \leqslant M$, we have

$$\Delta(u^i, e_\omega, 1) < M\delta.$$

By our choice of x_n, for every generator $s \in S'$ (of the group Γ'),

$$\Delta(\rho_\epsilon(s), e_\omega, 1) \leqslant \epsilon$$

and for one of the generators the inequality is the equality. Assume there exists an $M \in \mathbb{N}$ such that the order $|\rho_\epsilon(\Gamma')|$ is at most M for all $\epsilon \in (0, 1]$. Therefore, for every $g \in \Gamma'$,

$$\Delta(\rho_\epsilon(g), e_\omega 1) \leqslant M\epsilon.$$

Choosing ϵ such that $M\epsilon < \delta$. Since U_δ contains no nontrivial subgroups, we conclude that $\rho_\epsilon(\Gamma') = \{1\}$, which is a contradiction. Therefore,

$$\limsup_{\epsilon \to 0} |\rho_\epsilon(\Gamma')| = \infty\,.$$

This means that Γ' admits epimorphisms to finite abelian groups of arbitrarily large order. All such homomorphisms have to factor through the abelianization $(\Gamma')_{ab}$ of the group Γ', therefore, the group $(\Gamma')_{ab}$ is infinite. Since $(\Gamma')_{ab}$ is finitely generated we conclude that Γ' admits an epimorphism to $\mathbb{Z}$. We apply Proposition 12.39 and the induction hypothesis, and conclude that Γ' is virtually nilpotent. Thus Γ is also virtually nilpotent, and we are done. This concludes the proof of Theorem 14.3 and, hence, of Theorem 14.1. □

14.7. Quasi-isometric rigidity of nilpotent and abelian groups

Gromov's theorem has several spectacular corollaries, proving that certain algebraic properties of groups can be recovered from the coarse geometric information.

THEOREM 14.25 (M. Gromov). *Suppose that Γ_1, Γ_2 are quasiisometric finitely generated groups and Γ_1 is virtually nilpotent. Then Γ_2 is virtually nilpotent.*

PROOF. Being virtually nilpotent, Γ_1 has polynomial growth of degree d (Theorem 12.26). Since growth is invariant under quasiisometry, Γ_2 also has polynomial growth of degree d. By Theorem 14.1, Γ_2 is virtually nilpotent. □

Note that an alternative proof of this theorem (which does not use Gromov's theorem) was given by Y. Shalom [**Sha04**].

THEOREM 14.26 (P. Pansu). *Suppose that Γ_1, Γ_2 are quasiisometric finitely generated groups and Γ_1 is virtually abelian. Then Γ_2 is virtually isomorphic to Γ_1.*

PROOF. Without loss of generality, we may assume that Γ_1 is abelian. Let d denote the rank of Γ_1. Then $\mathfrak{G}_{\Gamma_1}(t) \asymp t^d$. Furthermore, d is the rational cohomological dimension of Γ_1. Then, by quasiisometry invariance of growth, Γ_2 also growth $\asymp t^d$. As we just saw above, Γ_2 is virtually nilpotent. Let $\Gamma_3 \leqslant \Gamma_2$ denote a nilpotent subgroup of finite index in Γ_2. Let $\Gamma := \Gamma_3/\mathrm{Tor}\,\Gamma_3$. By Bass–Guivarc'h Theorem (Theorem 12.26),

$$d = d(\Gamma) = \sum_{i=1}^{k} i\ m_i,$$

where m_i is the rank of $C^i\Gamma/C^{i+1}\Gamma$. Recall that the rational cohomological dimension is a quasiisometry invariant, see Theorem 6.62.

Therefore,

$$d = \dim(\Gamma) = \sum_{i=1}^{k} m_i,$$

and

$$\sum_{i=1}^{k} i\ m_i = \sum_{i=1}^{k} m_i.$$

The latter implies that $k = 1$, i.e., Γ is abelian. Virtual isomorphism of the groups Γ_1 and Γ (and, hence, of Γ_2 as well) follows from the equality of their ranks. □

14.8. Further developments

The following version of Gromov's theorem was proved by M. Sapir [**Sap15**], using the recent work of E. Hrushovsky [**Hru12**] on *approximate groups.* Sapir's theorem answered affirmatively a question posed by van den Dries and Wilkie in [**dDW84**].

THEOREM 14.27. *If* Γ *is a finitely generated group such that* some *asymptotic cone of* Γ *is locally compact, then* Γ *is virtually nilpotent.*

Note that a weaker version of this theorem was proven earlier by F. Point [**Poi95**], who was assuming, in addition, that the asymptotic cone has finite Minkowski dimension.

Below we review some properties of asymptotic cones of nilpotent groups.

Let (Γ, dist) be a finitely generated nilpotent group endowed with a word metric, let $\mathrm{Tor}\,(\Gamma)$ be the torsion subgroup of Γ and let H be the torsion-free nilpotent group $\Gamma/\mathrm{Tor}\,(\Gamma)$. Recall that according to Mal'cev's Theorem 11.40, the nilpotent group H is isomorphic to a uniform lattice in a connected nilpotent Lie group N.

With every k-step nilpotent Lie group N with Lie algebra $\mathfrak{n}$ one associates the *associated graded Lie algebra* $\bar{\mathfrak{n}}$ obtained as the direct sum

$$\oplus_{i=1}^{k} \mathfrak{c}^i\mathfrak{n}/\mathfrak{c}^{i+1}\mathfrak{n},$$

where $\mathfrak{c}^i\mathfrak{n}$ is the Lie algebra of C^iN. Every finite-dimensional Lie algebra is the Lie algebra of a connected Lie group; thus, consider the connected nilpotent Lie group $\overline{N}$ with the Lie algebra $\bar{\mathfrak{n}}$. The group $\overline{N}$ is called the *associated graded Lie group* of the group Γ and of the Lie group N. We refer to Pansu's paper [**Pan83**] for the definition of the Carnot-Caratheodory metric appearing in the following theorem:

THEOREM 14.28 (P. Pansu, [**Pan83**]).
- **(a)** *All the asymptotic cones of the finitely generated nilpotent group* Γ *are bilipschitz homeomorphic to the graded Lie group* $\overline{N}$ *endowed with a* Carnot-Caratheodory metric dist_{CC}.
- **(b)** *For every sequence* $\varepsilon_j > 0$ *converging to* 0 *and every word metric* dist *on* Γ*, the sequence of metric spaces* $(\Gamma, \varepsilon_j \cdot \mathrm{dist})$ *converges in the modified Hausdorff metric to* $(\overline{N}, \mathrm{dist}_{CC})$.
- **(c)** *The sub-bundle in* $\overline{N}$ *defining the Carnot-Caratheodory metric is independent of the word metric on* Γ*, only the norm on this subbundle depends on the word metric.*

(d) *The dimension of* $\overline{N}$ *equals the rational cohomological dimension of* Γ, *which, in turn, equals*

$$\dim(\Gamma) = \sum_{i=1}^{k} m_i,$$

where m_i *is the rank of the abelian quotient* $C^i\Gamma/C^{i+1}\Gamma$.

e) *The Hausdorff dimension of* $(\overline{N}, \mathrm{dist}_{CC})$ *equals to the degree of polynomial growth of* Γ, *that is, to*

$$d(\Gamma) = \sum_{i=1}^{k} i\ m_i$$

Note that, according to Theorem 7.46, (a) implies (b) in Pansu's theorem. We further note that $\overline{N}$, treated as a Lie group, is also a quasiisometry invariant of Γ, see [**Pan89**].

REMARK 14.29. One says that two metric spaces are *asymptotically bi-Lipschitz* if their asymptotic cones are bi-Lipschitz homeomorphic. Pansu's theorem above has lead to examples of asymptotically bi-Lipschitz nilpotent groups, which are not quasiisometric. Indeed, by Pansu's theorem, every two finitely generated nilpotent groups $\Gamma_i, i = 1, 2$, with isomorphic associated graded Lie groups $\overline{N}_i$, are asymptotically bi-Lipschitz. Y. Benist constructed two nilpotent groups Γ_1, Γ_2 with isomorphic associated graded Lie groups $\overline{N}_1, \overline{N}_2$, but distinct *virtual Betti numbers*. Y. Shalom proved that for of finitely generated nilpotent groups virtual Betti numbers are quasiisometry invariant. Therefore, Benoist's groups Γ_1, Γ_2 are asymptotically bi-Lipschitz but not quasiisometric. We refer to Shalom's paper [**Sha04**, p. 151-152] for the details.

CHAPTER 15

The Banach–Tarski paradox

In this chapter we discuss the *Banach–Tarski Paradox*, which relies upon existence of free nonabelian subgroups in orthogonal groups $O(n), n \geq 3$ and also connects to the notion of amenability, which will be discussed in detail in the next chapter.

15.1. Paradoxical decompositions

The Banach–Tarski Paradox deals with decompositions of subsets in the Euclidean space into congruent pieces and rearranging them via isometries. We begin with discussing the concepts involved in this process for general sets and group actions. In what follows, X is a set and $G \leqslant Bij(X)$ is a group of bijections. From the geometric viewpoint, the most interesting case is that of $X = \mathbb{E}^n$ (the Euclidean n-space) and G a subgroup of the group of isometries of $\mathbb{E}^n$. Another useful example which we discuss in the next chapter is when $X = G$ is a group and G acts on itself *via* left multiplication.

DEFINITION 15.1. Two subsets A, B in X are *G-congruent* if there exists $g \in G$ such that $g(A) = B$. The restriction $g\big|_A$ is called a *G-congruence* from A to B.

DEFINITION 15.2. A bijection $\phi : A \to B$ between two subsets of X is called a *piecewise G-congruence* if the subsets A, B admit partitions into nonempty parts,

$$A = A_1 \sqcup ... \sqcup A_k, \quad B = B_1 \sqcup ... \sqcup B_k$$

such that the restrictions

$$\phi_i = \phi\big|_{A_i} : A_i \to B_i, \quad i = 1, \ldots, k, \tag{15.1}$$

are G-congruences. Accordingly, two subsets A, B are called *piecewise G-congruent* if there exists a piecewise G-congruence $A \to B$.

EXERCISE 15.3. Prove that piecewise G-congruence is an equivalence relation.

DEFINITION 15.4. A subset $E \subset X$ is *G-paradoxical* if it is nonempty and admits a partition $E = E' \sqcup E''$ such that E' and E'' are both piecewise G-congruent to E. In detail: There exist partitions

$$E' = A_1' \sqcup \ldots, A_k', \quad E'' = A_1'' \sqcup \ldots, A_l''$$

and bijections $\phi' : E' \to E, \phi'' : E'' \to E$ which restrict to congruences

$$\phi_i' : E_i' \to \phi_i'(E_i') \subset E, \quad \phi_j'' : E_j'' \to \phi_j''(E_j'') \subset E,$$

$i = 1, \ldots, k$, $j = 1, \ldots, l$. The subsets E_i', E_j'', $1 \leqslant i \leqslant k, 1 \leqslant j \leqslant l$, are called *pieces* of the *G-paradoxical decomposition*

$$E_1' \sqcup \ldots \sqcup E_k' \sqcup E_1'' \sqcup \ldots \sqcup E_l''$$

of the subset $E \subset X$.

EXERCISE 15.5. If $A, B \subset X$ are piecewise G-congruent and A is G-paradoxical, then so is B.

A group action $G \curvearrowright X$ is called *paradoxical* if the set X is paradoxical with respect to this action. A group action $G \curvearrowright X$ is called *weakly paradoxical* if there exists a G-paradoxical subset $E \subset X$. Thus, every paradoxical action is also weakly paradoxical.

In the context of groups, considering the G-action on itself via left multiplication

$$L : G \times G \to G, \quad L(g, x) = L_g(x) = gx,$$

we arrive to the following definition:

DEFINITION 15.6. A group G is *paradoxical* (resp. *weakly paradoxical*) if the action $L : G \times G \to G$ is paradoxical (resp. weakly paradoxical).

Next, we prove several facts about piecewise congruences and paradoxical decompositions.

LEMMA 15.7. *Suppose that $S \subset G$ is a paradoxical subset, $G \curvearrowright X$ is an action such that for some $x \in X$ the orbit map*

$$f : G \to X, \quad g \mapsto g(x)$$

restricts to an injective map on S. Then X is G-paradoxical.

PROOF. Let

$$S = S' \sqcup S'', \quad S' = S'_1 \sqcup \ldots \sqcup S'_k, \quad S'' = S''_1 \sqcup \ldots \sqcup S''_l$$

be a G-paradoxical decomposition of S with the piecewise-congruences $\phi' : S' \to S, \phi'' : S'' \to S$, where

$$\phi'_i\big|_{S'_i} = g_i\big|_{S'_i}, \quad \phi''_j\big|_{S''_j} = g_j\big|_{S''_j}, \quad 1 \leqslant i \leqslant k, \quad 1 \leqslant j \leqslant l.$$

Define the partitioned subset $E = E' \sqcup E'' \subset X$ as $E = f(S), E' = f(S'), E'' = f(S'')$. Furthermore, define bijections

$$\psi' : E' \to E, \quad \psi'' : E'' \to E$$

by

$$\psi'(f(s)) = f(\phi'(s)), \quad \psi''(f(s)) = f(\phi''(s)).$$

It follows that the restriction of ψ' to $f(S'_i)$ is given by g'_i and the restriction of ψ'' to $f(S''_j)$ is given by g''_j, $1 \leqslant i \leqslant k, \quad 1 \leqslant j \leqslant l.$. Therefore, ψ', ψ'' are piecewise G-congruences. $\square$

LEMMA 15.8. *Suppose that $H \leqslant G$ is an infinite cyclic subgroup preserving a subset $A \subset X$. Suppose that $E \subset A$ is such that $hE \cap E = \emptyset$ for all $h \in H \setminus \{1\}$. Then A is piecewise H-congruent to $A \setminus E$. In particular, if H acts freely on its orbit $Hx \subset A$, then A is piecewise H-congruent to $A \setminus \{x\}$.*

PROOF. Let g be a generator of H. Define the partition $A = A_1 \sqcup A_2$,

$$A_1 = \bigcup_{n \in \mathbb{Z}_+} g^n E, \quad A_2 := A \setminus A_1.$$

Now consider the map $\phi : A \to A \setminus E$ which is the identity on A_2 and is $g\big|_{A_1}$ on A_1. Then ϕ is a piecewise H-congruence. $\square$

COROLLARY 15.9. *Let $G = SO(n)$. Then for each $n \geqslant 2$, the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{E}^n$ is piecewise G-congruent to $\mathbb{S}^{n-1} \setminus \{p\}$, where $p \in \mathbb{S}^{n-1}$ is any point.*

PROOF. As usual, we identify $\mathbb{E}^n$ with the vector space $\mathbb{R}^n$ equipped with the standard Euclidean metric. Without loss of generality, we may assume that p belongs to $\mathbb{R}^2 \cap \mathbb{S}^{n-1}$. Let $h \in O(2)$ (the orthogonal group of $\mathbb{R}^2$) be an infinite order rotation. Then no power $h^k, k \neq 0$, fixes p. Extending h by the identity to the orthogonal complement of $\mathbb{R}^2$ in $\mathbb{R}^n$, we obtain an isometry $g \in SO(n)$ such that no power $g^k, k \neq 0$, fixes p. Now claim follows from Lemma 15.8. □

The next lemma shows how to "double" paradoxical subsets:

LEMMA 15.10. *Suppose that $A \subset X$ is a G-paradoxical subset. Then A is piecewise G-congruent to any subset $B \subset X$ of the form*

$$B = B_1 \sqcup \ldots \sqcup B_k,$$

where each B_i is G-congruent to A.

PROOF. It suffices to consider $k = 2$ as the general case follows by induction. Let $A = A_1 \sqcup A_2$ be a G-paradoxical decomposition and $\phi_i : A_1 \to A$ are piecewise G-congruences. Then composing ϕ_i with a G-congruence $\psi_i : A \to B_i$ $(i = 1, 2)$, we obtain the required piecewise G-congruence $A \to B$. □

REMARK 15.11. We note that instead of taking finite partitions, one can also take countable partitions; this leads to the notion of *countable* G-congruence (and countably paradoxical decompositions), but we will not discuss it in the book as its relation to the geometric group theory is only tangential. We refer the reader to [**Wag85**] for the details.

We now specialize to the case $X = \mathbb{E}^n$, which is the n-dimensional Euclidean space and G the group of isometries of X. Building upon earlier work of Vitali [**Vit05**] and F. Hausdorff [**Hau14**], S. Banach and A. Tarski proved in [**BT24**] the following:

THEOREM 15.12 (Banach-Tarski paradox). *For $n \geqslant 3$, any two bounded subsets with nonempty interior in $\mathbb{E}^n$ are piecewise–congruent.*

A corollary of this theorem is much better known:

COROLLARY 15.13. *Let n be at least 3 and let G denote the group $\operatorname{Isom}(\mathbb{E}^n)$ of isometries of the Euclidean n-space.*

1. *Every closed ball in $\mathbb{E}^n$ is G-paradoxical.*
2. *For $m \in \mathbb{N}$, every closed ball in $\mathbb{E}^n$ is piecewise G-congruent to the disjoint union of m isometric copies of this ball in $\mathbb{E}^n$ (one can "double" the ball).*
3. *Any two round n-balls in $\mathbb{E}^n$ are piecewise G-congruent.*

We note that Part 2 of the corollary follows from Part 1 and Lemma 15.10.

REMARK 15.14. The Banach-Tarski paradox implies that there are no finitely-additive measures defined on *all* subsets of the Euclidean space of dimension at least 3 which are invariant with respect to isometries and take positive value on the unit cube. In particular, the congruent pieces A_i, B_i are not *Lebesgue measurable.*

Remark 15.15 (Banach-Tarski paradox and the Axiom of Choice). The Banach-Tarski paradox is neither provable nor disprovable with Zermelo-Fraenkel axioms (ZF) only: It is impossible to prove that the unit ball in $\mathbb{E}^3$ is paradoxical in ZF, it is also impossible to prove it is not paradoxical. An extra axiom is needed, e.g., the Axiom of Choice (AC). In fact, work of M. Foreman & F. Wehrung [**FW91**] and J. Pawlikowski [**Paw91**] shows that the Banach–Tarski paradox can be proved assuming ZF and the Hahn–Banach theorem (which is a weaker axiom than AC, see Section 7.1).

In this book we will prove only Parts 1 (and, hence, Part 2) of Corollary 15.13; we refer the reader to [**Wag85**] for a proof of Theorem 15.12. We only note here that Theorem 15.12 is derived from the doubling of a ball (Part 2 of Corollary 15.13) by using the Banach–Bernstein–Schroeder theorem (see [**Wag85**]).

Remark 15.16. Inspired by the Hausdorff's argument, R. M. Robinson, answering a question of von Neumann, proved in [**Rob47**] that five is the minimal number of pieces in a paradoxical decomposition of the unit 3-dimensional ball. See Section 16.7 for a discussion on the minimal number of pieces in a paradoxical decomposition.

Remark 15.17. It turns out that the sphere $\mathbb{S}^{n-1}$ can be partitioned into $2^{\aleph_0}$ pieces, so that each piece is piecewise congruent to $\mathbb{S}^{n-1}$, see [**Wag85**].

15.2. Step 1: A paradoxical decomposition of the free group F_2

Let F_2 be the free group of rank 2 with generators a, b. Given u, a reduced word in a, b, a^{-1}, b^{-1}, we denote by $\mathcal{W}_u$ the set of reduced words in a, b, a^{-1}, b^{-1} with the prefix u. Every $x \in F_2$ defines a map $L_x : F_2 \to F_2$, $L_x(y) = xy$ (left translation by x).

Then

$$F_2 = \{1\} \sqcup \mathcal{W}_a \sqcup \mathcal{W}_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{15.2}$$

but also $F_2 = L_a\mathcal{W}_{a^{-1}} \sqcup \mathcal{W}_a$, and $F_2 = L_b\mathcal{W}_{b^{-1}} \sqcup \mathcal{W}_b$. We slightly modify the above partition in order to include $\{1\}$ into one of the other four subsets. Consider the following modifications of $\mathcal{W}_a$ and $\mathcal{W}_{a^{-1}}$:

$$\mathcal{W}_a' = \mathcal{W}_a \setminus \{a^n \,;\, n \in \mathbb{Z}\} \text{ and } \mathcal{W}_{a^{-1}}' = \mathcal{W}_{a^{-1}} \sqcup \{a^n \,;\, n \in \mathbb{Z}\}\,.$$

Then

$$F_2 = \mathcal{W}_a' \sqcup \mathcal{W}_{a^{-1}}' \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{15.3}$$

and

$$F_2 = L_a\mathcal{W}_{a^{-1}}' \sqcup \mathcal{W}_a'\,.$$

Therefore, (15.3) is a G-paradoxical decomposition (with four pieces) of the group F_2 with $G \leqslant Bij(F_n)$ the group F_2 acting on itself *via* left multiplication, i.e. G is the image of $L : F_2 \to Bij(F_2)$, $L(u) = L_u$.

We, thus, proved:

Lemma 15.18. *The free group F_2 is paradoxical.*

Exercise 15.19. Prove that every free group $F_n, n \geqslant 2$, is paradoxical.

Lemma 15.20. *Suppose that X is a nonempty set and $F_2 \times X \to X$ is a free action of F_2. Then X is F_2-paradoxical, with a paradoxical decomposition consisting of four pieces.*

PROOF. *According to the axiom of choice* there exists a subset $D \subset X$ which intersects every F_2-orbit in X exactly once. For subsets $R \subset F_2$ and $S \subset \mathbb{S}^2$ we set

$$R \cdot S := \{\rho(g)(x) : g \in R, x \in S\}.$$

We now partition E as:

$$E_1' = \mathcal{W}_a' \cdot D, \quad E_2' = \mathcal{W}_{a^{-1}}' \cdot D, \quad E_1'' := \mathcal{W}_b \cdot D, \quad E_2'' := \mathcal{W}_{b^{-1}} \cdot D,$$

where

$$F_2 = \mathcal{W}_a' \sqcup \mathcal{W}_{a^{-1}}' \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}}$$

is the F_2-paradoxical decomposition of F_2 defined on on the Step 1. Then we have piecewise F_2-congruences

$$\phi' : E_1' \sqcup E_2' \to E, \phi'\big|_{E_1'} = \rho(a)\big|_{E_1'}, \quad \phi'\big|_{E_2'} = Id,$$

$$\phi'' : E_1'' \sqcup E_2'' \to E, \phi''\big|_{E_1''} = \rho(b)\big|_{E_2'}, \quad \phi'\big|_{E_2''} = Id.$$

□

CONVENTION 15.21. For the rest of the chapter, for simplicity of the notation we will refer to G-congruences simply as congruences and G-paradoxical decompositions simply as paradoxical decompositions, with $G = \mathrm{Isom}(\mathbb{E}^n)$.

15.3. Step 2: The Hausdorff paradox

In this section we prove the Hausdorff Paradox (and its generalization in higher dimensions), a historic precursor to the Banach–Tarski theorem. Recall that $\mathbb{S}^2$ is the unit sphere in $\mathbb{E}^3$ centered at the origin.

THEOREM 15.22 (Hausdorff Paradox). *There exists a countable subset $C \subset \mathbb{S}^2$ such that $\mathbb{S}^2 \setminus C$ is paradoxical.*

PROOF. Recall that there exists a isomorphism $\rho : F_2 \to H \leqslant SO(3)$: This can be viewed as a corollary of the Tits Alternative (Corollary 13.24), it was also proven in more directly Corollary 4.66. Let $C \subset \mathbb{S}^2$ denote the set of fixed points of elements of $H \setminus \{1\}$, this set is clearly H-invariant. The action of H on $E = \mathbb{S}^2 \setminus C$ is free and, hence, Lemma 15.20 implies that E is H-paradoxical. Since H acts isometrically on $\mathbb{E}^3$, theorem follows. □

We next extend Hausdorff Paradox in higher dimensions. Given a representation $\eta : SU(2) \to GL(n, \mathbb{R})$ we let $\mathrm{F}ix_\eta$ denote the union of linear subspaces in $\mathbb{R}^n$ fixed by all nontrivial elements of $SU(2)$ and $\mathrm{F}ix_\eta'$ denote the union of subspaces in $\mathbb{R}^n$ fixed by all noncentral elements of $SU(2)$.

LEMMA 15.23. *For each $n \geqslant 3$ there exists a representation $\eta : SU(2) \to SO(n)$, such that:*

1. $\mathrm{F}ix_\eta = \{0\}$*, if n is divisible by* 4.
2. $\mathrm{F}ix_\eta$ *is a line or a plane, if n is congruent to* 1 *or* 2 *mod* 4.
3. $\mathrm{F}ix_\eta'$ *is a countable union of real lines, if n is congruent to* 3 *mod* 4.

PROOF. 1. If $n = 4k$, we let

$$\eta_{4k} : SU(2) \to (SU(2))^k = \underbrace{SU(2) \times \ldots SU(2)}_{k \text{ times}}$$

denote the diagonal embedding. Viewing $(SU(2))^k$ as a subgroup of $SU(2k) < SO(n)$ we obtain a representation $\eta : SU(2) \to SO(n)$ with $Fix_\eta = \{0\}$.

2. If $n = 4k+1$ or $n = 4k+2$, we identify $SO(4k)$ with a subgroup of $SO(n)$ preserving a $4k$-dimensional subspace V and fixing pointwise its orthogonal complement $V^\perp$. Then for the representation

$$\eta : SU(2) \xrightarrow{\eta_{4k}} SO(4k) < SO(n).$$

we have $\mathrm{F}ix_\eta = V^\perp$.

3. Lastly, if $n = 4k+3$, we use the product representation

$$\eta_{4k} \times \zeta : SU(2) \to SO(4k) \times SO(3) < SO(4k+3),$$

where $\zeta : SU(2) \to SO(3)$ is the universal cover (whose kernel is the center of $SU(2)$). The group $SO(4k)$ fixes a 3-dimensional subspace $W \subset \mathbb{R}^n$ and $\mathrm{F}ix'_\eta$ is a countable union of lines in W. □

THEOREM 15.24. *For each $n \geqslant 3$, there exists a proper subset $C \subset \mathbb{S}^{n-1}$ whose complement $\mathbb{S}^{n-1} \setminus C$ is paradoxical. Furthermore, C is either empty (if n divisible by 4), or is countable (if n is odd) or is a single great circle (if n is even, not divisible by 4).*

PROOF. Given a monomorphism $\iota : F_2 \to SU(2)$ we let $\rho : F_2 \to SO(n), n \geqslant 4$, denote the composition of ι with the representation $\eta : SU(2) \to SO(n)$ constructed in Lemma 15.23. Note that ρ is a monomorphism since kernel of η can only contain central elements of $SU(2)$, hence, only the identity element of $\iota(F_2)$. Define the subset

$$C \subset \mathbb{S}^{n-1}$$

as the (countable) union of fixed-point set of nontrivial elements of $\rho(F_2)$. If n is divisible by 4, the subset C is empty. Now, the group $\rho(F_2)$ acts freely on $E := \mathbb{S}^{n-1} \setminus C$ theorem follows from Lemma 15.20. □

15.4. Step 3: Spheres of dimension $\geqslant 2$ are paradoxical

LEMMA 15.25. *For each $n \geqslant 2$, every subset $E \subset \mathbb{S}^{n-1}$ with countable complement is piecewise-congruent to $\mathbb{S}^{n-1}$.*

PROOF. Let C denote the complement of E in $\mathbb{S}^{n-1}$. We claim that there exists a codimension 2 subspace $F \subset \mathbb{R}^n$, which is disjoint from C. Indeed, for each $c \in C$ the set $L_c \subset Lin(\mathbb{R}^n, \mathbb{R}^2)$ of linear maps $\lambda : \mathbb{R}^n \to \mathbb{R}^2$ with $c \in \mathrm{Ker}(\lambda)$, is nowhere dense in $Lin(\mathbb{R}^n, \mathbb{R}^2)$. Therefore, by the Baire Theorem, there exists a linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^2$ whose kernel is disjoint from C. Then F is the subspace we needed. We identify $SO(2)$ with the subgroup of $SO(n)$ fixing F pointwise. Then for any two elements $c_1, c_2 \in C$ there exists at most one $g \in SO(2)$ such that $g(c_1) = c_2$. Since $SO(2)$ is uncountable, we conclude that there exists an element $g \in SO(2) < SO(n)$ such that

$$g^k(C) \cap C = \emptyset, \quad \forall k \in \mathbb{Z} \setminus \{0\}.$$

Now, the assertion follows from Lemma 15.8. □

EXERCISE 15.26. Suppose that $F \subset \mathbb{R}^n$ is a 2-dimensional subspace and $n \geqslant 2$. Then there exists $g \in SO(n)$ such that for all $k \in \mathbb{Z} \setminus \{0\}$, $g^k F \cap F = \{0\}$.

It now follows from Lemma 15.8 that each subset $E \subset \mathbb{S}^{n-1}$ whose complement is a great circle, is piecewise-congruent to $\mathbb{S}^{n-1}$. According to Theorem 15.24, for each $n \geqslant 3$ there exists a paradoxical subset $E \subset \mathbb{S}^{n-1}$ whose complement is either countable or is a great circle. Since E is piecewise-congruent to $\mathbb{S}^{n-1}$, we obtain:

THEOREM 15.27. *The sphere $\mathbb{S}^{n-1}$ is $SO(n)$-paradoxical for all $n \geqslant 3$.*

15.5. Step 4: Euclidean unit balls are paradoxical

According to Theorem 15.27, for each $n \geqslant 3$, there exists a partition

$$S = \mathbb{S}^{n-1} = E' \sqcup E''$$

and piecewise $SO(n)$-congruences $\phi' : E' \to S$ and $\phi'' : E'' \to S$. We define *radial extensions*

$$\hat{E}' = \{\lambda x : \lambda \in (0,1], x \in E'\}, \quad \hat{E}'' = \{\lambda x : \lambda \in (0,1], x \in E''\}$$

of the subsets E', E''. Accordingly, we let $\hat{\phi}', \hat{\phi}''$ be the radial extensions of the piecewise $SO(n)$-congruences ϕ', ψ''. Both maps $\hat{\phi}', \hat{\phi}''$ are piecewise $SO(n)$-congruences

$$\hat{\phi}' : \hat{E}' \to \mathbb{B}^n \setminus \{0\}, \quad \hat{\phi}'' : \hat{E}'' \to \mathbb{B}^n \setminus \{0\}.$$

Therefore, the *punctured ball* $\mathbb{B}^n \setminus \{0\}$ is $SO(n)$-paradoxical.

LEMMA 15.28. *The punctured unit ball $\mathbb{B}^n \setminus \{0\}$ is piecewise congruent to $\mathbb{B}^n$.*

Let $\Sigma \subset \mathbb{B}^n$ be a round sphere containing the origin 0. According to Lemma 15.25, there exists a piecewise congruence $\psi : \Sigma \setminus \{0\} \to \Sigma$. We then define a piecewise congruence

$$\phi : \mathbb{B}^n \setminus \{0\} \to \mathbb{B}^n$$

as the identity on $\mathbb{B}^n \setminus \Sigma$ and ψ on $\Sigma \setminus \{0\}$. □

Therefore, since the punctured ball $\mathbb{B}^n \setminus \{0\}$ is paradoxical, so is the ball $\mathbb{B}^n$. This concludes the proof of Corollary 15.13, Parts 1 and 2, for $n \leqslant 3$. □

CHAPTER 16

Amenability and paradoxical decomposition.

In this chapter we discuss in detail two important concepts behind the Banach-Tarski paradox: Amenability and paradoxical decompositions. Although both properties were first introduced for groups (of isometries), it turns out that amenability can be defined in purely metric terms, in the context of graphs of bounded geometry. We shall begin by discussing amenability for graphs, then we will turn to the case of groups, and after that, to the opposite property of being *paradoxical*.

CONVENTION 16.1. Throughout the chapter, all the graphs are assumed to be nonempty. In the case of a non-connected graph $\mathcal{G}$, we declare the distance between vertices in different components of $\mathcal{G}$ to be infinite.

16.1. Amenable graphs

We refer the reader to Definition 1.41 for definitions of various boundaries of subgraphs of a graph.

DEFINITION 16.2. A (nonempty) graph $\mathcal{G}$ is called *amenable* if there exists a sequence Φ_n of finite subsets of V such that

$$\lim_{n\to\infty} \frac{|\partial^V \Phi_n|}{|\Phi_n|} = 0\,. \tag{16.1}$$

Such sequence Φ_n is called a *Følner sequence for the graph* $\mathcal{G}$.

Note that if $\mathcal{G}$ has finite valence C, then the vertex boundaries $\partial_V \Phi_n$ and the exterior vertex boundaries $\partial^V \Phi_n$ satisfy

$$|\partial_V \Phi_n| \leqslant |E(\Phi_n, F_n^c)| \leqslant C|\partial_V \Phi_n|$$

$$|\partial^V \Phi_n| \leqslant |E(\Phi_n, F_n^c)| \leqslant C|\partial^V \Phi_n|.$$

Therefore in this case, (16.1) is equivalent to

$$\lim_{n\to\infty} \frac{|\partial_V \Phi_n|}{|\Phi_n|} = 0$$

and

$$\lim_{n\to\infty} \frac{|E(\Phi_n, \Phi_n^c)|}{|\Phi_n|} = 0.$$

It is immediate from the definition that every finite graph is amenable (take $\Phi_n = V$).

An infinite connected graph is amenable if and only if its Cheeger constant, as described in Definition 5.86, is zero. This equivalent characterization of amenability does not extend to finite graphs, which all have positive Cheeger constants.

We describe in what follows various metric properties equivalent to non-amenability. Our arguments are adapted from **[dlHGCS99]**. The only tool that will be needed is Hall–Rado Marriage Theorem from graph theory, stated below.

Let $Bip(Y, Z; E)$ denote the bipartite graph with vertex set V split as $V = Y \sqcup Z$, and the edge-set E. Given two integers $k, l \geqslant 1$, a *perfect* (k, l)*–matching* of $Bip(Y, Z; E)$ is a subset $M \subset E$ such that each vertex in Y is the endpoint of exactly k edges in M, while each vertex in Z is the endpoint of exactly l edges in M.

THEOREM 16.3 (Hall-Rado **[Bol79]**, §$III.2$). *Let $Bip(Y, Z; E)$ be a locally finite bipartite graph and let $k \geqslant 1$ be an integer such that:*

- *For every finite subset $A \subset Y$, its exterior vertex-boundary $\partial^V A$ contains at least $k|A|$ elements.*
- *For every finite subset B in Z, its exterior vertex-boundary contains at least $|B|$ elements.*

Then $Bip(Y, Z; E)$ has a perfect $(k, 1)$–matching.

Given a discrete metric space (X, dist), two (not necessarily disjoint) subsets Y, Z in X, and a real number $C \geqslant 0$, one defines a bipartite graph $Bip_C(Y, Z)$, with the vertex set $Y \sqcup Z$, where two vertices $y \in Y$ and $z \in Z$ are connected by an edge in $Bip_C(Y, Z)$ if and only if $\text{dist}(y, z) \leqslant C$. (The reader will recognize here a version of the Rips complex of a metric space.) We will use this construction in the case when $Y = Z = X$, then the vertex set of $Bip(X, X)$ will consist of two copies of the set X.

In what follows, given a subset $\Phi \subset V$ of the vertex set of a graph $\mathcal{G}$, we will use the notation $\overline{\mathcal{N}}_C(\Phi)$ and $\mathcal{N}_C(\Phi)$ to denote the "closed" and "open" C-neighborhoods of Φ in V:

$$\overline{\mathcal{N}}_C(\Phi) = \{v \in V : \text{dist}(v, \Phi) \leqslant C\}, \quad \mathcal{N}_C(\Phi) = \{v \in V : \text{dist}(v, \Phi) < C\}.$$

THEOREM 16.4. *Let $\mathcal{G}$ be a connected graph of bounded geometry, with vertex set V and edge set E, endowed, as usual, with the standard metric. The following conditions are equivalent:*

(a) *$\mathcal{G}$ is non-amenable.*

(b) *$\mathcal{G}$ satisfies the following* expansion condition*: There exists a constant $C > 0$ such that for every finite non-empty subset $\Phi \subset V$, the set $\overline{\mathcal{N}}_C(\Phi) \subset V$ contains at least twice as many vertices as Φ.*

(b') *For some (equivalently, every) $\beta > 1$ there exists $C > 0$ such that $\overline{\mathcal{N}}_C(\Phi) \cap V$ has cardinality at least β times the cardinality of Φ.*

(c) *There exists a constant $C > 0$ such that the graph $Bip_C(V, V)$ has a perfect $(2, 1)$–matching.*

(d) *There exists a map $f \in \mathcal{B}(V)$ (see Definition 5.20) such that for every $v \in V$ the preimage $f^{-1}(v)$ contains exactly two elements.*

(d') (Gromov's condition) *there exists a map $f \in \mathcal{B}(V)$ such that for every $v \in V$ the pre-image $f^{-1}(v)$ contains at least two elements.*

PROOF. Let $m \geqslant 1$ denote the valence of the graph $\mathcal{G}$.
(b) $\iff$ (b'). Observe that for every $\alpha > 1$, $C > 0$,

$$\forall \Phi, |\overline{\mathcal{N}}_C(\Phi)| \geqslant \alpha|\Phi| \Rightarrow \forall k \in \mathbb{N}, \quad |\overline{\mathcal{N}}_{kC}(\Phi)| \geqslant \alpha^k |\Phi|.$$

Therefore, (b) $\iff$ (b').

(a) $\Rightarrow$ (b'). The graph $\mathcal{G}$ is non-amenable if and only if its Cheeger constant is positive. In other words, there exists $\eta > 0$ such that for every finite set of vertices F, $|E(F, F^c)| \geqslant \eta|\Phi|$. This implies that $\overline{\mathcal{N}}_1(\Phi)$ contains at least $(1+\frac{\eta}{m})|\Phi|$ vertices.

(b) $\Rightarrow$ (c). Let C be the constant as in the expansion property. We form the bipartite graph $Bip_C(Y, Z)$, where Y, Z are two copies of V. Clearly, the graph $Bip_C(Y, Z)$ is locally finite. For any finite subset A in V, since $|\overline{\mathcal{N}}_C(A) \cap V| \geqslant 2|A|$, it follows that the edge–boundary of A in $Bip_C(Y, Z)$ has at least $2|A|$ elements, where we embed A in either one of the copies of V in $Bip_C(Y, Z)$. It follows by Theorem 16.3 that $Bip_C(Y, Z)$ has a perfect $(2, 1)$–matching.

(c) $\Rightarrow$ (d). The matching in (c) defines a map $f : Z = V \to Y = V$, so that $\operatorname{dist}_{\mathcal{G}}(z, f(z)) \leqslant C$. Hence, $f \in \mathcal{B}(V)$ and $|f^{-1}(y)| = 2$ for every $y \in V$.

The implication (d) $\Rightarrow$ (d') is obvious. We show that (d') $\Rightarrow$ (b). According to (d'), there exists a constant $M > 0$ and a map $f : V \to V$ such that for every $x \in V$, $\operatorname{dist}(x, f(x)) \leqslant M$, and $|f^{-1}(y)| \geq 2$ for every $y \in V$. For every finite nonempty set $F \subset V$, $f^{-1}(\Phi)$ is contained in $\mathcal{N}_M(\Phi)$ and it has at least twice as many elements. Thus, (b) is satisfied.

Thus, we proved that the properties (b) through (d') are equivalent.

It remains to be shown that (b) $\Rightarrow$ (a). By hypothesis, there exists a constant C such that for every finite non-empty subset $\Phi \subset V$, $|\overline{\mathcal{N}}_C(\Phi) \cap V| \geqslant 2|\Phi|$. Without loss of generality, we may assume that C is a positive integer. Recall that $\partial_V \Phi$ is the vertex–boundary of the subset $\Phi \subset V$. Since $\overline{\mathcal{N}}_C(\Phi) = \Phi \cup \mathcal{N}_C(\partial_V \Phi)$, it follows that $|\mathcal{N}_C(\partial_V \Phi) \setminus \Phi| \geqslant |\Phi|$.

Recall that the graph $\mathcal{G}$ has finite valence $m \geqslant 1$. Therefore,

$$|\overline{\mathcal{N}}_C(\partial_V \Phi)| \leqslant m^C |\partial_V \Phi|\,.$$

We have, thus, obtained that for every finite nonempty set $\Phi \subset V$,

$$|E(\Phi, \Phi^c)| \geqslant |\partial_V \Phi| \geqslant \frac{1}{m^C} |\mathcal{N}_C(\partial_V \Phi)| \geqslant \frac{1}{m^C} |\Phi|.$$

Therefore, the Cheeger constant of $\mathcal{G}$ is at least $\frac{1}{m^C} > 0$, and the graph is non-amenable. □

EXERCISE 16.5. Show that a sequence $\Phi_n \subset V$ is Følner if and only if for every $C \in \mathbb{R}_+$

$$\lim_{n \to \infty} \frac{|\mathcal{N}_C(\Phi_n)|}{|\Phi_n|} = 1.$$

Some graphs with bounded geometry admit Følner sequences which consist of metric balls. A proof of the following property (in the context of Cayley graphs) first appeared in [**AVS57**].

PROPOSITION 16.6. *A graph $\mathcal{G}$ of bounded geometry and sub-exponential growth (in the sense of Definition 5.75) is amenable and has the property that for every basepoint $v_0 \in V$ (where V is the vertex set of $\mathcal{G}$) there exists a Følner sequence consisting of metric balls with center v_0.*

PROOF. Let v_0 be an arbitrary vertex in $\mathcal{G}$. We equip the vertex set V of $\mathcal{G}$ with the restriction of the standard metric on $\mathcal{G}$ and set

$$\mathfrak{G}_{v_0,V}(n) = |\bar{B}(v_0, n)|,$$

here and in what follows $\bar{B}(x, n)$ is the ball of center x and radius x in V. Our goal is to show that for every $\varepsilon > 0$ there exists a radius R_ε such that $\partial_V \bar{B}(v_0, R_\varepsilon)$ has cardinality at most $\varepsilon\, |\bar{B}(v_0, R_\varepsilon)|$.

We argue by contradiction and assume that there exists $\varepsilon > 0$ such that for every integer $R > 0$,

$$|\partial_V \bar{B}(v_0, R)| \geqslant \varepsilon\, |\bar{B}(v_0, R)|\,.$$

(Since $\mathcal{G}$ has bounded geometry, considering vertex–boundary is equivalent to considering the edge-boundary.) This inequality implies that

$$|\bar{B}(v_0, R+1)| \geqslant (1+\varepsilon)|\bar{B}(v_0, R)|\,.$$

Applying the latter inequality inductively we obtain

$$\forall n \in \mathbb{N}, \quad |\bar{B}(v_0, n)| \geqslant (1+\varepsilon)^n\,,$$

whence

$$\limsup_{n\to\infty} \frac{\ln \mathfrak{G}_{v_0,V}}{n} \geqslant \ln(1+\varepsilon) > 0\,.$$

This contradicts the assumption that $\mathcal{G}$ has sub-exponential growth. □

LEMMA 16.7 (K. Whyte, [**Why99**]). *Suppose that $\mathcal{G}$ is a nonamenable graph of finite valence. Then for each net $V' \subset V = V(\mathcal{G})$, there exists a bijection $f : V' \to V$ which has bounded displacement: There exists $D < \infty$ such that*

$$\operatorname{dist}(x, f(x)) \leqslant D$$

for all $x \in V'$.

PROOF. Since V is not amenable, there exists C such that for every finite subset $A \subset V'$

$$|\partial^V A| \geqslant |A|,$$

where the exterior boundary is understood in the bipartite graph $Bip_C(V', V)$. Therefore, by the Hall–Rado Marriage Theorem, there exists a bijection $f : V' \to V$ sending each $v' \in V$ to a vertex $v = f(v')$ within distance $\leqslant C$ from v'. This map f is the required bijection of bounded displacement. □

The next theorem is also due to K. Whyte, although it was implicit in [**DSS95**]:

THEOREM 16.8 (K. Whyte [**Why99**]). *Let $\mathcal{G}_i, i = 1, 2$, be two non-amenable graphs of bounded geometry. Then every quasiisometry $h : \mathcal{G}_1 \to \mathcal{G}_2$ is at bounded distance from a bi-Lipschitz map $g : V(\mathcal{G}_1) \to V(\mathcal{G}_2)$.*

PROOF. Since h is a quasiisometric embedding, there exists a net $V_1' \subset V(\mathcal{G}_1)$, such that h restricts to a bi-Lipschitz bijection

$$h' := V_1' \to V_2' \subset V(\mathcal{G}_2).$$

The fact that h is a quasiisometry implies that V_2' is also a net in $V(\mathcal{G}_2)$. Now, Lemma 16.7 implies existence of bijections

$$f_i : V_i \to V(\mathcal{G}_i), \quad i = 1, 2$$

of bounded displacement. The composition

$$g := f_2 \circ h' \circ f_1^{-1} : V(\mathcal{G}_1) \to V(\mathcal{G}_2)$$

is the required bi-Lipschitz map. □

For the sake of completeness we mention without proof two more properties equivalent to those in Theorem 16.4.

The first will turn out to be relevant to a discussion later on between non-amenability and existence of free sub-groups (the von Neumann-Day Question 16.69).

THEOREM 16.9 (Theorem 1.3 in [**Why99**]). *Let $\mathcal{G}$ be an infinite connected graph of bounded geometry. The graph $\mathcal{G}$ is non-amenable if and only if there exists a free action of a free group of rank two on $\mathcal{G}$ by bi-Lipschitz maps which are at finite distance from the identity.*

The second property is related to probability on graphs. Let $\mathcal{G}$ be an infinite locally finite connected graph with set of vertices V and set of edges E. For every vertex x of $\mathcal{G}$ we denote by $\mathrm{val}(x)$ the valency at the vertex X. We refer the reader to [**Bre92, DS84, Woe00**] for the definition of Markov chains and detailed treatment of random walks on graphs and groups.

A *simple random walk* on $\mathcal{G}$ is a Markov chain with random variables

$$X_1, X_2, \dots, X_n, \dots$$

on V, with the transition probability $p(x,y) = \frac{1}{\mathrm{val}(x)}$ if x and y are two vertices joined by an edge, and $p(x,y) = 0$ if x and y are not joined by an edge.

We denote by $p_n(x,y)$ the probability that a random walk starting in x will be at y after n steps. The *spectral radius* of the graph $\mathcal{G}$ is defined by

$$\rho(\mathcal{G}) = \limsup_{n\to\infty} \left[p_n(x,y)\right]^{\frac{1}{n}} .$$

It can be easily checked that the spectral radius does not depend on x and y.

THEOREM 16.10 (J. Dodziuk, [**Dod84**]). *A graph of bounded geometry is non-amenable if and only if $\rho(\mathcal{G}) < 1$.*

Note that in the case of countable groups the corresponding theorem was proved by H. Kesten [**Kes59**].

COROLLARY 16.11. *In a non-amenable graph of bounded geometry, the simple random walk is transient, that is, for every $x, y \in V$,*

$$\sum_{n=1}^{\infty} p_n(x,y) < \infty .$$

16.2. Amenability and quasiisometry

THEOREM 16.12 (Graph amenability is QI invariant). *Suppose that $\mathcal{G}$ and $\mathcal{G}'$ are quasiisometric graphs of bounded geometry. Then $\mathcal{G}$ is amenable if and only if $\mathcal{G}'$ is.*

PROOF. We will show that non-amenability is a quasiisometry invariant. We will assume that both $\mathcal{G}$ and $\mathcal{G}'$ are infinite, otherwise the assertion is clear. Note that according to Theorem 16.4, Part (b), nonamenability is equivalent to existence of a constant $C > 0$ such that for every finite non-empty set F of vertices, its closed neighborhood $\bar{\mathcal{N}}_C(\Phi)$ contains at least $2|\Phi|$ vertices.

Let V and V' be the vertex sets of graphs $\mathcal{G}$ and $\mathcal{G}'$ respectively. We assume that V, V' are endowed with the metrics obtained by restriction of the standard metrics on the respective graphs. Let $m < \infty$ be an upper bound on the valence of graphs $\mathcal{G}, \mathcal{G}'$. Let $f : V \to V'$ and $g : V' \to V$ be L–Lipschitz maps that are quasiinverse to each other:

$$\mathrm{dist}(f \circ g, Id) \leqslant A, \quad \mathrm{dist}(g \circ f, Id) \leqslant A.$$

Assume that $\mathcal{G}'$ is amenable. Given a finite set F in V, consider

$$F \xrightarrow{f} F' = f(\Phi) \xrightarrow{g} F'' = g(F').$$

Since F'' is at Hausdorff distance $\leqslant A$ from F, it follows that $|\Phi| \leqslant b|F''|$, where $b = m^L$. In particular,

$$|f(\Phi)| \geqslant b^{-1}|\Phi|.$$

Likewise, for every finite set Φ' in V' we obtain

$$|g(\Phi')| \geqslant b^{-1}|\Phi'| \ .$$

By Theorem 16.4 (Part (b')), for every number $\alpha > b^2$, there exists $C \geqslant 1$ such that for an arbitrary finite set $F' \subset V'$, its neighborhood $\bar{\mathcal{N}}_C(F')$ contains at least $\alpha|F'|$ vertices. Therefore, for such C, the set $g\left(\bar{\mathcal{N}}_C(F')\right)$ contains at least

$$\frac{1}{b}|\mathcal{N}_C(F')| \geqslant \frac{\alpha}{b}|F'|$$

elements.

Pick a finite nonempty subset $\Phi \subset V$ and set $\Phi' := f(\Phi), F'' = gf(\Phi)$. Then $|F'| \geq b^{-1}|\Phi|$ and, therefore,

$$|g\left(\bar{\mathcal{N}}_C(F')\right)| \geqslant \frac{\alpha}{b^2}|\Phi|.$$

Since g is L–Lipschitz,

$$g\left(\bar{\mathcal{N}}_C(F')\right) \subset \bar{\mathcal{N}}_{LC}(F'') \subset \bar{\mathcal{N}}_{LC+A}(\Phi).$$

We conclude that

$$|\bar{\mathcal{N}}_{LC+A}(\Phi)| \geqslant \frac{\alpha}{b^2}|\Phi|.$$

Setting $C' := LC + A$, and $\beta := \frac{\alpha}{b^2} > 1$, we conclude that $\mathcal{G}$ satisfies the expansion property (b') in Theorem 16.4. Hence, $\mathcal{G}$ is also non-amenable. □

We will see below that this theorem generalizes in the context connected Riemannian manifolds M of bounded geometry and graphs $\mathcal{G}$ obtained by discretization of M, and, thus, quasiisometric to M. More precisely, we will see that non-amenability of the graph is equivalent to positivity of the Cheeger constant of the manifold (see Definition 2.21). This may be seen as a version within the setting of amenability/isoperimetric problem of the Milnor–Efremovich–Schwartz Theorem 5.78 stating that the growth functions of M and $\mathcal{G}$ are equivalent.

In what follows we use the terminology in Definitions 2.26 and 2.33 for the bounded geometry of a Riemannian manifold, respectively of a simplicial graph.

THEOREM 16.13. *Let M be a complete connected n-dimensional Riemannian manifold and $\mathcal{G}$ a simplicial graph, both of bounded geometry. Assume that M is quasiisometric to $\mathcal{G}$. Then the Cheeger constant of M is positive if and only if the graph $\mathcal{G}$ is non-amenable.*

REMARKS 16.14. (1) Theorem 16.13 was proved by R. Brooks [**Bro82a**], [**Bro81a**] in the special case when M is the universal cover of a compact Riemannian manifold and $\mathcal{G}$ is the a Cayley graph of the fundamental group of this compact manifold.

(2) A more general version of Theorem 16.13 requires a weaker condition of bounded geometry for the manifold than the one used in this book. See for instance [**Gro93**], Proposition $0.5.A_5$. A proof of that result can be obtained by combining the main theorem in [**Pan95**] and Proposition 11 in [**Pan07**].

PROOF. Since M has bounded geometry it follows that its sectional curvature is at least a and at most b, for some $b \geqslant a$. It also follows that the injectivity radius at every point of M is at least ρ, for some $\rho > 0$.

As in Theorem 2.23, we let $V_\kappa(r)$ denote the volume of ball of radius r in the n-dimensional space of constant curvature κ.

Choose ε so that $0 < \varepsilon < 2\rho$. Let N be a maximal ε-separated set in M.

It follows that $\mathcal{U} = \{B(x, \varepsilon) \mid x \in N\}$ is a covering of M, and by Lemma 2.31, (2), its multiplicity is at most

$$m = \frac{V_a\left(\frac{3\varepsilon}{2}\right)}{V_b\left(\frac{\varepsilon}{2}\right)}.$$

We now consider the restriction of the Riemannian distance function on M to the subset N. Define the Rips complex $Rips_{8\varepsilon}(N)$ (with respect to this metric on N), and the 1-dimensional skeleton of the Rips complex, the graph $\mathcal{G}_\varepsilon$. According to Theorem 5.50, the manifold M is quasiisometric to $\mathcal{G}_\varepsilon$. Furthermore, $\mathcal{G}_\varepsilon$ has bounded geometry as well. This and Theorem 16.12 imply that $\mathcal{G}_\varepsilon$ has positive Cheeger constant if and only if $\mathcal{G}$ has. Thus, it suffices to prove the equivalence in Theorem 16.13 for the graph $\mathcal{G} = \mathcal{G}_\varepsilon$.

Assume that M has positive Cheeger constant. This means that there exists $h > 0$ such that for every open submanifold $\Omega \subset M$ with compact closure and smooth boundary,

$$Area(\partial\Omega) \geqslant h\, Vol(\Omega)\,.$$

Our goal is to show that there exist uniform positive constants B and C such that for every finite subset $F \subset N$ there exists an open submanifold with compact closure and smooth boundary Ω, such that (with the notation in Definition 1.41),

$$|E(F, F^c)| \geqslant B\, Area(\partial\Omega) \text{ and } C Vol(\Omega) \geqslant |F|\,. \tag{16.2}$$

Then, it would follow that

$$|E(F, F^c)| \geqslant \frac{Bh}{C}|F|,$$

i.e., $\mathcal{G}$ would be non-amenable. Here, as usual, $F^c = N \setminus F$.

Since M has bounded geometry, the open cover $\mathcal{U}$ admits a smooth partition of unity $\{\varphi_x\ ;\ x \in N\}$ in the sense of Definition 2.7, such that all the functions φ_x are L–Lipschitz for some constant $L > 0$ independent of x, see Lemma 2.30. Let $F \subset N$ be a finite subset. Consider the smooth function $\varphi = \sum_{x\in F} \varphi_x$. By hypothesis and since $\mathcal{U}$ has multiplicity at most m, the function Φ is Lm–Lipschitz. Furthermore, in view of Sard's Theorem, since the map φ has compact support, the set Θ of singular values of φ is compact and has Lebesgue measure zero.

For every $t \in (0,1)$, the preimage

$$\Omega_t = \varphi^{-1}((t,\infty)) \subset M$$

is an open submanifold in M with compact closure. If we choose t to be a regular value of Φ, that is $t \notin \Theta$, then the hypersurface $\varphi^{-1}(t)$, which is the boundary of Ω_t, is smooth (Theorem 2.4).

Since N is ϵ-separated, the balls $B\left(x, \frac{\varepsilon}{2}\right), x \in N$, are pairwise disjoint. Therefore, for every $x \in N$ the function φ_x restricted to $B\left(x, \frac{\varepsilon}{2}\right)$ is identically equal to 1. Hence, the union

$$\bigsqcup_{x\in F} B\left(x, \frac{\varepsilon}{2}\right)$$

is contained in Ω_t for every $t \in (0,1)$, and in view of Part 2 of Theorem 2.23 we get

$$Vol(\Omega_t) \geqslant \sum_{x\in F} Vol\left(x, \frac{\varepsilon}{2}\right) \geqslant |F| \cdot V_b\left(\varepsilon/2\right) .$$

Therefore, for every $t \notin \Theta$, the domain Ω_t satisfies the second inequality in (16.2) with $C^{-1} = V_b\left(\varepsilon/2\right)$. Our next goal is to find values of $t \notin \Theta$ so that the first inequality in (16.2) holds.

Fix a constant η in the open interval $(0,1)$, and consider the open set $U = \Phi^{-1}((0,\eta))$.

Let F' be the set of points x in F such that $U \cap \overline{B(x,\varepsilon)} \neq \emptyset$. Since for every $y \in U$ there exists $x \in F$ such that $\varphi_x(y) > 0$, it follows that the set of closed balls centered in points of F' and of radius ε cover U.

Since $\{\varphi_x : x \in N\}$ is a partition of unity for the cover $\mathcal{U}$ of M, it follows that for every $y \in U$ there exists $z \in N \setminus F$ such that $\varphi_z(y) > 0$, whence $y \in \overline{B(z,\varepsilon)}$. Thus,

$$U \subset \left(\bigcup_{x\in F'} \overline{B(x,\varepsilon)} \right) \cap \left(\bigcup_{z\in N\setminus F} \overline{B(z,\varepsilon)} \right). \tag{16.3}$$

In particular, for every $x \in F'$ there exists $z \in N\setminus F$ such that $\overline{B(x,\varepsilon)} \cap \overline{B(z,\varepsilon)} \neq \emptyset$, whence x and z are connected by an edge in the graph $\mathcal{G}$.

Thus, every point $x \in F'$ belongs to the vertex-boundary $\partial_V F$ of the subset F of the vertex set of the graph $\mathcal{G}$. We conclude that $\operatorname{card} F' \leqslant \operatorname{card} E(F, F^c)$.

Since $|\nabla\Phi| \leqslant mL$, by the Coarea Theorem 2.15, with $g \equiv 1$, $f = \Phi$ and $U = \Phi^{-1}(0,\eta)$, we obtain:

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{d}t = \int_U |\nabla\varphi| \mathrm{d}V \leqslant mLVol(U) \leqslant mL \sum_{x\in F'} Vol(B(x,\varepsilon)).$$

The last inequality follows from the inclusion (16.3). At the same time, by applying Theorem 2.23, we obtain that for every $x \in M$

$$V_a(\varepsilon) \geqslant Vol(B(x,\varepsilon)).$$

By combining these inequalities, we obtain

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{d}t \leqslant mLV_a(\varepsilon)\,|F'| \leqslant mLV_a(\varepsilon)\,|E\left(F, F^c\right)|.$$

Since Θ has measure zero, it follows that for some $t \in (0, \eta) \setminus \Theta$,

$$Area(\partial\Omega_t) \leqslant 2\frac{m}{\eta} L V_a(\varepsilon)\, |E\,(F, F^c)\,| = B|E\,(F, F^c)\,|.$$

This establishes the first inequality in (16.2) and, hence, shows that nonamenability of M implies nonamenability of the graph $\mathcal{G}$.

We now prove the converse implication. To that end, we assume that for some δ satisfying $2\rho > \delta > 0$, some maximal δ-separated set N and the corresponding graph (of bounded geometry) $\mathcal{G} = \mathcal{G}_\delta$ are constructed as above, so that $\mathcal{G}$ has a positive Cheeger constant. Thus, there exists $h > 0$ such that for every finite subset F in N

$$\operatorname{card} E(F, F^c) \geqslant h \operatorname{card} F\,.$$

Let Ω be an arbitrary open bounded subset of M with smooth boundary. Our goal is to find a finite subset Φ_k in N such that for two constants P and Q independent of Ω, we have

$$Area(\partial\Omega) \geqslant P\,|E(\Phi_k, \Phi_k^c)| \;\text{ and }\; |\Phi_k| \geqslant Q Vol(\Omega)\,. \tag{16.4}$$

This would imply positivity of Cheeger constant of M. Note that, since the graph $\mathcal{G}$ has finite valence, in the first inequality of (16.4) we may replace the edge boundary $E(\Phi_k, \Phi_k^c)$ by the vertex boundary $\partial_V \Phi_k$ (see Definition 1.41).

Consider the finite subset F of points $x \in N$ such that $\Omega \cap B(x, \delta) \neq \emptyset$. It follows that $\Omega \subseteq \bigcup_{x \in F} B(x, \delta)$. We split the set F in two parts:

$$F_1 = \left\{ x \in F \;:\; Vol[\Omega \cap B(x, \delta)] > \frac{1}{2} Vol[B(x, \delta)] \right\} \tag{16.5}$$

and

$$F_2 = \left\{ x \in F \;:\; Vol[\Omega \cap B(x, \delta)] \leqslant \frac{1}{2} Vol[B(x, \delta)] \right\}.$$

Set

$$v_k := Vol\left(\Omega \cap \bigcup_{x \in F_k} B(x, \delta)\right), k = 1, 2.$$

Thus,

$$\max\,(v_1, v_2) \geqslant \frac{1}{2} Vol(\Omega).$$

Case 1: $v_1 \geqslant \frac{1}{2} Vol(\Omega)$. In view of Theorem 2.23, this inequality implies that

$$\frac{1}{2} Vol(\Omega) \leqslant \sum_{x \in F_1} Vol\,(B(x, \delta)) \leqslant |F_1|\, V_a(\delta)\,. \tag{16.6}$$

This gives the second inequality in (16.4). A point x in $\partial_V F_1$ is then a point in N satisfying (16.5), such that within distance 8δ of x there exists a point $y \in N$ satisfying the inequality opposite to (16.5). The (unique) shortest geodesic $[x, y] \subset M$ will, therefore, intersect the set of points

$$\text{Half} = \left\{ x \in M\;;\; Vol\,[B(x, \delta) \cap \Omega] = \frac{1}{2} Vol[B(x, \delta)] \right\}.$$

This implies that $\partial_V F_1$ is contained in the 8δ-neighborhood of the set Half $\subset M$. Given a maximal δ–separated subset H_δ of Half (with respect to the restriction of the Riemannian distance on M), $\partial_V F_1$ will then be contained in the 9δ-neighborhood of H_δ. In particular,

$$\bigsqcup_{x\in\partial_V F_1} B\left(x,\frac{\delta}{2}\right) \subseteq \bigcup_{y\in H_\delta} B(y,10\delta)\,,$$

whence

$$V_b\left(\delta/2\right)\,|\partial_V F_1| \leqslant Vol\left[\bigsqcup_{x\in\partial_V F_1} B\left(x,\frac{\delta}{2}\right)\right] \leqslant$$

$$\sum_{y\in H_\delta} Vol\left[B(y,10\delta)\right] \leqslant V_b(10\delta)\,|H_\delta|. \tag{16.7}$$

Since H_δ extends to a maximal δ–separated subset H' of M, Lemma 2.31, (2), implies that the multiplicity of the covering $\{B(x,\delta) \mid x \in H'\}$ is at most $\frac{V_a\left(\frac{3\delta}{2}\right)}{V_b\left(\frac{\delta}{2}\right)}$.

It follows that

$$m \cdot Area(\partial\Omega) \geqslant \sum_{y\in H_\delta} Area(\partial\Omega \cap B(y,\delta))\,.$$

We now apply Buser's Theorem 2.25 and deduce that there exists a constant $\lambda = \lambda(n,a,\delta)$ such that for all $y \in H_\delta$, we have,

$$\lambda Area(\partial\Omega \cap B(y,\delta)) \geqslant Vol\left[\Omega \cap B(y,\delta)\right] = \frac{1}{2} Vol[B(y,\delta)]\,.$$

It follows that

$$Area(\partial\Omega) \geqslant \frac{1}{2\lambda m}\sum_{y\in H_\delta} Vol[B(y,\delta)] \geqslant \frac{1}{2\lambda m} V_b(\rho)\,|H_\delta|\,.$$

Combining this estimate with the inequality (16.7), we conclude that

$$Area(\partial\Omega) \geqslant P|\partial_V F_1|\,,$$

for some constant P independent of Ω.

This establishes the first inequality in (16.4) and, hence, proves positivity of the Cheeger constant of M in the Case 1.

Case 2. Assume now that v_2 is at least $\frac{1}{2}Vol(\Omega)$.

We obtain, using Buser's Theorem 2.25 for the second inequality below, that

$$mArea(\partial\Omega) \geqslant \sum_{y\in F_2} Area\left(\partial\Omega \cap B(y,\delta)\right) \geqslant \frac{1}{\lambda}\sum_{y\in F_2} Vol\left[\Omega \cap B(y,\delta)\right] \geqslant \frac{1}{2\lambda} Vol(\Omega)\,.$$

Thus, in the Case 2 we obtain the required lower bound on $Area(\partial\Omega)$ directly. □

Corollary 16.15. *Let M and M' be two complete connected Riemann manifolds of bounded geometry which are quasiisometric to each other. Then M has positive Cheeger constant if and only if M' has positive Cheeger constant.*

Proof. Consider graphs of bounded geometry $\mathcal{G}$ and $\mathcal{G}'$ that are quasiisometric to M and M' respectively. Then $\mathcal{G}, \mathcal{G}'$ are also quasiisometric to each other. The result now follows by combining Theorem 16.13 with Theorem 16.12. □

An interesting consequence of Corollary 16.15 is quasiisometric invariance of positivity of the *spectral gap* of Riemannian manifolds. Recall that $h(M) = 0 \iff \lambda_1(M) = 0$ (Theorem 2.52).

COROLLARY 16.16. *If M and M' are complete connected Riemann manifolds of bounded geometry which are quasiisometric to each other, then $\lambda_1(M) = 0 \iff \lambda_1(M') = 0$.*

16.3. Amenability of groups

Motivated by the Banach-Tarski Paradox. John von Neumann [**vN29**] studied properties of group actions that make paradoxical decompositions possible and impossible. He defined the notion of *amenable group G*, based on the existence of a mean/finitely additive measure invariant under the action of the group on itself[1], and equivalent to the *nonexistence* of a G-paradoxical decomposition for any space on which the group acts. One can ask furthermore that no subset has a paradoxical decomposition, for any space endowed with an action of the group. This defines a strictly smaller class, that of *supramenable groups*; such groups will be discussed in Section 16.6.

In this section we define amenable actions and amenable groups, and prove that paradoxical behavior is equivalent to non-amenability. For simplicity of the discussion (and since it is most relevant for the geometric group theory), we only consider amenability in the context of discrete groups and group actions on sets (rather than continuous group actions on topological spaces). We refer the reader to §1.2 for the discussion on finitely additive probability measures (f.a.p. measures) on sets and finitely additive integrals. Later in the chapter we relate amenability and paradoxical decompositions and prove (among other things) that, for finitely generated groups, amenability is equivalent to amenability of its Cayley graph (Theorem 16.48).

Let $\mu : G \times X \to X$ be a left group action, $\mu(g, x) = g(x)$ on the a set X (for right group actions the discussion is very similar).

DEFINITION 16.17. (1) A group action $G \curvearrowright X$ on a set X is *amenable* if there exists a G–invariant f.a.p. measure μ on $\mathcal{P}(X) = 2^X$, the set of all subsets of X.

(2) A group G is *amenable* if the action of G on itself *via* left multiplication is amenable.

Yet another (more common) equivalent definition for amenability can be formulated using the concept of an *invariant mean*, which is responsible for the terminology 'amenable':

DEFINITION 16.18. A *mean* on a set X is a linear functional

$$m : B(X) \to \mathbb{R}$$

defined on the vector space $B(X)$ of bounded real-valued functions on X, satisfying the following properties:

(M1) If $f \geqslant 0$ on X, then $m(f) \geqslant 0$.
(M2) $m(\mathbf{1}_X) = 1$.

[1]Von Neumann called these groups *measurable*.

Assume, moreover, that X is endowed with an action of a group G, $G \times X \to X, (g, x) \mapsto g \cdot x$. This induces an action of G on the vector space $B(X)$ defined by $g \cdot f(x) = f(g^{-1} \cdot x)$. A mean m on X is called *invariant* if $m(g \cdot f) = m(f)$ for every $f \in B(X)$ and $g \in G$.

We now specialize to the case $X = G$. The group G admits two actions on itself, the action L by left multiplication and the action R by right multiplication

$$R : G \times G \to G, \quad R(g, x) = R_g(x) = xg^{-1}.$$

PROPOSITION 16.19. *A group action $G \curvearrowright X$ is amenable (in the sense of Definition 16.17) if and only if it admits a left-invariant mean.*

PROOF. According to Theorem 1.10 each G-invariant f.a.p. measure μ on X defines a G-invariant integral

$$m : B(X) \to \mathbb{R}, \quad m(f) = \int_X f \, d\mu.$$

Since the integral $\int_X$ is a linear functional nonnegative on nonnegative functions and satisfying

$$\int_X \mathbf{1}_X \, d\mu = 1,$$

the functional m is a G-invariant mean on X. Conversely, each G-invariant mean m on X, one defines a G-invariant f.a.p. measure μ on X by $\mu(A) = m(\mathbf{1}_A)$. $\square$

EXAMPLE 16.20. If X is a finite set, then every group action $G \curvearrowright X$ is amenable. In particular, every finite group is amenable. Indeed, for a finite set X define $\mu : \mathcal{P}(X) \to [0, 1]$ by $\mu(A) = \frac{|A|}{|X|}$, where $| \cdot |$ denotes the cardinality of a subset.

DEFINITION 16.21. A *left-invariant mean* on G is a mean invariant under the action L; a *right-invariant mean* on G is a mean invariant under the action R.

The following lemma shows that different notions of invariance for measures and means leads to the same class of groups:

PROPOSITION 16.22. *The following are equivalent:*

(a) *G is amenable.*
(b) *G has a right-invariant f.a.p. measure.*
(c) *G has a right-invariant mean.*

PROOF. (a) $\iff$ (b). Given an f.a.p. measure μ_L on G, we define a measure μ_R on G by

$$\mu_R(A) := \mu_L(A^{-1}), \quad A^{-1} = \{a^{-1} : a \in A\}.$$

Then μ_L is left-invariant iff μ_R is right-invariant. $\square$

In view of this proposition, by default, an invariant mean on a group G will mean a left-invariant mean.

LEMMA 16.23. *Every action $G \curvearrowright X$ of an amenable group is also amenable.*

PROOF. Let μ be an invariant measure on G. Given an action $G \curvearrowright X$, choose a point $x \in X$ and define a function $\nu : \mathcal{P}(X) \to [0, 1]$ by

$$\nu(A) = \mu(\{g \in G \; ; \; gx \in A\}).$$

We leave it to the reader to verify that ν is a G-invariant f.a.p. measure. $\square$

QUESTION 16.24. Suppose that G is a group which admits a mean $m : B(G) \to \mathbb{R}$ that is quasiinvariant, i.e., there exists a constant κ such that

$$|m(f \circ g) - m(f)| \leqslant \kappa$$

for all functions $f \in B(G)$ and all group elements g. Is it true that G is amenable?

We refer the reader to §15.1 for the definitions of paradoxical sets, group actions and groups.

LEMMA 16.25. *A paradoxical action $G \curvearrowright X$ cannot be amenable.*

PROOF. Suppose to the contrary that X admits a G-invariant f.a.p. measure μ and

$$X = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

is a G-paradoxical decomposition, i.e., for some $g_1, \ldots, g_k, h_1, \ldots, h_m \in G$,

$$g_1(X_1) \sqcup \ldots \sqcup g_k(X_k) = X \text{ and } h_1(Y_1) \sqcup \ldots \sqcup h_m(Y_m) = X\,.$$

Then

$$\mu(X_1 \sqcup \ldots \sqcup X_k) = \mu(Y_1 \sqcup \ldots \sqcup Y_k) = \mu(X),$$

which implies that $2\mu(X) = \mu(X)$, contradicting the fact that $\mu(X) = 1$. $\square$

COROLLARY 16.26. *A paradoxical group cannot be amenable.*

EXAMPLE 16.27. The free group of rank two F_2 is non-amenable since F_2 is paradoxical (F_2 acts paradoxically on itself), as explained in Section 15.2.

We will prove in Theorem 16.48 that a finitely generated group is amenable if and only if it is non-paradoxical.

The next theorem summarizes basic properties of amenable groups:

THEOREM 16.28. (1) *Each subgroup of an amenable group is amenable.*

(2) *Let N be a normal subgroup of a group G. The group G is amenable if and only if both N and G/N are amenable.*

(3) *The direct limit G (see Section 1.1) of a directed system $(H_i)_{i \in I}$ of amenable groups H_i, is amenable.*

PROOF. (1) Let μ be a f.a.p. measure on an amenable group G, and let H be a subgroup. By Axiom of Choice, there exists a subset D of G intersecting each right coset Hg in exactly one point. Then $\nu(A) := \mu(AD)$ defines a left-invariant f.a.p. measure on H.

(2) "$\Rightarrow$" Assume that G is amenable and let μ be a f.a.p. measure on G. The subgroup N is amenable according to (1). Amenability of G/N follows from Lemma 16.23, since G acts on G/N by left multiplication.

(2) "$\Leftarrow$" Let ν be a left-invariant f.a.p. measure on G/N, and λ a left-invariant f.a.p. measure on N. On every left coset gN one defines a f.a.p. measure by $\lambda_g(A) = \lambda(g^{-1}A)$. The H–left-invariance of λ implies that λ_g is independent of the representative g, i.e. $gN = g'N \Rightarrow \lambda_g = \lambda_{g'}$.

For every subset B in G define

$$\mu(B) = \int_{G/N} \lambda_g(B \cap gN) d\nu(gN)\,.$$

Then μ is an invariant f.a.p. measure on G.

(3) Let $h_{ij} : H_i \to H_j, i \leqslant j$, be the homomorphisms defining the direct system of groups (H_i) and let G be the direct limit. Let $h_i : H_i \to G$ be the homomorphisms to the direct limit, as defined in §1.5. The set of functions

$$\{f : \mathcal{P}(G) \to [0,1]\} = \prod_{\mathcal{P}(G)} [0,1]$$

is compact according to Tychonoff's theorem (see Remark 7.2, Part 5).

Note that each group H_i acts naturally on G by left multiplication *via* the homomorphism $h_i : H_i \to G$. For each $i \in I$ let $\mathcal{M}_i$ be the set of H_i–left-invariant f.a.p. measures μ on $\mathcal{P}(G)$. Since H_i is amenable, Lemma 16.23 implies that the set $\mathcal{M}_i$ is nonempty.

We claim that the subset $\mathcal{M}_i$ is closed in $\prod_{\mathcal{P}(G)}[0,1]$. Let $f : \mathcal{P}(G) \to [0,1]$ be an element of $\prod_{\mathcal{P}(G)}[0,1]$ in the closure of $\mathcal{M}_i$. Then, for every finite collection $A_1, \ldots, A_n$ of subsets of G and every $\epsilon > 0$ there exists μ in $\mathcal{M}_i$ such that

$$|f(A_j) - \mu(A_j)| \leqslant \epsilon$$

for every $j \in \{1, 2, ..., n\}$. This implies that for every $\epsilon > 0$, $|f(G) - 1| \leqslant \epsilon$,

$$|f(A \sqcup B) - f(A) - f(B)| \leqslant 3\epsilon$$

and

$$|f(gA) - f(A)| \leqslant 2\epsilon,$$

$\forall A, B \in \mathcal{P}(G)$ and $g \in H_i$. By letting $\epsilon \to 0$ we obtain that $f \in \mathcal{M}_i$. Thus, the subset $\mathcal{M}_i$ is indeed closed.

By the definition of compactness, if $\{V_i : i \in I\}$ is a family of closed subsets of a compact space X such that $\bigcap_{j \in J} V_j \neq \emptyset$ for every finite subset $J \subseteq I$, then $\bigcap_{i \in I} V_i \neq \emptyset$.

Consider a finite subset J of I. Since I is a directed set, there exists $k \in I$ such that $j \leqslant k, \forall j \in J$. Hence, we have homomorphisms $h_{jk} : H_j \to H_k, \forall j \in J$, and all homomorphisms $h_j : H_j \to G$ factor through $h_k : H_k \to G$. Thus, $\bigcap_{j \in J} \mathcal{M}_j$ contains $\mathcal{M}_k$, in particular, this intersection is non-empty. It follows from the above that $\bigcap_{i \in I} \mathcal{M}_i$ is non-empty. Every element μ of this intersection is clearly a f.a.p. measure, and μ is also G–left-invariant because

$$G = \bigcup_{i \in I} h_i(H_i). \quad \square$$

Below are several corollaries of this theorem.

Corollary 16.29. *Let G_1 and G_2 be two groups that are co-embeddable in the sense of Definition 3.13. Then G_1 is amenable if and only if G_2 is amenable.*

Corollary 16.30. *Any group containing a nonabelian free subgroup is non-amenable.*

The next corollary follows immediately from Part (2) of Theorem 16.28:

Corollary 16.31. *A semidirect product $N \rtimes H$ is amenable if and only if both N and H are amenable.*

Corollary 16.32. *1. If G_i, $i = 1, \ldots, n$, are amenable groups, then the finite Cartesian product $G = G_1 \times \ldots \times G_n$ is also amenable.*

2. Any direct sum $G = \oplus_{i \in I} G_i$ of amenable groups is again amenable.

Proof. 1. The statement follows from inductive application of Corollary 16.31. 2. This is a combination of Part 1 and the fact that G is a direct limit of finite direct products of the groups G_i. $\square$

Example 16.33 (Infinite direct products of amenable groups need not be amenable). Let $F = F_2$ be the free group of rank 2. Recall, Corollary 4.107, that F is residually finite, hence, for every $g \in F \setminus \{1\}$ there exists a homomorphism $\varphi_g : F \to \Phi_g$ such that $\varphi_g(g) \neq 1$ and Φ_g is a finite group. Each Φ_g is, of course, amenable. Consider the direct product of these finite groups:

$$G = \prod_{g \in F} \Phi_g.$$

Then the product of homomorphisms $\varphi_g : F \to \Phi_g$, defines a homomorphism $\varphi : F \to G$. This homomorphism is injective since for every $g \neq 1$, $\varphi_g(g) \neq 1$. Thus, G cannot be amenable since it contains a nonamenable subgroup, namely, $\varphi(F)$.

Corollary 16.34. *Amenability is preserved by virtual isomorphisms of groups.*

Proof. Suppose that $G/N \cong Q$ with finite normal subgroup $N \lhd G$. Since finite groups are amenable, Part (2) of Theorem 16.28 implies that G is amenable if and only if Q is.

Suppose that H is a finite index subgroup of a group G. Then H contains a subgroup $N \lhd G$ which has finite index in G. Therefore, G is amenable if and only if N is. If G is amenable, so is H. If H is amenable, then N is amenable, which implies that G is amenable if H is. $\square$

Corollary 16.35. *A group G is amenable if and only if all finitely generated subgroups of G are amenable.*

Proof. The direct implication follows from Theorem 16.28, Part (1). The converse implication follows from (3), where, given the group G, we let I be the directed set of all finite subsets in G (ordered by the inclusion), and for each $i \in I$, H_i is the subgroup of G generated by the elements in i. We define the directed system of groups (H_i) by letting $h_{ij} : H_i \to H_j$ be the natural inclusion map whenever $i \subset j$. Then G is the direct limit of the system (H_i) and the assertion follows from Theorem 16.28, Part 3. $\square$

Corollary 16.36. *Direct limits of direct systems of finite groups are amenable.*

Proof. Since each finite group is amenable, Corollary follows from Part 3 of Theorem 16.28.

In order to get more examples of amenable groups, we have to bring geometry into the discussion; this is done by introducing the *Følner sequence criterion of amenability* of groups discussed in the next section and thereby connecting amenability of groups with amenability of graphs.

16.4. Følner property

Suppose that $R : X \times G \to X$ is a *right action* of a group G on a set X. Given subsets $E \subset X$, $K \subset G$ we let EK denote the subset

$$EK = \{R(x, g) = xg \ : \ x \in E\, , \, k \in K\} \subset X.$$

DEFINITION 16.37. A sequence of nonempty subsets $\Omega_n \subset X$ is called a *Følner sequence* for the right action $X \times G \to X$ if for every $g \in G$

$$\lim_{n\to\infty} \frac{|\Omega_n g \bigtriangleup \Omega_n|}{|\Omega_n|} = 0\,. \tag{16.8}$$

A sequence of subsets $\Omega_n \subset G$ is called a *Følner sequence* in G if it is a *Følner sequences* with respect to the right action of G on itself *via* the right multiplication:

$$R_g(x) = xg, g, x \in G.$$

For instance, suppose that $G \cong \mathbb{Z} = X$ and G acts on itself via left multiplication. Then the sequence of intervals $\Omega_n \subset \mathbb{Z}$ of length diverging to infinity is a *Følner sequence* for this group action.

EXERCISE 16.38. Prove that the subsets $\Omega_n = \mathbb{Z}^k \cap [-n, n]^k$ form a Følner sequence for $\mathbb{Z}^k$.

EXERCISE 16.39. The following are equivalent for a sequence of nonempty subsets $\Omega_n \subset X$:

(1) Ω_n is a Følner sequence.
(2) For every finite subset $K \subset G$

$$\lim_{n\to\infty} \frac{|\Omega_n K \bigtriangleup \Omega_n|}{|\Omega_n|} = 0. \tag{16.9}$$

(3) For every symmetric finite subset $K \subset G$, (16.8) holds. Here K is symmetric if $K = K^{-1}$.

EXERCISE 16.40. A countable group G admits a Følner sequence if and only if G admits a Følner sequence Φ_n such that

$$\bigcup_{n\in\mathbb{N}} \Phi_n = G.$$

LEMMA 16.41. *Suppose that G is finitely generated with symmetric finite generating set S and $\mathcal{G}$ is the Cayley graph of G with respect to this generating set. Then the following are equivalent any sequence of nonempty subsets $\Omega_n \subset G$:*

(1)

$$\lim_{n\to\infty} \frac{|\Omega_n S \bigtriangleup \Omega_n|}{|\Omega_n|} = 0. \tag{16.10}$$

(2) *Ω_n is a Følner sequence in the graph $\mathcal{G}$ in the sense of Definition 16.1.*
(3) *Ω_n is a Følner sequence in G.*

PROOF. Let Ω be a subset G, i.e., Ω is a subset of the vertex set of $\mathcal{G}$. Then

$$\partial_V \Omega = \bigcup_{s\in S} \Omega \setminus \Omega s$$

$$\partial^V \Omega = \bigcup_{s\in S} \Omega s \setminus \Omega$$

and, hence,

$$\Omega S \bigtriangleup \Omega = \partial^V \Omega \cup \partial_V \Omega.$$

Therefore (since $\mathcal{G}$ has finite valence)

$$\lim_{n\to\infty} \frac{|\Omega_n S \triangle \Omega_n|}{|\Omega_n|} = 0$$

is equivalent to

$$\lim_{n\to\infty} \frac{|\partial^V \Omega_n|}{|\Omega_n|} = 0,$$

and the latter is the definiton of a Følner sequence in the graph $\mathcal{G}$.

It remains to show that (1) implies that (16.8) holds for each $g \in G$. In view of Exercise 16.5, if Ω_n is a Følner sequence in $\mathcal{G}$ for one finite generating set of G, the sequence Ω_n is also Følner in Cayley graph $\mathcal{G}'$ for another finite generating set of G. By taking a finite generating set of G which contains the generating set $S' = S \cup \{g\}$, we obtain the desired conclusion. □

Definition 16.42. 1. A group action $X \times G \to X$ is said to satisfy the *Følner Property* if it admits a Følner sequence $\Omega_n \subset X$.

2. A group G is said to have the *Følner Property* if G contains a Følner sequence.

Lemma 16.43. *A group G has the Følner Property if and only if for each $\epsilon > 0$ and each finite subset $K \subset G$ there exists a finite subset $F \subset G$ such that*

$$\frac{|KF \triangle F|}{|F|} \leqslant \epsilon\,. \tag{16.11}$$

Proof. Applying the anti-automorphism $G \to G$ given by the inversion $g \mapsto g^{-1}$, we obtain:

$$\frac{|KF \triangle F|}{|F|} = \frac{|F^{-1}K^{-1} \triangle F^{-1}|}{|F^{-1}|}.$$

Lemma follows. □

In view of the lemma, instead of defining the Følner Property for G *via* the right action of G on itself (by right multiplication), we can equivalently define it *via* the left action (by left multiplication).

Exercise 16.44. Show that the following are equivalent for a right action $X \times G \to X$:

(1) $X \times G \to X$ satisfies the *Følner Property*.
(2) For every $K \subset G$ there exists a sequence $\Omega_n \subset X$ such that

$$\lim_{n\to\infty} \frac{|\Omega_n K \triangle \Omega_n|}{|\Omega_n|} = 0.$$

Even though, as we will prove in the next section, the Følner Property is equivalent to amenability and the latter is inherited by subgroups, it is instructive to describe a construction of Følner sequences for a subgroup directly, in terms of Følner sequence on the ambient group.

Proposition 16.45. *Let H be a subgroup of a group G satisfying the Følner Property, and let $(\Omega_n)_{n\in\mathbb{N}}$ be a Følner sequence for G. For every $n \in \mathbb{N}$ there exists $g_n \in G$ such that the intersections $g_n^{-1}\Omega_n \cap H = \Phi_n$ form a Følner sequence for H.*

Proof. Consider a finite subset $K \subset H$, let s denote the cardinality of K. Since $(\Omega_n)_{n\in\mathbb{N}}$ is a Følner sequence for G, the ratios

$$\alpha_n = \frac{|\Omega_n K \bigtriangleup \Omega_n|}{|\Omega_n|} \tag{16.12}$$

converge to 0. We partition each subset Ω_n into intersections with left cosets of H:

$$\Omega_n = \Omega_n^{(1)} \sqcup \ldots \sqcup \Omega_n^{(k_n)},$$

where

$$\Omega_n^{(i)} = \Omega_n \cap g_i H, i = 1, \ldots, k_n, \quad g_i H \neq g_j H, \forall i \neq j.$$

Then $\Omega_n K \cap g_i H = \Omega_n^{(i)} K$. We have that

$$\Omega_n K \bigtriangleup \Omega_n = \left(\Omega_n^{(1)} K \bigtriangleup \Omega_n^{(1)}\right) \sqcup \cdots \sqcup \left(\Omega_n^{(k_m)} K \bigtriangleup \Omega_n^{(k_n)}\right).$$

The inequality

$$\frac{|\Omega_n K \bigtriangleup \Omega_n|}{|\Omega_n|} \leqslant \alpha_n$$

implies that there exists $i \in \{1, 2, \ldots, k_n\}$ such that

$$\frac{|\Omega_n^{(i)} K \bigtriangleup \Omega_n^{(i)}|}{|\Omega_n^{(i)}|} \leqslant \alpha_n.$$

In particular, $g_i^{-1}\Omega_n^{(i)} = \Phi_n$, with $\Phi_n \subseteq H$, and we obtain that

$$\frac{|\Phi_n K \bigtriangleup \Phi_n|}{|\Phi_n|} \leqslant \alpha_n. \quad \square$$

The following proposition complements Lemma 16.23.

Proposition 16.46. *Let G be a group acting on a nonempty set X. The group G is amenable if and only if the action $G \curvearrowright X$ is amenable and for every $p \in X$ the stabilizer G_p of the point p is amenable.*

Proof. The direct implication follows from Lemma 16.23 and from Part 1 of Theorem 16.28. Assume now that for every $p \in X$ its G-stabilizer G_p is amenable and let $m_X : B(X) \to \mathbb{R}$ and $m_p : B(G_p) \to \mathbb{R}$ be G-invariant and G_p-invariant means respectively. We define a left-invariant mean on $B(G)$ using a construction in the spirit of the construction of the product of two measures.

For each $p \in X$ and $F \in B(G)$ define a function F_p on the orbit Gp by

$$F_p(gp) = m_p\left(F|_{gG_p}\right).$$

Since m_p is G_p-invariant, $F_p(gp)$ depends only on $x = gp$ and not on g. Moreover, for each $q \in Gp$, the functions $F_p, F_q : Gp \to \mathbb{R}$ are equal. Therefore, we obtain a G-invariant function F_X on X whose restriction to each orbit Gp equals F_p. Since each m_p is a mean and F is bounded, the function F_X is bounded as well. We define

$$m(F) := m_X\left(F_X\right).$$

The linearity of m follows from the linearity of every m_p and of m_X, the properties properties (M1) and (M2) in Definition 16.18 follow from the fact that of m_X and m_p, $p \in X$. We will verify that m is G-invariant. Take an arbitrary element $h \in G$, and consider the pull-back function

$$h \cdot F : x \mapsto F(h^{-1} \cdot x)\, x \in X.$$

Then

$$(h\cdot F)_p(gp) = m_p\left((h\cdot F)|_{gG_p}\right) = m_p\left(F|_{h^{-1}gG_p}\right) = F_p(h^{-1}gp)\,.$$

We deduce from this that $(h\cdot F)_X = F_X \circ h^{-1} = h\cdot F_X$, whence

$$m(h\cdot F) = m_X\left((h\cdot F)_X\right) = m_X\left(h\cdot F_X\right) = m_X\left(F_X\right) = m(F)\,. \quad \square$$

16.5. Amenability, paradoxality and the Følner property

In this section we will show that, amenability of actions is equivalent to non-paradoxality and to the Følner property. According to Theorem 16.12, if one Cayley graph of a finitely generated group G is amenable then all the other Cayley graphs are. Thus, in what follows we fix a finite generating set S of G, the corresponding Cayley graph $\mathcal{G} = \mathrm{Cayley}(G,S)$, and the corresponding word metric on G.

We will use a construction of *Cayley graphs of group actions*, generalizing the usual notion of Cayley graphs for groups. Let G be a group with the generating set S, $X\times G\to X$ be a right action. We define the *Cayley graph* of this action (with respect to the generating set S) as the graph $\mathrm{Cayley}(X,G,S)$ whose vertex set is X and the edge set consists of unordered pairs $\{x,xs\}$, $x\in X, s\in S$.

Remark 16.47. This construction explains why did we choose to define Følner sequences using *mright actions* instead of *left actions*: One defines Cayley graphs using the *right action* of the generating sets S on the group G.

In the next theorem, given a (left) group action $L: G\times X\to X$, one we use the *right* group action $X\times G\to X$ defined by $R:(x,g)\mapsto L(g^{-1},x)$. We note that the equivalence (1) $\iff$ (2) in the next theorem is a special case of the Tarski's Alternative Theorem 16.49. Note that the proof of Theorem 16.48 (namely, the implication (4)$\Rightarrow$(1)) uses existence of ultrafilters on $\mathbb{N}$. One can show that ZF axioms of the set theory are insufficient to conclude that $\mathbb{Z}$ has an invariant mean.

Theorem 16.48. *The following three conditions are equivalent for a group action $G\curvearrowright X$:*

(1) *$G\curvearrowright X$ has an invariant mean.*
(2) *$G\curvearrowright X$ is not paradoxical.*
(3) *For every finitely generated subgroup $H\leqslant G$ (with the generating set S), the Cayley graph (of the associated right G-action) $\mathcal{G} = \mathrm{Cayley}(X,H,S)$ is amenable.*
(4) *The associated right action $X\times G\to X$ satisfies the Følner property.*

Proof. (1) $\Rightarrow$ (2). The implication (1) $\Rightarrow$ (2) is established in Corollary 16.26.

(2) $\Rightarrow$ (3). We will prove the contrapositive of the implication (2) $\Rightarrow$ (3). Assume that $\mathcal{G}$ is non-amenable for some finitely generated subgroup $H\leqslant G$. According to Theorem 16.4, this implies that there exists a map $f: X\to X$ which is at finite distance from the identity map, such that $|f^{-1}(y)| = 2$ for every $y\in X$. Repeating the proof of Lemma 5.33 verbatim, we conclude that there exists a finite subset $S=\{h_1,...,h_n\}\subset G$ and a decomposition $X = T_1\sqcup ...\sqcup T_n$ such that f restricted to T_i coincides with $h_i\big|_{T_i}$.

For every $y\in X$ we have that $f^{-1}(y)$ consists of two elements, which we label as $\{y_1,y_2\}$. This gives a decomposition of X into $Y_1\sqcup Y_2$. Now we decompose $Y_1 = A_1\sqcup ...\sqcup A_n$, where $A_i = Y_1\cap T_i$, and likewise $Y_2 = B_1\sqcup ...\sqcup B_n$, where

$B_i = Y_2 \cap T_i$. Clearly $h_1(A_1) \sqcup ... \sqcup h_n(A_n) = X$ and $B_1 h_1 \sqcup ... \sqcup h_n(B_n) = X$. We have, thus, proved that $G \curvearrowright X$ is paradoxical.

(3) $\iff$ (4) The proof of this equivalence is exactly the same as the one in Lemma 16.41. Let $K \subset G$ be a finite nonempty subset and let $H \leqslant G$ be the subgroup generated by K. It suffices to consider the case $K = K^{-1}$, see Exercise 16.39. Amenability of the Cayley graph $\mathcal{G} = \mathrm{Cayley}(X, H, K)$ implies that there exists a sequence of subsets $\Omega_n \subset X$ such that

$$\lim_{n\to\infty} \frac{|\partial^V \Omega_n|}{|\Omega_n|} = 0. \tag{16.13}$$

As in the proof of Lemma 16.41,

$$\Omega_n K \bigtriangleup \Omega_n = \partial^V(\Omega_n K) \cup \partial_V(\Omega_n K).$$

Therefore, (16.13) implies that

$$\lim_{n\to\infty} \frac{|\Omega_n K \bigtriangleup \Omega_n|}{|\Omega_n|} = 0.$$

Lastly, Exercise 16.44 shows that the action $X \times G \to X$ satisfies the Følner property.

The reverse implication (4)⇒(3) is proven similarly and we leave details to the reader.

(4) ⇒ (1). We first illustrate the proof in the case $G = \mathbb{Z} = X$ and the Følner sequence

$$\Omega_n = [-n, n] \subset \mathbb{Z},$$

since the proof is more transparent in this case and illustrates the general argument. Let ω be a non-principal ultrafilter ω on $\mathbb{N}$ (here we need a form of the Axiom of Choice). We define a function $\mu : 2^{\mathbb{Z}} \to [0, 1]$ by

$$\mu(A) := \omega\text{-lim} \frac{|A \cap \Omega_n|}{2n+1}, \quad A \subset \mathbb{Z}.$$

We leave it to the reader to check that A is a f.a.p. measure. Let us show that μ is invariant under the translation $g : z \mapsto z + 1$. Note that

$$||A \cap \Omega_n| - |gA \cap \Omega_n|| \leqslant 1.$$

Thus,

$$|\mu(A) - \mu(gA)| \leqslant \omega\text{-lim} \frac{1}{2n+1} = 0.$$

This implies that μ is $\mathbb{Z}$-invariant.

Consider now the general case. We use a Følner sequence (Ω_n) of subsets of G to construct a G-invariant f.a.p. measure on $2^X = \mathcal{P}(G)$. Following Remark 16.22, we can and will use the right action (*via right multiplication*) of G on itself in this part of the proof.

Let ω be a non-principal ultrafilter on $\mathbb{N}$. For every $A \subset X$ define

$$\mu(A) = \omega\text{-lim} \frac{|A \cap \Omega_n|}{|\Omega_n|}.$$

We claim that μ is a f.a.p. measure on X. Indeed, for any pair of disjoint subset $A, B \subset X$, we have

$$\mu(A \sqcup B) = \omega\text{-lim} \frac{|(A \cup B) \cap \Omega_n|}{|\Omega_n|} = \omega\text{-lim} \frac{|A \cap \Omega_n| + |B \cap \Omega_n|}{|\Omega_n|} =$$

$$\omega\text{-lim}\,\frac{|A\cap\Omega_n|}{|\Omega_n|}+\omega\text{-lim}\,\frac{|B\cap\Omega_n|}{|\Omega_n|}=\mu(A)+\mu(B).$$

The condition that $\mu(X)=1$ is equally clear. We will now verify that μ is G-invariant. Take an element $g\in G$. We have

$$|\mu(Ag)-\mu(A)|=\omega\text{-lim}\,\frac{\big||Ag\cap\Omega_n|-|A\cap\Omega_n|\big|}{|\Omega_n|}=\omega\text{-lim}\,\frac{\big||A\cap\Omega_n g^{-1}|-|A\cap\Omega_n|\big|}{|\Omega_n|}.$$

Furthermore,

$$\big||A\cap\Omega_n g^{-1}|-|A\cap\Omega_n|\big|=|A\cap(\Omega_n g^{-1}\bigtriangleup\Omega_n)|.$$

Since

$$\omega\text{-lim}\,\frac{|A\cap(\Omega_n g^{-1}\bigtriangleup\Omega_n)|}{|\Omega_n|}\leqslant\omega\text{-lim}\,\frac{|\Omega_n g^{-1}\bigtriangleup\Omega_n|}{|\Omega_n|}=0$$

(as (Ω_n) be a Følner sequence), it follows that

$$\mu(Ag)=\mu(A),$$

i.e., μ is G-invariant. □

In particular, Theorem 16.48 shows that the nonexistence of a G-paradoxical decomposition of X is equivalent to the existence of a G-invariant f.a.p. measure on X.

A. Tarski proved ([**Tar38**], [**Tar86**, pp. 599–643], see also [**Wag85**, Corollary 9.2]) the following stronger form of this equivalence:

THEOREM 16.49 (Tarski's alternative). *Let G be a group acting on a space X and let E be a subset in X. Then E is not G–paradoxical if and only if there exists a G–invariant finitely additive measure $\mu:\mathcal{P}(X)\to[0,\infty]$ such that $\mu(E)=1$.*

The equivalence in Theorem 16.48 gives another proof that the free group on two generators F_2 is paradoxical: Consider the map $f:F_2\to F_2$ which consists in deleting the last letter in every nonempty reduced word and $f(1)=1$. This map satisfies Gromov's condition in Theorem 16.4. Hence, the Cayley graph of F_2 is non-amenable; thus, F_2 is non-amenable as well.

COROLLARY 16.50. *Each group is either paradoxical or amenable.*

COROLLARY 16.51. *Amenability is QI invariant for finitely generated groups.*

PROOF. This follows from the fact that amenability of graphs of finite valence is QI invariant, see Theorem 16.12. □

Now that we know that the group $\mathbb{Z}$ is amenable, we can get a much larger class of amenable groups than direct limits of finite groups:

COROLLARY 16.52. *Every abelian group G is amenable.*

PROOF. Since every abelian group is a direct limit of finitely generated abelian subgroups, by Part (3) of Theorem 16.28, it suffices to prove amenability of finitely generated abelian groups. Since each $\mathbb{Z}^k$ satisfies the Følner Property (see Exercise 16.38), it is amenable. Each finitely generated abelian group A is a product of a finite group and a free abelian group of finite rank; therefore, A is amenable, e.g. by Part (2) of Theorem 16.28. □

COROLLARY 16.53. *Every solvable group is amenable.*

Proof. We argue by induction on the derived length. If $k = 1$ then G is abelian and, hence, are amenable by Corollary 16.52. Assume that the assertion holds for k and take a group G such that $G^{(k+1)} = \{1\}$ and $G^{(i)} \neq \{1\}$ for any $i \leqslant k$. Then $G^{(k)}$ is abelian and $\bar{G} = G/G^{(k)}$ is solvable with derived length equal to k. Whence, by the induction hypothesis, $\bar{G}$ is amenable. This and Theorem 16.28, Part (2), imply that G is amenable. □

Similarly, Theorem 16.28 proves amenability of a much larger class of groups, first introduced by M. Day [**Day57**]:

Definition 16.54. The class of *elementary amenable groups* $\mathcal{EA}$ is the smallest class of groups containing all finite groups, all abelian groups and closed under direct limits, taking subgroups, quotient groups and extensions

$$1 \to G_1 \to G_2 \to G_3 \to 1,$$

where both G_1, G_3 are elementary amenable.

This class contains all solvable groups and many groups which are not even virtually solvable. Some of these groups (elementary amenable but not virtually solvable) are finitely presented.

Example 16.55. Let H_n be the n-th *Houghton group* H_n, $n \geqslant 2$. Every finite group embeds in every H_n, hence each H_n is not virtually solvable. The group H_n is isomorphic to a semidirect product $F \ltimes \mathbb{Z}^n$, where the group F is locally finite, containing (up to an isomorphism) all finite groups. Finite presentability of Houghton groups (for $n \geqslant 3$) was proven by K. Brown [**Bro87**]. According to [**Lee12**], $H_n, n \geqslant 3$, has the following presentation:

$$\left\langle a, x_1, \ldots, x_{n-1} | a^2 = 1, (aa^{x_1})^3 = 1, [a, a^{x_1^2}] = 1, a = [x_i, x_j], a^{x_i^{-1}} = a^{x_j^{-1}}, 1 \leqslant i < j \leqslant n-1 \right\rangle.$$

Thus, each H_n, $n \geqslant 3$, is finitely presentable, elementary amenable but not virtually solvable.

Theorem 16.28 and Corollary 16.52 imply that all elementary amenable groups are amenable. There are finitely presented amenable groups which are not elementary amenable; the first such examples were constructed by R. Grigorchuk [**Gri98**]. Namely, Grigorchuk proves that the group with the following presentation

$$\langle a, c, d, t | a^2 = c^2 = d^2 = (ad)^4 = (adacac)^4 = 1, t^{-1}at = aca, t^{-1}ct = dc, t^{-1}dt = c \rangle$$

is amenable but not elementary amenable. None of the elementary amenable groups have intermediate growth according to the following theorem of C. Chou, [**Cho80**]:

Theorem 16.56. *A finitely generated elementary amenable group either is virtually nilpotent or contains a nonabelian free subsemigroup.*

We note that J. Rosenblatt [**Ros74**] earlier proved this alternative for solvable groups.

16.6. Supramenability and weakly paradoxical actions

The following definition is formally similar to the one of amenable actions and groups. In order to motivate this definition we note that in the Banach-Tarski paradox, we had a paradoxical decomposition of the *subset* $\mathbb{B}^n \subset \mathbb{E}^n$ rather than of the entire Euclidean space, i.e., the action of the Euclidean isometry group was weakly paradoxical. While amenability obstructs G-paradoxical decompositions of the entire set X on which the group is acting, supramenability obstructs G-paradoxical decompositions of subsets of X.

Definition 16.57. 1. A group action $G \curvearrowright X$ is said to be *supramenable* if for every nonempty subset $E \subset X$ there exists a f.a. G-invariant measure μ

$$\mu : \mathcal{P}(X) \to [0, \infty]$$

such that $\mu(E) = 1$.

2. A group G is said to be *supramenable* if the action $G \times G \to G$ of G on itself *via* left multiplication is supramenable.

Exercise 16.58. Show that in this definition it does not matter if G acts on itself *via* left or right multiplication.

It is immediate from the definition that each supramenable action is amenable and every supramenable group is amenable.

The following proposition is an analogue of Lemma 16.23 and Theorem 16.48 for supramenable groups.

Proposition 16.59. *The following are equivalent for a group G:*
1. G is not weakly paradoxical.
2. There are no weakly paradoxical actions $G \curvearrowright X$.
3. Every action $G \curvearrowright X$ is supramenable.
4. G is supramenable.

Proof. The implication (2)⇒(1) is immediate. The proof of the implication (3)⇒(2) is analogous to that of Lemma 16.25. Let $E \subset X$ be a nonempty subset and μ a f.a. G-invariant measure on X such that $\mu(E) = 1$. The existence of μ prevents G-paradoxical decompositions of E just as in the proof of Lemma 16.25.

The proof of the implication (4)⇒(3) is similar to the proof of Lemma 16.23. Consider an action $G \curvearrowright X$ and a non-empty subset E of X. Pick a point $x \in E$ and let G_E be the set of $g \in G$ such that $gx \in E$. This set is nonempty since $1 \in G_E$. Let μ be a G–left-invariant finitely additive measure $\mu_G : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(G_E) = 1$. For $A \subset X$ let

$$\mu(A) := \mu_G(\{g \in G : g(x) \in A\}).$$

Then $\mu(E) = 1$ and μ is G-invariant f.a. measure on X. Lastly, the implication (1)⇒(4) is a special case of Tarski's Theorem 16.49. □

Proposition 16.60. *Each finitely generated weakly paradoxical group has exponential growth.*

Proof. Let G be weakly paradoxical and let E be a G-paradoxical subset of G with the paradoxical decomposition

$$E = E' \sqcup E''$$

and bijections $\psi' : E \to E' \subset E, \psi'' : E \to E'' \subset E$ which are piecewise G-congruences:

$$E' = E'_1 \sqcup \ldots \sqcup E'_k, \quad E'' = E''_1 \sqcup \ldots \sqcup E''_l$$

and there exist $g'_1, \ldots, g'_k \in G, \quad g''_1, \ldots, g''_l \in G$ such that

$$\psi'\big|_{g'_i E'_i} = (g'_i)^{-1}\big|_{g'_i E'_i}, \quad \psi'\big|_{g''_j E''_j} = (g''_j)^{-1}\big|_{g''_j E''_j}, \quad 1 \leqslant i \leqslant k, \ 1 \leqslant j \leqslant l.$$

We now generate a semigroup H of (injective but not surjective) maps $E \to E$ by the maps ψ', ψ''. Since the images of these maps are disjoint, the semigroup H is free on the generating set $\Psi := \{\psi', \psi''\}$, see Lemma 4.59: We obtain an injective homomorphism $\rho : SF_2 \to Map(E, E)$ from the free semigroup of rank 2 on the generators s', s'', sending s' to ψ', s'' to ψ''. Moreover, according to Lemma 4.59, given any two distinct words elements $u, v \in SF_2$,

$$\rho(u)(E) \cap \rho(v)(E) = \emptyset.$$

In particular, given any $x \in E$, the subset

$$X_n := \{\rho(w)(x) : \ell_{SF_2}(w) \leqslant n\} \subset Gx$$

has cardinality

$$1 + 2 + 4 + \ldots + 2^n = 2^{n+1} - 1.$$

Here ℓ_{SF_2} is the word metric on the free semigroup SF_2. By the construction, for each $x \in G$ and every word $w \in SF_2$ of length m there exists a (positive) word $\bar{w}$ of the same length m in the alphabet $A = \{g'_1, \ldots, g'_k, g''_1, \ldots, g''_l\}$, such that

$$\rho(w)(x) = \bar{w}x.$$

Taking $x = 1 \in G$, we conclude that the subsemigroup of G generated by the set A has exponential growth. It follows that the group G has exponential growth as well. $\square$

The following corollary of Proposition 16.60 is a strengthening of Proposition 16.6 in the group-theoretic setting.

Corollary 16.61. *Every group of subexponential growth is supramenable.*

It is not known if the converse of Proposition 16.60 is true or if, to the contrary, there exist supramenable groups of exponential growth. A weaker form of the converse of Proposition 16.60 is known though, and it runs as follows.

Proposition 16.62. *A free two-generated subsemigroup S of a group G is always G–paradoxical, where the action $G \curvearrowright G$ is either by left or by right multiplication. In particular, a supramenable group G cannot contain a free two-generated subsemigroup.*

Proof. Let a, b be the two elements in G generating the free sub-semigroup S, let S_a and S_b be the subsets of elements in S represented by words beginning in a, respectively by words beginning in b. Then $S = S_a \sqcup S_b$, with $a^{-1}S_a = S$ and $b^{-1}S_b = S$. $\square$

Remark 16.63. The converse of Proposition 16.62, on the other hand is not true: a weakly paradoxical group does not necessarily contain a nonabelian free subsemigroup. There exist paradoxical torsion groups (see Remark 16.76 in the next section).

Below we discuss basic properties of supramenable groups which parallel those of amenable groups, given in Theorem 16.28.

PROPOSITION 16.64. (1) *Any subgroup of a supramenable group is supramenable.*

(2) *Any finite extension of a supramenable group is supramenable.*

(3) *Any quotient of a supramenable group is supramenable.*

(4) *Any direct limit of a directed system of supramenable groups is supramenable.*

PROOF. (1) Let $H \leqslant G$ with G supramenable and let E be a non-empty subset of H. Let $\mu_G : \mathcal{P}(G) \to [0, \infty]$ be a G–left-invariant finitely additive measure such that $\mu(E) = 1$. Restricting μ_G to $\mathcal{P}(H)$ we obtain the required measure on G.

(2) Let $H \leqslant G$ with H supramenable and $G = \bigsqcup_{i=1}^m Hx_i$. Let E be a non-empty subset of G. The group H acts on G by left multiplication, according to Proposition 16.59, there exists $\mu : \mathcal{P}(G) \to [0, \infty]$, an H–left-invariant f.a. measure such that

$$\mu\left(\cup_{i=1}^m x_i E\right) = 1.$$

Define a measure $\nu : \mathcal{P}(G) \to [0, \infty]$ by

$$\nu(A) = \frac{\sum_{i=1}^m \mu(x_i A)}{\sum_{i=1}^m \mu(x_i E)}.$$

Note that the denominator in this fraction is positive, and it is also clear that ν is finitely additive and satisfies $\nu(E) = 1$. We need to verify G-left-invariance of ν.

Let A be an arbitrary non-empty subset of G and g an arbitrary element in G. We have that

$$G = \bigsqcup_{i=1}^m Hx_i = \bigsqcup_{i=1}^m Hx_i g,$$

whence there exists a bijection $\varphi : \{1, \ldots, m\} \to \{1, \ldots, m\}$ depending on g such that $Hx_i g = Hx_{\varphi(i)}$.

We may then rewrite the numerator in the expression of $\nu(gA)$ as

$$\sum_{i=1}^m \mu(x_i gA) = \sum_{i=1}^m \mu(h_i x_{\varphi(i)} A) = \sum_{i=1}^m \mu(x_{\varphi(i)} A) = \sum_{j=1}^m \mu(x_j A).$$

For the second equality above we have used the H–invariance of μ. We conclude that ν is G–left-invariant.

(3) Let E be a non-empty subset of G/N. Proposition 16.59 applied to the action of G on G/N gives a G–left-invariant finitely additive measure $\mu : \mathcal{P}(G/N) \to [0, \infty]$ such that $\mu(E) = 1$. The same measure is also G/N–left-invariant.

(4) The proof is very similar to the one in Theorem 16.28, Part 4, and we present only a sketch. Let $h_{ij} : H_i \to H_j, i \leqslant j$, be the homomorphisms defining the direct system of groups (H_i) and let G be the direct limit. Let $h_i : H_i \to G$ be the homomorphisms to the direct limit, see Section 1.5. Consider a non-empty subset E of G. Without loss of generality we may assume that all intersections $h_i(H_i) \cap E$

are nonempty: There exists i_0 such that for every $i \geqslant i_0$, $h_i(H_i) \cap E \neq \emptyset$, and we then restrict to the set of indices $i \geqslant i_0$. The set of functions

$$\{f : \mathcal{P}(G) \to [0, \infty]\} = \prod_{\mathcal{P}(G)} [0, \infty]$$

is compact according to Tychonoff's theorem. For each $i \in I$ let $\mathcal{M}_i$ be the set of H_i–left-invariant f.a. measures μ on $\mathcal{P}(G)$ such that $\mu(E) = 1$. Since H_i is supramenable, Proposition 16.59 implies that the set $\mathcal{M}_i$ is non-empty. Then, as in the proof of Theorem 16.28 one verifies that each subset $\mathcal{M}_i$ is closed in $\prod_{\mathcal{P}(G)}[0, \infty]$ and that the intersection $\bigcap_{i \in I} \mathcal{M}_i$ is non-empty. Every element μ of this intersection is clearly a f.a. measure such that $\mu(E) = 1$; each μ in the intersection is also G–left-invariant because

$$G = \bigcup_{i \in I} h_i(H_i). \quad \square$$

REMARK 16.65. 1. Note that, unlike amenability, supramenability is not preserved by extensions: If a normal subgroup N in a group G is supramenable and $Q = G/N$ is supramenable then G might not be supramenable. Indeed, if G is metabelian but not virtually nilpotent, then both $G' = N$ and $G/G' = Q$ are abelian. However, each solvable group that is not virtually nilpotent contain a non-abelian free subsemigroup (Theorem 16.56) and, hence, cannot be supramenable according to Proposition 16.60.

2. Surprisingly, it is unknown if direct products of supramenable groups are supramenable.

As an example, the group of Euclidean isometries $G = \mathrm{Isom}_+(\mathbb{E}^2)$ is solvable but not virtually nilpotent, therefore, it is not supramenable. Specifically:

PROPOSITION 16.66. $\mathrm{Isom}_+(\mathbb{E}^2)$ *contains a free subsemigroup S on two generators. In particular, according to Proposition 16.62, the subset $S \subset G$ is G-paradoxical.*

PROOF. Let $\lambda \in \mathbb{C}$ be a transcendental number with $|\lambda| = 1$. Consider the rotation $g(z) = \lambda z$ in $\mathbb{E}^2$ (identified with the complex plane) and the translation $h : z \mapsto z + 1$. We claim that the semigroup $S \subset G$ generated by g and h is free two-generated. Indeed, consider the set X of all nonconstant integer polynomials with nonnegative coefficients in the variable λ. The semigroup S acts on X by the postcomposition

$$s \cdot p(\lambda) = s \circ p(\lambda), \quad s \in S.$$

Then for each $p \in X$, $h \cdot p$ has nonzero constant term, while $g \cdot p$ has zero constant term. Therefore, $g(X)$, $h(X)$ are disjoint subsets of X. Hence, according to Lemma 4.59, the semigroup S is free of rank 2. $\square$

COROLLARY 16.67 (Sierpinski–Mazurkiewicz paradox). *There exists a countable G-paradoxical subset $E \subset \mathbb{E}^2$. In particular, the action of $G = \mathrm{Isom}_+(\mathbb{E}^2)$ on $\mathbb{E}^2$ is weakly paradoxical.*

PROOF. Since the semigroup S is countable, for a generic choice of $z \in \mathbb{C}$ the orbit map $S \to \mathbb{C}, s \mapsto s(z)$ is injective. Now the claim follows from Lemma 15.7. $\square$

Neither one of the classes of supramenable and elementary amenable groups contains the other:

- solvable groups are all elementary amenable, while they are supramenable only if they are virtually nilpotent;
- there exist groups of intermediate growth that are not elementary amenable, see [**Gri84a**].

We are now able to relate amenability to the Banach–Tarski paradox.

PROPOSITION 16.68. (1) *The group of isometries* $\mathrm{Isom}(\mathbb{R})$ *is supramenable and, hence,* $\mathbb{R}$ *does not contain paradoxical subsets.*

(2) *The group of isometries* $G = \mathrm{Isom}(\mathbb{E}^2)$ *is amenable but not supramenable.*

(3) $\mathbb{E}^2$ *contains paradoxical subsets.*

(4) $\mathbb{E}^2$ *does not admit a paradoxical decomposition.*

(5) *The group of isometries* $\mathrm{Isom}(\mathbb{E}^n)$ *with* $n \geqslant 3$ *is non-amenable.*

PROOF. Part (1) follows from the fact that $\mathrm{Isom}(\mathbb{R})$ contains the abelian subgroup $\mathrm{Isom}_+(\mathbb{R})$ of index 2.

Part (2) follows from the fact that G is solvable (and, hence, amenable), but not supramenable since it contains a free subsemigroup of rank 2.

Part (3) is the Sierpinski–Mazurkiewicz paradox above.

Part (4) follows from amenability of $G = \mathrm{Isom}(\mathbb{E}^2)$, which implies amenability and, hence, nonparadoxality, of any action of G.

Part (5) follows from the fact that $SO(3) < \mathrm{Isom}(\mathbb{E}^n)$ contains rank 2 free subgroups. □

Since many examples and counterexamples display a connection between amenability and the algebraic structure of a group, it is natural to ask whether there exists a purely algebraic criterion of amenability. J. von Neumann made the observation that the existence of a free subgroup excludes amenability in the very paper where he introduced the notion of amenable groups, [**vN28**]. It is this observation that has led to the following question:

QUESTION 16.69 (the von Neumann-Day problem). Does every non-amenable group contain a nonabelian free subgroup?

The question is implicit in [**vN29**], and it was formulated explicitly by M. Day [**Day57**, §4].

When restricted to the class of subgroups of Lie groups with finitely many components (in particular, subgroups of $GL(n,\mathbb{R})$), Question 16.69 has an affirmative answer, since, in view of the Tits' alternative (see Chapter 13, Theorem 13.1), every such group either contains a nonabelian free subgroup or is virtually solvable. Since all virtually solvable groups are amenable, the claim follows. For the same reason, for all classes of groups satisfying the Tits Alternative (see section 13.7) Question 16.69 has an affirmative answer.

The first examples of finitely generated non-amenable groups with no (nonabelian) free subgroups were given by A. Ol′shanskiĭ in [**Ol′80**]. In [**Ady82**] it was shown that the free Burnside groups $B(n,m)$ with $n \geqslant 2$ and $m \geqslant 665$, m odd, are also non-amenable. The first examples of finitely presented non-amenable groups with no (non-abelian) free subgroups, were given by A. Ol′shanskiĭ and M. Sapir in [**OS02**].

Still, metric versions of the von Neumann-Day Question 16.69 have positive answers. One of these versions is Whyte's Theorem 16.9. Another metric version with positive answer was established by I. Benjamini and O. Schramm in [**BS97b**]. They proved

THEOREM 16.70.
- *An infinite locally finite simplicial graph $\mathcal{G}$ with positive Cheeger constant contains a tree with positive Cheeger constant.*[2]
- *If, moreover, the Cheeger constant of $\mathcal{G}$ is at least an integer $n \geqslant 0$, then $\mathcal{G}$ contains a spanning subgraph, where each connected component is a rooted tree with all vertices of valency n, except the root, which is of valency $n+1$.*
- *If X is either a graph or a Riemannian manifold with infinite diameter, bounded geometry and positive Cheeger constant (in particular, if X is the Cayley graph of a paradoxical group) then X contains a bi-Lipschitz embedding of the binary rooted tree.*

Related to the above, the following is asked in [**BS97b**]:

QUESTION 16.71. Is it true every Cayley graph of every finitely generated group with exponential growth contains a subtree with positive Cheeger constant?

We note that existence of such subtrees does not contradict amenability, for instance, each finitely generated elementary amenable group G which is not virtually nilpotent contains a rank 2 free subsemigroup and, hence one of the Cayley graphs of G contains a 2-adic subtree.

16.7. Quantitative approaches to non-amenability and weak paradoxality

One can measure "how paradoxical" a group or a group action is *via* the *Tarski numbers.*

DEFINITION 16.72.
1. Given an action of a group G on a set X, and a subset $E \subset X$, which admits a G–paradoxical decomposition, the *Tarski number of the paradoxical decomposition* is the number $k+l$ of pieces of that decomposition (see Definition 15.4).
2. The *Tarski number* $\mathrm{Tar}_G(X,E)$ is the infimum of the Tarski numbers taken over all G-paradoxical decompositions of E. We set $\mathrm{Tar}_G(X,E) = \infty$ in the case when there exists no G-paradoxical decomposition of the subset $E \subset X$. We use the notation $\mathrm{Tar}_G(X)$ for $\mathrm{Tar}_G(X,X)$.
3. We define the *lower Tarski number* $\mathrm{tar}(G)$ *of a group* G to be the infimum of the numbers $\mathrm{Tar}_G(X,E)$ for all the actions $G \curvearrowright X$ and all the non-empty subsets E of X.
4. When $X = G$ and the action is by left multiplication, we denote $\mathrm{Tar}_G(G,G)$ simply by $\mathrm{Tar}(G)$ and we call it the *Tarski number of* G.

[2]Note that in this result no uniform bound on valence is assumed. The definition of the Cheeger constant is considered with the edge boundary.

Thus, for any G, X, $E \subset X$ and an action $G \curvearrowright X$, we have:

$$\mathrm{tar}(G) \leqslant \mathrm{Tar}_G(X, E) \leqslant \mathrm{Tar}_G(X, X),$$

$$\mathrm{tar}(G) \leqslant \mathrm{Tar}(G)$$

Moreover, G is amenable if and only if $Tar(G) = \infty$, while G is supramenable if and only if $\mathrm{tar}(G) = \infty$. Thus, the number $\mathrm{tar}(G)$ measures *weak paradoxality* of G, i.e., the degree of its failure to be supramenable. Similarly, the number $\mathrm{Tar}(G)$ measures *paradoxality* of G, i.e., the degree of its failure to be amenable.

The following theorem was first proved by R. M. Robinson in [**Rob47**]:

THEOREM 16.73. *For $n \geqslant 3$, the Tarski number $\mathrm{Tar}_G(\mathbb{E}^n, \mathbb{B}^n)$ of the unit ball $\mathbb{B}^n \subset \mathbb{E}^n$ (with respect to the action of the group of isometries $G = \mathrm{Isom}(\mathbb{E}^n)$) is 5.*

REMARK 16.74. Of course, one could refine the discussion further and use other cardinal numbers besides the finite ones to quantify nonamenability. We do not pursue this direction here.

PROPOSITION 16.75. *Let G be a group, $G \curvearrowright X$ be an action and $E \subset X$ be a nonempty subset.*

(1) *For each subgroup $H \leqslant G$, $\mathrm{Tar}_G(X, E) \leqslant \mathrm{Tar}_H(X, E)$.*
(2) *The lower Tarski number $\mathrm{tar}(G)$ of every group is at least two. Moreover, $\mathrm{tar}(G) = 2$ if and only if G contains a free two-generated subsemigroup S.*

PROOF. (1) This inequality follows immediately from the fact that for every $E \subset X$, each H-paradoxical decomposition of E is also G-paradoxical.

(2) The fact that every $\mathrm{tar}_G(X, E)$ is at least two is clear from the definition of a paradoxical decomposition. We prove the direct part of the equivalence. Assume that $\mathrm{tar}(G) = 2$. Then there exists an action $G \curvearrowright X$, a subset $E \subset X$ with a decomposition $E = E' \sqcup E''$ and two elements $g', g'' \in G$ such that $g'(E') = E$ and $g''(E'') = E$. Setting $g := (g')^{-1}, h := (h')^{-1}$, we obtain injective maps

$$g : E \to E' \subset E, \quad h : E \to E'' \subset E$$

with disjoint images. Lemma 4.59 implies that g, h generate a rank 2 free subsemigroup in G.

Conversely, if x, y be two elements in G generating the free sub-semigroup S, let S_x be the set of words beginning with x and S_y be the set of words beginning with y. Then $S = S_x \sqcup S_y$, with $x^{-1}S_x = S$ and $y^{-1}S_y = S$. □

REMARK 16.76. R. Grigorchuk constructed in [**Gri87**] examples of finitely generated amenable torsion groups G which are weakly paradoxical, thus answering Rosenblatt's conjecture [**Wag85**, Question 12.9.b]. Every such amenable group G satisfies

$$3 \leqslant \mathrm{tar}(G) < \infty.$$

QUESTION 16.77. What are the possible values of $\mathrm{tar}(G)$ for weakly paradoxical group G? How different can it be from $\mathrm{Tar}(G)$?

We now move on to study values of Tarski numbers $\mathrm{Tar}_G(X)$ and $\mathrm{Tar}(G)$

PROPOSITION 16.78. *For every group action $G \curvearrowright X$ on a nonempty set X we have:*

(1) $\mathrm{Tar}_G(X) \geqslant 4$.

(2) *If G acts freely on X and G contains a free subgroup of rank two, then* $\mathrm{Tar}_G(X) = 4$.

PROOF. (1) Since G acts on X via bijections, in every G-paradoxical decomposition of X one must have $k \geqslant 2$ and $l \geqslant 2$. Thus, the Tarski number $\mathrm{Tar}_G(X)$ is always at least 4.

(2) This statement is the content of Lemma 15.20. □

The next proposition complements Part (2) of Proposition 16.78:

PROPOSITION 16.79. *1. If there exists a G-action $G \curvearrowright X$ for which X admits a G-paradoxical decomposition*

$$X = X' \sqcup X'', \quad X' = X'_1 \sqcup X'_2, \quad X'' = X''_1 \sqcup \ldots \sqcup X''_l, \quad l \geqslant 2, \tag{16.14}$$

then G contains an element of infinite order.

2. If there exists an action $G \curvearrowright X$ with $\mathrm{Tar}_G(X) = 4$, then G contains a non-abelian free subgroup.

PROOF. We let $\phi' : X' \to X$, $\phi'' : X'' \to X$ be piecewise G-congruences from (16.14); they restrict to X'_i as $g_i \in G$, $i = 1, 2$, and to X''_j as h_j, $j = 1, \ldots, l$. We define products

$$g := g_1^{-1} g_2, \quad h := h_1^{-1} h_2.$$

We leave it to the reader to verify that

$$g(X'_1 \sqcup X'') \subset X'_1 \tag{16.15}$$

and, therefore, by relabelling $1 \leftrightarrow 2$,

$$g^{-1}(X'_2 \sqcup X'') \subset X'_2,$$

1. Since X'' is nonempty, (16.15) implies that

$$g(X'_1) \subsetneq X'_1.$$

It now follows from Exercise 4.60 that $g \in G$ has infinite order.

2. Since $\mathrm{Tar}_G(X) = 4$, there exists a G-paradoxical decomposition as in (16.14) with $l = 2$. Since the number of pieces in X' and X'' is now the same, we obtain (by relabelling):

$$h(X''_1 \sqcup X') \subset X''_1$$

and

$$h^{-1}(X''_2 \sqcup X') \subset X''_2.$$

It follows that the pair of elements $g, h \in G$ and the subsets X'_1, X'_2, X''_1, X''_2 satisfies the assumption of the Ping-pong Lemma (Lemma 4.61). Hence, g and h generate a free subgroup of rank 2 in G. □

COROLLARY 16.80. *If G is a torsion group then for every G-action on a set X, $\mathrm{Tar}_G(X) \geqslant 6$.*

PROOF. Suppose that X admits a G-paradoxical decomposition with $k+l$ parts. Part 1 of Lemma 16.79 implies that if either k or l is $\leqslant 2$, then G contains an infinite order element, which contradicts our assumptions. Therefore, $k \geqslant 3, l \geqslant 3$ and $\mathrm{Tar}_G(X) \geqslant 6$. □

S. Wagon (Theorems 4.5 and 4.8 in [**Wag85**]) proved a stronger form of Proposition 16.79 and Proposition 16.78, part (2):

THEOREM 16.81 (S. Wagon). *Let G be a group acting on a nonempty set X. The Tarski number $\mathrm{Tar}_G(X)$ is four if and only if G contains a nonabelian free subgroup F such that the stabilizer in F of each point in X is abelian.*

Below we describe the behavior of the Tarski number of groups with respect to certain group operations.

PROPOSITION 16.82. (1) *If H is a subgroup of G then* $\mathrm{Tar}(G) \leqslant \mathrm{Tar}(H)$.

(2) *Every paradoxical group G contains a finitely generated subgroup H such that* $\mathrm{Tar}(G) = \mathrm{Tar}(H)$.

(3) $\mathrm{Tar}(G) \leqslant \mathrm{Tar}(Q)$ *for every quotient group Q of G.*

PROOF. (1) If H is amenable then there is nothing to prove. Consider an H-paradoxical decomposition of H:

$$H = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_l$$

such that

$$H = g_1 X_1 \sqcup ... \sqcup g_k X_k = h_1 Y_1 \sqcup ... \sqcup h_l Y_l,$$

where g_i, h_j are elements of H and $k + l = \mathrm{Tar}(H)$. Let $S \subset G$ denote the set of representatives of right H–cosets inside G: the product map

$$H \times S \to G, \quad (h, s) \mapsto hs,$$

is a bijection. Then the subsets

$$X_i' = X_i S,\ 1 \leqslant i \leqslant k$$

together with

$$Y_j' = Y_j S,\ 1 \leqslant j \leqslant l,$$

form a paradoxical decomposition of G.

(2) Given a decomposition

$$G = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_l$$

such that

$$G = g_1 X_1 \sqcup ... \sqcup g_k X_k = h_1 Y_1 \sqcup ... \sqcup h_l Y_l \tag{16.16}$$

and $k + l = \mathrm{Tar}(G)$, consider the subgroup $H \leqslant G$ generated by $g_1, ..., g_k, h_1, ..., h_l$. Then (16.16) is an H-paradoxical decomposition of G with respect to the action of H on G *via* left multiplication. Intersecting pieces of this decomposition with H, we obtain

$$G = (H \cap X_1) \sqcup ... \sqcup (H \cap X_k) \sqcup (H \cap Y_1) \sqcup ... \sqcup (H \cap Y_l),$$

an H-paradoxical decomposition of H. Thus $\mathrm{Tar}(H) \leqslant \mathrm{Tar}_H(G, G) \leqslant \mathrm{Tar}(G)$; the opposite inequality is proven in Part 1.

(3) Suppose that $\pi : G \to Q$ is a quotient map with a set-theoretic cross-section $\sigma : Q \to G$. As before, we may assume, without loss of generality, that Q is paradoxical. Let

$$Q = X_1' \sqcup ... \sqcup X_k' \sqcup X_1'' \sqcup ... \sqcup X_l''$$

be a paradoxical decomposition of Q with piecewise congruences given by restrictions of elements $q_1', \dots, q_k', q_1'', \dots, q_l''$ of Q. We assume that $\mathrm{Tar}(Q) = k + l$. Then

$$G = \pi^{-1}(X_1') \sqcup ... \sqcup \pi^{-1}(X_k') \sqcup \pi^{-1}(X_1'') \sqcup ... \sqcup \pi^{-1}(X_l'')$$

is a paradoxical decomposition of G with $k+l$ pieces and with piecewise congruences defined by restrictions of the elements $\sigma(q_i'), 1 \leqslant i \leqslant k$, $\sigma(q_j''), 1 \leqslant j \leqslant l$, of the group G . □

Corollary 16.83. *A group G contains a non-abelian free subgroup if and only if $Tar(G) = 4$.*

Proof. If a group G contains a non-abelian free subgroup then the result follows by Proposition 16.78, (1), (2), and Proposition 16.82, (1). If a group G has $\mathrm{Tar}(G) = 4$ then the claim follows from Proposition 16.79. □

Thus, the Tarski number helps to classify the groups that are non-amenable and do not contain nonabelian free subgroups. This class of groups is not very well understood, its only known members are (small) "infinite monsters". For torsion groups G as we noted above, $\mathrm{Tar}(G) \geqslant 6$. On the other hand, T. Ceccherini, R. Grigorchuk and P. de la Harpe proved:

Theorem 16.84 (Theorem 2, [**CSGdlH98**]). *The Tarski number of every free Burnside group $B(n; m)$ with $n \geqslant 2$ and $m \geqslant 665$, m odd, is at most 14.*

Part (1) of Proposition 16.82 implies the following quantitative version of Corollary 16.29:

Corollary 16.85. *If two groups G_1, G_2 are co-embeddable then they have the same Tarski number:* $\mathrm{Tar}(G_1) = \mathrm{Tar}(G_2)$.

It is proven in [**Šir76**], [**Ady79**, Theorem VI.3.7] that, for every odd $m \geqslant 665$, $n \geqslant 2, k \geqslant 2$, the two free Burnside groups $B(n; m)$ and $B(k; m)$ of exponent m are co-embeddable. Thus:

Corollary 16.86. *For every odd $m \geqslant 665$, and $n \geqslant 2$, the Tarski number of the free Burnside groups $B(n; m)$ is finite and independent of the number of generators n.*

Natural questions, in view of Corollary 16.86, are the following:

Question 16.87. How does the Tarski number of a free Burnside group $B(n; m)$ depend on the exponent m? What are its possible values?

The following natural question, appears as Question 22 [**dlHGCS99**] (asked also in [**CSGdlH98**]):

Question 16.88 (Question 22 [**dlHGCS99**], [**CSGdlH98**]). What are the possible values for the Tarski numbers of groups? Do they include 5 or 6? Are there groups which have arbitrarily large finite Tarski numbers?

The paper by Ershov, Golan and Sapir [**ME**] provided very illuminating answers. Among other things, they prove:

THEOREM 16.89 (Ershov, Golan and Sapir). *1. There exist nonamenable 2-generated torsion groups*[3] *with arbitrarily large Tarski numbers.*

2. There exist groups with Tarski numbers 5 and 6.

3. There exist a nonamenable group G and a subgroup $H < G$ of finite index such that $\mathrm{Tar}(H) \neq \mathrm{Tar}(G)$.

4. If $H < G$ has finite index then $\mathrm{Tar}(H)?2 \leqslant |G : H|(\mathrm{Tar}(G)?2)$.

It would be still interesting to understand how much of the Tarski number is encoded in the large scale geometry of a group. In particular:

QUESTION 16.90. Is it at least true that the existence of an (L, C)-quasiisometry between groups (with fixed finite generating sets) implies that their Tarski number differ at most by a constant $K = K(L, A)$?

DEFINITION 16.91. A group G is called *small* if it contains no nonabelian free subgroups. Thus, G is small iff $\mathrm{Tar}(G) > 4$. Accordingly, a group is called *large* if it contains a nonabelian free subgroup. Dually, a group G is *co-large* if it contains a finite index subgroup $\Gamma' \leqslant \Gamma$, which admits an epimorphism to a nonabelian free group.

QUESTION 16.92. Is smallness invariant under quasiisometries of finitely generated groups? Is co-largeness a QI invariant for hyperbolic groups?

Note that co-largeness is not a QI invariant for finitely generated (and even $CAT(0)$) groups. The simplest examples of this phenomenon are given by uniform lattices acting on $\mathbb{H}^2 \times \mathbb{H}^2$: Among such lattices there are product groups $\Gamma = \Gamma_1 \times \Gamma_2$, where both Γ_1, Γ_2 are surface groups, as well as *irreducible lattices* Γ'. The groups Γ, Γ' are quasiisometric to each other, but Γ is co-large, while Γ_2 is not (as follows from Margulis' Superrigidity Theorem).

16.8. Uniform amenability and ultrapowers

In this section we discuss a *uniform version of amenability* and its relation to ultrapowers of groups.

Recall (Definition 16.42) that a group G has *the Følner Property* if for every finite subset K of G and every $\epsilon \in (0, 1)$ there exists a finite non-empty subset $F \subset G$ satisfying:

$$|KF \triangle F| < \epsilon |F|.$$

DEFINITION 16.93. A group G has the *uniform Følner Property* if, in addition, one can bound the size of F in terms of ϵ and $|K|$, i.e. there exists a function $\phi : (0, 1) \times \mathbb{N} \to \mathbb{N}$ such that for each $\epsilon \in (0, 1)$ and a finite subset $K \subset G$, there exists a finite subset $F \subset G$ satisfying the inequality

$$|F| \leqslant \phi(\epsilon, |K|)\,.$$

EXAMPLES 16.94. (1) Nilpotent groups have the uniform Følner property, [**Boż80**].

[3] In fact, these groups even satisfy the Property T.

(2) A subgroup of a group with the uniform Følner Property also has this property, [**Boż80**].
(3) Let N be a normal subgroup of G. The group G has the uniform Følner Property if and only if both N and G/N have this property, [**Boż80**].

THEOREM 16.95 (G. Keller [**Kel72**], [**Wys88**]). *The following are equivalent:*
(1) *G has the uniform Følner Property.*
(2) *For some non-principal ultrafilter ω the ultrapower G^ω has the Følner Property.*
(3) *For every non-principal ultrafilter ω, the ultrapower G^ω also has the uniform Følner property.*

PROOF. The implication (3) $\Rightarrow$ (2) is clear, we prove the two other implications.

(2) $\Rightarrow$ (1). We identify the group G with the "diagonal" subgroup $\widehat{G}$ of G^ω, represented by constant sequences in G. It follows from Proposition 16.45 that G has the Følner property. Assume that G does not have the uniform Følner property. Then there exists $\varepsilon > 0$ and a sequence of subsets K_n in G of fixed cardinality k such that for every sequence of subsets $\Omega_n \subset G$

$$|K_n\Omega_n \bigtriangleup \Omega_n| < \epsilon|\Omega_n| \Rightarrow \lim_{n\to\infty} |\Omega_n| = \infty.$$

Let $K_\omega = (K_n)^\omega$. According to Lemma 7.35, K has cardinality k. Since G^ω satisfies the Følner property, there exists a finite subset $U \in G^\omega$ such that

$$|KU \bigtriangleup U| < \epsilon|U|.$$

Let c denote the cardinality of U. According to Lemma 7.35, Part (3), $U = (U_n)^\omega$, where each $U_n \subset G$ has cardinality c. Moreover, ω-almost surely

$$|KU_n \bigtriangleup U_n| < \epsilon|U_n|.$$

Contradiction. We, therefore, conclude that G has the uniform Følner Property.

(1) $\Rightarrow$ (3). Let $k \in \mathbb{N}$ and $\epsilon > 0$; define $m := \phi(\epsilon, k)$ where ϕ is the function coming from the uniform Følner property of G. Let K be a subset of cardinality k in G^ω. Lemma 7.35 implies that $K = (K_n)^\omega$, for some sequence of subsets $K_n \subset G$ of cardinality k. The uniform Følner Property of G implies that there exists Ω_n of cardinality at most m such that

$$|K_n\Omega_n \bigtriangleup \Omega_n| < \epsilon|\Omega_n|.$$

Let $F := (\Omega_n)^\omega$. The description of K and F given by Lemma 7.35, (1), implies that

$$KF \bigtriangleup F = (K_n\Omega_n \bigtriangleup \Omega_n)^\omega,$$

whence $|KF \bigtriangleup F| < \epsilon|\Phi|$. Since $|\Phi| \leqslant m$ according to Lemma 7.35, Part (1), the claim follows. □

COROLLARY 16.96 (G. Keller, [**Kel72**], Corollary 5.9). *Every group with the uniform Følner property satisfies a law.*

PROOF. Indeed, by Theorem 16.95, if G has the uniform Følner Property then every ultrapower G^ω has the uniform Følner Property. Assume that G does not satisfy any law, i.e., in view of Lemma 7.42, the group G^ω contains a subgroup isomorphic to the free group F_2. By Proposition 16.45 it would then follow that F_2 has the Følner Property, a contradiction. □

Example 16.97. Let $H = H_n, n \geqslant 3$ be the nth Houghton group, see Example 16.55. Then H is finitely presented, amenable and each finite group embeds into H. We claim that H cannot satisfy any law. Indeed, if H did satisfy a law $w(x_1, \ldots, x_n) = 1$ then all finite groups would satisfy this law. Then the direct product

$$G = \prod_{\Phi \in \mathcal{F}} \Phi$$

would satisfy the same law. (Here $\mathcal{F}$ denotes the set of isomorphism classes of all finite groups.) All subgroups of G would satisfy this law as well. However, since the free group F_2 is residually finite, it embeds in G. A contradiction. Therefore, H is amenable but not uniformly amenable.

16.9. Quantitative approaches to amenability

One quantitative approach to amenability (of finitely generated groups and of graphs of finite valence) is due to A.M. Vershik, who introduced in **[Ver82]** the notion of a *Følner function*. Given an amenable graph $\mathcal{G}$ of bounded geometry, its *Følner function* $\mathrm{F}_o^{\mathcal{G}} : (0, \infty) \to \mathbb{N}$ is defined by the condition that $\mathrm{F}_o^{\mathcal{G}}(t)$ is the minimal cardinality of a finite non-empty subset $F \subset V(\mathcal{G})$ of satisfying the inequality

$$|\partial_V F| \leqslant \frac{1}{t}|F|\,.$$

According to the inequality (1.4) relating the cardinalities of the vertex and edge boundaries, if one replaces in this definition $\partial_V F$ with $E(F, F^c)$ or $\partial^V F$, one obtains a function asymptotically equal to the first, in the sense of Definition 1.4.

The following is a quantitative version of Theorem 16.12 (which establishes quasiisometry invariance of amenability for graphs).

Proposition 16.98. *If two graphs of bounded geometry are quasiisometric then their Følner functions are asymptotically equal.*

Proof. Let $\mathcal{G}$ and $\mathcal{G}'$ be two graphs of bounded geometry, and let $f : \mathcal{G} \to \mathcal{G}'$ and $g : \mathcal{G}' \to \mathcal{G}$ be two (L, C)–quasiisometries such that $f \circ g$ and $g \circ f$ are at distance at most C from the respective identity maps (in the sense of the inequalities (5.3)). Without loss of generality we may assume that both f and g send vertices to vertices. Let m be the maximal valency of a vertex in both $\mathcal{G}$ and $\mathcal{G}'$.

We begin by some general considerations. We denote by α the maximal cardinality of $B(x, C) \cap V$, where $B(x, C)$ is an arbitrary ball of radius C in either $\mathcal{G}$ or $\mathcal{G}'$. Since both graphs have bounded geometry, it follows that α is finite.

Let A be a finite subset in $V(\mathcal{G})$, let $A' = f(A)$ and $A'' = g(A')$. It is obvious that $|A''| \leqslant |A'| \leqslant |A|$. The Hausdorff distance between A'' and A is at most C, and, therefore, $|A| \leqslant \alpha |A''|$. Thus we have the inequalities

$$\frac{1}{\alpha}|A| \leqslant |f(A)| \leqslant |A|\,, \tag{16.17}$$

and similar inequalities for finite subsets in $V(\mathcal{G}')$ and their images by g.

Assume now that both $\mathcal{G}$ and $\mathcal{G}'$ are amenable, and let $\mathrm{F}_o^{\mathcal{G}}$ and $\mathrm{F}_o^{\mathcal{G}'}$ be their respective Følner functions. Without loss of generality we assume that both Følner functions are defined using the vertex boundary.

Fix $t \in (0,\infty)$, and let A be a finite subset in $V(\mathcal{G})$ such that $|A| = \mathrm{F}_o^{\mathcal{G}}(t)$ and

$$|\partial_V A| \leqslant \frac{1}{t}|A|\,.$$

Let $A' = f(A)$ and $A'' = g(A')$. We fix the constant $R = L(2C+1)$, and consider the set $B = \mathcal{N}_R(A')$. The vertex-boundary $\partial_V(B)$ consists of vertices u such that $R \leqslant \mathrm{dist}(u, A') < R+1$.

It follows that

$$\mathrm{dist}(g(u), A) \geqslant \mathrm{dist}(g(u), A'') - C \geqslant \frac{1}{L}R - 2C = 1$$

and that

$$\mathrm{dist}(g(u), A) \leqslant L(R+1) + C\,.$$

In particular, every vertex $g(u)$ is at distance at most $L(R+1)+C-1$ from $\partial_V(A)$ and it is not contained in A. We have, thus, proved that

$$g\left(\partial_V(B)\right) \subseteq \mathcal{N}_{L(R+1)+C-1}\left(\partial_V A\right) \setminus A\,.$$

It follows that if we denote $m^{L(R+1)+C-1}$ by λ, then we can write, using (16.17),

$$|\partial_V B| \leqslant \alpha\,|g\left(\partial_V B\right)| \leqslant \alpha\lambda\,|\partial_V(A)| \leqslant \alpha\lambda\frac{1}{t}|A| \leqslant$$

$$\alpha^2\lambda\frac{1}{t}|A'| \leqslant \alpha^2\lambda\frac{1}{t}|B|\,.$$

We have thus obtained that, for $\kappa = \alpha^2\lambda$ and every $t > 0$, the value $\mathrm{F}_o^{\mathcal{G}'}\left(\frac{t}{\kappa}\right)$ is at most $|B| \leqslant m^R|A'| \leqslant m^R|A| = m^R\,\mathrm{F}_o^{\mathcal{G}}(t)$. We conclude that $\mathrm{F}_o^{\mathcal{G}'} \preceq \mathrm{F}_o^{\mathcal{G}}$.

The opposite inequality $\mathrm{F}_o^{\mathcal{G}} \preceq \mathrm{F}_o^{\mathcal{G}'}$ is obtained by relabelling. □

Proposition 16.98 implies that, given a finitely generated amenable group G, any two of its Cayley graphs have asymptotically equal Følner functions. We will, therefore, write F_o^{G}, for the equivalence class of all these functions.

DEFINITIONS 16.99. (1) We say that a sequence (Φ_n) of finite subsets in a group *realizes the Følner function* of that group if for some generating set S, $|\Phi_n| = \mathrm{F}_o^{\mathcal{G}}(n)$, where $\mathcal{G}$ is the Cayley graph of G with respect to S, and

$$|E(\Phi_n, \Phi_n^c)| \leqslant \frac{1}{n}\,|\Phi_n|\,.$$

(2) We say that a sequence (A_n) of finite subsets in a group *quasirealizes the Følner function* of that group if $|A_n| \asymp \mathrm{F}_o^{G}(n)$ and there exists a constant $a > 0$ and a finite generating set S such that for every n,

$$|E(A_n, A_n^c)| \leqslant \frac{a}{n}\,|A_n|\,,$$

where $|E(A_n, A_n^c)|$ is the edge boundary of A_n in the Cayley graph of G with respect to S.

LEMMA 16.100. *Let H be a finitely generated subgroup of a finitely generated amenable group G. Then $\mathrm{F}_o^{H} \preceq \mathrm{F}_o^{G}$.*

PROOF. Consider a generating set S of G containing a generating set T of H. We shall prove that the Følner functions defined with respect to these generating sets, satisfy the inequality

$$\mathrm{F}_o^H(t) \leqslant \mathrm{F}_o^G(t)$$

for every $t > 0$. Let F be a finite subset in G such that $|\Phi| = \mathrm{F}_o^G(t)$ and $|\partial_V F| \leqslant \frac{1}{t}|F|$.

The set F intersects finitely many cosets of H, $g_1H, \ldots, g_kH$. In particular $F = \bigsqcup_{i=1}^k \Phi_i$, where $\Phi_i = F \cap g_iH$. We denote by $\partial_V^i \Phi_i$ the set of vertices in $\partial_V \Phi_i$ joined to vertices in Φ_i by edges with labels in X. The sets $\partial_V^i \Phi_i$ are contained in g_iH for every $i \in \{1, 2, \ldots, k\}$, hence, they are pairwise disjoint subsets of $\partial_V F$. We, thus, obtain:

$$\sum_{i=1}^k \left|\partial_V^i \Phi_i\right| \leqslant |\partial_V F| \leqslant \frac{1}{x}|F| = \frac{1}{t}\sum_{i=1}^k |\Phi_i|.$$

It follows that there exists $i \in \{1, 2, \ldots, k\}$ such that

$$\left|\partial_V^i \Phi_i\right| \leqslant \frac{1}{x}|\Phi_i|.$$

By the construction, $\Phi_i = g_iK_i$ with $K_i \subset H$, and the previous inequality is equivalent to

$$|\partial_V K_i| \leqslant \frac{1}{t}|K_i|,$$

where the vertex-boundary $\partial_V K_i$ is considered in the Cayley graph of H with respect to the generating set T. We then conclude that

$$\mathrm{F}_o^H(t) \leqslant |K_i| \leqslant |\Phi| = \mathrm{F}_o^G(t).$$

□

One may ask how do the Følner functions relate to the growth functions, and when do the sequences of balls of fixed centre quasirealize the Følner function, especially under the extra hypothesis of subexponential growth, see Proposition 16.6.

THEOREM 16.101. *Let G be an infinite finitely generated group. Then:*

1. $\mathrm{F}_o^G(t) \succeq \mathfrak{G}_G(t)$.
2. *If a sequence of balls $B(1,t)$ quasirealizes the Følner function of G then G is virtually nilpotent.*

PROOF. (1) Consider a sequence (Φ_n) of finite subsets in G that realizes the Følner function of that group (for some generating set X). In particular

$$|E(\Phi_n, \Phi_n^c)| \leqslant \frac{1}{n}|\Phi_n|.$$

We let $\mathfrak{G}$ denote the growth function of G with respect to the generating set X.

The inequality (5.10) in Proposition 5.89 implies that

$$\frac{|\Phi_n|}{2dk_n} \leqslant \frac{1}{n}|\Phi_n|,$$

where $d = |S|$ and k_n is such that $\mathfrak{G}(k_n - 1) \leqslant 2|\Phi_n| < \mathfrak{G}(k_n - 1)$. Therefore,

$$k_n - 1 \geqslant \frac{n}{2d} - 1 \geqslant \frac{n}{4d},$$

whence,

$$2\,\mathrm{F}_o^G(n) \geqslant \mathfrak{G}\,(k_n-1) \geqslant \mathfrak{G}\left(\frac{n}{4d}\right).$$

(2) The inequality in Part (2) of Definition 16.99 implies that for some $a>0$,

$$|S(1,n+1)| \leqslant \frac{a}{n}\,|B(1,n)|\ ,$$

where $S(1,k)$ is the sphere of radius k centered at $1\in G$. In terms of the growth function, this inequality can be re-written as

$$\frac{\mathfrak{G}(n+1)-\mathfrak{G}(n)}{\mathfrak{G}(n)} \leqslant \frac{a}{n}\,. \tag{16.18}$$

Let $f(t)$ be the piecewise-linear function on $\mathbb{R}_+$ whose restriction to $\mathbb{N}$ equals $\mathfrak{G}$ and which is linear on every interval $[n,n+1], n\in\mathbb{N}$. Then the inequality (16.18) means that for all $t\notin\mathbb{N}$,

$$\frac{f'(t)}{f(t)} \leqslant \frac{a}{t}\,.$$

which, by integration, implies that

$$\log|f(t)| \leqslant a\log|t|+b$$

and, hence,

$$f(t)\leqslant bt^a.$$

In particular, it follows that $\mathfrak{G}(t)$ is bounded by a polynomial in t, whence, G is virtually nilpotent by Theorem 14.1. □

In view of Theorem 16.101, (1), one may ask if there is a general upper bound for the Følner functions of a group, same as the exponential function is a general upper bound for the growth functions; related to this, one may ask how much can the Følner function and the growth function of a group differ. This question is addressed in two papers of A. Erschler:

The first theorem shows that one cannot have an exponential upper bound for Følner functions:

<u>Theorem</u> 16.102 (A. Erschler, [**Ers03**]). *Let G and H be two amenable groups and assume that some representative F of F_o^H has the property that for every $a>0$ there exists $b>0$ so that $a\mathrm{F}(t)<\mathrm{F}(bt)$ for every $t>0$. Then the Følner function of the wreath product $G\wr H$ is asymptotically equal to $[\mathrm{F}_o^B(x)]^{\mathrm{F}_o^A(x)}$.*

The second theorem shows that there are no upper bounds for Følner functions whatsoever:

<u>Theorem</u> 16.103 (A. Erschler, [**Ers06**]). *For every function $f:\mathbb{N}\to\mathbb{N}$, there exists a finitely generated group G, which is a subgroup of a group of intermediate growth (in particular, G is amenable), whose Følner function satisfies $\mathrm{F}_o^G(n)\geqslant f(n)$ for all sufficiently large n.*

16.10. Summary of equivalent definitions of amenability

Below we present a (very much incomplete) list of equivalent definitions of amenability; most of these are theorems stated or proven earlier in this chapter, the exceptions are the characterization in terms of bounded cohomology groups and of measure-equivalence. In order to streamline the discussion, G is assumed to be an infinite finitely generated group equipped with a word metric.

(1) G is amenable iff it admits an invariant mean.
(2) G is amenable iff it admits a Følner sequence.
(3) A finitely presented group G is amenable iff G is the fundamental group of a closed Riemannian manifold whose universal comer $\tilde{M}$ satisfies $\lambda_1(\tilde{M}) = 0$ (Theorem 16.13).
(4) G is amenable iff for all Banach $\mathbb{Z}G$-modules V and all $n \geq 1$, $H_b^n(G, V) = 0 \iff \forall V, H_b^1(G, V) = 0$ (G. Noskov, [**Nos91**]).
(5) G is amenable iff it is non-paradoxical, i.e., if $\mathrm{Tar}(G) = \infty$ (Theorem 16.48).
(6) G is amenable iff it is measure-equivalent to $\mathbb{Z}$ (D. Ornstein and B. Weiss, [**OW80**]).
(7) G is nonamenable iff there exists a constant $C > 0$ such that for every finite non-empty subset $F \subset G$, the set $\overline{\mathcal{N}}_C(\Phi) \subset G$ contains at least twice as many vertices as F (Theorem 16.4).
(8) G is nonamenable iff there exists a map $f \in \mathcal{B}(G)$ such that for every $v \in V$ the pre-image $f^{-1}(v)$ contains at least two elements (Theorem 16.4).
(9) G is nonamenable iff there exists a map $f \in \mathcal{B}(G)$ such that for every $v \in V$ the pre-image $f^{-1}(v)$ contains exactly two elements (Theorem 16.4).
(10) G is nonamenable iff its Cayley graph $\mathcal{G}$ has spectral radius $\rho(\mathcal{G}) \geqslant 1$ (Theorem 16.10).

16.11. Amenable hierarchy

We conclude this chapter with the following diagram illustrating hierarchy of amenable groups:

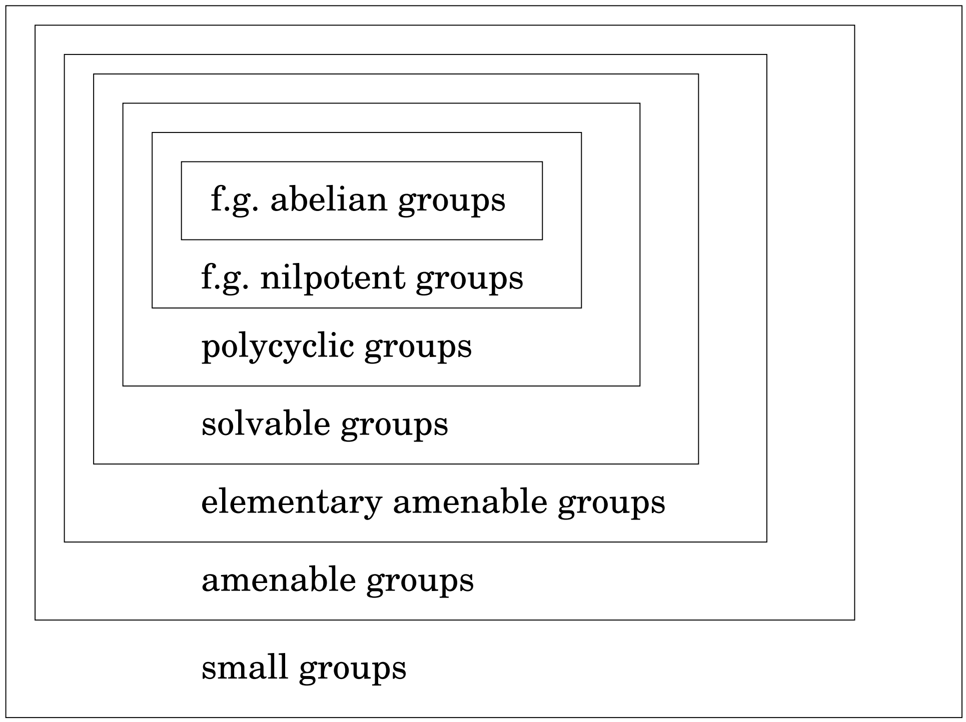


FIGURE 16.1. The hierarchy of amenable groups

CHAPTER 17

Ultralimits and fixed point properties

In this section we discuss groups satisfying various fixed-point properties, most notably, groups satisfying Kazhdan's Property T. The latter can be regarded as a strong negation of amenability for groups.

17.1. Classes of spaces stable with respect to (rescaled) ultralimits

DEFINITION 17.1. *Consider a class $\mathcal{C}$ of metric spaces. We say that $\mathcal{C}$ is* stable with respect to ultralimits *if for every set of indices I, every nonprincipal ultrafilter ω on I, every collection $(X_i, \mathrm{dist}_i)_{i\in I}$ of metric spaces in $\mathcal{C}$ and every sets of base-points $(e_i)_{i\in I}$ with $e_i \in X_i$, the ultralimit ω-lim$(X_i, e_i, \mathrm{dist}_i)$ is isometric to a metric space in $\mathcal{C}$.*

We say that $\mathcal{C}$ is stable with respect to rescaled ultralimits *if for every choice of I, ω, $(X_i, \mathrm{dist}_i)_{i\in I}$ and $(e_i)_{i\in I}$ as above, and, moreover, every indexed set of positive real numbers $(\lambda_i)_{i\in I}$, the ultralimit of rescaled spaces ω-lim$(X_i, e_i, \lambda_i \mathrm{dist}_i)$ is isometric to a metric space in $\mathcal{C}$.*

Note that in this definition we are not making any assumptions about the limits ω-$\lim \lambda_i$; in particular, they are allowed to be zero and ∞.

EXAMPLE 17.2. The class of $CAT(0)$ spaces is stable with respect to rescaled ultralimits.

Since in a normed vector space V the scaling $x \mapsto \lambda x$, $\lambda \in \mathbb{R}_+$, scales the metric by λ, the metric space $(V, \lambda\mathrm{dist})$ is isometric to (V, dist), where $\mathrm{dist}(u, v) = \|u - v\|$. Therefore, taking rescaled ultralimits of normed spaces is the same as taking their ultralimits.

In this section we show that certain classes of Banach spaces are stable with respect to ultralimits. It is easy to see that ultralimits of Banach spaces are Banach spaces. Below, we will see that within the class of Banach spaces, Hilbert spaces and L^p–spaces can be distinguished by properties that are preserved under ultralimits. The main references for this section are [**LT79**], [**Kak41**] and [**BDCK66**].

CONVENTION 17.3. 1. Unless otherwise stated, for every ultralimit of Banach spaces, the base-points are the zero vectors. This assumption is harmless since translations of Banach spaces are isometries.

2. We do not assume Hilbert spaces to be separable.

THEOREM 17.4 (Jordan–von Neumann [**JvN35**].). *A (real or complex) Banach space $(X, \| \ \|)$ is Hilbert (i.e., the norm $\| \ \|$ comes from an inner product) if and only if every pair of vectors $x, y \in X$ satisfies the* parallelogram identity*:*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 .$$

Proof. We claim that the inner/hermitian product on X is given by the formula:

$$(x,y) := \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) = \frac{1}{4}\sum_{k=0}^{1}(-1)^k\|x+(-1)^k y\|^2, \quad \text{if } X \text{ is real}$$

and

$$(x,y) := \frac{1}{4}\sum_{k=0}^{3} i^k\|x+i^k y\|^2, \quad \text{if } X \text{ is complex,}$$

where $i = \sqrt{-1}$.

Note that it is clear that $(x,x) = \|x\|^2$ (real case), $(x,\bar{x}) = \|x\|^2$ (complex case). We will verify that $(\cdot,\cdot)$ is a hermitian inner product in the complex case; the real case is similar and is left to the reader. We leave it to the reader to show that

$$(ix,y) = (x,-iy) = i(x,y), \quad (x,y) = \overline{(y,x)}$$

and that the parallelogram identity implies the equality

$$\|u+v\|^2 = -\|u\|^2 + \frac{1}{2}\|v\|^2 + 2\|u+\frac{1}{2}v\|^2. \tag{17.1}$$

By the definition of $(\cdot,\cdot)$, we have:

$$4(x/2,z) = \sum_{k=0}^{3} i^k\left\|\frac{x}{2}+i^k z\right\|^2 =$$

(by applying the equation (17.1) to each term of this sum)

$$\sum_{k=0}^{3} i^k(2\left\|\frac{x}{2}+i^k\frac{z}{2}\right\|^2 + \left\|i^k\frac{z}{2}\right\|^2) - \|x/2\|^2) =$$

(again, by the definition of $(\cdot,\cdot)$)

$$\sum_{k=0}^{3} i^k(2\left\|\frac{x}{2}+i^k\frac{z}{2}\right\|^2 + \left\|i^k\frac{z}{2}\right\|^2) = 2(x,z).$$

Thus, $(x/2,z) = (x,z)$ and, clearly,

$$(2x,z) = 2(x,z) \tag{17.2}$$

By the symmetry of $(\cdot,\cdot)$ it follows that

$$(x,2z) = 2(x,z). \tag{17.3}$$

Instead of proving the multiplicative property for $(\cdot,\cdot)$ for all scalars, we now prove the additivity property of $(\cdot,\cdot)$.

By the definition of $(\cdot,\cdot)$, we have

$$4(x+y,z) = \sum_{k=0}^{3}\left\|(x+y)+i^k z\right\|^2 =$$

(by applying the parallelogram to each term of this sum)

$$\sum_{k=0}^{3} i^k(2\left\|x+i^k(z/2)\right\|^2 + \left\|y+i^k(z/2)\right\|^2) - \|x-y\|^2) =$$

$$\sum_{k=0}^{3} i^k(2\left\|x+i^k(z/2)\right\|^2+\left\|y+i^k(z/2)\right\|^2)=8(x,z/2)+8(y,z/2)=$$

(by applying (17.3))

$$4(x,z)+4(y,z).$$

Thus, $(x+y,z)=(x,z)+(y,z)$.

By applying the additivity property of $(\cdot,\cdot)$ inductively, we obtain

$$(nx,y)=n(x,y), \forall n\in\mathbb{N}.$$

For every $n\in\mathbb{N}$ we also have

$$(x,y)=(n\frac{1}{n}x,y)=n(\frac{1}{n}x,y)\Rightarrow(\frac{1}{n}x,y)=\frac{1}{n}(x,y).$$

Combined with the additivity property, this implies that $(rx,y)=r(x,y)$ holds for all $r\in\mathbb{Q}, r\geqslant 0$. Since $(-x,y)=-(x,y)$, we have the same multiplicative identity for all $r\in\mathbb{Q}$. Note that so far we did not use the triangle inequality in X. Observe that the triangle inequality in X implies that for all $x,y\in X$ the function

$$t\mapsto(tx,y)=\frac{1}{4}\left(\|tx+y\|^2-\|tx-y\|^2\right)$$

is continuous. Continuity implies that the identity $(tx,y)=t(x,y)$ holds for all $t\in\mathbb{Q}$. Hence, by the symmetry of $(\cdot,\cdot)$, it follows that (x,y) is indeed an inner product on X. $\square$

COROLLARY 17.5. *Every ultralimit of a sequence of Hilbert spaces is a Hilbert space.*

EXERCISE 17.6. Every closed linear subspace of a Hilbert space is a Hilbert space.

A key feature of Banach function spaces is the existence of an order relation satisfying certain properties with respect to the algebraic operations and the norm.

DEFINITION 17.7. A *Banach lattice* $(X\,,\|\ \|,\leq)$ is a real Banach space $(X\,,\|\ \|)$ endowed with a partial order $\leqslant$ such that:

1. if $x\leqslant y$ then $x+z\leqslant y+z$ for every $x,y,z\in X$;
2. if $x\geqslant 0$ and λ is a non-negative real number then $\lambda x\geqslant 0$;
3. for every x,y in X there exists a *least upper bound (l.u.b)*, denoted by $x\vee y$, and a *greatest lower bound (g.l.b)*, denoted by $x\wedge y$; this allows to define *the absolute value* of a vector $|x|=x\vee(-x)$;
4. if $|x|\leqslant|y|$ then $\|x\|\leqslant\|y\|$.

REMARKS 17.8. (a) It suffices to require the existence of one of the two bounds in Definition 17.7, (3). Either the relation $x\vee y+x\wedge y=x+y$ or the relation $x\wedge y=-[(-x)\vee(-y)]$ allows to deduce the existence of one bound from the existence of the other.

(b) The conditions (1), (2) and (3) in Definition 17.7 imply that

$$|x-y|=|x\vee z-y\vee z|+|x\wedge z-y\wedge z|\,. \tag{17.4}$$

This and condition (4) imply that both operations $\vee$ and $\wedge$ on X are continuous.

(c) Condition (4) applied to $x = u$ and $y = |u|$, and to $x = |u|$ and $y = u$ imply that $\|u\| = \| \, |u| \, \|$.

DEFINITION 17.9. A *sublattice* in a Banach lattice $(X, \| \, \|, \leqslant)$ is a linear subspace Y of X such that if y, y' are elements of Y then $y \vee y'$ is in Y (hence $y \wedge y' = y + y' - y \vee y'$ is also in Y).

DEFINITION 17.10. Two elements $x, y \in X$ of a Banach lattice are called *disjoint* if $x \wedge y = 0$.

EXERCISE 17.11. Prove that:

(1) For every $p \in [1, \infty)$ and every measure space (X, μ), the space $L^p(X, \mu)$ with the order defined by
$$f \leqslant g \;\Leftrightarrow f(x) \leqslant g(x), \;\mu\text{-almost surely},$$
is a Banach lattice.

(2) For every compact Hausdorff topological space X, the space $C(K)$ of continuous functions on X with the pointwise partial order and the sup-norm is a Banach lattice.

(3) For both (1) and (2) prove that two functions are disjoint in the sense of Definition 17.10 if and only if both are non-negative functions with disjoint supports (up to a set of measure zero in the first case).

DEFINITION 17.12. Two Banach lattices X, Y are *order isometric* if there exists a linear isometry $T : X \to Y$ which is also an order isomorphism. Such T is called an *order isometry.*

PROPOSITION 17.13 (Ultralimits of Banach lattices). *Any ultralimit of Banach lattices is a Banach lattice.*

PROOF. Let $(X_i, \| \, \|_i, \leqslant_i), i \in I$, be a family of Banach lattices and let ω be a nonprincipal ultrafilter on I. Consider the ultralimit X_ω endowed with the limit norm $\| \, \|_\omega$. We will need:

LEMMA 17.14. *Suppose that $a_i, b_i \in X_i$ are such that $u = \omega\text{-}lim\, a_i = \omega\text{-}lim\, b_i$. Then $u = \omega\text{-}lim(a_i \vee b_i) = \omega\text{-}lim(a_i \wedge b_i)$.*

PROOF. Equation (17.4) and Definition 17.7, (4), imply that
$$|x - y| \geqslant |x \vee z - y \vee z| \text{ and } |x - y| \geqslant |x \wedge z - y \wedge z|.$$
These inequalities applied to $x = a_i$ and $y = z = b_i$ imply that $|a_i \vee b_i - b_i| \leqslant |a_i - b_i|$ and $|a_i \wedge b_i - b_i| \leqslant |a_i - b_i|$. Part (4) of Definition 17.7 concludes the proof. □

We define on X_ω a relation $\leqslant$ as follows:

Points $u, v \in X_\omega$ satisfy $u \leqslant v$ if and only if there exist representatives $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ of u and v (i.e., $u = \omega\text{-lim}\, x_i$ and $v = \omega\text{-lim}\, y_i$) such that $x_i \leqslant y_i$ ω-almost surely.

We now verify that $\leqslant$ is an order. Reflexivity of $\leqslant$ is obvious. Let us check anti-symmetry. If $u \leqslant v$ and $v \leqslant u$ then we can write these vectors as
$$u = \omega\text{-lim}\, x_i = \omega\text{-lim}\, x_i'$$
and
$$v = \omega\text{-lim}\, y_i = \omega\text{-lim}\, y_i',$$

so that ω-almost surely $x_i \leqslant y_i$ and $y'_i \leqslant x'_i$. The vectors $z_i = x_i - y'_i$ satisfy the inequalities $z_i \leqslant y_i - y'_i$ and $-z_i \leqslant x'_i - x_i$. This implies that

$$|z_i| \leqslant (y_i - y'_i) \vee (x'_i - x_i) \leqslant |y_i - y'_i| \vee |x'_i - x_i| \leqslant |y_i - y'_i| + |x'_i - x_i| .$$

Property (4) in Definition 17.7, the triangle inequality and Remark 17.8, (c), imply that

$$\|z_i\| \leqslant \|y_i - y'_i\| + \|x'_i - x_i\|.$$

It follows that $\omega\text{-lim}\, z_i = 0$, hence $u = v$.

We now check transitivity. Consider vectors

$$u = \omega\text{-lim}\, x_i \,,\ v = \omega\text{-lim}\, y_i = \omega\text{-lim}\, y'_i \,,\ w = \omega\text{-lim}\, z_i$$

such that ω-almost surely $x_i \leqslant y_i$ and $y'_i \leqslant z_i$. Then

$$x_i \leqslant z_i + y_i - y'_i \,.$$

Since $w = \omega\text{-lim}(z_i + y_i - y'_i)$, it follows that $u \leqslant w$.

We will now verify that X_ω is a Banach lattice with respect to the order $\leqslant$. Properties (1) and (2) in Definition 17.7 are immediate.

Given $u = \omega\text{-lim}\, x_i$ and $v = \omega\text{-lim}\, y_i$ define $u \vee v$ as $\omega\text{-lim}(x_i \vee y_i)$. We claim that $u \vee v$ is well–defined, i.e., does not depend on the choice of representatives for u and v. Indeed, assume that $u = \omega\text{-lim}\, x'_i$ and $v = \omega\text{-lim}\, y'_i$ and take $w = \omega\text{-lim}(x_i \vee y_i)$ and $w' = \omega\text{-lim}(x'_i \vee y'_i)$. Let $a_i = x_i \wedge x'_i$ and $A_i = x_i \vee x'_i$; likewise, $b_i = y_i \wedge y'_i$ and $B_i = y_i \vee y'_i$. Clearly,

$$\omega\text{-lim}(a_i \vee b_i) \leqslant w \leqslant \omega\text{-lim}(A_i \vee B_i) \,,$$

and the same for w'. The inequalities

$$a_i \vee b_i \leqslant A_i \vee B_i \leqslant a_i \vee b_i + A_i - a_i + B_i - b_i$$

imply that

$$\omega\text{-lim}(a_i \vee b_i) = \omega\text{-lim}(A_i \vee B_i)$$

and, therefore, $w = w'$. We conclude that the vector $u \vee v = \omega\text{-lim}(x_i \vee y_i)$ is well–defined. Clearly, $u \vee v \geqslant u$ and $u \vee v \geqslant v$. We need to verify that $u \vee v$ is the least upper bound for the vectors u, v.

Suppose that $z \geqslant u$, $z \geqslant v$, where $u = \omega\text{-lim}\, x_i$, $v = \omega\text{-lim}\, y_i$ and $z = \omega\text{-lim}\, z_i = \omega\text{-lim}\, z'_i$ such that ω-almost surely $z_i \geqslant x_i$ and $z'_i \geqslant y_i$. Lemma 17.14 implies that $z = \omega\text{-lim}(z_i \vee z'_i)$ and $z_i \vee z'_i \geqslant x_i \vee y_i$, whence, $z \geqslant (u \vee v)$.

Consider now $u, v \in X_\omega$ such that $|u| \leqslant |v|$. It follows that $u = \omega\text{-lim}\, x_i = \omega\text{-lim}\, x'_i$ and $v = \omega\text{-lim}\, y_i = \omega\text{-lim}\, y'_i$, where

$$x_i \vee (-x'_i) \leqslant y_i \vee (-y'_i).$$

Therefore:

$$|x_i| = x_i \vee (-x_i) \leqslant x_i \vee (-x'_i) + |x_i - x'_i| \leqslant |y_i| + |y_i - y'_i| + |x_i - x'_i| .$$

This inequality, part (4) of Definition 17.7 and the triangle inequality imply that

$$\|x_i\| \leqslant \|y_i\| + \|y_i - y'_i\| + \|x_i - x'_i\| .$$

In particular, $\|u\| \leqslant \|v\|$. □

It is a remarkable fact that L^p–spaces can be identified, up to order isometry, within the class of Banach lattices by a simple criterion that we will state below.

DEFINITION 17.15. Let $p \in [1, \infty)$. An *abstract L^p–space* is a Banach lattice such that for every pair of disjoint vectors $x, y \in X$,

$$\|x + y\|^p = \|x\|^p + \|y\|^p .$$

Clearly, every space $L^p(X, \mu)$, with (X, μ) a measure space, is an abstract L^p–space. S. Kakutani proved that the converse is also true:

THEOREM 17.16 (Kakutani representation theorem [**Kak41**], see also Theorem 3 in [**BDCK66**] and Theorem 1.b.2 in [**LT79**]). *For every $p \in [1, \infty)$ every abstract L^p–space is order isometric to some space $L^p(X, \mu)$ for some measure space (X, μ).*

COROLLARY 17.17. *For every $p \in [1, \infty)$ any closed sublattice of some space $L^p(X, \mu)$ is order isometric to some space $L^p(Y, \nu)$.*

COROLLARY 17.18. *Consider an indexed family of spaces $L^{p_i}(X_i, \mu_i)$, $i \in I$, such that $p_i \in [1, \infty)$. If ω is an ultrafilter on I such that ω-$\lim p_i = p$ then the ultralimit ω-$\lim L^{p_i}(X_i, \mu_i)$ is order isometric to some space $L^p(Y, \nu)$.*

COROLLARY 17.19. *For fixed p, the family of spaces $L^p(X, \mu)$, where (X, μ) are measure spaces, is stable with respect to (rescaled) ultralimits.*

REMARK 17.20. The measure space (Y, ν) in Corollary 17.18 can be identified with the ultralimit of measure spaces (X_i, μ_i). We refer to [**Cut01**] and [**War12**] for details of the construction of the *Loeb measure*, which is the ultralimit of the measure spaces (X_i, μ_i).

17.2. Limit actions and point selection theorem

In this section, all topological spaces as assumed to be Hausdorff. For compactly generated groups a *limit* of a family of actions may naturally occur in various settings, as noted in [**Gro03**] (see also [**BFGM07**], §3, c). In order to simplify the proofs, we will work with finitely generated groups (equipped with discrete topology). Let G be a finitely generated group, and let S be a finite generating subset of G; it will be convenient to assume that $1 \in S$.

DEFINITION 17.21. We say that a topological action $\rho : G \curvearrowright X$ of a discrete group G on a metric space X is an L-action if each $\rho(s), s \in S$ is an L-bi-Lipschitz transformation.

We note that changing the finite generating set amounts to changing the Lipschitz constant L and, therefore, the particular choice of S is irrelevant. We, therefore, refer to L-actions (with unspecified L) as *Lipschitz actions*.

Let $(X_i, \text{dist}_i), i \in I$, be a family of complete metric spaces and let

$$\rho_i : G \curvearrowright X_i$$

be nontrivial actions of the group G, such that the generators $s \in S$ act on X_i by L–bi-Lipschitz transformations $\rho_i(s)$ (with L independent of i). For each i we define $F_i \subset X_i$, the set of points fixed by $\rho_i(G)$. Nontriviality of the action means that $F_i \neq X_i$ for each i.

THEOREM 17.22 (Point-selection theorem). *Let $x_i \in X_i \setminus F_i$ be a family of base-points. Assume that for some ultrafilter ω on I, either*

$$\omega\text{-}lim \frac{\text{dist}(x_i, F_i)}{\text{diam}(S_i x_i)} = \infty \tag{17.5}$$

or $F_i = \emptyset$ *ω-almost surely. Let* δ_i *be either equal to* $\frac{\mathrm{dist}(x_i, F_i)}{2\,\mathrm{diam}(S_i x_i)}$ *in the former case, or we take* (δ_i) *to be any sequence of real numbers such that* $\omega\text{-}lim\,\delta_i = +\infty$ *in the latter case.*

Then there exists an action $\rho_\omega : G \curvearrowright X_\omega$ *by bi-Lipschitz transformations on some rescaled ultralimit of the form*

$$X_\omega = \omega\text{-}lim(X_i, y_i, \lambda_i \mathrm{dist}_i)\,,\ \text{with } \lambda_i \geqslant \frac{2}{\mathrm{diam}(S_i x_i)(1 + 2\delta_i(L+1))}\,, \tag{17.6}$$

such that for every point z_ω *in* X_ω *the diameter of* $\rho_\omega(S)z_\omega$ *is at least* 1. *Moreover, each generator* $s \in S$ *acts on* X_ω *as a bi-Lipschitz transformation.*

Proof. We first note that, according to the choice of δ_i, (17.5) implies that

$$\omega\text{-lim}\,\delta_i = +\infty$$

when $F_i \neq \emptyset$ ω-almost surely. In what follows, for simplicity of the notation, instead of writing $\rho_i(S)x$, we will write Sx for elements $x \in X_i, i \in I$.

Proposition 17.23. *ω-almost surely there exists a point*

$$y_i \in B(x_i, 2\delta_i\,\mathrm{diam}(Sx_i))$$

such that for every point z *in the ball* $B\left(y_i, \frac{\delta_i\,\mathrm{diam}(Sy_i)}{2}\right)$ *the diameter of* $S_i z$ *is at least* $\frac{\mathrm{diam}(Sy_i)}{2}$.

Proof. Assume to the contrary that ω-almost surely for every point y_i in $B(x_i, 2\delta_i\,\mathrm{diam}(Sx_i))$ there exists

$$z_i \in B\left(y_i, \frac{\delta_i\,\mathrm{diam}(Sy_i)}{2}\right)$$

such that the diameter of Sz is strictly less than $\frac{\mathrm{diam}(Sy_i)}{2}$. Let $J \subset I$ be the set of indices such that the above holds, $\omega(J) = 1$, and let i be a fixed index in J. In what follows the argument is only for the index i and for simplicity we suppress the index i in our notation.

Set

$$D := 2\delta\,\mathrm{diam}(Sx) = \mathrm{dist}(x, F), \quad R := \frac{D}{2}.$$

Then for every point y in the ball $B(x, D) \subset X_i$, there exists

$$z \in B\left(y, \frac{\delta\,\mathrm{diam}(Sy)}{2}\right)$$

such that $\mathrm{diam}(Sz) < \frac{\mathrm{diam}(Sy)}{2}$. Applied to $y = x$, it follows that there exists

$$u_1 \in B\left(x, \frac{R}{2}\right),$$

such that $\mathrm{diam}(Su_1) < \frac{\mathrm{diam}(Sx)}{2}$. Applied to u_1, the same statement implies that there exists

$$u_2 \in B\left(x_1, \frac{\delta\,\mathrm{diam}(Su_1)}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4}\right)$$

such that

$$\mathrm{diam}(Su_2) < \frac{\mathrm{diam}(Su_1)}{2} < \frac{\mathrm{diam}(Sx)}{2^2}.$$

Assume that we thus found points $u_1, u_2, \ldots, u_n \in X_i$ such that for every $j \in \{1, 2, \ldots, n\}$,

$$u_j \in B\left(x_{j-1}, \frac{\delta \operatorname{diam}(Su_{j-1})}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4} + \ldots + \frac{R}{2^j}\right) \tag{17.7}$$

and $\operatorname{diam}(Su_j) < \frac{\operatorname{diam}(Sx)}{2^j}$. Then, by taking $y = u_n$, we conclude that there exists

$$u_{n+1} \in B\left(u_n, \frac{\delta \operatorname{diam}(Su_n)}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4} + \ldots + \frac{R}{2^n} + \frac{R}{2^{n+1}}\right)$$

such that

$$\operatorname{diam}(Su_{n+1}) < \frac{\operatorname{diam}(Su_n)}{2} < \frac{\operatorname{diam}(Su)}{2^{n+1}}. \tag{17.8}$$

We thus obtain a Cauchy sequence (u_n) in a complete metric space X_i; therefore, (u_n) converges to a point u in X_i. By the inequalities (17.8), taking into account that the action of G on X_i is continuous, we conclude that

$$\operatorname{diam}(Su) = 0,$$

and, hence, u is fixed by S, thus by the entire group G (since S generates G). Furthermore, (17.7) implies that $\operatorname{dist}(u, x) \leqslant R$. On the other hand, $R = \frac{\operatorname{dist}(x,F)}{2}$, where F is the set of points fixed by G, a contradiction. □

Thus, ω-almost surely there exists y_i in $B(x_i, 2\delta_i \operatorname{diam}(S_i x_i))$ such that for every point

$$z \in B\left(y_i, \frac{\delta_i \operatorname{diam}(S_i y_i)}{2}\right),$$

the diameter of $S_i z$ is at least $\frac{\operatorname{diam}(S_i y_i)}{2}$. Define

$$\lambda_i := \frac{2}{\operatorname{diam}(S_i y_i)}. \tag{17.9}$$

Then triangle inequalities and the fact that each generator $s \in S$ acts on X_i as an L-bi-Lipschitz transformation, implies that

$$\lambda_i \geqslant \frac{2}{\operatorname{diam}(S_i x_i)(1 + 2\delta_i(L+1))}.$$

Furthermore, for each $s \in S$ we have

$$\lambda_i(sy_i, y_i) \leqslant 2.$$

We, therefore, obtain an L-bi-Lipschitz action ρ_ω of the group G on the ultralimit

$$X_\omega = \omega\text{-lim}\,(X_i, y_i, \lambda_i \operatorname{dist}_i)\,,$$

cf. Lemma 7.82. Since $\omega\text{-lim}_i\, \delta_i = \infty$, it follows that the natural inclusion map

$$\omega\text{-lim}\left(B\left(y_i, \frac{\delta_i \operatorname{diam}(S_i y_i)}{2}\right), y_i, \lambda_i \operatorname{dist}_i\right) \to \omega\text{-lim}\,(X_i, y_i, \lambda_i \operatorname{dist}_i)$$

is surjective. Hence, for every point z_ω in X_ω the diameter of $\overline{K} z_\omega$ is at least 1. □

Remark 17.24. Note that it could happen that the limit $\omega\text{-lim}\,\lambda_i$ is positive or even infinite. However, if

$$\omega\text{-lim} \inf_{x \in X_i} \operatorname{diam}(\rho_i(S)(x)) = \infty,$$

then (17.9) implies that $\omega\text{-}\lim \lambda_i = 0$. This situation appears frequently when one constructs group actions on real trees associated with divergent sequences of isometric actions $G \curvearrowright X$, X is a δ-hyperbolic metric space, see §9.23, as well as **[Kap01]** for applications to the theory of Kleinian groups.

Our next goal is to sharpen a bit the conclusion of the Point Selection Theorem. Suppose, in addition, that the family I is the poset $(\mathbb{N}, \leqslant)$ with the standard order. Assume also that $(N_i)_{i\in I}$ is a directed collection of normal subgroups in G, i.e., if $i \leqslant j$ then $N_i \leqslant N_j$. Accordingly, we obtain a collection of quotient groups $G_i = G/N_i$ and projection homomorphisms

$$p_i : G \to G_i, \quad f_{ij} : G_i \to G_j, \quad i \leqslant j.$$

The direct limit of the corresponding direct system of groups $(G_i, f_{ij})_{i,j\in I}$ is naturally isomorphic to the quotient group $\overline{G} = G/N$, where

$$N = \bigcup_{i\in I} N_i.$$

We let $p : G \to \overline{G}$ denote the quotient map.

For every nonpincipal ultrafilter ω on I and ω-large subset $J \subset I$, the natural homomorphism of inverse limits

$$\varinjlim_{j\in J} G_j \longrightarrow \varinjlim_{i\in I} G_i \cong \overline{G}$$

is an isomorphism, since the subset J is cofinal in $(I, \leqslant)$. We, thus, obtain the following addendum to Theorem 17.22:

Theorem 17.25. *Assume that each representation ρ_i in Theorem 17.22 factors through the projection $p_i : G \to G_i$. Then for each nonprincipal ultrafilter ω, each limit action ρ_ω in Theorem 17.22 factors through a Lipschitz action $\overline{G} \curvearrowright X_\omega$.*

The next corollary is an application of Theorem 17.25. Suppose that $G, (G_i)_{i\in I}$ are as in Theorem 17.25. In particular, G is a finitely generated group and S is its finite generating set. We equip each quotient group G_i and the direct limit group $\overline{G}$ with the finite generating sets $S_i, \overline{S}$, which are the images of S under the projections $p_i : G \to G_i, p : G \to \overline{G}$.

Corollary 17.26. *Let $\mathcal{C}$ be a class of complete metric spaces stable with respect to rescaled ultralimits taken using countably infinite index sets I. If the group $\overline{G}$ has the Property F$\mathcal{C}$, then there exists i_0 and $\varepsilon > 0$, such that for every $i > i_0$ the group G_i has the Property F$\mathcal{C}$. Furthermore, for every isometric action φ_i of G_i on some space $(X_i, \mathrm{dist}_i) \in \mathcal{C}$, if F_i is the set of points in X_i fixed by G_i, then for every point $x \in X_i$ the diameter of $\varphi_i(S_i)x$ is at least $\varepsilon\mathrm{dist}_i(x, F_i)$ (and, obviously, at most $2\mathrm{dist}_i(x, F_i)$).*

Proof. Assume to the contrary that for every $\varepsilon > 0$ and i_0 there exists an $i > i_0$ such that G_i has an isometric action ρ_i on some space $X_i \in \mathcal{C}$ either without fixed points or such that for some point $x_i \in X_i$ the diameter of $\rho_i(S_i)x_i$ is $<$ $\varepsilon\mathrm{dist}_i(x, F_i)$. Then there exists a strictly increasing sequence of indexes i_n and a sequence of actions φ_{i_n} of G_{i_n} on some $X_{i_n} \in \mathcal{C}$ either without fixed points or with points $x_n \in X_{i_n}$ satisfying

$$\mathrm{diam}(\varphi_{i_n}(S_{i_n})x_n) < \frac{1}{n}\mathrm{dist}_{i_n}(x_n, F_{i_n}).$$

We consider the case of nonempty fixed-point sets since the other case is analogous. Let ω be a nonprincipal ultrafilter on $\mathbb{N}$ containing the sequence (i_n). Then

$$\omega\text{-lim}\,\frac{\mathrm{dist}_{i_n}(x_n, F_{i_n})}{\mathrm{diam}(\varphi_{i_n}(S_{i_n})x_n)}.$$

Theorem 17.25 then yields a contradiction as it results in a fixed-point free action of $\overline{G}$ on some space $X_\omega \in \mathcal{C}$. ☐

17.3. Fixed point properties FL^p

Theorem 17.22 allows one to prove certain fixed point properties for actions of groups using ultralimits. We will discuss such applications in this and in the next section. Let $\mathcal{C}$ be a collection of metric spaces, let $L \geqslant 1$ and let G be a (discrete) group.

Definition 17.27. We say that a group G has *the fixed point property* $F\mathcal{C}$ if for every isometric action ρ of G on every space $X \in \mathcal{C}$, the group $\rho(G)$ fixes a point in X.

Remark 17.28. Note that one can further loosen the definition of the property $F\mathcal{C}$ by considering finitely generated groups G with a fixed generating set S, and *L-Lipschitz* actions $G \curvearrowright X$, where each generator $s \in S$ acts as an L-Lipschitz transformation of X. Some proofs in this and the following section go though in this setting without much change. However, we will not pursue this direction here.

Below we list several special cases of such fixed point properties which are important in group theory. Given a real Hilbert space $(\mathcal{H}, \langle\ ,\ \rangle)$, its *orthogonal group* $O(\mathcal{H})$ is the group of orthogonal linear invertible operators $U : \mathcal{H} \to \mathcal{H}$, i.e., $\langle Ux, Uy\rangle = \langle x, y\rangle$. By the Mazur-Ulam theorem for real Hilbert spaces (see e.g. [**FJ03**, p. 6] or [**Nic12**]), every isometry of $\mathcal{H}$ is an affine transformation. Thus, $\mathrm{Isom}(\mathcal{H})$, the isometry group of $\mathcal{H}$, has the form $\mathcal{H} \rtimes O(\mathcal{H})$, where the first factor $\mathcal{H}$ is identified to the group of translations on $\mathcal{H}$.

An *isometric affine action* of a (discrete) group G on a Hilbert space $\mathcal{H}$ is a homomorphism $\alpha : G \to \mathrm{Isom}(\mathcal{H})$.

Definition 17.29. A discrete group G has *Property* FH if every affine isometric continuous action of G on a Hilbert space has a fixed point.

We note that for topological groups G (with topologies other than the discrete one) one also defines the *topological* Property FH, by restricting to continuous affine isometric actions. We will discuss the relation of the Property FH and Kazhdan's Property (T) in the next section: We will see that for (discrete) groups, these two properties are equivalent.

Exercise 17.30. Show that a discrete group G has Property FH if and only if $H^1(G, \mathcal{H}_\pi) = 0$ for every unitary representation $\pi : G \to O(\mathcal{H})$. Hint: Use Lemma 3.122.

In view of the fixed-point theorem for isometric group actions on $CAT(0)$ spaces (Theorem 2.69), we obtain:

Corollary 17.31 (A. Guichardet). *A group G has the Property FH if and only if every affine isometric continuous action of G on a Hilbert space has a bounded orbit.*

Recall (see Definition 2.72) that a group G has Property FA if every isometric action of G on a real tree has a fixed point.[1] Here is a link between the two fixed-point properties:

THEOREM 17.32 (R. Alperin [**Alp82**] and Y. Watatani [**Wat82**]; see also [**BdlHV08**]). *$FH \Rightarrow FA$: Every (discrete) group with Property FH also has Property FA.*

We will prove this theorem in Section 17.5.

Another interesting connection is between the Property FH and isometric group actions on $\mathbb{H}^n$:

THEOREM 17.33 (J. Faraut and K. Harzallah [**FH74**]; see also [**BdlHV08**]). *If a group G has the Property FH, then every isometric action of G on a real-hyperbolic space $\mathbb{H}^n$ has a fixed point.*

Given these two examples, the reader might wonder if FH implies that every isometric action on a Rips-hyperbolic space has a bounded orbit. It turns out that the answer is negative, as there are infinite hyperbolic groups which have Property FH. The oldest example of infinite hyperbolic groups with Property FH comes from the theory of symmetric spaces. Let $X = \mathbf{H}\mathbb{H}^n, n \geqslant 2$ be the *quaternionic hyperbolic symmetric space* of quaternionic dimension $\geqslant 2$, or the octonionic hyperbolic plane, see Section 8.9. Let G denote the isometry group of X. We note that such G contains both uniform and nonuniform lattices; every uniform lattice $\Gamma < G$ acts isometrically properly discontinuosly cocompactly on X, which implies that Γ is and infinite hyperbolic group.

THEOREM 17.34. *(See e.g. [**BdlHV08**].) Every group G and lattice $\Gamma < G$ as above satisfies the Topological Property FH and the Property FH respectively.*

There are other examples of infinite hyperbolic groups satisfying Property FH, we will discuss them in more detail in Section 17.5. Furthermore, it turns out that in Gromov's model of randomness for groups, for a certain range of the parameter d called *density*, "a majority of groups" are infinite hyperbolic with the Property FH, see [**Żuk03**]. On the other hand, in for values of d varying in other intervals, a majority of groups are infinite, hyperbolic and without the Property FH, see [**OW11**].

Below is a generalization of the Property FH in the context of other Banach spaces.

For $p > 1$ each space $L^p(X, \mu)$ is reflexive and, hence, satisfies the following fixed-point property (that we already saw in the case $p = 2$), see e.g. [**Now15**, 3.4]:

- If $G \curvearrowright L^p(X, \mu)$ is an isometric affine action which has a bounded orbit, then G has a fixed point in $L^p(X, \mu)$.

DEFINITION 17.35. A (discrete) group G has *Property FL^p*, where p is a real number in $[1, +\infty)$, if for every set X equipped with a measure μ, every affine isometric action of G on $L^p(X, \mu)$ has bounded orbits.

We note that boundedness of G-orbits implies existence of a fixed point except for $p = 1$; in view of this exception, the Property FL^p should be probably called BL^p, but we follow the notation used in the literature.

[1]Some authors restrict to group actions on simplicial trees.

Thus, for $p \in (1, \infty)$, the Property FL^p is equivalent to the property $F\mathcal{C}$ in Definition 17.27, where $\mathcal{C}$ the class of L^p–spaces.

Below is yet another application of limits of actions. The next theorem was proved by Y. Shalom [**Sha00**, Theorem p. 5] in the case $p = 2$ (i.e., the Property FH), answering a question of R. Grigorchuk and A. Zuk.

THEOREM 17.36. *Let G be a finitely generated group satisfying the Property FL^p, for some $p > 1$. Then G can be written as H/N, where H is a finitely presented group with Property FL^p and N is a normal subgroup in H.*

PROOF. Consider an infinite presentation of G, $G = \langle S \mid r_1, \ldots, r_n, \ldots \rangle$, where S is a finite set generating G and (r_i) is a sequence of relators in S. Let $F(S)$ be the free group in the alphabet S and N_i the normal closure in $F(S)$ of the finite set $\{r_1, \ldots, r_i\}$. The groups $G_i = F(S)/N_i$ are all finitely presented, and form a direct system whose direct limit is G. Assume that none of these groups has Property FL^p. It follows that for each i there exists some space $L^p(Y_i, \mu_i)$ and an affine isometric action of G_i on $L^p(Y_i, \mu_i)$ without a fixed point. Theorem 17.22 and Corollary 17.18 imply that G acts by affine isometries and without a global fixed point on some space $L^p(Z, \nu)$, contradicting the hypothesis. $\square$

We now compare the Properties FL^p for various values of p. Exercise 1.143 shows that for each isometric action $G \curvearrowright L^p(X, \mu), p \in [1, 2]$, there exists an isometric embedding $\phi : L^p(X, \mu) \to \mathcal{H}$ into a Hilbert space $\mathcal{H}$, such that ϕ is equivariant with respect to a homomorphism $G \to \mathrm{Isom}(\mathcal{H})$. If G acted on $L^p(X, \mu)$ with unbounded orbits, the action of G on $\phi(L^p(X, \mu))$ also has unbounded orbits. We, thus obtain:

THEOREM 17.37 ([**Del77**], [**AW81**], [**WW75**]). *If a discrete group G satisfies the Property FH, then it also satisfies the Property FL^p for each $p \in [1, 2]$.*

The converse to this theorem holds as well, see e.g. [**BFGM07**], [**CDH10**], and we obtain:

THEOREM 17.38. *For every $p \in [1, 2]$,*

$$FL^p \iff FH.$$

The implication FH $\Rightarrow FL^p$ extends a little bit beyond the interval $[1, 2]$ according to the following theorem:

THEOREM 17.39 (D. Fisher and G. Margulis, see [**BFGM07**], §3.c). *For every discrete group G with Property FH there exists $\varepsilon = \varepsilon(G)$ such that G has Property FL^p for every $p \in [1, 2 + \varepsilon)$.*

We will prove a bit stronger form of this result below. For each discrete group G define the subset $\mathcal{FP}_G \subset [1, \infty)$ consisting of those p such that G satisfies the Property FL^p.

THEOREM 17.40. *For each finitely generated group G the set $\mathcal{FP}_G$ is open.*

PROOF. In view of Theorem 17.38, it suffices to show that the set of $p \in [2, \infty)$ such that G does not satisfy the Property FL^p is closed. Let $p_n \in [2, \infty)$ be a seqeunce converging to $p < \infty$, such that for every n, G has an isometric action on some space $L^{p_n}(X_n, \mu_n)$ without a fixed point. Theorem 17.22 and Corollary

17.18 imply that G, for some set Y and a measure m on Y, the group G also acts isometrically on the space $L^p(Y,\nu)$ without a fixed point. $\square$

For p much larger than 2, Properties FH and FL^p are no longer equivalent. Examples for which FH does not imply FL^p.

Theorem 17.41 (P. Pansu, [**Pan95**] and Y. de Cornulier, R. Tessera and A. Valette [**dCTV08**].). *Let G denote the isometry group of the quaternionic-hyperbolic space $\mathbb{HH}^n$. Then every lattice $\Gamma < G$ admits a proper isometric action on some $L^p(X,\mu)$ for every $p > 4n+2$.*

Thus, each lattice in G has the Property FH and does not satisfy the Property FL^p for all $p > 4n+2$. Furthermore:

Theorem 17.42. *1. For every infinite hyperbolic group G there exists $p = p(G)$ such that G admits an isometric action on $\ell^p(G)$ without a fixed point (M. Bourdon, H. Pajot* [**BP03**]*).*

2. Moreover, for every hyperbolic group G there exists $p = p(G)$ such that G admits a proper isometric action on $L^p(X,\mu)$ (G. Yu [**Yu05**]*).*

We refer the reader to [**Bou16**] and [**Nic13**] for alternative proofs of Part 2 of this theorem.

On the other hand, lattices in Lie groups of higher rank exibit different behavior with respect to the Property FL^p:

Theorem 17.43 (U. Bader, A. Furman, T. Gelander and N. Monod [**BFGM07**]). *Let G be a semisimple Lie group with all noncompact factors of rank $\geqslant 2$. Then every lattice $\Gamma < G$ satisfies the Property FL^p for all $p \in [1,\infty)$.*

17.4. Kazhdan's Property T

In this section we specialize the discussion of the previous section to the case $p = 2$ and isometric group actions on Hilbert spaces. This leads us to the *Kazhdan's Property (T)*. Our discussion of the Property (T) is limited, we refer the reader to [**BdlHV08**] for the in-depth treatment. Recall that finitely generated groups, by default, are endowed with the discrete topology. A unitary representation of a topological group G in a Hilbert space $\mathcal{H}$ is a homomorphism $\pi : G \to U(\mathcal{H})$ such that for every $x \in \mathcal{H}$, the map from G to $\mathcal{H}$ defined by $g \mapsto gx$, is continuous.

Definition 17.44. Let $(\pi, \mathcal{H})$ be a unitary representation of a topological group G.

(1) Given a subset $S \subseteq G$ and a number $\varepsilon > 0$, a unit vector x in $\mathcal{H}$ is (S,ε)-*invariant* if
$$\sup_{g\in S} \|\pi(g)x - x\| \leqslant \varepsilon \|x\| .$$
(2) The representation $(\pi, \mathcal{H})$ *almost has invariant vectors* if it has (K,ε)–invariants vectors for every compact subset K of G and every $\varepsilon > 0$.
(3) The representation $(\pi, \mathcal{H})$ has *invariant vectors* if there exists a unit vector x in $\mathcal{H}$ such that $\pi(g)x = x$ for all $g \in G$.

Clearly, existence of invariant vectors implies existence of almost invariant vectors. It is a remarkable fact that there are many groups for which the converse holds as well.

Definition 17.45. A topological group G has *Kazhdan's Property* (T) if for every unitary representation π of G, if π has an almost invariant vector, then it also has an invariant vector.

In order to simplify the discussion, we will primarily limit ourselves to groups equipped with discrete topology.

Theorem 17.46 (D. Kazhdan). *Each discrete group G satisfying Property (T) is finitely generated.*

Proof. For each finitely generated subgroup $H \leqslant G$ we define the quotient G/H; the group G acts on this space by the left multiplication. Accordingly, G has a unitary representation $\pi_{G/H}$ on the Hilbert space $\ell^2(G/H)$. Let $\mathbf{1}_H \in \ell^2(G/H)$ denote the indicator function of the coset H in G/H; this is a unit vector fixed by the representation $\pi_{G/H}$. Now, consider the infinite direct sum

$$V := \bigoplus_{H \leqslant G} \ell^2(G/H)$$

taken over all finitely generated subgroups $H \leqslant G$. This is a pre-Hilbert space, we let $\mathcal{H}$ denote its completion, a Hilbert space. The unitary representations $\pi_{G/H}$ yield a unitary representation π of G on $\mathcal{H}$. We regard each $\ell^2(G/H)$ as a subspace in $\mathcal{H}$. Then each vector $\mathbf{1}_H \in \ell^2(G/H) \subset \mathcal{H}$ is fixed by the subgroup H. Therefore, for every finite subset $K \subset G$ generating the subgroup $H = \langle K \rangle$, the vector $\mathbf{1}_H$ is fixed by the action of K. It follows that the representation $(\pi, \mathcal{H})$ almost has invariant vectors. Since G satisfies the Property (T), π has a fixed unit vector $x \in \mathcal{H}$. Let x^* denote the (nonzero) linear functional on $\mathcal{H}$ dual to x:

$$x^*(v) = \langle x, v \rangle .$$

This functional is G-invariant and has nonzero restriction to the dense subspace V, hence, nonzero G-invariant restriction to one of the subspaces $\ell^2(G/H)$. Therefore, G has a fixed nonzero vector u in $\ell^2(G/H)$. Since G acts transitively on G/H, the function $u \in \ell^2(G/H)$ has to be constant. It follows that the set G/H is finite. Thus, the finitely generated group H has finite index in G. Therefore, G itself is finitely generated. □

Corollary 17.47. *For each discrete group G satisfying the Property (T) and equipped with a finite generating set S, there exists a number $\epsilon > 0$ such that whenever a unitary representation π has an (S, ε)–invariant vector, π has a nonzero invariant vector.*

Proof. Suppose that this assertion fails and consider a sequence of positive numbers $\epsilon_i \to 0$ and unitary representations $\pi_i : G \to O(\mathcal{H}_i)$, such that π_i has an (S, ϵ_i)-invariant unit vector in $\mathcal{H}_i$ but no invariant vectors. Then, as in the proof of Theorem 17.46, consider the natural action π of G on the completion $\mathcal{H}$ of the direct sum of Hilbert spaces

$$\bigoplus_i \mathcal{H}_i.$$

Then, as in the proof of Theorem 17.46, π has almost invariant vectors but no invariant vectors. This is a contradiction. □

Definition 17.48. Each pair $(S, \epsilon) \subset G \times \mathbb{R}$ satisfying this corollary is called a *Kazhdan pair* for G. A number $\epsilon > 0$ for which there exists $S \subset G$, such that (S, ϵ) is a Kazhdan pair, is called a *Kazhdan constant* of G.

The next theorem is due to Delorme [**Del77**] (who proved the implication T⇒FH) and Guichardet [**Gui77**] (who proved the opposite implication), see also [**dlHV89**] and [**CCJ$^+$01**]. In these references the theorem is proven in greater generality, namely, in the setting of second countable, locally compact, Hausdorff topological groups; we limit ourselves to finitely generated groups with discrete topology.

THEOREM 17.49. *Let G be a finitely generated group. The the following are equivalent:*

1. G has the Property FH.
2. G has the Property T.
3. Every conditionally seminegative G-invariant kernel is bounded.

PROOF. We will deduce the implication (1) ⇒ (2) from Theorem 17.22. Our proof follows [**Sil**] and [**Gro03**].

Let G be a finitely generated group with Property FH and assume that it does not satisfy (2). Fix a compact generating set S of G. Then, for every $n \in \mathbb{N}$ there exists a unitary representation $\pi_n : G \to U(\mathcal{H}_n)$ with an $(S, \frac{1}{n})$–invariant (unit) vector x_n and no invariant vectors. Let X_n be the unit sphere $\{u \in \mathcal{H}_n \ : \ \|u\| = 1\}$ with the induced path metric dist_n. Theorem 17.22 applied to the sequence of isometric actions of G on X_n and a choice of δ_n such that $\omega\text{-lim}\, \delta_n = +\infty$ and $\omega\text{-lim}[\delta_n \,\mathrm{diam}(Sx_n)] = 0$, implies that G acts by isometries on a rescaled ultralimit

$$X_\omega = \omega\text{-lim}(X_n, x_n, \lambda_n \mathrm{dist}_n)\,, \text{ with } \lambda_n \geqslant \frac{2}{(1+2\delta_n)\,\mathrm{diam}(Sx_n)}\,.$$

Note that $\omega\text{-lim}\, \lambda_n = +\infty$. Moreover, for every point $z_\omega \in X_\omega$ the diameter of Sz_ω is at least 1. Since $\omega\text{-lim}\, \lambda_n = +\infty$, Example 7.64 shows that the ultralimit X_ω is isometric to a Hilbert space $\mathcal{H}$. We, thus, obtain an isometric action of G on a Hilbert space $\mathcal{H}$ without a global fixed point, contradicting Property FH.

In order to prove the implication (2) ⇒ (1), we will need two lemmas. In both lemmas, $\mathcal{H}$ is a Hilbert space.

LEMMA 17.50. *For each $s > 0$ there exists a Hilbert space $\mathcal{H}_s$, a representation $\rho_s : \mathrm{Isom}(\mathcal{H}) \to O(\mathcal{H}_s)$ and a ρ-equivariant map $F = F_s$ from $\mathcal{H}$ to the unit sphere in $\mathcal{H}_s$. Moreover, this map satisfies the property:*

$$\lim_{\|x-y\|\to\infty} \langle F(x), F(y)\rangle = 0.$$

PROOF. Let $\mathcal{H}'_s$ denote the space of finitely-supported real-valued functions on $\mathcal{H}$. Define the inner product on $\mathcal{H}'_s$ by the formula:

$$\langle f, g\rangle = \sum_{x,y\in\mathcal{H}} e^{-s||x-y||} f(x) g(y).$$

We set $F(x) := \mathbf{1}_x$. Lastly, we let $\mathcal{H}_s$ denote the completion of $\mathcal{H}'_s$ with respect to the norm defined by this inner product. □

LEMMA 17.51. *Suppose that G is a (discrete) group acting isometrically on $\mathcal{H}$ with unbounded orbits. Then $\rho(G)$ has no nonzero fixed vectors in $\mathcal{H}_s$, where $\mathcal{H}_s$ is as in the previous lemma.*

PROOF. Suppose that $v \in \mathcal{H}_s$ is a vector fixed by $\rho(G)$. Consider a sequence $g_n \in G$ such that $\|g_n(x)\| \to \infty$ for (one/all) $x \in \mathcal{H}$. For $F = F_s$,

$$\lim_{n\to\infty} \langle v, F(g_n x)\rangle = 0,$$

$$\langle v, F(g_n x)\rangle = \langle g_n v, F(g_n x)\rangle = \langle g_n v, g_n F(x)\rangle = \langle v, F(x)\rangle .$$

Hence, $\langle v, F(x)\rangle = 0$ for all $x \in \mathcal{H}$. Since the vectors $F(x), x \in \mathcal{H}$, span a dense subset in $\mathcal{H}_s$, it follows that $v = 0$. □

We now return to the proof of the theorem. Suppose that G is finitely generated, satisfies the Property (T) and let ϵ denote the Kazhdan constant of G with respect to a finite generating set S. Then for each affine isometric action $G \curvearrowright \mathcal{H}$, the parameter s can be chosen so that for one (each) unit vector $u \in \mathcal{H}_s$, the diameter of the set $\rho_s(S)$ can be made $< \epsilon$.

Since G satisfies the Property T, $\rho_s(G)$ has to fix a nonzero vector in $\mathcal{H}_s$. According to Lemma 17.51, the action $G \curvearrowright \mathcal{H}$ has to have bounded orbits.

Lastly, we will prove equivalence of (3) and (1). If ψ is an unbounded conditionally nonpositive G-invariant kernel on G, then, according to Theorem 1.137, there exists a Hilbert space $\mathcal{H}$ and a representation

$$\rho : G \to \mathrm{Isom}(\mathcal{H})$$

such that

$$\psi(g, h) = \|\rho(g)(0) - \rho(h)(0)\|^2. \tag{17.10}$$

(Here in order to apply Theorem 1.137 we let $X = G$ and let G act on itself *via* the left multiplication.) Since ψ is unbounded, taking $h = 1 \in G$, we conclude that the action of G on $\mathcal{H}$ has unbounded orbits. This contradicts the Property FH. Conversely, given a representation $\rho : G \to \mathrm{Isom}(\mathcal{H})$ of G to the isometry group of a Hilbert space, we define the kernel ψ on G by the formula (17.10). This kernel is clearly G-invariant; it is also conditionally nonpositive according to Theorem 1.137. □

We can now prove the implication (3)⇒(1). If G acts by isometries on a Hilbert space $\mathcal{H}$ then for any $v \in \mathcal{H}$ the map $\psi(g, h) = \|g\cdot v - h\cdot v\|^2$ is a left invariant conditionally negative definite kernel on G, which, therefore, has to be bounded. Taking $h = 1$, we conclude that the action $G \curvearrowright \mathcal{H}$ has bounded orbits and, therefore, has a fixed point (see Corollary 17.31). □

REMARK 17.52. Yves de Cornulier constructed in **[dC06]** examples of uncountable discrete groups with Property FH that do not satisfy Property (T).

Note that amenability and Property (T) are incompatible for infinite groups according to the following result:

THEOREM 17.53 (See Theorem 1.1.6, **[BdlHV08]**). *Let G be a finitely generated group. The following properties are equivalent:*

(1) *G is both amenable and has Property (T);*

(2) *G is finite.*

PROOF. We will use yet another characterization of amenable groups. Below, for a discrete group G, $\ell^2(G) = \ell^2(G, \mu)$ where μ is the counting measure (measure of each finite subset equals its cardinality).

Proposition 17.54. *A finitely generated group G is amenable if and only if the action of G on $\mathcal{H} = \ell^2(G)$ via left multiplication has almost invariant vectors.*

Proof. We will prove only the direct implication (needed for the proof of Theorem 17.53). Let $F_i \subset G$ be a Følner sequence. Let $f_i = \frac{1}{N_i}\mathbf{1}_{\Omega_i}$, where $N_i := |\Omega_i|^{1/2}$ and $\mathbf{1}_{\Omega_i}$ denotes the characteristic function of Ω_i. Then for every $g \in G$

$$\|g(f_i) - f_i\|^2 \leqslant \frac{|g\Omega_i \triangle \Omega_i|}{|\Omega_i|},$$

which converges to zero by the definition of a Følner sequence. □

Suppose that G satisfies the Property (T). Thus, there exists a nonzero G-invariant vector $f \in \ell^2(G)$; the function f, hence, is constant. Since $f \in \ell^2(G)$, it follows that G has finite total measure, i.e. is finite. □

Further properties of groups with Property (T).

In view of equivalence of Property (T) and Property FH, it is clear that every quotient group of a group with Property (T) also has Property (T). Since a (discrete) amenable group has property (T) if and only if such group is finite, it follows that every amenable quotient of a group with Property (T) has to be finite. In particular, every discrete group with Property (T) has finite abelianization. For instance, free groups and surface groups never have Property (T). On the other hand, we will see below that, unlike amenability, Property (T) is not inherited by subgroups.

Lemma 17.55. *Property (T) is a VI-invariant.*

Proof. 1. Suppose that a group H has Property (T) and G is a group containing H as a finite index subgroup. Suppose that $G \curvearrowright \mathcal{H}$ is an isometric affine action of G on a Hilbert space. Since H has Property (T), there exists $x \in hh$ fixed by H. Therefore, the G-orbit of x is finite. Therefore, by Theorem 2.69, G fixes a point in $\mathcal{H}$ as well.

2. Suppose that $H \leqslant G$ is a finite index subgroup and G has Property (T). Let $H \curvearrowright \mathcal{H}$ be an isometric affine action. Define the *induced* action Ind_H^G of G on the space V:

$$V = \{\phi : G \to \mathcal{H} : \phi(gh^{-1}) = h\phi(g), \forall h \in H, g \in G\}.$$

Every such function is, of course, determined by its values on $\{g_1, \ldots, g_n\}$, coset representatives for G/H. The group G acts on V by the left multiplication $g : \phi(x) \mapsto \phi(gx)$. Therefore, as a vector space, V is naturally isomorphic to the n-fold sum of $\mathcal{H}$. We equip V with the inner product

$$\langle \phi, \psi \rangle := \sum_{i=1}^{n} \langle \phi(g_i), \psi(g_i) \rangle ,$$

making it a Hilbert space. We leave it to the reader to verify that the action of G on V is affine and isometric. The initial Hilbert space $\mathcal{H}$ embeds diagonally in V; this embedding is a H-equivariant, linear and isometric. Since G has Property (T), it has a fixed vector $\psi \in V$. Therefore, the orthogonal projection of ψ to the diagonal in V is fixed by H. Hence, H also has Property (T).

3. Consider a short exact sequence

$$1 \to F \to G \to H \to 1.$$

If G has property (T), then so does H (as a quotient of G).

Conversely, suppose that H and F both have Property (T) (we will use it in the case where F is a finite group). Consider an affine isometric action $G \curvearrowright \mathcal{H}$ on a Hilbert space. Since F has Property (T), it has nonempty fixed-point set $V \subset \mathcal{H}$. Then V is a closed affine subspace in $\mathcal{H}$, which implies that V (with the restriction of the metric from $\mathcal{H}$) is isometric to a Hilbert space. The group G preserves V and the affine isometric action $G \curvearrowright V$ factors through the group H. Since H has Property (T), it has a fixed point $v \in V$. Thus, v is fixed by the entire group G. In particular, every co-extension of a group with Property (T) with finite kernel, also has Property (T).

Putting all these facts together, we conclude that Property (T) is invariant under virtual isomorphisms. □

Moreover (see e.g. [**BdlHV08**]):

<u>THEOREM</u> 17.56. *Let G be a locally compact group and $\Gamma < G$ is a lattice, equipped with discrete topology. Then G has the (topological) Property T if and only if Γ does.*

17.5. Examples of groups with and without Property T

THEOREM 17.57 (R. Alperin and Y. Watatani). *FH⇒FA: Each group with the property FH satisfies the Property FA.*

PROOF. We will prove the contrapositive. Let G be a group and $G \curvearrowright X$ is an isometric action on a real tree with unbounded orbits. We claim that the function $\text{dist}(x, y)$ is a conditionally nonpositive definite G-invariant kernel on X. The only statement which is not immediate is that dist is conditionally nonpositive definite. Since this statement needs to be verified for each finite subset $Y = \{y_1, \ldots, y_n\}$ of X, it suffices to prove it for the finite metric subtree $T \subset X$ which is the convex hull hull of Y in X. We will assume that Y consists of at least two points since the statement is clear otherwise. Being a finite metric tree, T is a finite simply-connected graph equipped with a path-metric. We orient each edge e of this graph in arbitrary fashion. Let $V \subset T$ denote the vertex set of the tree. For each point $p \in T \setminus V$ we define a function $f_p : T \to \{0, 1\}$ as follows. The point p separates T in two connected components. Let $e = uv$ be the oriented edge of T containing p. If $x \in T$ is contained in the same connected component of $T \setminus \{p\}$ as the vertex v, we set $f_p(x) = 1$. For all other points $x \in T$ we set $f_p(x) = 0$. We equip the tree T with the measure μ without atoms, whose restriction to each edge e of T is the Lebesgue measure, so that $\mu(e)$ equals the length of e. Define the function

$$\psi(x, y) = \int_T (1 - f_p(x)) f_p(y) d\mu(p).$$

This function can be regarded as a *nonsymmetric pseudometric* on T: It is non-negative and satisfies the triangle inequality, but, in general, $\psi(x, y) \neq \psi(y, x)$. We leave it to the reader to verify that for all points $x, y \in T$,

$$\text{dist}(x, y) = \psi(x, y) + \psi(y, x).$$

We are now ready to verify that dist is conditionally nonpositive definite. Let $Y = \{y_1, \ldots, y_n\} \subset T \subset X$ be as above (T is the convex hull of Y in X) and take a vector $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ such that

$$\lambda_1 + \ldots + \lambda_n = 0.$$

We have:

$$\sum_{i,j=1}^{n} \lambda_i \lambda_j \psi(y_i, y_j) = \sum_{i,j=1}^{n} \lambda_i \lambda_j \int_T (1 - f_p(y_i)) f_p(y_j) d\mu(p) =$$

$$\sum_{i,j=1}^{n} \lambda_i \lambda_j \int_T f_p(y_j) d\mu(p) - \sum_{i,j=1}^{n} \lambda_i \lambda_j \int_T f_p(y_i) f_p(y_j) d\mu(p) =$$

$$(\lambda_1 + \ldots + \lambda_n) \sum_{j=1}^{n} \lambda_j \int_T f_p(y_j) d\mu(p) - \int_T \left(\sum_{i=1}^{n} \lambda_i f_p(y_i) \right)^2 d\mu(p) =$$

$$- \int_T \left(\sum_{i=1}^{n} \lambda_i f_p(y_i) \right)^2 d\mu(p) \leqslant 0.$$

It follows that $\operatorname{dist}(x, y)$ is conditionally nonpositive definite. Since the action $G \curvearrowright X$ has unbounded orbits, the kernel $\operatorname{dist}(x, y)$ is unbounded. Therefore, according to Theorem 17.49, the group G does not have the Property FH. □

Corollary 17.58. *Each group G which admits a nontrivial amalgamated free product decomposition $G \cong G_1 \star_{G_3} G_2$ or a nontrivial HNN-decomposition $G \cong G_1 \star_{G_3}$, does not have the Property FH.*

As an application of this result, we obtain:

Corollary 17.59. *If M is a connected 3-dimensional manifold with infinite fundamental group G, then G does not satisfy the Property T.*

Proof. First of all, if G is not finitely generated, it cannot satisfy the Property T, see Theorem 17.46. Thus, we will assume that G is finitely generated. According to the Scott Compact Core Theorem (see **[Sco73]**), there exists a compact submanifold (possibly with boundary) $M_1 \subset M$ such that the inclusion map $M_1 \to M$ induces an isomorphism of fundamental groups. Thus, the problem reduces to the case of compact 3-dimensional manifolds. Since Property T is a virtual isomorphism invariant, we can assume that the manifold M_1 is oriented. Attach 3-dimensional balls to each spherical boundary component of M_1; this results in a compact 3-dimensional manifold M_2 such that each boundary component of M_2 has genus $\geqslant 1$. The Euler characteristic $\chi(M_2)$ of the manifold M_2 equals

$$\frac{1}{2} \chi(\partial M_2) \leqslant 0.$$

if one of the boundary components of M_2 has genus $\geqslant 2$, then

$$\chi(M_2) = b_0(M_2) - b_1(M_2) + b_2(M_2) < 0.$$

It follows that $b_1(M_2) \geqslant 1$, i.e., there exists an epimorphism

$$\pi_1(M_2) \to \mathbb{Z}.$$

Since $\mathbb{Z}$ does not satisfy the Property T, the group $\pi_1(M_2)$ does not satisfy it either.

Suppose, therefore, that each boundary component of $N = M_2$ is a torus (this includes the case $\partial M_2 = \emptyset$). Since $\pi_1(N)$ is assumed to be infinite, N is not homeomorphic to S^3 (this is the 3-dimensional Poincaré Conjecture, proven by Perelman). We now apply *Thurston's Geometrization Conjecture/Perelman's Theorem* to the manifold N. According to this theorem, N admits a 2-step decomposition as follows. First of all, N splits as a connected sum

$$N = N_1 \# \ldots \# N_n, \quad n \geqslant 1,$$

where each manifold N_i is *prime*, i.e., does not have a nontrivial connected sum decomposition, and is not simply-connected. If $n \geqslant 2$, then the group $\pi_1(N)$ admits a nontrivial free product decomposition and, hence, cannot have the Property T. Assume, therefore, that the manifold N itself is prime, $n = 1$. Then N admits a splitting along a system of pairwise disjoint π_1-injective tori T^2 into submanifolds $K_1, \ldots, K_m$ with toral boundary (which could be empty if $N = K_1$ and $\partial N = \emptyset$). Each piece K_i of this decomposition is *geometric*. If the secondary decomposition of N is nontrivial, then $\pi_1(N)$ is isomorphic to the fundamental group of a nontrivial graph of groups and, therefore, again, $\pi_1(N)$ admits a nontrivial action on a simplicial tree. We are thus, left with the case when N itself is geometric, i.e., admits a geometric structure modeled on one of Thurston's eight 3-dimensional geometries. By looking at these geometries one-by-one, it is clear that either:

1. $\pi_1(N)$ is virtually solvable and, thus amenable (this happens in the case of the geometries $\mathbb{E}^3$, Nil, Sol and $\mathbb{S}^2 \times \mathbb{R}$). The group $\pi_1(N)$ cannot satisfy the Property T in these cases.

2. $\pi_1(N)$ admits an isometric action on the hyperbolic space $\mathbb{H}^3$ with unbounded orbits (this happens in the case of the geometries $\mathbb{H}^3$, $\mathbb{H}^2 \times \mathbb{R}$ and $\widetilde{SL}(2, \mathbb{R})$). In these cases $\pi_1(N)$ cannot satisfy the Property T according to Theorem 17.33. □

REMARK 17.60. Actually, according to the very recent solution of the *Virtual 1st Betti number Problem* by Agol and Wise [], if M is a compact 3-dimensional manifold with infinite fundamental group, then M has a finite cover $N \to M$ such that $b_1(N) > 0$. Since the Property T is a virtual isomorphism invariant, one again concludes that $\pi_1(M)$ cannot satisfy the Property T. However, the proof given above is much more elementary than the solution of the Virtual 1st Betti number Problem.

Groups satisfying the Property T. Recall that the Property FH (equivalently, Property T) for a discrete group G can be reformulated as vanishing $H^1(G, \mathcal{H}_\pi) = 0$ for all unitary representation $\pi : G \to U(\mathcal{H})$. There is an old technique (going back to Bochner) for proving vanishing theorems of this type; namely, if $G \cong \pi_1(M)$, where M is a closed Riemannian manifold,

$$H^1(G, \mathcal{H}_\pi) \cong H^1_{DR}(M, \mathcal{V}),$$

where the right hand side the *de Rham cohomology* of M with coefficients in a flat vector bundle over M with fibers isometric to $\mathcal{H}$. Then one uses the Hodge Theorem to represent de Rham cohomology classes $\omega \in H^1_{DR}(M, \mathcal{V})$ by *harmonic 1-forms*. Lastly, one uses some geometric properties of M to show that each harmonic form as above has to vanish. Harmonic 1-forms on M (with coefficients in $\mathcal{V}$) lift to G-invariant forms on the universal cover $\tilde{M}$ of M and can be interpreted as *G-equivariant harmonic maps* $\tilde{M} \to \mathcal{H}$, where G acts on $\tilde{M}$ *via* covering transformations and on $\mathcal{H}$ *via* the an affine isometric action ρ whose linear part is the representation π and the translational part is given by the 1-cocycle $c : G \to \mathcal{H}$ representing the cohomology class ω. We note that the existence of equivariant harmonic maps with respect to group actions on CAT(0) spaces was established by Korevaar and Schoen in much greater generality than the one of Hilbert space targets, see [**KS97**].

This line of reasoning was extended by H. Garland [**Gar73**] and, later, in greater generality, A. Żuk [**Żuk96, Żuk03**] and by W. Ballmann and J. Swiatkowski, [**BS97a**], to the setting when one replaces the free action $G \curvearrowright \tilde{M}$ with a properly discontinuous, isometric and cocompact action on piecewise-Euclidean simplicial complexes, $G \curvearrowright X$. (The latter actions are not required to be free.) We now describe the combinatorial conditions on X (replacing geometric conditions on M in the classical setting of Bochner technique), which lead to vanishing of H^1 and, hence, to combinatorial examples of groups satisfying the Property T.

Below is we describe a combinatorial replacement of the Bochner technique; our discussion follows the paper by A. Żuk, [**Żuk03**]. Let S be a symmetric finite generating set of a group G, such that $1 \notin S$. Define the following graph $L = L(S)$: Vertices of L are the elements $s \in S$. Two vertices s, s' are connected by an (oriented) edge $[s, s']$ iff

$$s^{-1}s' \in S.$$

The *Laplacian* Δ on the vector space $\ell^2(S)$ is defined by the formula

$$\Delta f(s) = f(s) - \frac{1}{val(s)} \sum_{s' \sim s} f(s'),$$

where $s' \sim s$ iff there exists an edge $[s, s'] \in E(L)$. Let $\lambda_1(L)$ denote the smallest positive eigenvalue of Δ. For instance, suppose that L is the incidence graph of the finite projective plane $P^2 F_q$, where F_q is the field of the order q. Then

$$\lambda_1(L) = 1 - \frac{\sqrt{q}}{q+1}.$$

THEOREM 17.61 (A. Żuk, [**Żuk03**]). *If L is connected and $\lambda = \lambda_1(L) > \frac{1}{2}$ then G has the Property T. Moreover,*

$$\frac{2}{\sqrt{3}} \left(2 - \frac{1}{\lambda}\right)$$

is a Kazhdan constant with respect to S.

17.6. Failure of quasi-isometric invariance of the Property T

THEOREM 17.62. *The Property T is not QI invariant.*

PROOF. This theorem should be probably attributed to S. Gersten and M. Ramachandran; the example below is a variation on the *Raghunathan's example* discussed in [**Ger92**].

Let Γ be a hyperbolic group which satisfies Property (T) and such that $H^2(\Gamma, \mathbb{Z})$ is nontrivial. To construct such a group, start for instance with an infinite hyperbolic group F satisfying Property (T) which has an aspherical presentation complex (see for instance [**BS97a**] for the existence of such groups). Then $H^1(F, \mathbb{Z}) = 0$ (since F satisfies (T)), if $H^2(F, \mathbb{Z}) = 0$, we add more random relations to F, keeping the resulting groups F' hyperbolic, infinite, 2-dimensional. Then $H^1(F', \mathbb{Z}) = 0$ since F' also satisfies (T). For large number of relators we get a group $\Gamma = F'$ such that $\chi(\Gamma) > 0$ (the number of relators is larger than the number of generators), hence $H^2(\Gamma, \mathbb{Z}) \neq 0$. Now, pick a nontrivial element $\omega \in H^2(\Gamma, \mathbb{Z})$ and consider a central extension

$$1 \to \mathbb{Z} \to G \to \Gamma \to 1$$

with the extension class ω. Since the group Γ is hyperbolic, theorem Theorem 9.150 implies that the groups G and $G' := \mathbb{Z} \times \Gamma$ are quasi-isometric (see also [**Ger92**] for a more general version of this argument in the case of central co-extensions defined by *bounded cohomology classes*). The group G' does not satisfy (T), since it surjects to $\mathbb{Z}$. On the other hand, the group G satisfies (T), see [**dlHV89**, 2.c, Theorem 12]. □

In a very recent paper [**CAPV14**], M. Carette, S. Arnt, T. Pillon and A. Valette proved:

THEOREM 17.63. *The Haagerup property is not QI invariant.*

We note that the Haagerup Property is yet another negation of the Property T: A topological group G is said to satisfy the Haagerup Property if there is a proper continuous isometric affine action of G on a Hilbert space.

Examples and non-examples of groups with Property T:

Groups with Property (T)	**Groups without Property (T)**
All Lie groups with simple factors of rank $\geqslant 2$	$O(n,1)$ and $U(n,1)$
Lattices in simple Lie groups of rank $\geqslant 2$	Unbounded subgroups of $O(n,1)$ and $U(n,1)$
$SL(n,\mathbb{Z}), n \geqslant 3$	$SL(2,\mathbb{Z})$
Lattices in the isometry group of $\mathbb{HH}^n, n \geqslant 2$	Lattices in $O(n,1)$ and $U(n,1)$
$SL(n,\mathbb{Z}[t]), n \geqslant 3$	All Mapping class groups
	Thompson group
	All finitely generated infinite Coxeter groups
	Infinite 3-manifold groups
Some hyperbolic groups	Some hyperbolic groups
	Groups which admit nontrivial splittings as amalgams
	Infinite amenable groups
	Infinite fundamental groups of closed conformally-flat manifolds

Property (T) is unclear for the following groups:

- $Out(F_n)$, $n \geqslant 4$.
- Infinite Burnside groups.
- Shephard groups. (Property (T) fails at least for some of these groups.)
- Generalized van Dyck groups. (Property (T) fails at least for some of these groups.)
- Hyperbolic Kähler groups. (Property (T) fails at least for some of these groups.)

17.7. Map of the world of finitely generated groups

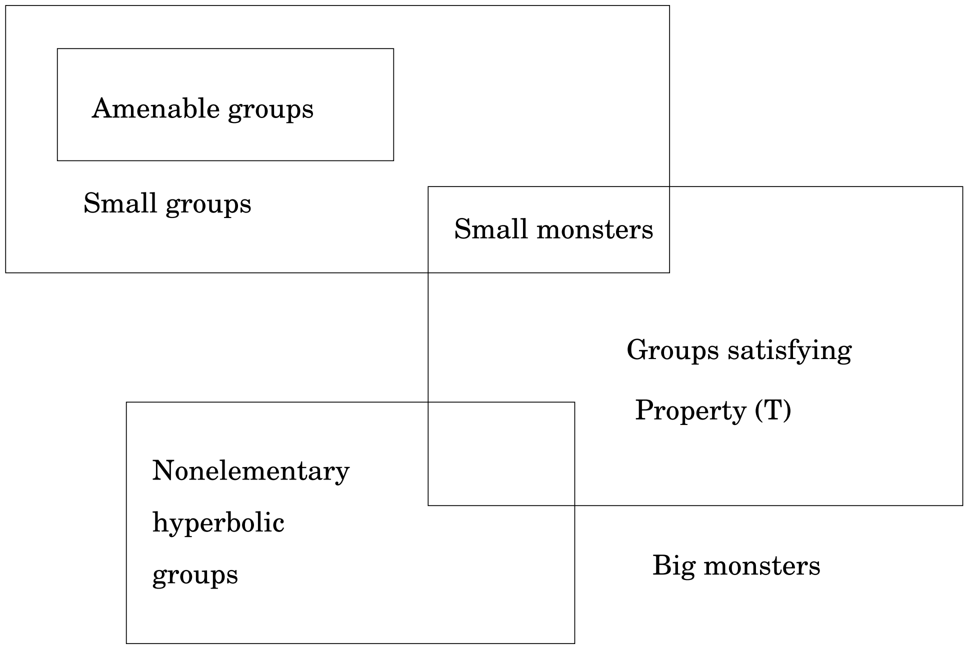


FIGURE 17.1. The world of infinite finitely generated groups.

CHAPTER 18

Stallings Theorem and accessibility

The goal of this chapter is to prove Stallings Theorem (Theorem 6.24) on ends of groups in the class of (almost) finitely presented groups and Dunwoody's Accessibility Theorem for finitely presented groups. As a corollary we obtain QI rigidity of the class of virtually free groups. Our proofs are a geometric combination of arguments due to Dunwoody [**Dun85**], Swarup [**Swa93**] and Jaco and Rubinstein [**JR88**], which are inspired by the theory of minimal surfaces. One advantage of this approach is that in the process we fill in some of the details the theory of *PL minimal surfaces* developed by Jaco and Rubinstein. The definition of *almost finitely presented groups* (abbreviated as afp groups) will be given in Definition 18.27, for now it suffices to note that the class of afp groups contains all finitely presented groups.

THEOREM 18.1 (The Stallings ends of groups theorem for afp groups). *Let G be an afp group with at least 2 ends. Then G splits as the fundamental group of a finite graph of finitely generated groups with finite edge-groups.*

The Stallings theorem allows one to start the decomposition process (using graphs of groups with finite edge groups) of groups with at least two ends. A group is called *accessible* if any such decomposition process terminates after finitely many steps:

THEOREM 18.2 (Dunwoody accessibility theorem). *Every afp group is accessible.*

As a combination of these two fundamental theorems one obtains:

COROLLARY 18.3. *Suppose that G is an afp group with at least 2 ends. Then G splits as the fundamental group of a finite graph of finitely generated groups with finite edge-groups, such that each vertex group is either finite or 1-ended.*

The Stallings theorem, unlike the one by Dunwoody, holds for all finitely generated groups. In the next chapter we prove the Stallings theorem for finitely generated groups using harmonic functions following and idea proposed by Gromov.

18.1. Maps to trees and hyperbolic metrics on 2-dimensional simplicial complexes

Collapsing maps. Let Δ be a 2-dimensional simplex with the vertices $x_i, i = 1, 2, 3$. Our goal is to define a class of maps $\Delta \to Y$, where Y is a simplicial tree with the standard metric (the same could be done when targets are arbitrary real trees but we will not need it). The construction of f is, as usual, by induction on skeleta. This construction is analogous to the construction of collapsing maps κ in

9.8. (The difference with the maps κ is that the maps f will not be isometric on edges, only linear.) Let $f : \Delta^{(0)} = \{x_1, x_2, x_3\} \to Y$ be given. If the image of this map is contained in a geodesic segment α in Y, then we extend f to be a linear map $f : \Delta \to \alpha$. Otherwise, the points of $f(\Delta^{(0)})$ span a tripod T in Y with the centroid o and extreme vertices $y_i := f(x_i)$. We extend f to the map $f : \Delta^{(1)} \to Y$ by sending edges $[x_i, x_{i+1}]$ of Δ to the geodesics $[y_i, y_{i+1}] \subset T$ by linear maps. The preimage $f^{-1}(o)$ consists of three interior points x_{ij} of the edges $[x_i, x_j]$ of Δ, called *center points* of Δ (with respect to f). The 1-dimensional triangle $T(x_{12}, x_{23}, x_{31})$ (called *middle triangle*) splits Δ in four solid sub-triangles $\blacktriangle_i, i = 0, 1, 2, 3$ ($\blacktriangle_0$ is spanned by the center points while each $\blacktriangle_i$ contains x_i as a vertex). Then f sends the vertices of each $\blacktriangle_i$ to points in one of the legs of T. We then extend f to a linear map on each of these four sub-triangles; clearly, $f(\blacktriangle_0) = \{o\}$.

DEFINITION 18.4. The resulting map $f : \Delta \to \tau \subset Y$ is called a *canonical collapsing* map.

It is clear that if X is a simplicial complex and $f : X^{(0)} \to Y$ is a map, then f admits a unique extension to $f : X \to Y$ which is linear on every edge of X and is a canonical collapsing map on each 2-simplex. We refer to the map $f : X \to Y$ as a *canonical map* $X \to Y$ (it depends, of course, on the initial map $f : X^{(0)} \to Y$). Suppose that G is a group acting simplicially on X and isometrically on Y. By uniqueness of the extension f from $X^{(0)}$ to X, if $f : X^{(0)} \to Y$ is a G-equivariant map, then its extension $f : X \to Y$ is also G-equivariant. Such an equivariant map $f : X \to Y$ is called a *canonical resolution* of the G-tree Y.

Existence of resolutions of simplicial G-trees. Recall that every finite group acting isometrically on a real tree T has a fixed point (Corollary 2.70 and Exercise 2.71). If T is a simplicial tree with the standard metric and the action is without inversions, then G has to fix a vertex of T (since a fixed point in the interior of an edge implies that the edge is fixed pointwise).

Let T be a simplicial tree and $G \curvearrowright T$ be a cocompact simplicial action (without inversions). Let X be a connected simplicial 2-dimensional complex on which G acts properly discontinuously and cocompactly (possibly non-freely). We construct a resolution $f : X \to T$ as follows. Let $v \in X^{(0)}$ be a vertex. This vertex has finite stabilizer G_v in G, therefore, this stabilizer fixes a vertex w in T. We then set $f(v) := w$. (If the fixed vertex is not unique then we choose it arbitrarily.) We then extend this map to the orbit $G \cdot v$ by equivariance. Repeating this for each vertex-orbit we obtain an equivariant map $f : X^{(0)} \to T^{(0)}$. Note that without loss of generality, by subdividing X barycentrically if necessary, we may assume that $f : X^{(0)} \to T^{(0)}$ is onto (all that we need for this is that X/G has more vertices than T/G). We then extend f to the rest of X by the canonical collapsing map, therefore obtaining the resolution.

Piecewise-canonical maps. In the proof of Theorem 18.39 we will need a mild generalization of the canonical maps and resolutions. Suppose that in the 2-simplex Δ we are given a subdivision into the solid triangles $\blacktriangle_i, i = 0, ..., 3$ with vertices x_i, x_{jk}. Suppose we are also given a structure of a polygonal cell complex P on Δ such that:

(1) Every vertex belongs to the boundary of Δ.
(2) Every edge is geodesic.
(3) Every segment $[x_{ij}, x_{jk}]$ is an edge.

(4) Every vertex has valence 3 except for $x_{jk}, x_i, i, j, k = 1, 2, 3$.

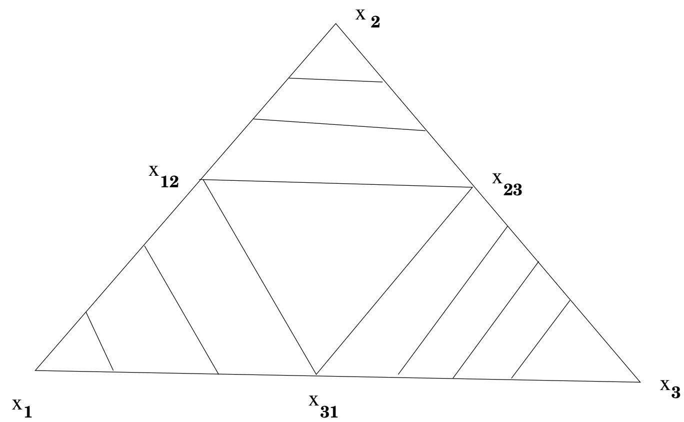


FIGURE 18.1. *P*olygonal subdivision of a simplex.

Edges of P not contained in the boundary of Δ are called *interior edges*.

DEFINITION 18.5. A map $f : \Delta \to Y$ is called *piecewise-canonical* (PC) if it is constant on every interior edge and linear on each 2-cell. Note that the map f could be constant on some 2-faces of P (for instance, it is always constant on the solid middle triangle).

Clearly, a map f of the 1-skeleton of P which is constant on every interior edge, admits a unique PC extension to Δ. A map $f : X \to Y$ from a simplicial complex to a tree is *piecewise-canonical* (PC) if it is PC on every 2-simplex of X and piecewise-linear on every edge not contained in a 2-face. Every canonical map $f : \Delta \to Y$ is also PC: The vertices of P are the points x_i, x_{jk}.

Let X be a simplicial complex, Y a simplicial tree and $f : X \to Y$ be an PC map. We say that a point $y \in Y$ is a *regular value* of f if for every 2-simplex Δ in X we have:
a. $f^{-1}(y)$ is disjoint from the vertex set of Δ.
b. $f^{-1}(y)$ is either empty or is a single topological arc (which necessarily connects distinct edges of Δ).

A point $y \in Y$ which is not a regular value of f is called a *critical value* of f. The following is an analogue of Sard's Theorem in the context of PC maps.

LEMMA 18.6. *Let X be a countable simplicial complex, Y a simplicial tree and $f : X \to Y$ be a PC map Then almost every point $y \in Y$ is a regular point of f.*

PROOF. Let $\Delta \subset X$ be a 2-simplex and P be its polygonal cell complex structure. Then there are only finitely many critical values of f, namely the images of the vertices of Δ and of all the 2-faces of P where f is constant. Since X is countable, this means that the set of critical values of f is at most countable. □

Complete hyperbolic metrics on punctured 2-dimensional simplicial complexes. Our next goal is to introduce a path metric on $X' := X \setminus X^{(0)}$, such that each 2-simplex (minus vertices) is isometric to a solid ideal hyperbolic triangle.

Proposition 18.7. *Let X be a locally finite 2-dimensional simplicial complex. Then there exists a proper path-metric on $X' := X \setminus X^{(0)}$ such that each 2-simplex in X is isometric to the ideal hyperbolic triangle. Moreover, this metric is invariant under all automorphisms of X.*

Proof. We identify each 2-simplex s in X with the solid ideal hyperbolic triangle $\blacktriangle$ (so that vertices of s correspond to the ideal vertices of the hyperbolic triangle). We now would like to glue edges of the solid triangles isometrically according to the combinatorics of the complex X. However, this identification is not unique since each complete geodesic in $\mathbb{H}^2$ is invariant under a group of translations. Moreover, some of the identifications will yield incomplete hyperbolic metrics. (Even if we glue two ideal triangles along their boundaries!) Therefore, we have to choose gluing isometries appropriately.

The ideal triangle $\blacktriangle$ admits a unique inscribed circle; the points of tangency of this circle and the sides τ_{ij} of $\blacktriangle$ are the *central points* $x_{ij} \in \tau_{ij}$, see Section 9.8.

Now, given two solid ideal triangles $\blacktriangle_i, i = 1, 2$ and oriented sides $\tau_i, i = 1, 2$ of these triangles, there is a unique isometry $\tau_1 \to \tau_2$ which sends center-point to center-point and preserves orientation. We use these gluings to obtain a path-metric on X'. Clearly, this metric is invariant under all automorphisms of X in the following sense:

If $g \in Aut(X)$ then the restriction of g to $X^{(0)}$ admits a unique extension $\widehat{g} : X \to X$ which is an isometry of X'.

We claim that X' is proper. The proof relies upon a certain collection of functions b_ξ on X' defined below, $\xi \in X^{(0)}$.

We first define three functions b_1, b_2, b_3 on the ideal triangle $\blacktriangle$. Let ξ_i, $i = 1, 2, 3$, be the ideal vertices of $\blacktriangle$, τ_k the ideal edge connecting ξ_i to ξ_j; $\xi_{ij} \in \tau_k$ be the central point ($k = i + j \bmod 3$).

Each pair of central points ξ_{ij}, ξ_{jk} belongs to a unique horocycle H_j in $\mathbb{H}^2$ with the ideal center ξ_j. One can see this using the upper half-plane model of $\mathbb{H}^2$ so that $\xi_j = \infty$. Then H_j is the horizontal line passing through the points ξ_{ij} and ξ_{jk}.

Consider circular arcs $\alpha_i := H_i \cap \blacktriangle$. The arcs $\alpha_1, \alpha_2, \alpha_3$ cut out a solid triangle $\blacktriangledown$ (with horocyclic arcs α_i's as it edges) from $\blacktriangle$. We refer to the complementary components C_i of $\blacktriangle \setminus \tau$ as *corners* of $\blacktriangle$ with the ideal vertices ξ_i, $i = 1, 2, 3$. Their closures in Δ are the *closed corners* $\overline{C}_i$. We then define a 1-Lipschitz function $b_i : \overline{C}_i \to \mathbb{R}_+$ by

$$b_i(x) = \operatorname{dist}(x, \alpha_i), \quad i = 1, 2, 3.$$

The level sets of $b_i : C_i \to \mathbb{R}_+$ are arcs of horocycles in C_i. (The functions b_i are the negatives of *Busemann functions*, see [**Bal95**].) We extend each b_i by zero to $\blacktriangle \setminus C_i$.

For each vertex ξ of X we define $\overline{C}_\xi$ to be the union of closed corners $\overline{C}_i$ (with the vertex $\xi = \xi_i$) of 2–simplices $s \subset X$ which have ξ as a vertex. Then the functions $b_i : \overline{C}_i \to \mathbb{R}_+$ match on intersections of their domains (since the central points do), thus, we obtain a collection of 1-Lipschitz functions $b_\xi : X' \to \mathbb{R}_+$. It is clear from the construction that each b_ξ is proper on $\overline{C}_\xi$.

Set

$$\blacktriangle_r := \{x \in \blacktriangle : \forall i \quad b_i(x) \leqslant r\}, \quad X_r := \{x \in X' : \forall \xi \in X^{(0)}, b_\xi(x) \leqslant r\}.$$

Since each b_ξ is 1-Lipschitz, for every path $\mathfrak{p}$ in X' of length $\leqslant r$, if $Image(\mathfrak{p}) \cap X_0 \neq \emptyset$ then $p \subset X_r$.

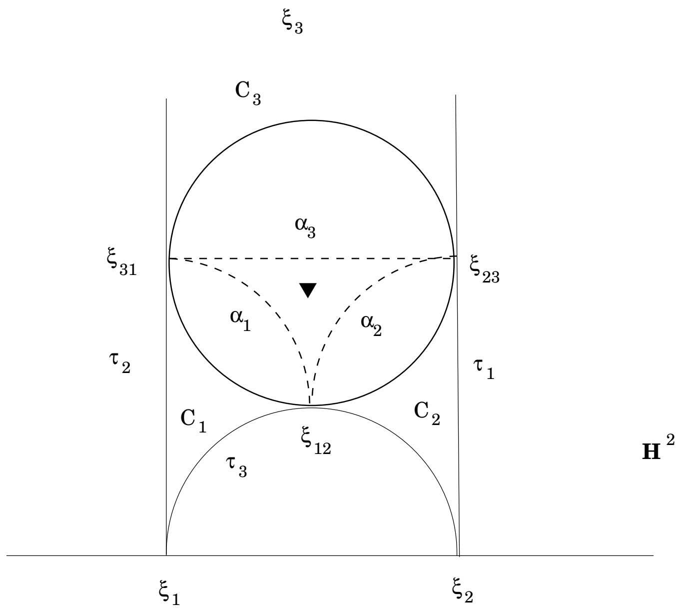


FIGURE 18.2. *G*eometry of the ideal hyperbolic triangle.

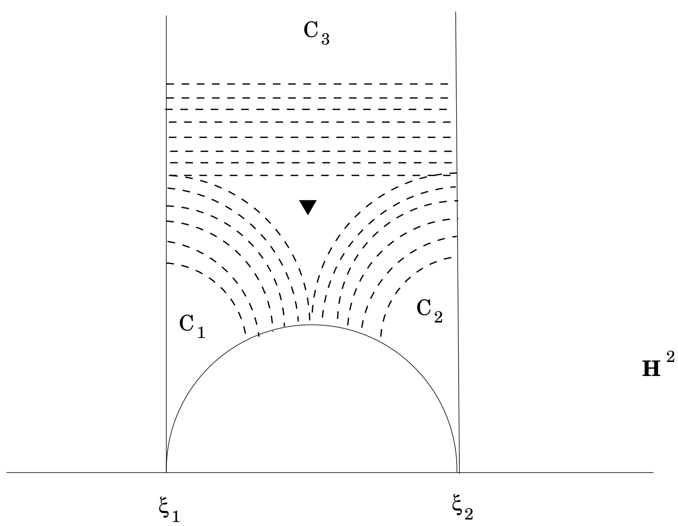


FIGURE 18.3. *P*artial foliation of corners of an ideal triangle by level sets of the functions b_i.

For $x, y \in X'$ we let $\rho(x, y)$ be the minimal number of edges that a path in X' from x to y has to intersect. Since X is locally finite, for every $x \in X$, $k \in \mathbb{N}$, the set $\{y \in X' : \rho(x, y) \leqslant k\}$ is a union of finitely many cells.

Every ideal side τ_i of $\blacktriangle$ intersects $\blacktriangle_r$ in a compact subset. Thus, there exists $D(r) > 0$ such that the minimal distance between the geodesics

$$\tau_i \cap \blacktriangle_r, \tau_j \cap \blacktriangle_r \quad (i \neq j)$$

is at least $D(r)$. Therefore, if $\mathfrak{p}$ is a path in X_r connecting x to y, then its length is at least

$$D(r)\rho(x,y).$$

Thus, for every $x \in X_0$ the metric ball $B_r(x)$ intersects only finitely many cells in X and is contained in X_r. Since intersection of X_r with any finite subcomplex in X is compact, it is now immediate that X' is a proper metric space. ☐

18.2. Transversal graphs and Dunwoody tracks

We continue with the notation of the previous section.

Our goal is to introduce for X' notions analogous to transversality in the theory of smooth manifolds. We define the *vertex-complexity* of a finite graph Γ, denoted $\nu(\Gamma)$, to be cardinality of the vertex set $V(\Gamma)$. We say that a properly embedded graph $\Gamma \subset X'$ is *transversal* if the following hold:

1. $\Gamma \cap X^{(1)} = V(\Gamma) = \Gamma^{(0)}$.
2. Γ intersects every 2-face of X along a (possibly empty) disjoint union of edges.

Transversal graphs generalize the concept of properly embedded *smooth codimension 1 submanifolds* in a smooth manifold. Note that every transversal graph satisfies the property:

For every edge $e \subset X^{(1)}$, every pair of distinct 2-faces $s_1, s_2 \subset X$ containing e and every vertex $v \in \Gamma \cap e$, there are exactly two edges $\gamma_1 \subset \Gamma \cap s_1, \gamma_2 \subset \Gamma \cap s_2$ which have v as a vertex.

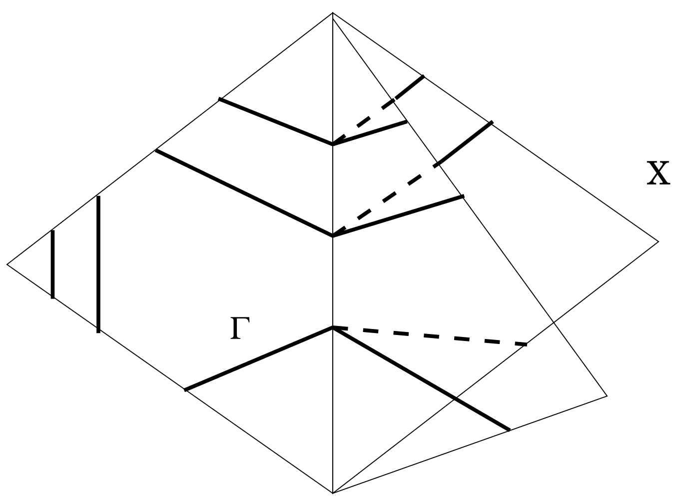


FIGURE 18.4. Dunwoody track.

If a transversal graph Γ satisfies the property:

3. For every edge γ of Γ, the end-points of γ belong to distinct edges of $X^{(1)}$,

then Γ is called a *Dunwoody track*, or simply a *track*.

EXERCISE 18.8. Let $f : X \to Y$ be a PC map from a simplicial 2-complex X to a simplicial tree Y and $y \in Y$ be a regular value of f. Then $f^{-1}(y)$ is a Dunwoody track in X.

The following lemma is left as an exercise to the reader, it shows that every Dunwoody track in X behaves like a codimension one smooth submanifold in a differentiable manifold.

LEMMA 18.9. *Let Γ be a Dunwoody track. Then for every $x \in \Gamma$ there exists a neighborhood U of x which is naturally homeomorphic to the product*

$$\Gamma_U \times [-1, 1],$$

where $\Gamma_U = \Gamma \cap U$ and the above homeomorphism sends Γ_U to $\Gamma_U \times \{0\}$. We will refer to the neighborhoods U as product neighborhoods *of points of Γ.*

Note that the entire track need not have a product neighborhood. For instance, let Γ be a non-separating loop in the Moebius band X. Triangulate X so that Γ is a track. Then every regular neighborhood of Γ in X is again a Moebius band. However, the neighborhoods Γ_U combine in a neighborhood N_Γ of Γ in X which is an interval bundle over Γ, where the product neighborhoods $U \cong \Gamma_U \times [-1, 1]$ above serve as coordinate neighborhoods in the fibration.

We say that the track Γ is *1-sided* if the interval bundle $N_\Gamma \to \Gamma$ is non-trivial and *2-sided* otherwise.

EXERCISE 18.10. Suppose that Γ is connected and 1-sided. Then $N_\Gamma \setminus \Gamma$ is connected.

For each edge transversal graph $\Gamma \subset X$ we define the *counting function* m_Γ : $Edges(X) \to \mathbb{Z}$:

$$m_\Gamma(e) := |\Gamma \cap e|.$$

The $\mathbb{Z}_2$-cocycle of a transversal graph. Recall that, by the Poincaré duality, every proper codimension k embedding of smooth manifolds

$$N \hookrightarrow M$$

defines an element $[N]$ of $H^k(M, \mathbb{Z}_2)$. Our goal is to introduce a similar concept for transversal graphs Γ in X. Observe that for every 2-face s in X

$$\sum_{i=1}^{3} m_\Gamma(e_i) = 0, \quad \mod 2,$$

where e_1, e_2, e_3 are the edges of s (since every edge γ of $\Gamma \cap s$ contributes zero to this sum. Therefore, m_Γ determines an element of $Z^1(X, \mathbb{Z}_2)$. If Γ is finite, then $m_\Gamma \in Z^1_c(X, \mathbb{Z}_2)$ since the cocycle m_Γ is supported only on the finitely many edges which cross Γ. We let $[\Gamma]$ denote the cohomology class in $H^1_c(X, \mathbb{Z}_2)$ determined by m_Γ. It is clear that $[\Gamma]$ depends only on the isotopy class of Γ.

LEMMA 18.11. *Suppose that Γ is 1-sided. Then m_Γ represents a nontrivial class in $H^1(X, \mathbb{Z}_2)$.*

PROOF. We first subdivide X so that N_Γ is a subcomplex in X and will compute $H^1(X, \mathbb{Z}_2)$ using the new cell complex denoted X'. We will use the notation m for the cocycle on X' defined by Γ. Since Γ is 1-sided, there are vertices u, v of X' which belong to $\partial N_\Gamma \cap X^{(1)}$, such that:

1. The edge $\tau = [u, v]$ of X' connecting u and v is contained in an edge e of X.
2. The edge τ intersects Γ in exactly one point.

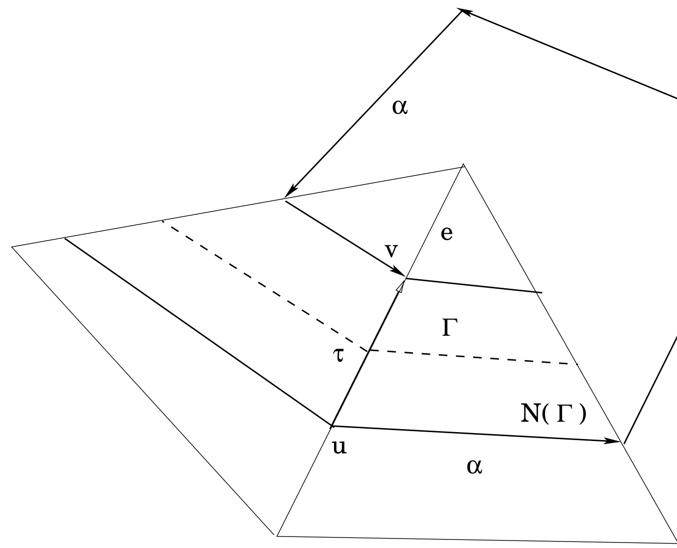


FIGURE 18.5. Nontriviality of a cocycle.

Hence,

$$m(u) + m(v) = 1 \pmod 2 \tag{18.1}$$

Since $N_\Gamma \setminus \Gamma$ is connected, there exists a path $\alpha \subset \partial N_\Gamma$ connecting u to v. Suppose that $m = \delta\eta$, where $\eta \in C^0(X', \mathbb{Z}_2)$. In other words, $\eta : (X')^{(0)} \to \mathbb{Z}_2$ and for every pair of vertices p, q of X' connected by the edge pq, we have:

$$\eta(p) - \eta(q) = m([p, q]).$$

In particular, if p, q are connected by an edge-path in X' which is disjoint from Γ, then $\eta(p) = \eta(q)$. Since the path α connecting u to v is disjoint from Γ, we obtain

$$\eta(u) = \eta(v).$$

On the other hand, in view of (18.1), we also have

$$\eta(u) + \eta(v) = 1$$

Contradiction. □

LEMMA 18.12. *Suppose that $H^1(X, \mathbb{Z}_2) = 0$. Then a connected finite transversal graph Γ separates X into at least two unbounded components if and only if $[\Gamma]$ is a nontrivial class in $H^1_c(X, \mathbb{Z}_2)$. Such a track is said to be* essential.

PROOF. The proof is similar to the argument of Lemma 18.11.

1. Suppose that $X \setminus \Gamma$ contains at least two unbounded complementary components U and V, but $[\Gamma] = 0$ in $H^1_c(X, \mathbb{Z}_2)$. Then there exists a compactly-supported function $\sigma : X^{(0)} \to \mathbb{Z}_2$ such that $\delta(\sigma) = m_\Gamma$, mod 2. Since σ is compactly supported, there exists a compact subset $C \subset X$ such that $\sigma = 0$ on $U \setminus C, V \setminus C$. Let $\alpha \subset X^{(1)}$ be a path connecting $u \in U \cap X^{(0)}$ to $v \in V \cap X^{(0)}$. We leave it to the reader to verify that if an edge $e = [x, y]$ of X crosses Γ in an even number of points then x, y belong to the same connected component of $X \setminus \Gamma$ (this is the only place

where we use the assumption that Γ is connected). Therefore, the path α crosses Γ in an odd number of points, which implies that

$$\langle m_\Gamma, \alpha \rangle = 1 \in \mathbb{Z}_2.$$

However,

$$\langle m_\Gamma, \alpha \rangle = \langle \sigma, \partial\alpha \rangle = \sigma(u) + \sigma(v) = 0.$$

Contradiction.

2. Suppose that $[\Gamma] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$. Since $H^1(X, \mathbb{Z}_2) = 0$, there exists a 0-cochain $\sigma : X^{(0)} \to \mathbb{Z}_2$ such that

$$\delta\sigma = m_\Gamma.$$

Since m_Γ takes nonzero value on some edge $e = [u, v]$ of X, we obtain $\sigma(u) = 0, \sigma(v) = 1$. If the set $\sigma^{-1}(1) \subset X^{(0)}$ is finite, then $\sigma \in C^0_c(X, \mathbb{Z}_2)$ and, hence $[\Gamma] = 0$, which is a contradiction. Therefore, the set of such vertices is unbounded. Consider another 0-cochain $\sigma+1$ (which equals to $\sigma(x)+1$ on every vertex $x \in X^{(0)}$). Then $\delta(\sigma + 1) = \delta\sigma = m$ and

$$\{w \in X^{(0)} : \sigma(w) = 0\} = \{w \in X^{(0)} : \sigma(u) + 1 = 1\}.$$

Therefore, by the above argument, the set $\sigma^{-1}(0) \subset X^{(0)}$ is also unbounded. Thus, since Γ is a finite graph, there are unbounded connected subsets $U, V \subset X \setminus \Gamma$ such that

$$\forall u \in U \cap X^{(0)}, \sigma(u) = 0, \quad \forall v \in V \cap X^{(0)}, \sigma(v) = 1.$$

These are the required unbounded complementary components of Γ. □

EXERCISE 18.13. If $H^1(X, \mathbb{Z}_2) = 0$, then every connected essential Dunwoody track $\Gamma \subset X$ has exactly two complementary components, both of which are unbounded. We will use the notation $\Gamma^\pm$ for these components.

The following key lemma due to Dunwoody is a direct generalization of the Kneser–Haken finiteness theorem for triangulated 3-dimensional manifolds, see e.g. [**Hem78**].

LEMMA 18.14 (M. Dunwoody). *Suppose that X has F faces and $H^1(X, \mathbb{Z}_2) \cong \mathbb{Z}_2^r$. Suppose that $\Gamma_1, ..., \Gamma_k$ are pairwise disjoint pairwise non-isotopic connected tracks in X. Then*

$$k \leqslant 6F + r.$$

PROOF. The union Γ of tracks Γ_i cuts each 2-simplex s in X in triangles, quadrilaterals and hexagons. Note that some of the complementary rectangles might contain vertices of X. In what follows, we regard such rectangles as *degenerate hexagons* (and not as rectangles). The boundary of each complementary rectangle has two disjoint edges contained in $X^{(1)}$, we call these edges *vertical*. Consider an edge of Γ which is contained in the boundary of a complementary triangle or a (possibly degenerate) hexagon. The number of such edges is at most $6F$. Thus, the number of tracks Γ_i containing such edges is at most $6F$ as well. We now remove from X the union of closures of all components of $X \setminus \Gamma$ which contain complementary triangles and (possibly degenerate) hexagons.

The remainder R is a union of rectangles Q_j glued together along their vertical edges. Therefore, R is homeomorphic to an open interval bundle over a track $\Lambda \subset X$: The edges of Λ are geodesics connecting midpoints of vertical edges of Q_j's.

If a component R_i of R is a trivial interval bundle then the boundary of R_i is the union of tracks Γ_j, Γ_k which are therefore isotopic. This contradicts our assumption on the tracks Γ_i. Therefore, each component of R is a nontrivial interval bundle. For each R_i we define the cohomology class $[\Lambda_i] \in H^1(X, \mathbb{Z}_2) = H^1_c(X, \mathbb{Z}_2)$ (using the counting function m_Λ). We claim that these classes are linearly independent. Suppose to the contrary that

$$\sum_{i=1}^{\ell} [\Lambda_i] = 0.$$

This means that the track $\Lambda' := \Lambda_1 \cup ... \cup \Lambda_\ell$ determines a trivial cohomology class $[\Lambda'] = 0$. Since Λ' is 1-sided, we obtain a contradiction with Lemma 18.11. Therefore, the number of components of R is at most r, the dimension of $H^1(X, \mathbb{Z}_2)$. Each component of R is bounded by a track Γ_i. Therefore, the total number of tracks Γ_i is at most $6F + r$. □

18.3. Existence of minimal Dunwoody tracks

Our next goal is to deform finite transversal graphs to Dunwoody tracks, so that the cohomology class is preserved and so that the counting function $m_\Gamma : Edges(X) \to \mathbb{Z}$ decreases as the result of the deformation. To this end, we define the operation *pull* on transversal graphs $\Gamma \subset X$.

Pull. Suppose that v_1, v_2 are distinct vertices of Γ which belong to a common edge e of X and which are not separated by any vertex of $\Gamma \cap e$ on e. We call such vertex pair $\{v_1, v_2\}$ *innermost.* Then for every 2-face s of X containing e and every pair of distinct edges $\gamma_i = [u_i, v_i], i = 1, 2$ of Γ we perform the following operation. We replace $\gamma_1 \cup \gamma_2 \subset \Gamma$ by a single edge $\gamma = [u_1, u_2] \subset s$, keeping the rest of $\Gamma' = \Gamma \setminus \gamma_1 \cup \gamma_2$ unchanged, so that γ intersects Γ' only at the end-points u_1, u_2. In case $\gamma_1 = \gamma_2$ we simply eliminate this edge from Γ. Let $pull(\Gamma)$ denote the resulting graph. It is clear that $\nu(pull(\Gamma)) < \nu(\Gamma)$ and $pull(\Gamma)$ is again a transversal graph. Note that *a priori*, $pull(\Gamma)$ need not be connected even if Γ is.

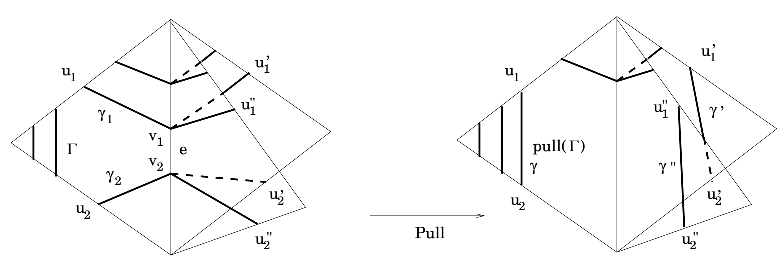


FIGURE 18.6. *P*ull.

EXERCISE 18.15. Verify that $[pull(\Gamma)] = [\Gamma]$; actually, the functions m_Γ and $m_{pull(\Gamma)}$ are equal as $\mathbb{Z}_2$-cochains.

LEMMA 18.16. *Given a finite transversal graph $\Gamma \subset X$, there exists a finite sequence of pull-operations which transforms Γ to a new graph Γ'; the graph Γ' is a track such that for every edge e, $m_{\Gamma'}(e) \in \{0, 1\}$. Moreover, $[\Gamma] = [\Gamma']$.*

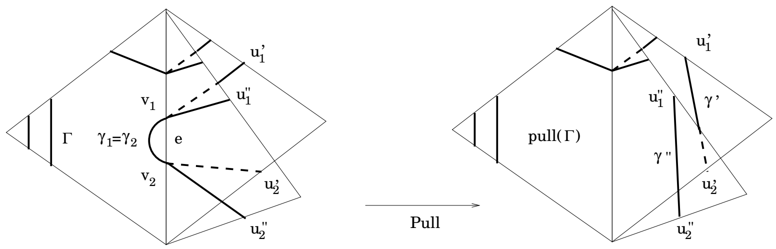


FIGURE 18.7. Eliminating edge $\gamma_1 = \gamma_2$. In this example, $u_1 = v_2, u_2 = v_1$

PROOF. We apply the pull-operation to Γ as long as possible; since the vertex-complexity under *pull* is decreasing, this process terminates at some transversal graph Γ'. If $m_{\Gamma'}(e) \geqslant 2$ for some edge e of X, we can again perform *pull* using an innermost pair of vertices of Γ' on e, which is a contradiction. Since *pull* preserves the cohomology class of a transversal graph, $[\Gamma] = [\Gamma']$. $\square$

LEMMA 18.17. *Assume that* $H^1(X, \mathbb{Z}_2) = 0$. *If* $|Ends(X)| > 1$ *then there exists a connected essential transversal graph* $\Gamma \subset X$.

PROOF. Let ϵ_+, ϵ_- be distinct ends of X. We claim that there exists a proper 1-Lipschitz function $h : X \to \mathbb{R}$ such that

$$\lim_{x \to \epsilon_\pm} = \pm\infty.$$

Indeed, let K be a compact which separates the ends ϵ_+, ϵ_-. We define h to be constant on K. We temporarily re-metrize X by equipping it with the standard metric (every simplex is isometric to the standard Euclidean simplex with unit edges). Let $U_\pm$ be the unbounded components of $X \setminus K$ which are neighborhoods of the ends $\epsilon_\pm$. We then set

$$h|_{U_\pm} := \pm \mathrm{dist}(\cdot, K).$$

For every other component V of $X \setminus K$ we set

$$h|_V := \mathrm{dist}(\cdot, K).$$

It is immediate that this function satisfies the required properties. We give $\mathbb{R}$ the structure of a simplicial tree T, where integers serve as vertices. We next approximate h by a proper canonical map $f : X \to T$. Namely, for every vertex v of X we let $f(v)$ be a vertex of T within distance $\leqslant 1$ from $f(v)$. We extend $f : X^{(0)} \to T$ to a canonical map $f : X \to T$. Then $\mathrm{dist}(f, h) \leqslant 3$ and, hence, f is again proper. By Lemma 18.6, $\Gamma := f^{-1}(y)$ is a finite transversal graph separating ϵ_+ from ϵ_- for almost every y. Hence, $[\Gamma] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$. The graph Γ need not be connected, let $\Gamma_1, ..., \Gamma_n$ be its connected components: They are still transversal graphs. Thus,

$$[\Gamma] = \sum_{i=1}^{n} [\Gamma_i],$$

which implies that at least one of the graphs Γ_i is essential. $\square$

Note, that the graph Γ constructed in the above proof need not be a Dunwoody track. However, Lemma 18.16 implies that we can replace Γ with a essential Dunwoody track Γ' which intersects every edge in at most one point. The graph Γ' need not be connected, but it has an essential connected component. Therefore:

COROLLARY 18.18. *Assume that $H^1(X, \mathbb{Z}_2) = 0$ and $|Ends(X)| > 1$. Then there exists a connected essential Dunwoody track $\Gamma \subset X$. Moreover,*

$$m_\Gamma : Edges(X) \to \{0, 1\}.$$

We define the *complexity* of a transversal graph $\Gamma \subset X$, denoted $c(\Gamma)$, to be the pair $(\nu(\Gamma), \ell(\Gamma))$, where $\nu(\Gamma)$ is the number of vertices in Γ and $\ell(\Gamma)$ is the total length of Γ. We give the set of complexities the lexicographic order. It is clear that $c(\Gamma)$ is preserved by isometric actions $G \curvearrowright X'$.

An essential connected Dunwoody track $\Gamma \subset X$ is said to be *minimal* if it has minimal complexity among all connected essential Dunwoody tracks in X.

DEFINITION 18.19. A vertex v of X is said to be a *cut-vertex* if $X \setminus \{v\}$ contains at least two unbounded components. (Note that our definition is slightly stronger than the usual definition of a cut-vertex, where it is only assumed that $X \setminus \{v\}$ is not connected.)

LEMMA 18.20. *Suppose that X admits a cocompact simplicial action $G \curvearrowright X$, $H^1(X, \mathbb{Z}_2) = 0$, $|Ends(X)| > 1$ and X has no cut-vertices. Then there exists a (connected and essential) minimal track Γ_{min}.*

PROOF. By Corollary 18.18, the set of connected essential tracks in X is nonempty. Let Γ_i be a sequence of such graphs whose complexity converges to the infimum. Without loss of generality, we can assume that each Γ_i has minimal vertex-complexity $\nu = \nu(\Gamma)$ among all connected essential tracks in X. Since X is a simplicial complex, it is easy to see that each Γ_i is also a simplicial complex. Therefore, the number of edges of the graphs Γ_i is also uniformly bounded (by $\frac{\nu(\nu-1)}{2}$). In particular, there are only finitely many combinatorial types of these graphs; therefore, after passing to a subsequence, we can assume that the graphs Γ_i are combinatorially isomorphic to a fixed graph Γ.

Replace each edge of Γ_i with the hyperbolic geodesic (in the appropriate 2-simplex of X). This does not increase the complexity of Γ_i, keeps the graph embedded and preserves all the properties of Dunwoody tracks. Therefore, we will assume that each edge of Γ_i is geodesic. We let $h_i : \Gamma \to \Gamma_i$ be graph isomorphisms. Since $\ell(\Gamma_i)$ are uniformly bounded from above, there exists a path-metric on Γ such that all the maps h_i are 1-Lipschitz. We let $\bar{h}_i$ denote the composition of h_i with the quotient map $X \to Y = X/G$. If there exists a compact set $C \subset Y' := X'/G$ such that $\bar{h}_i(\Gamma) \cap C \neq \emptyset$, then the Arzela–Ascoli Theorem implies that the sequence $(\bar{h}_i)$ subconverges to a 1-Lipschitz map $\Gamma \to Y'$. On the other hand, if such compact does not exist, then, since edges of Y' have infinite length and Γ is connected, the sequence of maps $\bar{h}_i$ subconverges to a constant map sending Γ to one of the vertices of Y. Hence, in this case, by post-composing the maps h_i with $g_i \in G$, we conclude that the sequence $g_i \circ h_i$ subconverges to a constant map whose image is one of the vertices of X. Recall that, by our assumption, X has no cut-vertices. Therefore, every sufficiently small neighborhood of a vertex v of X does not separate X into several unbounded components. This contradicts the assumption that each Γ_i is essential.

Therefore, by replacing h_i with $g_i \circ h_i$ (and preserving the notation h_i for the resulting maps), we conclude that the maps h_i subconverge to a 1-Lipschitz map $h : \Gamma \to X'$. In view of Lemma 18.16 (and the fact that Γ_i's have minimal vertex-complexity), for every face s and edge $e \subset s$ of X there exists at most one edge of Γ_i contained in s and intersecting e. Therefore, the map h is injective and $\Gamma_{min} = h(\Gamma)$ is a track in X. Moreover, for each sufficiently large i, the graph Γ_{min} is isotopic to Γ_i as they have the same counting function $m_\Gamma = m_{\Gamma_i}$. Thus, Γ_{min} is essential. Therefore, it is the required minimal track. ☐

18.4. Properties of minimal tracks

18.4.1. Stationarity. The following discussion is local and does not require any assumptions on $H^1(X, \mathbb{Z}_2)$.

We say that a transversal graph Γ is *stationary* if for every small smooth isotopy Γ_t of Γ (through transversal graphs), with $\Gamma_0 = \Gamma$, we have

$$\frac{d}{dt}\ell(\Gamma_t)|_{t=0} = 0.$$

In particular, every edge of Γ is geodesic.

EXAMPLE 18.21. Every minimal essential Dunwoody track is stationary.

Let Γ be a Dunwoody track with geodesic edges. Let v be a vertex of Γ which belongs to an edge e of X and γ be an edge of Γ incident to v. We assume that $\gamma = \gamma(t)$ is parameterized by its arc-length so that $\gamma(0) = v$. We define $\pi_e(\gamma')$ to be the orthogonal projection of the vector $\gamma'(0) \in T_e\mathbb{H}^2$ to the tangent line of e at v.

LEMMA 18.22. *If Γ is stationary then for every vertex v as above we have*

$$\sum_{\gamma} \pi_e(\gamma') = 0 \qquad (18.2)$$

where the sum is taken over all edges $\gamma_1, ..., \gamma_k$ of Γ incident to v.

PROOF. We construct a small isotopy Γ_t of Γ by fixing all the vertices and edges of Γ except for the vertex v which is moved along e, so that $v(t), t \in [0, 1]$, is a smooth function. We assume that all edges of Γ_t are geodesic. This variation of v uniquely determines Γ_t. It is clear that

$$0 = \frac{d}{dt}\ell(\Gamma_t)|_{t=0} = \sum_{i=1}^{k} \frac{d}{dt}\ell(\gamma_i(t))|_{t=0}.$$

By the first variation formula (8.11), we conclude that

$$0 = \sum_{i=1}^{k} \frac{d}{dt}\ell(\gamma_i(t))|_{t=0} = \sum_{i=1}^{k} \pi_e(\gamma'). \quad ☐$$

REMARK 18.23. The proof of the above lemma also shows the following. Suppose that Γ fails the stationarity condition (18.2) at a vertex v. Orient the edge e and assume that the vector

$$\sum_{\gamma} \pi_e(\gamma')$$

points to the "right" of zero. Construct a small isotopy Γ_t, $\Gamma_0 = \Gamma, t \in [0,1)$, so that all edges of Γ_t are geodesic, vertices of Γ_t except for v stay fixed, while the vertex $v(t)$ moves to the "right" of $v = v(0)$. Then

$$\ell(\Gamma_t) < \ell(\Gamma)$$

for all small $t > 0$.

LEMMA 18.24 (The Maximum Principle). *Let Λ_1, Λ_2 be stationary Dunwoody tracks. Then in a small product neighborhood U of every common vertex u of these graphs, either the graphs Λ_1, Λ_2 coincide, or one "crosses" the other. The latter means that*

$$\Lambda_1 \cap U_+ \neq \emptyset, \Lambda_2 \cap U_- \neq \emptyset.$$

Here $U_\pm = \Lambda_{1,U} \times (0, \pm 1]$ where we identify the product neighborhood U with $\Lambda_{1,U} \times [-1,1]$, $\Lambda_{1,U} = \Lambda_1 \cap U$. In other words, if $h : U = \Lambda_{1,U} \times [-1,1] \to [-1,1]$ is the projection to the second factor, then $h|\Lambda_2$ cannot have maximum or minimum at u, unless $h|\Lambda_2$ is identically zero.

PROOF. Let e be the edge of X containing u. Since Λ_1, Λ_2 are tracks, every 2-simplex s of X adjacent to e contains (unique) edges $\gamma_{i,s} \subset \Lambda_i, i = 1, 2$, which are incident to u. Suppose that Λ_2 does not cross Λ_1. Then either for every $s, \gamma_i = \gamma_{i,s}, i = 1, 2$ as above,

$$\pi_e(\gamma_1') \geqslant \pi_e(\gamma_2')$$

or for every s, γ_1, γ_2

$$\pi_e(\gamma_1') \geqslant \pi_e(\gamma_2').$$

Since, by the previous lemma,

$$\sum_s \pi_e(\gamma_{i,s}') = 0, \quad i = 1, 2,$$

we conclude that $\pi_e(\gamma_{1,s}') = \pi_e(\gamma_{2,s}')$. Therefore, since any geodesic is uniquely determined by its derivative at a point, it follows that $\gamma_{1,s} = \gamma_{2,s}$ for every 2-simplex s containing e. Thus, $\Lambda_1 \cap U = \Lambda_2 \cap U$. □

18.4.2. Disjointness of essential minimal tracks. The following proposition is the key for the proof of Stallings Theorem presented in the next section:

PROPOSITION 18.25. *If $H^1(X, \mathbb{Z}_2) = 0$ then any two (connected, essential) minimal tracks in X are either equal or disjoint.*

PROOF. Our proof follows [**JR89**]. The central ingredients in the proof are the exchange and round-off arguments as well as the Meeks–Yau trick. All three come from the theory of least area surfaces in 3-dimensional Riemannian manifolds.

Suppose that Λ, M are distinct connected essential minimal tracks which have nonempty intersection.

Step 1: Transversal case. We first present an argument that this is impossible under the assumption that the graphs Λ and M are *transverse to each other*, i.e., $\Lambda \cap \mathrm{M}$ is disjoint from the 1-skeleton of X. Since both Λ, M are essential and connected, each of them separates X into exactly two components, denoted $\Lambda^\pm, M^\pm$; all of these components are unbounded, see Exercise 18.13. We consider the four sets

$$\Lambda^+ \cap \mathrm{M}^+, \Lambda^+ \cap \mathrm{M}^-, \Lambda^- \cap \mathrm{M}^+, \Lambda^- \cap \mathrm{M}^-.$$

Since both Λ, M separate ends of X, at least two of the above sets are unbounded. After relabeling, we obtain that the intersections

$$\Lambda^+ \cap \mathrm{M}^+, \quad \Lambda^- \cap \mathrm{M}^-$$

are unbounded. Observe that

$$\Lambda \cup \mathrm{M} = \partial(\Lambda^+ \cap \mathrm{M}^+) \cup \partial(\Lambda^- \cap \mathrm{M}^-).$$

Set

$$\Gamma_+ := \partial(\Lambda^+ \cap \mathrm{M}^+), \Gamma_- := \partial(\Lambda^- \cap \mathrm{M}^-).$$

It is immediate that both graphs are transversal (here and below we disregard valency 2 vertices of $\Gamma_\pm$ contained in the interiors of 2-simplices of X). Note that, at this point, we do not yet know if the graphs $\Gamma_\pm$ are connected.

We now compare complexity of Λ (which is the same as the complexity of M) and complexities of the graphs Γ_+, Γ_-. After relabeling, $\ell(\Gamma_+) \leqslant \ell(\Gamma_-)$. We leave it to the reader to verify that for both Γ_+, Γ_-, the number of edges is the same as the number of edges of Λ. Clearly, the total length of $\Gamma_+ \cup \Gamma_-$ is the same as $2\ell(\Lambda) = 2\ell(\mathrm{M})$. Therefore,

$$\ell(\Gamma_+) \leqslant \ell(\Lambda).$$

Hence, $c(\Gamma_+) \leqslant c(\Lambda)$. The transition from Λ to the graph Γ_+ is called the *exchange argument*: We replaced parts of Λ with parts of M in order to get Γ_+.

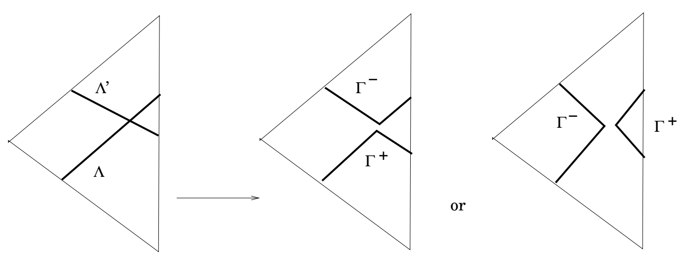


FIGURE 18.8. *E*xchange argument.

By the assumption, the intersection $\Lambda^+ \cap \mathrm{M}^+$ is unbounded. The complement to this intersection contains $\Lambda^- \cap \mathrm{M}^-$, which is also assumed to be unbounded. Therefore, the graph Γ_+ separates ends of X and, hence, $[\Gamma_+] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$, see Lemma 18.12. It is then immediate that at least one connected component of Γ_+ represents a nontrivial element of $H^1_c(X, \mathbb{Z}_2)$. Since $H^1(X, \mathbb{Z}_2) = 0$, this component is essential. By minimality of the graph Λ, it follows that Γ_+ is connected (otherwise we replace it with the above essential component thereby decreasing the vertex-complexity). Since Λ, M cross at a point $x \notin X^{(1)}$, there exists an edge γ of Γ_+ which is a broken geodesic containing x in its interior. (Recall that $X \setminus X^{(0)}$ is equipped with a certain path-metric which is hyperbolic on each 2-dimensional simplex.) Replacing the broken edge γ with a shorter path (and keeping the end-points) we get a new graph $\widehat{\Lambda}$ whose total length is strictly smaller than the one of Γ_+. (This part of the proof is called the “round-off” argument.) We obtain a contradiction with minimality of Λ. This finishes the proof in the case of transversal intersections of Λ and M.

Step 2: Weakly transversal case: Meeks–Yau trick. We assume now that $\Lambda \cap \mathrm{M}$ contains at least one point p of transversal intersection which is not in the 1-skeleton of X. We say that in this situation Λ, M are *weakly transversal* to each other. Note that by doing "exchange and round-off" at p we have some definite reduction in the complexity of the tracks, which depended only on the intersection angle α between Λ, M at p. Then, the weakly transversal case is handled *via* the "original Meeks-Yau trick" [**MY81**], which reduces the proof to the transversal case. This trick was introduced in the work by Meeks and Yau in the context of minimal surfaces in 3-dimensional manifolds and generalized by Jaco and Rubinstein in the context of PL minimal surfaces in 3-manifolds, see [**JR88**]. The idea is to isotope Λ to a (non-minimal) geodesic graph Λ_t, whose total length is slightly larger than Λ but which is transversal to M: $\ell(\Lambda_t) = \ell(\Lambda) + o(t)$.

The the intersection angle α_t between Λ_t and M near p can be made arbitrarily close to the original angle α. Therefore, by taking t small, one can make the complexity loss ϵ to be higher than the length gain $\ell(\Lambda_t) - \ell(\Lambda)$. Then, as in Case 1, we obtain a contradiction with minimality of Λ and M.

Step 3: Non-weakly transversal case. We, thus assume that $\Lambda \cap \mathrm{M}$ contains no points of transversal intersection. (This case does not occur in the context of minimal surfaces in 3-dimensional Riemannian manifolds.) The idea is again to isotope Λ to Λ_t, so that $\ell(\Lambda_t) = \ell(\Lambda) + o(t)$. One then repeats the arguments from Step 1 (exchange and round off) and verifies that the new graph $\widehat{\Lambda}_t$ satisfies

$$\ell(\Lambda_t) - \ell(\widehat{\Lambda}_t) \geqslant O(t).$$

It will then follow that $\ell(\widehat{\Lambda}_t) < \ell(\Lambda)$ when t is sufficiently small, contradicting minimality of Λ.

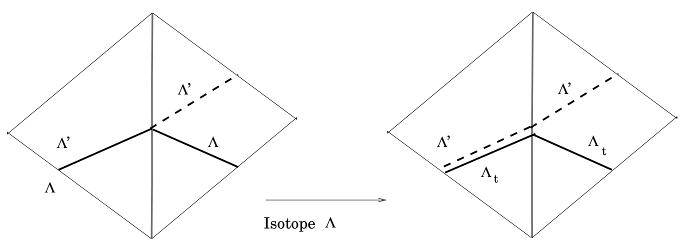


FIGURE 18.9. *M*eeks-Yau trick: Initially, graphs Λ, M had a common edge. After isotopy of Λ, this edge is no longer common. The isotopy Λ_t is through geodesic graphs, which no longer satisfy the balancing condition.

We now provide the details of the Meeks–Yau trick in this setting. We first push the graph Λ in the direction of Λ^+, so that the result is a smooth family of isotopic Dunwoody tracks $\Lambda_t, t \in [0, t_0], \Lambda_0 = \Lambda$, where each Λ_t has geodesic edges and so that each vertex of Λ_t is within distance t from the corresponding vertex of Λ. Since Λ was stationary, we have

$$\ell(L_t) = \ell(\Lambda) + ct^2 + o(t^2).$$

It follows from the Maximum Principle (Lemma 18.24) that the graphs Λ_t and M have to intersect. For sufficiently small values of $t \neq 0$, the intersection is necessarily disjoint from $X^{(1)}$. We now apply the exchange argument and obtain a graph Γ_{t+}, such that

$$\ell(\Gamma_{t+}) \leqslant \ell(\Lambda) + ct^2 + o(t^2).$$

Let $\widehat{\Lambda}_t$ be obtained from Γ_{t+} by the round-off argument (straightening the broken edges). Lastly, we need to estimate from below the difference

$$\ell(\widehat{\Lambda}_t) - \ell(\Gamma_{t+}).$$

It suffices to analyze what happens within a single 2-simplex s of X where the graphs Λ_t and M intersect. We will consider only the most difficult case:

The geodesic segments λ and μ in s, which are components of $\Lambda \cap s$ and $\mathrm{M} \cap \mathrm{s}$ *respectively and such that λ and μ share only their end-point A.*

In particular, the point A belongs to an edge e of s. Furthermore, a component $\lambda_t \subset \Lambda_t \cap s$, such that

$$\lim_{t \to 0} \lambda_t = \lambda_0 := \lambda,$$

has nonempty intersection with M for small $t > 0$.

REMARK 18.26. If such face s and segments λ, μ do not exist, then $\Lambda = \mathrm{M}$. In this situation, the geodesic segments λ_t and μ will be disjoint for small $t > 0$ and nothing interesting happens during the exchange and round-off argument.

We introduce the notation:

$$\lambda = [A, B], \quad \mu = [A, C], \quad \lambda_t = [A_t, B_t], \quad t \in [0, t_0].$$

By construction, $\operatorname{dist}(A, A_t) = t, \operatorname{dist}(B, B_t) = t$. Set $D_t := \mu \cap \lambda_t$. There are several possibilities for the intersection $\Gamma_{t+} \cap s$. If this intersection contains the broken geodesics AD_tA_t or B_tD_tC, then the round-off of Γ_{t+} will result in reduction of the number of edges, contradicting minimality of Λ. We, therefore, consider the case when $\Gamma_{t+} \cap s$ contains the broken geodesic A_tD_tC, as the case of the path AD_tB_t is similar.

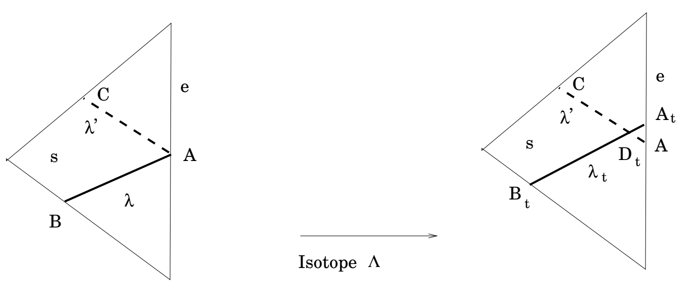


FIGURE 18.10. Meeks-Yau trick: Isotopying edge λ so that $D_t = \lambda_t \cap \mu$ no longer in the edge e.

Consider the triangle $\Delta(A, D_t, A_t)$. We note that the angles of this triangle are bounded away from zero and π if t_0 is sufficiently small. Therefore, the Sine Law for hyperbolic triangles (8.4) implies that $\operatorname{dist}(A_t, D_t) \sim t$ as $t \to 0$. Consider then the triangle $\Delta(A_t, D_t, C)$ (see Figure 18.10). Again, the angles of this triangle are bounded away from zero and π if t_0 is sufficiently small. Therefore, Lemma 8.42 implies that

$$\operatorname{dist}(A_t, D_t) + \operatorname{dist}(D_t, C) - \operatorname{dist}(A_t, C) \geqslant c_1 \operatorname{dist}(A_t, D_t) \geqslant c_2 t = O(t)$$

if t_0 is sufficiently small. Here c_1, c_2 are positive constants. Observe that when we replace Γ_{t+} with $\widehat{\Lambda}_t$ (the round-off), the path $[A_t, D_t] \cup [D_t, C]$ is replaced with the geodesic $[A_t, C]$. Therefore,

$$\ell(\Gamma_{t+}) - \ell(\widehat{\Lambda}_t) = O(t).$$

Since

$$\ell(\Gamma_{t+}) - \ell(\Lambda) = o(t),$$

we conclude that

$$\ell(\widehat{\Lambda}_t) - \ell(\Lambda) < 0$$

if t is sufficiently small. This contradicts minimality of Λ. □

18.5. Stallings Theorem for almost finitely presented groups

DEFINITION 18.27. A group G is said to be *almost finitely presented* (afp) if it admits a properly discontinuous cocompact simplicial action on a 2-dimensional simplicial complex X such that $H^1(X, \mathbb{Z}_2) = 0$.

Note that avery free similicial action is properly discontinuous. Furthermore, in view of Lemma 3.92, in the definition of an afp group one can replace a complex X with a new simplicial complex $\widehat{X}$ which is 2-dimensional, has $H^1(\widehat{X}, \mathbb{Z}_2) = 0$, and the action $G \curvearrowright \widehat{X}$ is free and cocompact.

LEMMA 18.28. *Every finitely presented group G is also afp.*

PROOF. Let Y be a finite presentation complex of G, subdivide it to obtain a triangulated complex W, then let X be the universal cover of W. □

We are now ready to prove

THEOREM 18.29. *Let G be an almost finitely presented group with at least 2 ends. Then G splits as the fundamental group of a finite graph of finitely generated groups with finite edge-groups.*

PROOF. Since G is afp, it admits a properly discontinuous cocompact simplicial action on a (locally finite) 2-dimensional simplicial complex X with $H^1(X, \mathbb{Z}_2) = 0$. We give $X' := X \setminus X^{(0)}$ the piecewise-hyperbolic path metric as in Section 18.1.

DEFINITION 18.30. A subset $Z \subset X$ is called *precisely-invariant* (under its G-stabilizer) if for every $g \in G$ either $gZ = Z$ or $gZ \cap Z = \emptyset$.

PROPOSITION 18.31. *There exists an finite subgraph $\Lambda \subset X$ which separates ends of X and is precisely-invariant.*

PROOF. If X has a cut-vertex, then we take Λ to be this vertex. Suppose, therefore, that X contains no such vertices. Then, by Lemma 18.20, X contains a minimal (essential) Dunwoody track $\Lambda \subset X$. By Proposition 18.25, for every $g \in G$, the track $\mathrm{M} := \mathrm{g}\Lambda$ (which is also minimal) is either disjoint from Λ or equal to Λ. □

The proof of Theorem 18.29 then reduces to:

PROPOSITION 18.32. *Every finite subgraph $\Lambda \subset X$ as in Proposition 18.31 gives rise to a nontrivial action of G on a simplicial tree T with finite edge-stabilizers and finitely generated vertex groups.*

Proof. Let Λ be either a cut-vertex of X or a finite connected essential precisely-invariant track $\Gamma \subset X$ (see Definition 18.30). We first consider the more interesting case of when Γ is a Dunwoody track.

We partition of X into components of $G \cdot \Gamma$ and components of $X \setminus G \cdot \Gamma$, which we will refer to as *complementary regions*. Each complementary region C_v is declared to be a *vertex* v of the partition and each $\Gamma_e := g \cdot \Gamma$ is declared to be an *edge* e. Since Γ is a Dunwoody track and $H_1(X, \mathbb{Z}_2) = 0$, the complement $X \setminus \Gamma_e$ consists of exactly two components $\Gamma_e^{\pm}$; therefore, each edge of the partition is incident to exactly two (distinct) complementary regions. These regions represent vertices incident to e. Thus, we obtain a graph T. Since the action of G preserves the above partition of X, the group G acts on the graph T.

Lemma 18.33. *The group G does not fix any vertices of T and does not stabilize any edges.*

Proof. Suppose that G fixes a vertex v of T. Let E_v denote the set of edges of T incident to v. By relabeling, we can assume that the corresponding component C_v of $X \setminus G \cdot \Gamma$ equals

$$\bigcap_{e \in E_v} \Gamma_e^{+}.$$

Therefore, for every $x \in C_v$, $g \in G$, and $e \in E_v$ we have

$$g(x) \notin \Gamma_e^{-}.$$

Recall that the action $\Gamma \curvearrowright X$ is cocompact. Therefore, there exists a finite subcomplex $K \subset X$ whose G-orbit is the entire X. Clearly, $x \in K$ for some $x \in C_v$. On the other hand, by the above observation, the intersection

$$G \cdot K \cap \Gamma_e^{-}$$

is a finite subcomplex. This contradicts the fact that Γ_e^{-} is unbounded. Thus, G does not fix any vertex in T. Similarly, we see that G does not preserve any edge of T. □

Lemma 18.34. *The graph T is a tree.*

Proof. Connectedness of T immediately follows from connectedness of X. If T were to contain a circuit, it would follow that some Γ_e did not separate X, which is a contradiction. □

Lastly, we observe that compactness of Γ_e's and proper discontinuity of the action $G \curvearrowright X$ imply that the stabilizer G_e of every edge e in G is finite. Note that, a prori, G acts on T with inversions since $g \in G$ can preserve Γ_e and interchange $\Gamma_e^{+}, \Gamma_e^{-}$.

Since the closure $\overline{C}_v$ of each vertex-space C_v is connected and $\overline{C}_v/G_v$ is compact, it follows that the stabilizer G_v of each vertex $v \in T$ is finitely generated (this is a special case of the Milnor-Schwartz Lemma).

Suppose now that Λ is a single vertex v. If Λ were to separate X into exactly two components, we would be done by repeating the arguments above. Otherwise, we modify X by replacing the vertex v with an edge e whose mid-point m separates X into exactly two components both of which are unbounded. We repeat this for every point in $G \cdot v$ in G-equivariant fashion. The result is a new complex $\widehat{X}$ with a cocompact action $G \curvearrowright \widehat{X}$. Clearly, $\Lambda := \{m\}$ is precisely-invariant, and, hence, we are done as above. Proposition 18.32 follows. $\square$

In both cases, the quotient graph $\Gamma = T/G$ is finite since the action $G \curvearrowright X$ is cocompact.

We can now finish the proof of Theorem 18.29. In view of Proposition 18.32, Bass–Serre correspondence (Section 4.5.6), implies that the group G is the fundamental group of a nontrivial finite graph of groups $\mathcal{G}$ with finite edge groups and finitely generated vertex groups. $\square$

18.6. Accessibility

Let G be a finitely generated group which splits nontrivially as an amalgam $G_1 \star_H G_2$ or $G_1 \star_H$ with finite edge-group H. Sometimes, this decomposition process can be iterated, by decomposing the groups G_i as amalgams with finite edge groups, etc. The key issue that we will be addressing in this section is:

Does the decomposition process terminate after finitely many steps?

If this decomposition process of G terminates then the group G is isomorphic to the fundamemtal group of a graph of groups, where all edge groups are finite and all vertex groups have at most one end. This leads to

DEFINITION 18.35. A group G is said to be *accessible* if it admits a (finite) graph of groups decomposition with finite edge groups and 1-ended vertex groups.

C. Thomassen and W. Woess in [**TW93**] prove:

THEOREM 18.36. *A finitely generated group G is accessible if and only if one (equivalently, every) Cayley graph Γ of G satisfies the following property:*

There exists a number D so that every two ends of Γ_G can be separated by a bounded subset of Γ of diameter $\leqslant D$.

In particular, accessibility is QI invariant.

Our first goal is to show that all finitely generated torsion-free groups are accessible. Recall that the rank of a finitely generated group is the least number of its generators.

THEOREM 18.37 (Grushko's theorem). *Suppose that G is finitely generated and $G = G_1 \star G_2$ is a nontrivial free product decomposition. Then*

$$\text{rank}\,(G_i) < \text{rank}\,(G), \quad i = 1, 2.$$

We refer the reader to the paper by Scott and Wall [**SW79**] for a topological proof of this classical theorem in group theory. We can now prove:

THEOREM 18.38. *All torsion-free finitely generated groups are accessible.*

PROOF. If G is torsion-free, then all (inductively constructed) decompositions $G_1 \star_H G_2$ or $G_1 \star_H$ are just free products $G_1 \star G_2$ and $G_1 \star \mathbb{Z}$ respectively. Then, by Grushko's theorem, $\text{rank}\,(G_i) < \text{rank}\,(G), i = 1, 2$ and, hence, the decomposition process terminates after at most $\text{rank}\,(G)$ steps. $\square$

M. Dunwoody ([**Dun93**]) constructed an example of a finitely generated group which is not accessible. The main result of this section is

THEOREM 18.39 (M. Dunwoody, [**Dun85**]). *Every almost finitely presented group is accessible.*

Our goal below is to give a proof of Dunwoody's theorem, mostly following papers by M. Dunwoody [**Dun85**] and G. Swarup [**Swa93**]. Before proving Dunwoody's theorem, we will establish several technical facts.

Refinements of graphs of groups. Let $\mathcal{G}$ be a graph of groups with the underlying graph Γ, let $H = G_v$ be one of its vertex groups. Let $\mathcal{H}$ be a graph-of-groups decomposition of H with the underlying graph Λ. Suppose that:

ASSUMPTION 18.40. For every edge $e \subset \Gamma$ incident to v, the subgroup $G_e \subset H$ is conjugate in H to a subgroup of one of the vertex groups H_w of $\mathcal{H}$, $w = w(e)$ (this vertex need not be unique). For instance, if every G_e is finite, then, in view of Property FA for finite groups, G_e will fix a vertex in the tree corresponding to $\mathcal{H}$. Thus, our assumption will hold in this case.

Under this assumption, we can construct a new graph of groups decomposition $\mathcal{F}$ of G as follows. Cut Γ open at v, i.e. remove v from Γ and then replace each open or half-open edge of the resulting space with a closed edge. The resulting graph Γ' could be disconnected. We have the natural map $r : \Gamma' \to \Gamma$. Let Φ denote the graph obtained from the union $\Gamma' \sqcup \Lambda$ by identifying each vertex $v_i' \in r^{-1}(v) \in e_i' \subset \Gamma'$ with the vertex $w(e) \in \Lambda$ as in the above assumption. Then Φ is connected. We retain for Φ the vertex and edge groups and the group homomorphisms coming from Γ and Φ. The only group homomorphisms which need to be defined are for the edges $e = [e^-, e^+]$, where $e^- = w(e) = w$. In this case, the embedding $G_e \to G_w$ is the one given by the conjugation of G_e to the corresponding subgroup of G_w.

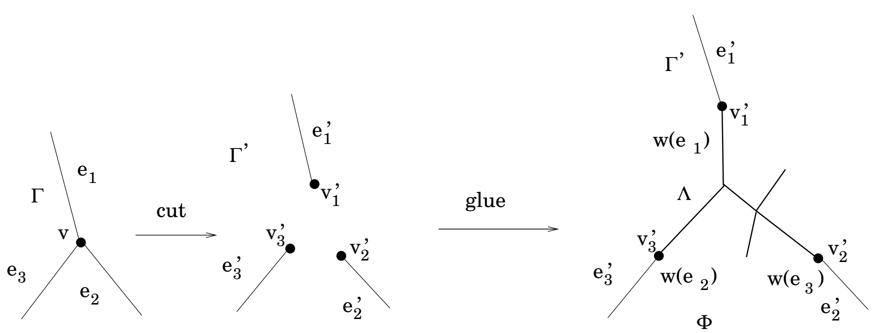


FIGURE 18.11. *C*ut Γ open and glue Φ from Γ' and Λ.

We leave it to the reader to verify (using Seifert – Van Kampen theorem) that $\pi_1(\Phi) = G$.

DEFINITION 18.41. The new graph of groups $\mathcal{F}$ is called the *refinement* of $\mathcal{G}$ *via* $\mathcal{H}$. A refinement is said to be *trivial* if $\mathcal{H}$ is a trivial graph of groups. We use the notation $\mathcal{G} \prec \mathcal{F}$ for a refinement.

PROPOSITION 18.42. *Let $\mathcal{G}$ be a finite graph of finitely generated groups with finite edge-subgroups (with the underlying graph Λ). Then:*

1. Every vertex subgroup G_v is QI embedded in G. (Note that finite generation of the vertex groups implies that G is itself finitely generated.)

2. If, in addition, the group G is finitely presented, then every vertex group of $\mathcal{G}$ is also finitely presented.

PROOF. The proofs of both statements are very similar. We first construct (as in Section 4.5.4) a tree of graphs $\mathcal{Z}$, corresponding to $\mathcal{G}$.

Namely, let $G \curvearrowright T$ be the action of G on a tree corresponding to the graph of groups $\mathcal{G}$ (see Section 4.5.4). For every edge-group G_e in $\mathcal{G}$ we take $S_e := G_e \setminus \{1\}$ as the generating set of G_e. For every vertex-group G_v in $\mathcal{G}$ we pick a finite generating set S_v of G_v, such that for every edge $e = [v, w]$, the sets S_v, S_w contain the images of S_e under the embeddings $G_e \to G_v, G_e \to G_w$.

Then, using the projection $p : T \to \Lambda = T/G$, we define generating sets $S_{\tilde{v}}, S_{\tilde{e}}$ of $G_{\tilde{v}}$, $G_{\tilde{e}}$ using isomorphisms $G_{\tilde{v}} \to G_v, G_{\tilde{e}} \to G_e$, where $\tilde{v}, \tilde{e}$ project to v, e under the map $T \to \Lambda$. Let $Z_{\tilde{v}}, Z_{\tilde{e}}$ denote the Cayley graphs of the groups $G_{\tilde{v}}, G_{\tilde{e}}$ ($\tilde{v} \in V(T), \tilde{e} \in E(T)$) with respect to the generating sets $S_{\tilde{v}}, S_{\tilde{e}}$.

Now, to simplify the notation, we will use the notation v and e for vertices and edges of T. Let Z_v, Z_e denote the Cayley graphs of the groups G_v, G_e ($v \in V(T), e \in E(T)$) with respect to the generating sets S_v, S_e defined above. We have natural injective maps of graphs $f_{ev} : Z_e \hookrightarrow Z_v$, whenever v is incident to e. The resulting collection of graphs Z_v, Z_e and embeddings $Z_e \hookrightarrow Z_v$, defines a *tree of graphs* $\mathcal{Z}$ with the underlying tree T. For each Z_e we consider the product $Z_e \times [-1, 1]$ with the standard triangulation of the product of simplicial complexes. Let $\tilde{Z}_e$ be the 1-skeleton of this product. Lastly, let Z denote the graph obtained by identifying vertices and edges of each $\tilde{Z}_e$ with their images in Z_v under the maps $f_{ev} \times \{\pm 1\}$. We endow Z with the standard metric.

Clearly, the group G acts on Z freely. The quotient graph Z/G has only finitely many vertices; this quotient graph is finite if each G_v is finitely generated.

For every $v \in V(T)$ define the map $\rho_v : Z^0 \to Z_v^0 = G_v$ to be the G_v-equivariant nearest-point projection. For every edge $e = [v, w]$, the subgraph $f_v(Z_e) \subset Z_v$ separates Z, and every two distinct vertices in $f_v(Z_e)$ are connected by an edge in this graph. It follows that for $x, y \in Z^0$ within unit distance from each other,

$$\text{dist}(\rho_v(x), \rho_v(y)) \leqslant 1.$$

Hence, the map ρ_v is 1-Lipschitz. We then extend, G_v-equivariantly, each ρ_v to the entire graph Z.

We now can prove the assertions of the proposition.

1. Since each G_v is finitely generated, the action $G \curvearrowright Z_v$ is geometric and, hence, the action $G \curvearrowright Z$ is geometric as well. Thus, the space Z is QI to the group G and Z_v is QI to G_v for every vertex v. Let $x, y \in Z_v$ be two vertices and $\alpha \subset Z$ be the shortest edge-path connecting them. Then $\rho(\alpha) \subset Z$ still connects x to y and its length is at most the length of α. It follows that Z_v is isometrically embedded in Z. Hence, each G_v is QI embedded in G. This proves (1).

2. Since G is finitely presented and $G \curvearrowright Z$ is geometric, the space Z is coarsely simply-connected by Corollaries 6.35 and 6.54. Our goal is to show that each vertex space Z_v of Z is also coarsely simply-connected. Let α be a loop in a vertex space Z_v. Since Z is coarsely simply-connected, there exists a constant C (independent of α) and a collection of (oriented) loops α_i in the 1-skeleton of Z such that

$$\alpha = \prod_i \alpha_i$$

and each α_i has length $\leqslant C$. We then apply the retraction ρ to the loops α_i. Then

$$\alpha = \prod_i \rho(\alpha_i)$$

and $\mathrm{length}(\rho(\alpha_i)) \leqslant \mathrm{length}(\alpha_i) \leqslant C$ for each i. Thus, Z_v is coarsely simply-connected and, therefore, G_v is finitely presented. □

We are now ready to prove Dunwoody's accessibility theorem.

Proof of Theorem 18.39. We will construct inductively a finite chain of refinements

$$\mathcal{G}_1 \prec \mathcal{G}_2 \prec \mathcal{G}_3 \ldots$$

which are graph-of-groups with finite edge groups, afp vertex groups, and so that the terminal graph of groups has only finite and 1-ended vertex groups.

We need a notion of *complexity* (which will be denoted $\sigma(G)$) for afp groups G which generalizes the notion of *rank* used for the proof of accessibility in the torsion-free case. Note that if we drop the assumotion $H^1(X, \mathbb{Z}_2) = 0$ below and assume instead that X is a graph, then $\sigma(G)$ equals $\mathrm{rank}\,(G) + 1$.

Definition 18.43 (σ-complexity). Suppose that X be a 2-dimensional simplicial complex with $H^1(X, \mathbb{Z}_2) = 0$, such that X admits a simplicial properly discontinuous, cocompact action $G \curvearrowright X$. We let $\sigma(G, X)$ denote the total number of cells in the cell-complex X/G (the quotient need not be a simplicial complex). Accoringly, we will use the notation $\sigma(G)$ for the minimum of the numbers $\sigma(G, X)$ where the minimum is taken over all complexes X and group actions $G \curvearrowright X$ as above. If $\mathcal{G}$ is a finite graph with afp vertex groups, then $\sigma(\mathcal{G})$ is defined to be the maximum of complexities $\sigma(G_v)$, taken over all vertex groups G_v of $\mathcal{G}$.

We will show that *some* process of refinement results in the strict reduct of the σ-complexity. Such refinement process necessarily terminates.

Let G be an afp group and X be a G-complex which realizes the complexity $\sigma(G)$.

First, suppose that such X has a cut-vertex v (see Definition 18.19). Then, as in the proof of Theorem 18.29, we split G as a graph of groups (with the edge-groups stabilizing v) so that each vertex-group G_i acts on a subcomplex $X_i \subset X$, where the frontier of X_i in X is contained in $G \cdot v$. It follows from the Mayer-Vietoris sequence that $H^1(X_i, \mathbb{Z}_2) = 0$ for each i. Thus, $\sigma(G_i) < \sigma(G)$ for every i.

Hence, without loss of generality, we may assume that X has no cut-vertices. If the group G has at most one end, we are done. Suppose that G has at least 2 ends. Then, by Propositions 18.31, 18.32, there exists a (connected) finite precisely-invariant track $\tilde{\tau}_1 \subset X_1 := X$ which determines a nontrivial graph of groups decomposition $\mathcal{G}_1$ of $G_1 := G$ with finite edge groups. Our assumption that $H^1(X, \mathbb{Z}_2) = 0$ implies that τ_1 is 2-sided in $X_1 := X$. Let X_2 be the closure of a connected component of $X \setminus G \cdot \tilde{\tau}_1$. By compactness of X/G and Milnor-Schwartz Lemma, the stabilizer G_2 of X_2 in G is finitely generated. Since $H^1(X, \mathbb{Z}_2) = 0$, it follows by excision and Mayer-Vietoris sequence that

$$H^1(X_2, \partial X_2; \mathbb{Z}_2) = 0,$$

where ∂X_2 is the frontier of X_2 in X_1.

Therefore, if define W_2 by pinching each boundary component of X_2 to a point, then $H^1(W_2, \mathbb{Z}_2) = 0$. The stabilizer G_2 of X_2 in G acts on W_2 properly discontinuously cocompactly. Therefore, each vertex group of $\mathcal{G}_1$ is again afp.

If each vertex group of $\mathcal{G}_1$ is 1-ended, we are again done. Suppose therefore that the closure X_2 of some component $X_1 \setminus G \cdot \tilde{\tau}_1$ as above has stabilizer $G_2 < G_1$ which has at least two ends. According to Theorem 6.24, G_2 splits (nontrivially) as a graph of groups with finite edge groups. Let $G_2 \curvearrowright T_2$ be a nontrivial action of G_2 on a simplicial tree (without inversions) which corresponds to this decomposition. Since each edge-group of $\mathcal{G}_1$ is finite, if such group is contained in G_2, it has to fix a vertex in T_2, see Corollary 2.70. Recall that the edge-groups of $\mathcal{G}_1$ are conjugate to the stabilizers of components of $G \cdot \tilde{\tau}_1$ in G. Therefore, every such stabilizer fixes a vertex in T_2. We let X_2^+ denote the union X_2 with all simplices in X which have nonempty intersection ∂X_2. Clearly, G_2 still acts properly discontinuously cocompactly on X_2^+. The stabilizer of each component of $X_2^+ \setminus X_2$ is an edge group of $\mathcal{G}_1$.

We then construct a (G_1-equivariant) PC map (see Definition 18.5) $f_2 : X_2^+ \to T_2$ such that:

1. f_2 sends components of $X_2^+ \setminus int(X_2)$ to the corresponding fixed vertices in T_2.
2. f_2 sends vertices to vertices of T.
3. f_2 is linear on each edge of the cell-complex X_2.

If the image of the map f_2 is unbounded then the action $G_2 \curvearrowright T_2$ has a bounded orbit. By Cartan's Theorem (Theorem 2.69), G_2 fixes a point in T_2, which contradicts the assumption that the action $G_2 \curvearrowright T_2$ is nontrivial.

Therefore, the image of f_2 contains an edge $e \subset T_2$ which separates T_2 into at least two unbounded subsets.

Then, by Lemma 18.6, there exists a point $t \in e$ which is a regular value of f_2. Thus, by Exercise 18.8, $f^{-1}(t)$ is a track. It is immediate that $f^{-1}(t)$ is precisely-invariant in X_2 with finite G_2-stabilizer. By the choice of e, the graph $f^{-1}(t)$ separates X into at least two unbounded components. Let $\tilde{\tau}_2$ be an essential component of $f^{-1}(t)$.

Thus, by Proposition 18.32, the track τ_2 determines a decomposition of G_2 as a graph of groups $\mathcal{G}_3$ with finite edge groups. We continue this decomposition inductively. We obtain a collection of pairwise disjoint connected tracks $\tau_1, \tau_2, ... \subset Y = X/G$ which are projections of the tracks $\tilde{\tau}_i \subset X$.

Suppose that the number of tracks τ_i is $> 6F + r$, where F is the number of faces in X and r is the dimension of $H^1(X, \mathbb{Z}_2)$. Then, by Lemma 18.14, every τ_k, $k > 6F + r$ is isotopic to some graph $\tau_{i(k)}, i = i(k) \leqslant 6F + r$. Let R be the product region in Y bounded by such tracks. Lifting this region in X we again obtain a product region $\tilde{R}$ bounded by tracks $g_i\tilde{\tau}_i, g_k\tilde{\tau}_k$, $g_i, g_k \in G$. Therefore, the stabilizers of $g_i\tilde{\tau}_i, g_k\tilde{\tau}_k$ and R in G have to be the same. It follows that every X_k, $k > 6F + r$ is a product region whose stabilizer fixes its boundary components. The corresponding tree T_k is just a union of two edges which are fixed by the entire group G_k. This contradicts the fact that each graph of groups $\mathcal{G}_k$ is nontrivial. Therefore, the decomposition process of G terminates after $6F + r$ steps and G is accessible. □

18.7. QI rigidity of virtually free groups and free products

THEOREM 18.44. *If G is virtually free, then every group G' which is QI to G is also virtually free.*

Proof. If the group G is finite, the assertion is clear. If G is virtually cyclic, then G' and G are 2-ended, which, by Part 3 of Theorem 6.22, implies that G' is also virtually cyclic.

Suppose now that G has infinitely many ends. Since G is finitely presented, by Corollary 6.54, the group G' is finitely presented as well. The group G acts geometrically on a locally finite simplicial tree T with infinitely many ends, therefore, the groups G and G' are QI to T. Since G' is finitely presented, by Theorem 18.39, the group G' splits as a graph of groups where every edge group is finite, every vertex group is finitely generated and each vertex group has 0 or 1 ends.

By Proposition 18.42, every vertex group G'_v of this decomposition is QI embedded in G'. In particular, every G'_v is quasi-isometrically embedded in a simplicial tree of finite valence. The image of such an embedding is coarsely-connected (with respect to the restriction of the metric on T) and, therefore, is within finite distance from a subtree $T'_v \subset T$. Thus, each G'_v is QI to a locally-finite simplicial tree (embedded in T).

Lemma 18.45. *T'_v cannot have one end.*

Proof. Suppose that T'_v has one end. The group G'_v is infinite and finitely generated. Therefore, its Cayley graph contains a bi-infinite geodesic (see Exercise 4.85). Such geodesic γ cannot embed quasi-isometrically in a 1-ended tree (since both ends of γ would have to map to the same end of T'_v). □

Thus, every vertex group G'_v has 0 ends and, hence, is finite. By Theorem 4.52, the group G is virtually free. □

We conclude this chapter by generalizing Theorem 18.44 to quasi-isometries of arbitrary finitely presented groups. The following theorem is due to P. Papasoglu and K. Whyte, [**PW02**], who proved it for finitely generated groups. We refer the reader to C. Cashen's paper [**Cas10**] for further analysis.

Theorem 18.46 (QI rigidity for graphs of groups with finite edge groups). *Suppose that G, G' are quasi-isometric finitely presented groups and $f : G' \to G$ is a quasi-isometry. Assume also that the group G has a decomposition as a finite graph of groups $\mathcal{G}$ with finite edge groups and finitely generated vertex groups which have at most one end. Then G' also admits a decomposition $\mathcal{G}'$, such that for every 1-ended vertex group G'_v in $\mathcal{G}'$, $f(G'_v)$ is Hausdorff-close to a conjugate of a 1-ended vertex group of $\mathcal{G}$. In particular, every 1-ended vertex group of $\mathcal{G}$ is QI to a vertex group of $\mathcal{G}'$.*

Proof. Finite presentability implies that both groups G and G' are accessible. As in the proof of Theorem 18.44, we conclude that:

The group G' also splits as a finite graph of groups $\mathcal{G}'$ with finite edge groups, with finitely generated vertex groups each of which has at most two ends. We associate with the graph $\mathcal{G}$ a tree of spaces $\mathfrak{X}$ as in §4.5.6, with the underlying cell complex X. Recall that there exists a G-equivariant projection $p : X \to T$ to the associated Bass-Serre tree of $\mathcal{G}$.

Each edge e of T splits the tree T in two subtrees $T_e^{\pm}$, which are the two components of the subgraph of T obtained by erasing the edge e. (We will see below how to assign the $\pm$ labels to these subtrees, depending on a vertex v of $\mathcal{G}'$.) We define two subspaces $X_e^{\pm} \subset X$, the unions of all vertex spaces X_v with $v \in V(T_e^{\pm})$. Consider the Cayley graph Γ'_v of a 1-ended vertex group G'_v of $\mathcal{G}'$.

Since the edge-spaces of $\mathfrak{X}$ have bounded diameter, they cannot coarsely separate the image $Y := f(\Gamma'_v)$ in X. Thus, there exists a number D such that for every edge e in $E := E(T)$, exactly one of the two subspaces X_e^+, X_e^- *coarsely contains* Y; we denote this subspace X_e^+:

$$Y \subset \mathcal{N}_D(X_e^+).$$

REMARK 18.47. Note that since G'_v has infinitely many end, Y cannot be coarsely contained in X_e^-. Thus, the choice of the vertex v defines a label $\pm$ for each edge e of $\mathcal{G}$. Accordingly, we obtain a collection of subtrees T_e^+, $e \in E$.

We claim that the intersection of all these subtrees,

$$T^+ := \bigcap_{e\in E} T_e^+,$$

consists of a single vertex w. It then would immediately follow that

$$Y \subset \mathcal{N}_D(\bigcap_{e\in E} X_e^+).$$

Note that it is not even clear that the intersection T^+ is nonempty. We note that, since no edge-space X_e coarsely separates Y, the projection of Y to T is contained in a subtree $A \subset T$ of finite diameter (otherwise, there exists an edge $e \in E$ such that Y contains points in X_e^+, X_e^- arbitrarily far from X_e). Since the graph Γ'_v is connected, and has infinite diameter, it follows that Γ'_v contains an embedded half-line H, see Exercise 1.65. Lemma 1.66 then shows that there exists a vertex $w \in V(T)$ such that $p^{-1}(w) \cap Y$ is unbounded. If w were a vertex of some T_e^-, then the intersection

$$X_w \cap Y \subset p^{-1}(T_e^-) \cap Y$$

would be bounded. Therefore, w belongs to the intersection of all subtrees T_e^+:

$$w \in T^+.$$

Lastly, if T^+ contains another vertex $u \neq w$, then it also contains an edge e separating these two vertices. Since $w \in T_e^+$, $u \in T_e^-$, which means that u is not in T^+. This is a contradiction. Therefore, $T^+ = \{w\}$.

To summarize: There is $D \in \mathbb{R}$ such that for each 1-ended vertex subgroup G'_v of $\mathcal{G}'$, there exists a vertex $w \in T$ such that $f(\Gamma'_v) \subset \mathcal{N}_D(X_v)$. Applying the coarse inverse map $\bar{f} : G \to G'$, we conclude that $f(G'_v)$ and G_w are Hausdorff-close to each other. □

CHAPTER 19

Proof of Stallings' Theorem using harmonic functions

In this chapter we will prove Stallings' theorem in full generality:

THEOREM 19.1. *If G is a finitely generated group with infinitely many ends then G is admits a nontrivial decomposition as a graph of groups with finite edge groups.*

In his essay [**Gro87**, Pages 228–230], Gromov outlined a proof of Stallings' theorem using harmonic functions. The goal of this chapter is to provide details for Gromov's arguments. One advantage is that this proof works for finitely generated groups without the finite presentability assumption. However, the geometry behind the proof is not as transparent as in Chapter 18. The proof that we present in this chapter is a simplified form of the arguments which appear in [**Kap14**]. The simplification presented here (in the proof of Theorem 19.5) is due to Bruce Kleiner.

We refer the reader to the material of §2.9 for the definition of harmonic functions on Riemannian manifolds and to §2.4 for the discussion of the coarea formula. Both will be key ingredients in the proof.

Every finitely generated group G admits an isometric free properly discontinuous cocompact action $G \curvearrowright M$ on a Riemannian manifold M, which, then, necessarily has bounded geometry (since it covers a compact Riemannian manifold).

EXAMPLE 19.2. If G is k–generated, and N is a Riemann surface of genus k, we have an epimorphism

$$\phi : \pi_1(N) \to G.$$

Then G acts isometrically and cocompactly on the covering space M of N with $\pi_1(M) = \mathrm{Ker}(\phi)$.

The space $\epsilon(M)$ of ends of M is naturally homeomorphic to the space of ends of G, see Proposition 6.18. Let $\bar{M} := M \cup \epsilon(M)$ denote the compactification of M by its space of ends; the action of the group G extends to a topological action of G on $\bar{M}$.

We will see in §19.3 that every continuous function $\chi : \epsilon(M) \to \{0, 1\}$, admits a unique continuous extension

$$h = h_\chi : \bar{M} \to [0, 1],$$

such that the function $h\big|_M$ is harmonic (the function h is the energy minimizer among all extensions of χ lying in a suitable Sobolev space). Furthermore, unless χ is constant, the extension h restricts to a proper function $h : M \to (0, 1)$.

REMARK 19.3. This theorem fails if M has one or two ends. For instance, every harmonic function $h : \mathbb{R} \to (0, 1)$ has to be linear and, hence, constant.

We will refer to the function h as the *harmonic extension* of χ (even though, only its restriction to M is harmonic). Let $H(M)$ denote the space of functions $M \to \mathbb{R}$ which are harmonic extensions nonconstant functions $\chi : \epsilon(M) \to \{0,1\}$. We equip $H(M)$ with the topology of uniform convergence on compacts in M. For each $h \in H(M)$ we define its *energy*

$$E(h) := E(h\big|_M),$$

see §2.9 for the definition of energy of functions $M \to \mathbb{R}$. We will see in §19.3 that $E(h)$ is always finite.

DEFINITION 19.4. We define the *energy gap* $e(M)$ of M as

$$e(M) := \inf\{E(h) : h \in H(M)\}.$$

The isometric group action $G \curvearrowright M$ yields a linear action $G \curvearrowright H(M)$

$$g \cdot h = h \circ g^{-1}$$

which preserves the functional E. Therefore E projects to a lower semi-continuous (see Theorem 2.40) functional $E : H(M)/G \to \mathbb{R}_+$, where we give $H(M)/G$ the quotient topology. The main technical result needed for the proof of Stallings' theorem is:

THEOREM 19.5. *1. $e(M) > 0$.*
2. The functional $E : H(M)/G \to \mathbb{R}_+$ is proper in the sense that the preimage

$$E^{-1}([0,T])$$

is compact for every $T \in \mathbb{R}_+$. In particular, $e(M)$ is attained.

We now sketch our proof of Stallings' theorem. Let $h \in H(M)$ be an energy-minimizing harmonic function guaranteed by Theorem 19.5, $E(h) = e(M)$. We then verify that the set $\{h(x) = \frac{1}{2}\}$ is *precisely-invariant* with respect to the action of G (see Definition 18.30).

By choosing t sufficiently close to $\frac{1}{2}$ we obtain a smooth hypersurface $S = \{h(x) = t\}$ which is still precisely-invariant under G and separates the ends of M. If this hypersurface is connected, we can use the standard construction of a dual simplicial tree T using the tiling of M by the components of $M \setminus G \cdot S$, cf. §4.5.6. The edges of T in this case are the "walls", i.e., hypersurfaces $g(S)$, $g \in G$. Every wall then lies in the boundary of exactly two components of $M \setminus G \cdot S$, which are the adjacent vertices.

This construction does not work in the case when S is not connected. We still use hypersurfaces $g(S)$ as the edges, but the definition of vertices has to be modified. Instead of the topological separation (which is meaningless if we separate disconnected subsets of M), we define separation via the functions $g \cdot h$. A union U of components of $M \setminus G \cdot S$ is called *indecomposable* if U cannot be separated by any function $g \cdot h$. These indecomposable sets are the vertices of T. We then show that each $g(S)$ lies in the boundary of exactly two indecomposable subsets of $M \setminus G \cdot S$, thereby defining a graph on which G is acting. Since the action of G on M is proper and S is compact, the edge stabilizers for the action $G \curvearrowright T$ are finite. Lastly, we verify that the graph T is a tree and that the action of G on T is without inversions and is nontrivial, i.e., G does not fix a vertex in T.

19.1. Proof of Stallings' theorem

The goal of this section is to prove Stallings' theorem assuming results of §19.3 and Theorem 19.5.

Let $H(M)$ denote the space of harmonic functions $h : M \to [0,1]$ as above. According to Theorem 19.5, there exists a function $h = h_\chi \in H(M)$ with minimal energy $E(h) = e(M) > 0$. We will refer to the harmonic function h as *minimal*. Then, for every $g \in G$, the function

$$g^* h := h \circ g$$

has the same energy as h and (in view of uniqueness of the harmonic extension) equals

$$h_{g^*(\chi)}.$$

Following Gromov, for $g \in G$, define two new functions

$$g_+(h) := \max(h, g^*(h)), \quad g_-(h) := \min(h, g^*(h)).$$

Clearly,

$$g_-(h)(x) = g_+(h)(x) \iff h(x) = g^*(x).$$

We will see (Lemma 19.20) that

$$E(g_+(h)) + E(g_-(h)) = 2E(h).$$

Both functions $g_+(h), g_-(h)$ admit continuous extension to $\bar{M}$: The function $g_+(h)$ extends to $\chi_+ := \max(\chi, g^*(\chi))$ and $g_-(h)$ extends to $\chi_- := \min(\chi, g^*(\chi))$. The functions $\chi_\pm$ take only the values 0 and 1 on $\epsilon(M)$. Define

$$h_\pm := h_{\chi_\pm},$$

the harmonic extensions of $\chi_\pm$. Since harmonic functions are energy-minimizers,

$$E(h_\pm) \leqslant E(g_\pm(h)),$$

and, hence,

$$E(h_+) + E(h_-) \leqslant E(g_+(h)) + E(g_-(h)) = 2E(h) = 2e(M). \tag{19.1}$$

REMARK 19.6. The functions h are *functional analogues* of the minimal tracks in Chapter 18. The definition of the functions $g_\pm(h)$ is an analogue of the "exchange" argument and the definition of the functions $h_\pm$ is an analogue of the "round off" argument.

Note that it is, *a priori*, possible that χ_- or χ_+ is constant. Set

$$G^c := \{g \in G : \chi_- \text{ or } \chi_+ \text{ is constant}\}.$$

We first analyze the set $G \setminus G^c$. For $g \notin G^c$, both h_- and h_+ belong to $H(M)$ and, hence, by (19.1),

$$E(h_+) = E(h_-) = E(h) = e(M),$$

and

$$E(g_+(h)) = E(h_+), \quad E(g_-(h)) = E(h_-).$$

It follows that both functions $g_\pm(h)$ are harmonic. Since

$$g_-(h) \leqslant g_+(h),$$

the maximum principle (see Corollary 2.46) implies that either $g_-(h) = g_+(h)$ or $g_-(h) < g_+(h)$. Indeed, if $g_-(h)(x) = g_+(h)(x)$ at some $x \in M$, then the difference

$$g_+(h) - g_-(h)$$

is harmonic and attains its absolute maximum at $x \in M$. The maximum principle then implies that the difference $g_+(h) - g_-(h)$ is constant, hence, equal to zero.

Hence, the set $\{h = g^*h\}$ is either empty or equals the entire M, in which case $g^*(h) = h$. Therefore, for every $g \in G \setminus G^c$ one of the following holds:

1. $g^*h = h$.
2. $g^*h(x) < h(x),\ \forall x \in M$.
3. $g^*h(x) > h(x),\ \forall x \in M$.

In particular, the level set

$$\Sigma := h^{-1}\left(\frac{1}{2}\right)$$

is *precisely–invariant* under the elements of $G \setminus G^c$: For every $g \in G \setminus G^c$, either

$$g(\Sigma) = \Sigma$$

or

$$g(\Sigma) \cap \Sigma = \emptyset.$$

(The equality case occurs iff $g^*h = h$.)

We now consider elements $g \in G^c$: For such g's either χ_- is constant or χ_+ is constant. Since $\chi_- \leqslant \chi_+$ and both functions only take the values 0 and 1, either $\chi_- \equiv 0$ or $\chi_+ \equiv 1$.

Case 1: $g \in G^c$ is such that $\chi_- \equiv 0$. In other words, whenever $\chi(\xi) = 1$, we also have $g^*(\xi) = 0$, $\xi \in \epsilon(M)$. It follows that

$$g^*\chi \leqslant 1 - \chi.$$

We claim that

$$g^*h \leqslant 1 - h.$$

Indeed, suppose that $h_1 = h_{\chi_1}, h_2 = h_{\chi_2} \in H(M)$ are such that $\chi_1 \leqslant \chi_2$, but

$$h_{\chi_1}(x) > h_{\chi_2}(x)$$

for some $x \in M$. Then the difference

$$f = h_{\chi_1} - h_{\chi_2}$$

is harmonic, positive at x and $\lim_{n\to\infty} f(x_n) \leqslant 0$ for each sequence (x_n) in M diverging to infinity. This means that there exists an open bounded subset $\Omega \subset M$ with smooth boundary, which contains x and such that $f\big|_\Omega$ attains its maximum at the point x (and not on the boundary). This contradicts the Maximum Principle for harmonic functions. Applying this to the harmonic functions $h_1 = g^*h$ and $h_2 = 1 - h$, we conclude that

$$g^*h \leqslant 1 - h.$$

The same argument shows that if $g^*h(x) = 1 - h(x)$ for some $x \in M$, then $g^*h = 1 - h$. The latter implies that

$$g(\Sigma) = \Sigma,$$

where $\Sigma = \{h = \frac{1}{2}\}$ as before.

This leaves us with the case

$$g^*h < 1 - h.$$

Then, clearly, $g(\Sigma) \cap \Sigma = \emptyset$.

Case 2: $\chi_+ \equiv 1$. We then replace χ with $\chi' = 1 - \chi$, h with $h_{\chi'}$, and conclude that $\chi'_- \equiv 0$. Then, by appealing to the Case 1, we see that either $g^*h = h$ or $g^*h > h$.

We thus proved an analogue of the Proposition 18.31 in the proof of Stallings' theorem for almost finitely presented groups in Chapter 18:

LEMMA 19.7. *Each minimal harmonic function h satisfies:*
1. For every $g \in G$ one of the following occurs:

$$g^*h = h, \quad g^*h < h, \quad g^*h > h, \quad g^*h = 1 - h, \quad g^*h < 1 - h, \quad g^*h > 1 - h. \tag{19.2}$$

2. The subset $h^{-1}(1/2) = \Sigma \subset M$ is compact and precisely–invariant under the action of the group G. Moreover, for each $g \in G$, if $g(\Sigma) = \Sigma$ then either $g^*h = h$ or $g^*h = 1 - h$.

We let G_Σ denote the stabilizer of Σ in G. Since Σ is compact and the action $G \curvearrowright M$ is properly discontinuous, the group G_Σ is finite.

By the construction, the subset Σ separates M into at least two unbounded components: One where $h > 1/2$ and the other where $h < 1/2$. By Sard's Theorem, the subset $R \subset (0, 1)$ of regular values of $h : M \to (0, 1)$ has full measure in $(0, 1)$. Since h extends to a continuous function on $\bar{M}$, which takes only the values 0 and 1 in $\epsilon(M)$, the function $h : M \to (0, 1)$ is proper. This immediately implies that the subset R is open in $(0, 1)$.

LEMMA 19.8. *For $t \in (0, 1)$ sufficiently close to $\frac{1}{2}$ the subspace $h^{-1}(t)$ is still precisely–invariant under G.*

PROOF. The subset

$$K = \left\{ \frac{1}{4} \leqslant h \leqslant \frac{3}{4} \right\}$$

is compact in M. Therefore, since the action of G on M is properly discontinuous, the subset $G_K \subset G$ consisting of all $g \in G$ for which $gK \cap K \neq \emptyset$ is finite. Since Σ is precisely-invariant for each $g \in G_K$ either $g^*h = h$ or there exists an interval $(\frac{1}{2} - \epsilon_g, \frac{1}{2} + \epsilon_g)$ such that for all $t \in (\frac{1}{2} - \epsilon_g, \frac{1}{2} + \epsilon_g) \setminus \frac{1}{2}$,

$$g\{h = t\} \cap \{h = t\} = \emptyset.$$

Taking the minimum of ϵ_g's, $g \in G_K$, we obtain the desired open interval about $\frac{1}{2}$. $\square$

We next pick $t > \frac{1}{2}$ sufficiently close to $\frac{1}{2}$. The hypersurface $S = \{h = t\}$ is smooth, compact and precisely–invariant under G. Let $G_S \subset G$ denote the stabilizer of S in G.

EXERCISE 19.9. G_S is a subgroup of G_Σ.

We now show that G splits nontrivially over a subgroup of G_S. (As we noted in the beginning of the chapter, the proof is straightforward under the assumption that S is connected, but requires extra work in general.) We proceed by constructing a simplicial G–tree T on which G acts nontrivially, without inversions and with finite edge–stabilizers.

Construction of T. Given a minimal harmonic function h, define the set of minimal functions

$$\mathcal{M} = \{g^* h : g \in G\}.$$

Each function $f \in \mathcal{M}$ defines the *wall* $W_f = \{x : f(x) = t\}$ and the *half-spaces* $W_f^+ := \{x : f(x) > t\}$, $W_f^- := \{x : f(x) < t\}$ (these spaces are not necessarily connected). Note that for $g = 1, f = h$ and $W_f = S$.

Let $\mathcal{E}$ denote the set of walls. We say that a wall W_f *separates* $x, y \in M$ if

$$x \in W_f^+, \quad y \in W_f^-.$$

Maximal subsets V of

$$M^o := M \setminus \bigcup_{f \in \mathcal{M}} W_f$$

consisting of points which cannot be separated from each other by a wall, are called *indecomposable* subsets of M^o. Note that such sets need not be connected. Set

$$\mathcal{V} := \{\text{indecomposable subsets of } M^o\}.$$

We will refer to the elements of $\mathcal{V}$ as *vertex spaces* and to the walls W_f as *edge-spaces*. We say that a wall W is *adjacent* to $V \in \mathcal{V}$ if $W \cap cl(V) \neq \emptyset$.

Exercise 19.10. Each $f \in \mathcal{M}$ yields the indecomposable set

$$\{1 - t < f < t\}.$$

It is adjacent to the walls W_f, W_{1-f}. We will refer to such indecomposable sets as *special*.

Lemma 19.11. *If t is taken to be sufficiently close to $\frac{1}{2}$ then the following hold:*

1. No wall W_f separates points of another wall W_h.

2. If a vertex space V is adjacent to distinct edge-spaces W_f, W_h, then either $f < h$ or $h > f$ or $f + h = 1$. The latter occurs if and only if V is special.

Proof. 1. The proof of Part 1 is analogous to that of Lemma 19.8, we leave it as an exercise for the reader.

2. Suppose first that $f + h \neq 1$. Without loss of generality, $V \subset W_f^- \cap W_h^+$. Then there exist points $x, x' \in V$ such that

$$h(x') > t > (1 - f)(x) > f(x)$$

(here we are using the fact that $t > 1/2$). Since V is indecomposable,

$$h(x) > (1 - f)(x) > f(x).$$

In particular, the option $h < f$ is impossible. We need to rule out the inequalities $f + h < 1$ and $f + h > 1$. Pick points $y \in \{f = 1/2\}, y' \in \{h = 1/2\}$. If t is sufficiently small, then

$$(f + h)(y) = \frac{1}{2} + h(y) > 1, \quad (f + h)(y') = f(y') + \frac{1}{2} < 1.$$

Hence both $f + h < 1, f + h > 1$ are ruled out and we are left with $f < h$.

Lastly, if V is special then it is defined as the intersection of the spaces

$$\{f < t\}, \quad \{f > 1 - t\},$$

in which case $h = 1 - f$. The converse is clear as well. □

Lemma 19.12. *Each wall $W = W_f$ is adjacent to exactly two elements V^+, V^- of $\mathcal{V}$ (contained in W^+, W^- respectively).*

PROOF. We first construct two vertex spaces adjacent to W. Pick a point $x \in W$. Since t is a regular value of f, there exist sequences $x_k^{\pm} \in W^{\pm}$ which converge to x. The inequalities (19.2) imply that for large k the points x_k^+ belong to the same vertex space, which we denote by V^+ and the points x_k^- belong to the same vertex space, which we denote by V^-.

Now, let us prove that $V^{\pm}$ are the only vertex spaces adjacent to W. Suppose that $V \subset W^+$ is adjacent to W. Considering a sequence $y_k^+ \in V$ converging to some $y \in W$, we conclude that no wall separates y_k^+ from x_k^+ for large k (since this holds for the limit points y, x). □

We define a graph T with the vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, where $W \in \mathcal{E}$ connects V^+ and V^- if and only if W is adjacent to both. From now on, we abbreviate W_{f_i} to W_i.

LEMMA 19.13. *The graph T is a tree.*

PROOF. By the construction, every point of M belongs to a wall or to an indecomposable set. Hence, connectedness of T follows from connectedness of M. Suppose that T contains a circuit with the consecutive vertices $V_1, V_2, \ldots, V_k, V_1$. Every other vertex in this circuit has to correspond to a special vertex set, in particular, $k = 2n$. After relabeling, we can assume that each V_{2i} is special and each V_{2i-1} is non-special. In view of Lemma 19.12 (after reversing the orientation of the circuit if necessary), we obtain:

$$f_1 < f_3 < f_5 < \ldots < f_{k-1} < f_1,$$

which is absurd. □

We next note that G acts naturally on T since the sets $\mathcal{M}$, $\mathcal{E}$ and $\mathcal{V}$ are G–invariant and G preserves adjacency. If $g(W_f) = W_f$, then $g^* f = f$, which implies that g preserves both W_f^+ and W_f^- (here we are using the assumption that $t \neq \frac{1}{2}$). Hence, g fixes both vertices of the edge corresponding to W, which means that G acts on T without inversions. The stabilizer of an edge in T corresponding to a wall W is finite, since W is compact and G acts on M properly discontinuously.

It remains to verify nontriviality of the action of G on T. Suppose that $G \curvearrowright T$ has a fixed vertex. This means that the indecomposable subset $V \subset M$ defining this vertex is G–invariant. Since G acts cocompactly on M, it follows that $M = \mathcal{N}_r(V)$ for some $r \in \mathbb{R}_+$. The indecomposable subset V is contained in the half-space W^+ for some wall W. Since W is compact and W^- is not, the subset W^- is not contained in $\mathcal{N}_r(W)$. Thus $W^- \setminus \mathcal{N}_r(V) \neq \emptyset$. Contradiction.

Therefore T is a nontrivial G–tree and we obtain a nontrivial graph of groups decomposition of G where the edge groups are conjugate to subgroups of the finite group G_S. □

19.2. Nonamenability

The goal of the section is to show that each group G with infinitely many ends is nonamenable; Accordingly, the manifold M (on which G acts geometrically) has $\lambda_1(M) > 0$, equivalently, M has positive Cheeger constant. This property will be used in constructing harmonic extensions h_χ of functions $\chi : \epsilon(M) \to \{0, 1\}$ and proving Part 1 of Theorem 19.5.

Let X be a metric space. A metric ball $B(x, r) \subset X$ is a *neck* (more precisely, an r-neck) if

$$X \setminus B(x, r)$$

has at least three unbounded components. The point x is the *center* of the r-neck. The following theorem was proven by C. Pittet in [**Pit98**]:

THEOREM 19.14. *Let X be a connected graph (equipped with the standard metric) such that there exists $r > 0$ for which every vertex $x \in V(X)$ is the center of an r-neck in X. Then X is nonamenable.*

PROOF. Let m be an integer such that $m > 4r + 2$. Define $V \subset V(X)$ as a maximal m-separated subset of $V(X)$. We will prove the theorem by constructing a map

$$f : V \to V$$

such that $d(v, f(v)) \leqslant 2m + 1$ and $f^{-1}(u)$ has cardinality $\geqslant 2$ for every $u \in V$. Then the theorem will follow from Theorem 16.4.

The construction of f is somewhat reminiscent of the proof of Milnor–Schwartz lemma. Fix a vertex $v_0 \in V$. For $v \in V$ such that $d(v_0, v) \leqslant m$ set $f(v) = v_0$. Otherwise, take a geodesic $g \subset X$ connecting v_0 to v and let $x \in g$ be the vertex of X with $d(x, v) = m + 1$. Then let $f(v)$ be a point $w \in V$ closest to x. (If there are several such points, pick one at random.) By maximality of V, $d(x, w) \leqslant m$. Therefore,

$$d(v, f(v)) \leqslant 2m + 1.$$

Before proving the statement about cardinality of $f^{-1}(u)$, we will need a technical lemma.

LEMMA 19.15. *Pick $u \in V$ and consider the ball $B = B(u, r) \subset X$. Let C, C' be distinct components of $X \setminus B$. Then for $v \in V \cap C, v' \in V \cap C'$ we have*

$$d(v, v') > m + 1.$$

PROOF. Every geodesic g connecting v to v' is the union

$$\overline{vx} \cup \overline{xx'} \cup \overline{x'v'}$$

where $x \in g \cap B, x' \in g \cap B$ and $\overline{xx'} \subset B$. Then

$$d(v, v') \geqslant d(v, x) + d(v', x') \geqslant d(v, u) - r + d(v', u) - r \geqslant 2m - 2r > m + 1$$

by our choice of m. $\square$

We will now proceed to proving the inequality on the cardinality of $f^{-1}(u)$. For $u \in V$ let C be one of the (at least two) unbounded components of $X \setminus B(u, r)$ which does not contain v_0. Let $v \in C \cap V$ be a point closest to v_0. We claim that $f(v) = u$. Since there are at least two such components C, we will then conclude that $f^{-1}(u)$ contains at least two points.

We let g denote a geodesic in X connecting v_0 to v. This geodesic necessarily passes through the ball $B = B(u, r)$.

Case 1: $d(v, v_0) \leqslant m$ (which implies that $f(v) = v_0$ by the definition of f). If v_0, v belong to distinct components of $X \setminus B$, contradicting Lemma 19.15. This means that v_0 has to belong to the ball B, i.e., $v_0 = u$. Thus, in this case, $f(v) = u$, as required.

Case 2: $d(v_0, v) \geqslant m + 1$. Let $x \in g$ be the vertex with $d(x, v) = m + 1$.

Subcase 2a: $x \notin C \cup B$. Pick $y \in g \cap B$. Then

$$m + 1 = d(x, v) \geqslant d(x, x') + (d(u, v) - d(y, u)) \geqslant d(x, y) + m - r,$$

which implies that

$$d(x, y) \leqslant r + 1.$$

Therefore,

$$d(x, u) \leqslant d(x, y) + r \leqslant 2r + 1.$$

If $w \in V$ is a vertex with $d(x, w) \leqslant 2r + 1$ then

$$d(u, w) \leqslant 4r + 2 < m$$

implying that $u = w$, as the set V is m-separated. Therefore, in this case, $f(v) = u$.

Subcase 2b: $x \in B$. Then $d(x, u) \leqslant r$ and, hence, for every $u' \in V \setminus \{u\}$, $d(x, u') \geqslant 2m - r > 2$. Therefore, in this case, again, $f(v) = u$.

Subcase 2c: $x \in C$. We leave it to the reader to verify that for every component C' of $X \setminus (B \cup C)$, if $v' \in C' \cap V$ then $d(u, x) < d(v', x)$, implying that

$$f(v) \in \{u\} \cup C.$$

Suppose, that $f(v) = w \in V \cap C$. Then $d(x, w) \leqslant m$ and, hence,

$$d(v_0, w) \leqslant d(v_0, x) + m < d(v_0, v)$$

as $d(v_0, x) = m$. This, however, contradicts the choice of v as the point in $V \cap C$ closest to v_0. This leaves only one possibility: $f(v) = u$. □

COROLLARY 19.16. *Suppose that M is a Riemannian manifold which admits an isometric properly discontinuous cocompact action of a group G with infinitely many ends. Then G is nonamenable and $\lambda_1(M) > 0$.*

PROOF. Since amenability is QI invariant, G is amenable if and only if its Cayley graph X is. Since the graph X is nonamenable by Theorem 19.14, its Cheeger constant is positive, $h(M) > 0$ (see Theorem 16.13), equivalently, $\lambda_1(M) > 0$, see Theorem 2.52. □

19.3. An existence theorem for harmonic functions

Theorem 19.17 below was originally proven by Kaimanovich and Woess in Theorem 5 of [**KW92**] using probabilistic methods (they also proved it for arbitrary continuous functions with values in $[0, 1]$). At the same time, an analytical proof of this result was given by Li and Tam [**LT92**], see also [**Li12**, Chapter 21] for a detailed and more general treatment.

Let M be a Riemannian manifold as in the beginning of this chapter (M admits a geometric action of a group G with infinitely many ends). We owe the following proof to Mohan Ramachandran:

THEOREM 19.17. *Let $\chi : \epsilon(M) \to \{0, 1\}$ be a continuous function. Then:*
1. χ admits a continuous harmonic extension to M.
2. This harmonic extension h has finite energy.

Proof. We let $C_c^\infty(M)$ denote the space of C^∞ functions $M \to \mathbb{R}$ with compact support. For $u, v \in C_c^\infty(M)$ define the inner product

$$\langle u, v\rangle = \int_M uv \, \mathrm{d}V$$

where $\mathrm{d}V$ is the Riemannian volume density on M. In what follows we will use this volume density to define a measure (of the Lebesgue class) on the sigma-algebra of Borel subsets of M. We let

$$\|u\|_{L_2} = \langle u, v\rangle^{1/2}$$

denote the norm of u with respect to this inner product. Since the differential of each function $u \in C_c^\infty(M)$ is also compactly supported, the energy $E(u)$ of the function u is finite.

We leave it to the reader to verify that the quantity

$$\|u\| := \|u\|_{L_2} + \sqrt{E(u)}.$$

is also a norm on $C_c^\infty(M)$. We define a *Sobolev space* $W_o^{1,2}(M)$ as the completion of $C_c^\infty(M)$ with respect to the norm $\|u\|$. The space $W_o^{1,2}(M)$ sits naturally in the Hilbert space $L^2(M)$. Furthermore, by construction, the energy functional extends continuously to $W_o^{1,2}(M)$.

Recall that every function χ extends to a continuous function $\varphi : \bar{M} \to \mathbb{R}$ which is smooth on M, see Lemma 6.26.

We let $L^2_{loc}(M)$ denote the space of functions of M which are *locally* in L^2, i.e., functions whose restrictions to compact subsets $K \subset M$ are in $L^2(K)$. By continuity, every extension φ above, belongs to $L^2_{loc}(M)$. Thus, for a fixed function φ we define the affine subspace of functions

$$\mathcal{G} := \varphi + W_o^{1,2}(M) \subset L^2_{loc}(M).$$

Then the energy is a continuous (nonlinear) functional on $\mathcal{G}$ and we set $E := \inf_{f\in\mathcal{G}} E(f)$.

Note that, since $\mathcal{G}$ is affine, for $u, v \in \mathcal{G}$ we also have

$$\frac{u+v}{2} \in \mathcal{G},$$

in particular,

$$E\left(\frac{u+v}{2}\right) \geqslant E.$$

We set

$$E(u, v) := 2E\left(\frac{u+v}{2}\right) - \frac{E(u) + E(v)}{2}.$$

The latter equals

$$E(u, v) := \int_M \langle \nabla u, \nabla v\rangle \, \mathrm{d}V \tag{19.3}$$

in the case when u, v are smooth. We thus obtain,

$$E(u, v) \geqslant 2E - \frac{E(u) + E(v)}{2}$$

for all $u, v \in \mathcal{G}$. Hence, in view of (19.3), by continuity of E,

$$E(u - v) = E(u) + E(v) - 2E(u, v) \leqslant 2E(u) + 2E(v) - 4E. \tag{19.4}$$

Pick a sequence $u_n \in \mathcal{G}$ such that

$$\lim_{n\to\infty} E(u_n) = E.$$

Then, according to (19.4),

$$E(u_m - u_m) \leqslant 2E(u_n) + 2E(u_m) - 4E = 2(E(u_n) - E) + 2(E(u_m) - E).$$

We now come to the first, and only, point where the assumption that the number of ends M is infinite (and not 2) is used:

Since $\lambda = \lambda_1(M) > 0$ (Theorem 19.14), we obtain, by the definition of the bottom of the spectrum (2.5),

$$\lambda \int_M f^2 \mathrm{d}V \leqslant E(f) \tag{19.5}$$

for all $f \in C_c^\infty(M)$ and, hence, by continuity, for all $f \in W_o^{1,2}(M)$. Therefore, the functions $v_n := u_n - \varphi \in W_o^{1,2}(M)$ satisfy

$$\|v_n - v_m\| \leqslant (2 + \lambda^{-1})(E(u_n) - E + E(u_m) - E).$$

Hence, the sequence (v_n) is Cauchy in $W_o^{1,2}(M)$. Set

$$v := \lim_n v_n, u := \varphi + v \in \mathcal{F}.$$

By semicontinuity of energy (Theorem 2.40), $E(u) = E$. Since u minimizes energy among all functions in $\mathcal{G}$, it is necessarily harmonic and, hence, u is smooth (see §2.9). Since $d\varphi$ is compactly supported (its support K is contained in the support of φ), the function v is also harmonic away from the compact subset $K \subset M$. By the inequality (19.5), we have

$$\int_M v^2 \mathrm{d}V \leqslant \lambda^{-1} E(v) < \infty. \tag{19.6}$$

Let $r > 0$ denote the injectivity radius of M. Pick a base-point $o \in M$. Then (19.6) implies that there exists a function $\rho : M \to \mathbb{R}_+$ which converges to 0 as $d(x, o) \to \infty$, such that

$$\int_{B(x,r)} v^2(x) \mathrm{d}V \leqslant \rho(x)$$

for all $x \in M$. By the mean value inequality (Corollary 2.48 in §2.9), there exists $C_1 < \infty$, such that

$$\sup_{x\in B(x,r)} v^2(x) \leqslant C_1 \inf_{B(x,r)} v^2$$

provided that $d(x, K) \geqslant r$. Therefore,

$$v^2(x) \leqslant \frac{C_1}{Vol(B(x,r))} \int_{B(x,r)} v^2 \leqslant C_2 \rho(x),$$

and, thus,

$$\lim_{d(x,o)\to\infty} v(x) = 0.$$

We conclude that the harmonic function u extends to the function χ on $\epsilon(M)$. $\square$

Properties of harmonic extensions.

PROPOSITION 19.18. *1. For each continuous function $\chi : \epsilon(M) \to \{0,1\}$ its harmonic extension $h : M \to \mathbb{R}$ is unique.*

2. The unique harmonic extension takes values in the interval $[0,1]$.

3. If $h(x) = 0$ or $h(x) = 1$, for some $x \in M$, then h is constant. In other words, the harmonic extension $h : M \to (0,1)$ is either constant or proper.

PROOF. We prove all three properties by appealing to the Maximum Principle for harmonic functions. Since the proofs are analogous to the ones which appear in Lemma 19.7, our arguments will be somewhat brief.

1. Suppose that $h_1, h_2 : \bar{M} \to \mathbb{R}$ are harmonic extensions of $\chi : \epsilon(M) \to \{0,1\}$. Then the difference $h = h_1 - h_2$ is harmonic and for every sequence $x_n \in M$ diverging to infinity,

$$\lim_{n\to\infty} h(x_n) = 0.$$

Hence, h attains its maximum or minimum at a point $x \in M$. By the Maximum Principle, h is constant, implying that $h_1 = h_2$.

2. Let h be the unique harmonic extension of χ. Suppose that there exists $x \in M$ such that $h(x) \geqslant 1$. Then h again attains its maximum at a point of M, implying that h is constant. This is impossible if $h(x) > 1$. The same argument, with $1-h$ replacing h, handles the case $h(x) \leqslant 0$. This proves (2) as well as (3). □

19.4. Energy of minimum and maximum of two smooth functions

The arguments here are again due to Mohan Ramachandran.

Let M be a smooth manifold and f be a C^1-smooth function on M. Define the function $f^+ := \max(f, 0)$ and the closed set

$$\Gamma := \{x \in M : f(x) = 0, df(x) = 0\}.$$

Set

$$\Omega := \{x \in M : f(x) = 0, df(x) \neq 0\} = f^{-1}(0) \setminus \Gamma.$$

By the implicit function theorem, Ω is a smooth submanifold in M and, hence, has measure zero. Clearly, f^+ is smooth on $M \setminus \{f = 0\}$.

LEMMA 19.19. *Under the above conditions, a.e. on M we have: $df^+(x) = df(x)$ if $f(x) > 0$ and $df^+(x) = 0$ if $f(x) \leqslant 0$.*

PROOF. The claim is clear at the points x where $f(x) \neq 0$. Since Ω has measure zero, it suffices to prove the assertion for points $x_0 \in \Gamma$. Choose local coordinates on M at a point $x_0 \in \Gamma$, so that $x_0 = 0$. Since f has zero derivative at 0, we have:

$$\lim_{v\to 0} \frac{|f(v)|}{\|v\|} = 0.$$

Since $0 \leqslant |f^+| \leqslant |f|$, it follows that

$$\lim_{v\to 0} \frac{|f^+(v)|}{\|v\|} = 0.$$

Therefore, f^+ is differentiable at x_0 and $df^+(x_0) = 0$. □

Consider now two C^1-smooth functions f_1, f_2 on M. Define

$$f_{max} := \max(f_1, f_2), \quad f_{min} := \min(f_1, f_2), \quad f := f_1 - f_2.$$

LEMMA 19.20. $E(f_1) + E(f_2) = E(f_{max}) + E(f_{min})$.

Proof. Set

$$M_1 := \{f_1 > f_2\}, M_2 := \{f_2 > f_1\}, M_0 := \{f_1 = f_2\}.$$

Since

$$f_{max} = f_2 - f^+, \quad f_{min} = f_1 - f^+,$$

by the above lemma we have:

$$\nabla f_{max} = \nabla f_2, \quad \nabla f_{min} = \nabla f_1 \quad \text{a.e. on} \quad M_0.$$

Clearly,

$$\nabla f_{max} = \nabla f_i \big|_{M_i}, \nabla f_{min} = \nabla f_{i+1} \big|_{M_{i+1}}, i = 1, 2.$$

Hence,

$$\int_{M_i} \left(|\nabla f_{max}|^2 + |\nabla f_{min}|^2\right) \mathrm{d}V = \int_{M_i} \left(|\nabla f_1|^2 + |\nabla f_2|^2\right) \mathrm{d}V, i = 0, 1, 2.$$

Therefore,

$$E(f_1) + E(f_2) = E(f_{max}) + E(f_{min}). \quad \square$$

19.5. A compactness theorem for harmonic functions

19.5.1. Positive energy gap implies existence of an energy minimizer. Let M be a bounded geometry Riemannian manifold with infinitely many ends and positive Cheeger constant $\geqslant c > 0$, and let $\bar{M} = M \cup \epsilon(M)$ be the end compactification of M.

We state several definitions and notations used in what follows. For an m-dimensional Riemannian manifold N (possibly with boundary), we let $|N|$ denote the m-dimensional volume of N. Given a function $f : N \to \mathbb{R}$, we set $Var(f) := \sup(f) - \inf(f)$, the *variation* of f on N. For a function f on N, we define the *average* of f,

$$⨍_N f = \frac{\int_N f \mathrm{d}V}{\mathrm{V}ol(N)}.$$

In order to simplify the notation, we will frequently omit dV in the notation for integrals. Let $U \subset M$ be a smooth codimension 0 submanifold with compact boundary K. The *capacitance* $cap(U, K)$ of the pair (U, K) is the infimum of energies of compactly supported functions $u : U \to [0, 1]$, which are equal to 1 on K. We refer to §19.2 for the definition of R-necks in M. Note that each proper function $f : M \to (t_1, t_2)$ admits a continuous extension to $\bar{M}$: We will always retain the name f for this extension. The same convention will be used for functions defined on subsets with compact boundary in M.

Let $\mathcal{F}$ denote the collection of continuous functions u on $\bar{M}$, whose restriction to $\epsilon(M)$ is nonconstant, and takes values in $\{0, 1\}$, while u is differentiable almost everywhere on M.

In §19.5.4 we will prove Part 1 of Theorem 19.5:

Theorem 19.21. *There exists $\mu > 0$ such that every $u \in \mathcal{F}$ has energy at least μ, i.e., $e(M) > 0$, M has positive energy gap.*

Our goal below is to derive Part 2 of Theorem 19.5 from Part 1. We first state several corollaries of Theorem 19.21.

Corollary 19.22. *For each $U \subset M$ and K as above, $cap(U, K) \geqslant \mu$.*

PROOF. Given a function $u : U \to [0,1]$ which equals 1 on K, we extend u by 1 to the rest of M. Then, clearly, the extension $\tilde{u}$ has the same energy as u (since $\nabla\tilde{u}$ vanishes on $M \setminus U$) and $u \in \mathcal{F}$. Therefore, $E(u) = E(\tilde{u}) \geqslant \mu$. □

As an application, we prove:

PROPOSITION 19.23. *Assume that every point in M belongs to an R-neck. Then for all $0 < a < b < 1$, $E \in [0,\infty)$, there is an $r = r(a,b,E) \in (0,\infty)$ with the following property. If $u : M \to (0,1)$ is a proper a.e. smooth map, and $p \in M$, then either:*

1. *$u(B(p,r))$ is not contained in the interval $[a,b]$, or*
2. *the energy of u is at least E.*

PROOF. Define $s = \min(a^2, 1-b^2)$. Since every point of M belongs to an R-neck, there exists r_0, such that the complement of every ball $B(p,r_0)$ has more than

$$k = \frac{E}{s\mu}$$

unbounded components.

We claim that the desired property holds for $r = r_0$ (and, hence, for all greater values of r as well). Suppose that this fails. For a point $p \in M$ the distance function $d(p,\cdot)$ is smooth away from p and a.e. $r \in \mathbb{R}_+$ is its regular value. Thus, for generic $r \geqslant r_0$ (which we fix from now on), the metric ball $B(p,r) \subset M$ has smooth boundary. We let $\mathcal{C}$ denote the collection of unbounded components of $M \setminus B(p,r)$. Let $u : M \to (0,1)$ be a proper a.e. smooth map such that $u(B(p,r)) \subset [a,b]$, while u has energy $\leqslant E$. For each $U \in \mathcal{C}$, the function u takes the values in $[a,b]$ on $K = \partial U$. Consider the two functions $u^+ = \max\{b,u\}$ and $u^- = \min\{a,u\}$ on U. Then

$$E(u^\pm) \leqslant E(u|_U)$$

and $u^+|_K \equiv b, u^-|_K \equiv a$. Let $\tilde{u}^\pm$ denote the extension of $u^\pm$ to the rest of M such that

$$\tilde{u}^\pm|_{M\setminus U} \equiv u^\pm|_K.$$

Then

$$E(\tilde{u}^\pm) = E(u^\pm) \leqslant E(u|_U).$$

Note that $u^-, \tilde{u}^-$ and $u^+, \tilde{u}^+$ are proper functions to intervals $(0,a)$ and $(b,1)$ respectively.

Consider the function $\tilde{u}^-$: Its values on $\epsilon(M)$ belong to $\{0,a\}$ and a is one of its values. If $\tilde{u}^-$ does not take zero value on $\epsilon(M)$, then $u|_{\epsilon(U)}$ takes only the value 1.

Case 1: The function $\tilde{u}^-$ takes both values 0 and a on $\epsilon(M)$. Then the rescaled function $\frac{1}{a}\tilde{u}^-$ belongs to $\mathcal{F}$ and, hence,

$$E(u|_U) \geqslant E(\tilde{u}^-) \geqslant a^2\mu$$

by Theorem 19.21.

Case 2: The function $u|_{\epsilon(U)}$ takes only the value 1. (Then $\tilde{u}^-$ is constant, equal to a, on $\epsilon(M)$ and we obtain no lower energy bound from the above arguments.) Since u is nonconstant on $\epsilon(M)$, it has to take the zero value somewhere on $\epsilon(M\setminus U)$, which means that the function $\tilde{u}^+$ takes both values b and 1 on $\epsilon(M)$.

Consider the function

$$\tilde{v} := 1 - \tilde{u}^+$$

and, similarly to the Case 1 argument, obtain

$$E(u\big|_U) \geqslant E(\tilde{v}) \geqslant (1-b^2)\mu.$$

In either case, we conclude that

$$E(u\big|_U) \geqslant s\mu > 0,$$

where, as we recall, $s = \min(a^2, 1-b^2)$.

By the definition of r_0, the number of unbounded components of $M \setminus B(p,r)$ is greater than

$$k = \frac{E}{s^2\mu}.$$

The restriction of u to each of these ends is at least $s^2\mu$, which implies that the energy of u is greater than E. This is a contradiction. □

COROLLARY 19.24. *If $u : M \to (0,1)$ is a proper a.e. smooth function of energy $\leqslant E$ and u is nearly constant on a large ball $B = B(p,r)$, then it nearly equals to 0 or 1 on B. More precisely, the supremum-norm of $u\big|_B$ or of $(u-1)\big|_B$ converges to zero as $Var(u\big|_B) \to 0$.*

We next prove that harmonic functions of bounded energy have small variation on "most" balls in M.

LEMMA 19.25. *Suppose that $h : M \to [0,1]$ is a harmonic function of finite energy. Fix $r > 0$ and let $x_i \in M$ be a sequence diverging to infinity, i.e.,*

$$\lim_{i\to\infty} d(x_1, x_i) = \infty.$$

Then

$$\lim_{i\to\infty} Var(h\big|_{B(x_i,r)}) = 0.$$

PROOF. Suppose to the contrary that there exists a sequence (x_i) such that the variation of h on $B(x_i, r)$ does not converge to zero. After passing to a subsequence, we can assume that the balls $B(x_i, r)$ are pairwise disjoint and there exist $\delta > 0$ and points $y_i \in B(x_i, r)$ such that $|h(x_i) - h(y_i)| \geqslant \delta$ for all i. For each i we pick a geodesic segment $\gamma_i \subset B(x_i, r)$ of length $\leqslant r$ connecting x_i to y_i.

By the Mean Value Theorem, for each i there exists $z_i \in \gamma_i$ such that

$$|\nabla u(z_i)| \geqslant \frac{|h(x_i) - h(y_i)|}{r} \geqslant \frac{\delta}{r}.$$

Hence, we obtain a lower energy-density bound at one point:

$$|\nabla h(z_i)|^2 \geqslant \frac{\delta^2}{r^2}.$$

We next promote this to a lower energy bound for h. According to Theorem 2.50, there exists a constant L depending only on the geometry bounds of M, such that

$$|\nabla|\nabla h(x)|^2| \leqslant L$$

at each $x \in M$ where $\nabla h(x) \neq 0$. By appealing to the Mean Value Theorem again, for all $x \in B(z_i, \rho)$, we obtain:

$$|\nabla h(x)|^2 \geqslant \eta := \frac{\delta^2}{r^2} - L\rho.$$

We fix $\rho > 0$ such that $\eta > 0$ and observe that there exists $V > 0$ for which

$$\mathrm{V}ol(B(z_i, \rho)) \geqslant V,$$

for all i. Therefore,

$$E(h\big|_{B(z_i,\rho)}) \geqslant \eta V.$$

Since the balls $B(z_i, \rho)$ are pairwise disjoint, we conclude that h has infinite energy. A contradiction. □

We can now prove Part 2 of Theorem 19.5. Recall that $H = H(M)$ is the space of functions $f \in \mathcal{F}$ which are harmonic on M. Consider a sequence $f_n \in H$ whose energy is bounded above by a number E. Since each f_n takes both values 0 and 1 on $\epsilon(M)$, there exist a sequence $x_n \in f_n^{-1}(1/2)$. By applying elements of the group G, we can assume that the points x_n belong to a fixed compact $K \subset M$. After passing to a subsequence, we may assume that $\lim_{n\to\infty} x_n = x \in K$. In view of Compactness Theorem 2.51, the sequence of functions f_n subconverges uniformly on compacts to a harmonic function f which attains the value 1/2 at $x \in K$. We have to show that $f \in \mathcal{F}$.

1. By lower semicontinuity of energy (Theorem 2.40), f has energy $\leqslant E$.

2. Suppose that f is constant on M. Then for each $\delta > 0$ and $r > 0$ there exists n such that

$$Var(f_n\big|_{B(x,r)}) < \delta.$$

By taking r sufficiently large and taking into account Corollary 19.24, we conclude that f_n approximately equals to 0 or 1 on $B(x, r)$. This contradicts the assumption that $f_n(x_n) = 1/2$. Therefore, f cannot be constant.

3. Suppose now that f either does not extend continuously to a point $\xi \in \mathrm{Ends}(M)$ or that the extension $f(\xi)$ exists but $f(\xi)$ is different from 0 and 1.

Then there exist $a, b \in (0, 1)$ and a sequence $p_i \in M$ converging to ξ in the topology of $\bar{M}$ such that for all i,

$$0 < a \leqslant f(p_i) \leqslant b < 1.$$

Remark 19.26. Note that this also includes the case when there are sequences $x_i \to \xi, y_i \to \xi$ with

$$\lim_{i\to\infty} f(x_i) = 0, \quad \lim_{i\to\infty} f(y_i) = 1.$$

In this case, for large i, f takes the value $\frac{1}{2}$ at a point p_i in a path connecting x_i to y_i and close to ξ in the topology of $\bar{M}$.

Take $r = r(a, b, E)$ as in Proposition 19.23. Since $E(f) < \infty$, $Var(f\big|_{B_r(p_i)})$ converges to 0 as $i \to \infty$, see Lemma 19.25. Since for each fixed i

$$\lim_{n\to\infty} f_n\big|_{B(p_i,r)} = f\big|_{B(p_i,r)},$$

we conclude that (for large n and i) the function f_n contradicts Proposition 19.23. □

Remark 19.27. One could remove the cocompactness assumption by saying that any sequence $u_i \in H$ has a pointed limit living in a pointed Gromov-Hausdorff limit of a sequence (M, x_n) (which will be another bounded geometry manifold with a linear isoperimetric inequality and ubiquitous R-necks).

Thus, it remains to prove Theorem 19.21; the proof of occupies the rest of the chapter.

19.5.2. Some coarea estimates. Recall that if $u : M \to \mathbb{R}$ is a smooth function on a Riemannian manifold M, then for a.e. $t \in \mathbb{R}$, the level set $u^{-1}(\{t\}) = \{u = t\}$ is a smooth hypersurface, and for any measurable function $\phi : M \to \mathbb{R}$ such that $\phi|\nabla u|$ is integrable, we have the *coarea formula*

$$\int_M \phi|\nabla u| = \int_{\mathbb{R}} \left(\int_{\{u=t\}} \phi \right) dt \,, \tag{19.7}$$

where the integration $\int_{\{u=t\}} \phi$ is with respect to the Riemannian measure on the hypersurface, see Theorem 2.15.

The two applications of this appearing below are:

$$\int_{\{t_1 \leqslant u \leqslant t_2\}} |\nabla u|^2 = \int_{t_1}^{t_2} \left(\int_{\{u=t\}} |\nabla u| \right) dt \,, \tag{19.8}$$

where we take $\phi = |\nabla u|$ on $\{t_1 \leqslant u \leqslant t_2\}$ and zero otherwise, and

$$|\{t_1 \leqslant u \leqslant t_2\}| = \int_{\{t_1 \leqslant u \leqslant t_2\}} 1 = \int_{t_1}^{t_2} \left(\int_{\{u=t\}} \frac{1}{|\nabla u|} \right) dt \,, \tag{19.9}$$

where we take $\phi = \frac{1}{|\nabla u|}$ under the assumption that $\nabla u \neq 0$ a.e. on M.

We first combine these in the following general inequality:

Lemma 19.28. *Suppose that $u : M \to [t_1, t_2]$ is a smooth function on a compact Riemannian manifold with boundary, such that $A(t) = |\{u = t\}| \geqslant A_0 > 0$ for a.e. t. Then*

$$\int_M |\nabla u|^2 \geqslant \frac{A_0^2 (t_2 - t_1)^2}{|M|} \,. \tag{19.10}$$

Proof. The argument combines (19.8), (19.9), and Jensen's inequality. We decompose M as $M = M_0 \sqcup M_+$, where $M_0 = \{x \in M : \nabla u(x) = 0\}$. Of course,

$$\int_M |\nabla u|^2 = \int_{M_+} |\nabla u|^2.$$

By Sard's Theorem, a.e. $t \in [t_1, t_2]$, is a regular value of u. Furthermore, since u is proper, the set of critical values of u is a closed nowhere dense subset of $[t_1, t_2]$. For a.e. $t \in [t_1, t_2]$ we have

$$\begin{aligned} \int_{\{u=t\}} |\nabla u| &= A(t) ⨍_{\{u=t\}} |\nabla u| \\ &\geqslant A(t) \frac{1}{⨍_{\{u=t\}} \frac{1}{|\nabla u|}} \quad \text{by Jensen's inequality} \\ &= \frac{A^2(t)}{\int_{\{u=t\}} \frac{1}{|\nabla u|}} \,, \end{aligned} \tag{19.11}$$

with the equality in the case when $|\nabla u|$ is constant on M.

Since ∇u is nonzero on almost every hypersurface $\{u = t\}$, in the following computation we can consider only nonzero values of ∇u:

$$\int_M |\nabla u|^2 = \int_{t_1}^{t_2} \left(\int_{\{u=t\}} |\nabla u| \right) dt \qquad \text{by (19.8)}$$

$$\geqslant \int_{t_1}^{t_2} \frac{A(t)}{\left(\fint_{\{u=t\}} \frac{1}{|\nabla u|}\right)} dt \qquad \text{by (19.11)}$$

$$\geqslant A_0(t_2-t_1) \fint_{t_1}^{t_2} \frac{dt}{\left(\fint_{\{u=t\}} \frac{1}{|\nabla u|}\right)}$$

$$\geqslant A_0(t_2-t_1) \frac{1}{\fint_{t_1}^{t_2} \left(\fint_{\{u=t\}} \frac{1}{|\nabla u|}\right) dt} \qquad \text{by Jensen's inequality}$$

$$\geqslant \frac{A_0^2(t_2-t_1)^2}{\int_{t_1}^{t_2} \left(\int_{\{u=t\}} \frac{1}{|\nabla u|}\right) dt} = \frac{A_0^2(t_2-t_1)^2}{|M_+|} \geqslant \frac{A_0^2(t_2-t_1)^2}{|M|} \tag{19.12}$$

by (19.9). □ □

We note, furthermore, that continuity of u implies that the volume function

$$V(t) = |\{u \geqslant t\}|$$

is continuous for each t. The fact that the set of critical of u is closed and has zero measure, in conjunction with (19.9), implies that the function $V(t)$ is differentiable a.e. in $[t_1, t_2]$: For every regular value $t \in [t_1, t_2]$ of u we have:

$$\frac{d}{dt} V(t) = \int_{\{u=t\}} \frac{1}{|\nabla u|}.$$

19.5.3. Energy comparison in the case of a linear isoperimetric inequality. Recall that in §2.11.1 for each κ we defined X_κ, the unique complete simply-connected surface of the curvature κ. In the case when

$$\kappa = -c^2 < 0,$$

this surface is the upper half-plane $\mathbf{U}^2$ equipped with the Riemannian metric

$$\frac{dx^2 + dy^2}{c^2 y^2}.$$

Consider now a cyclic parabolic subgroup $\Gamma < \mathrm{Isom}(X_\kappa)$ and its quotient surface

$$\hat{N} := X_\kappa / \Gamma.$$

The group Γ preserves horoballs in X_κ with a common center in $\partial_\infty X_\kappa$ fixed by Γ. We will denote projections of these horoballs to $\hat{N}$ by $D_s, s \in \mathbb{R}_+$, with the convention that the length of the boundary of D_s equals s.

Exercise 19.29. Show that each $D = D_s$ satisfies the isoperimetric inequality

$$s = \mathrm{length}(\partial D) = c\mathrm{A}rea(D).$$

A function $\hat{u} : \hat{N} \to \mathbb{R}$ is said to be *radial* if it is constant on the circles ∂D_s, $s \in \mathbb{R}_+$.

Consider a Riemannian manifold N with compact boundary ∂N. We will assume that N has Cheeger constant $\geqslant c > 0$, i.e., N satisfies the linear isoperimetric inequality

$$|\partial \Omega| \geqslant c|\Omega|, \tag{19.13}$$

where $\Omega \subset N$ is any compact domain with smooth boundary. Our goal is to estimate from below the energy of smooth proper functions $u : N \to (0, 1]$, satisfying

$u^{-1}(\{1\}) = \partial N$. We will do so by comparing the energy of u with that of a suitable proper radial function $\hat{u} : \hat{N} \to (0, 1)$.

Given u, we define a proper radial function

$$\hat{u} : \hat{N} \to (0, 1)$$

such that the superlevel sets of $\hat{u}$ have the same volume as the corresponding superlevel sets of u:

$$|u^{-1}([t, 1))| = |\{\hat{u} \geqslant t\}| = |\{u \geqslant t\}| \quad \text{for all } t \in (0, 1).$$

Since the function $V(t) = |\{\hat{u} \geqslant t\}|$ is continuous and differentiable a.e., so is the function $\hat{u}$. For $t \in (0, 1)$, define $\hat{V}(t) = |\{t \leqslant \hat{u} < 1\}|$, $\hat{A}(t) = |\{\hat{u} = t\}|$ and $A(t) = |\{u = t\}|$. As we noted above,

$$\hat{A}(t) = c\hat{V}(t)$$

for each t.

FIGURE 19.1. Functions u and $\hat{u}$.

LEMMA 19.30 (Energy comparison lemma). *Suppose that for some $T \in (0, 1]$, we have*

$$V(T) \geqslant \frac{2}{c} A(1) = \frac{2}{c} |\partial N|.$$

Then

$$\int_{\{0<u\leqslant T\}} |\nabla u|^2 \geqslant \frac{1}{4} \int_{\{0<\hat{u}\leqslant T\}} |\nabla \hat{u}|^2 .$$

PROOF. Since $V(t) = \hat{V}(t)$, differentiating

$$V(t) = \int_t^1 \int_{u=\tau} \frac{1}{|\nabla u|} d\tau$$

with respect to t, we get that (for a.e. $t \in [0, 1]$)

$$\int_{\{u=t\}} \frac{1}{|\nabla u|} = \int_{\{\hat{u}=t\}} \frac{1}{|\nabla \hat{u}|}. \tag{19.14}$$

For all $t \leqslant T$, in view of the isoperimetric inequality (19.13), we have

$$|\partial\{u \geqslant t\}| = |\partial N| + A(t) \geqslant cV(t),$$

while

$$|\partial N| \leqslant \frac{c}{2} V(T) \leqslant \frac{c}{2} V(t).$$

By combining these inequalities, we obtain:

$$A(t) \geqslant cV(t) - |\partial N| \geqslant cV(t) - \frac{c}{2} V(t) \geqslant \frac{c}{2} V(t) = \frac{c}{2} \hat{V}(t) = \frac{\hat{A}(t)}{2} . \tag{19.15}$$

Now, for each regular value t of u,

$$\begin{aligned} \int_{\{u=t\}} |\nabla u| &\geqslant \frac{A^2(t)}{\int_{\{u=t\}} \frac{1}{|\nabla u|}} \qquad \text{see (19.11)} \\ &\geqslant \frac{\hat{A}^2(t)}{4 \int_{\{\hat{u}=t\}} \frac{1}{|\nabla \hat{u}|}} \qquad \text{by (19.14) and (19.15)} \end{aligned}$$

$$= \frac{c^2}{2}\int_{\{\hat{u}=t\}} |\nabla \hat{u}| \tag{19.16}$$

because $|\nabla \hat{u}|$ is constant on $\{u = t\}$ and so the equality case of (19.11) applies.

The lemma now follows from (19.8) and (19.16). □

19.5.4. Proof of positivity of the energy gap. Suppose $v \in \mathcal{F}$ with the set of regular values $R \subset (0,1)$. Every regular level set of v defines a nontrivial homology class in M, since it separates positive and negative ends of M. As we saw in Theorem 2.38,

$$\inf_{\tau \in R} |\{v = \tau\}| \geqslant A_0 > 0,$$

for a certain constant A_0.

Remark 19.31. This is another place where the proof simplifies considerably in the case when M is a surface: Then every homologically nontrivial cycle in M has length $\geqslant A_0$, the injectivity radius of M.

Choose a regular value $t_1 \in (0,1)$ of v where $A(t), t \in R$, almost attains its infimum, i.e.,

$$A(t_1) \geqslant \inf_{\tau \in R} |\{v = \tau\}| \geqslant A(t_1)/2.$$

We may assume that $t_1 \geqslant \frac{1}{2}$ (otherwise, we use the function $1 - v$ instead) and we focus attention on the codimension 0 submanifold with boundary $N \subset M$ given by the sublevel set $\{v \leqslant t_1\}$. Replacing v with $u = \frac{1}{t_1} v$, we get a proper function $u : N \to (0, 1]$ which is 1 on ∂N, such that all the level sets $\{u = t\}$ have area at least $\frac{1}{2}\operatorname{Area}(\partial N)$. Clearly,

$$E(v) \geqslant t_1^2 E(u) \geqslant \frac{E(u)}{4}.$$

Thus, it suffices to get a lower bound on $E(u)$. We will see below that

$$E(u) \geqslant \frac{cA_0}{32}.$$

Since the volume $V(t) = |\{u \geqslant t\}|$ is a continuous function of t which vanishes at $t = 1$ and satisfies

$$\lim_{t \to 0} |\{v \geqslant t\}| = \infty,$$

there exists a superlevel set $\{u \geqslant T\} \subset N$ whose volume equals $\frac{2}{c}|\partial N|$, where c is the Cheeger constant of M.

Case 1: $T \leqslant \frac{1}{2}$. Applying Lemma 19.28, we get

$$E(u) \geqslant \int_{\{T \leqslant u \leqslant 1\}} |\nabla u|^2 \geqslant \frac{(1-T)^2 A_0^2}{\frac{2}{c}|\partial N|},$$

$$\geqslant \frac{\left((1-T)\frac{|\partial N|}{2}\right)^2}{\frac{2}{c}|\partial N|} = \frac{(1-T)^2}{8} c|\partial N| \geqslant \frac{(1-T)^2}{8} cA_0 \geqslant \frac{cA_0}{32}.$$

Therefore, we obtain a lower energy bound for u (and, hence, v) in this case.

Case 2: $T \leqslant \frac{1}{2}$. Lemma 19.30 shows that the energy of u is at least

$$\frac{1}{4}\int_{\{\hat{u} \geqslant T\}} |\nabla \hat{u}|^2,$$

where $\hat{u}$ is the radial comparison function defined on the surface $\hat{N} = X_\kappa/\Gamma$.

By our choice of T,

$$\hat{V}(T) = |\{T \leqslant \hat{u} \leqslant 1\}| = |\{T \leqslant u \leqslant 1\}| = \frac{2}{c}|\partial N| \geqslant \frac{2}{c}A_0.$$

LEMMA 19.32.

$$E\left(\hat{u}\big|_{\{\hat{u}\leqslant T\}}\right) = \int_{0<\hat{u}\leqslant T} |\nabla \hat{u}|^2 \geqslant \frac{cA_0}{2}.$$

PROOF. We will identify the subsurface $\{0 < \hat{u} \leqslant T\} \subset \hat{N}$ with the rectangle

$$Q = \{(x, y) : 0 < y \leqslant 1, 0 \leqslant x \leqslant a\},$$

whose vertical sides are identified via the translation $(x, y) \mapsto (x + a, y)$. Since energy is a conformal invariant, it suffices to do the computation of energy with respect to the Euclidean metric. Accordingly, below, $|\nabla \hat{u}|$ is the Euclidean norm of the Euclidean gradient. Since the function $\hat{u}$ is radial, $\hat{u}(x, y) = f(y)$ and, hence,

$$|\nabla \hat{u}|^2 = f'^2.$$

We obtain:

$$E\left(\hat{u}\big|_{\{0<\hat{u}\leqslant T\}}\right) = a\int_0^1 (f')^2 dy \qquad \text{(by Cauchy's inequality)} \tag{19.17}$$

$$\geqslant a\left(\int_0^1 f' dy\right)^2 = aT^2 \geqslant \frac{a}{4} \quad (\text{since } T \geqslant \tfrac{1}{2}).$$

In order to estimate the number a, note that the area $\hat{V}(T)$ equals the area of the strip

$$P = \{(x, y) : 1 < y < \infty, 0 \leqslant x \leqslant a\} \subset X_\kappa.$$

The latter equals $\frac{a}{c^2}$ and, hence,

$$\frac{a}{c^2} = \hat{V}(T) \geqslant \frac{2}{c}A_0,$$

$$a \geqslant 2cA_0. \tag{19.18}$$

Combining the inequalities (19.17) and (19.18), we obtain:

$$E\left(\hat{u}\big|_{\{\hat{u}\leqslant T\}}\right) = \int_{0<\hat{u}\leqslant T} |\nabla \hat{u}|^2 \geqslant \frac{a}{4} \geqslant \frac{cA_0}{2}. \quad \square$$

Lemmata 19.30 and 19.32 imply that, in the Case 2:

$$E(u) \geqslant \frac{1}{4}E(\hat{u}\big|_{\{\hat{u}\leqslant T\}}) \geqslant \frac{cA_0}{8},$$

which is a higher bound than in the Case 1.

Therefore, the energy of the function $v : M \to (0, 1)$ is at least $\frac{1}{4}E(u) \geqslant 2^{-7}cA_0$. This completes the proof of Theorem 19.21. $\square$

CHAPTER 20

Quasiconformal mappings

Quasiconformal and quasisymmetric maps play prominent role in geometric analysis and geometric group theory. (The classes of quasiconformal and quasisymmetric maps coincide in the case of maps $\mathbb{S}^n \to \mathbb{S}^n$, $n \geqslant 2$.) Their importance in geometric group theory comes from the fundamental fact that quasisymmetric maps appear as *boundary extensions* of quasiisometries between Gromov-hyperbolic spaces: Each quasiisometry

$$f : X \to Y$$

extends to a unique quasisymmetric homeomorphism

$$f_\infty : \partial_\infty X \to \partial_\infty Y$$

of the Gromov boundaries of the spaces X and Y. Conversely, each quasisymmetric homeomorphism $\partial_\infty X \to \partial_\infty Y$ extends to a quasiisometry $X \to Y$ and any two such quasiisometric extensions are within bounded distance from each other. In the case when $X = Y = \mathbb{H}^{n+1}$, the extensions f_∞ are the *classical quasiconformal maps*. This remarkable interaction between quasiconformal analysis and hyperbolic geometry is somewhat akin to fruitful relation between complex analysis and hyperbolic geometry.

The intuition of classical quasiconformal maps comes from the theory of holomorphic functions of one complex variable: Conformal maps are characterized (locally) by the property that they send infinitesimal circles to infinitesimal circles. Accordingly, *classical quasiconformal maps* are defined by the condition that they send infinitesimal spheres to infinitesimal ellipsoids of uniformly bounded eccentricity.

We refer the reader to the books [**Res89**], [**Vuo88**], [**Väi71**] and [**IM01**] for the detailed discussion of classical quasiconformal maps and to [**HK95**], [**HK98**] and [**Hei01**] for the treatment of quasiconformal and quasisymmetric maps between more general metric spaces.

In this book we will be using only classical quasiconformal maps, whose basic analytical and geometric properties will be established in this chapter. The main applications of classical quasiconformal maps in the book are Mostow Rigidity Theorem (for lattices in the isometry groups $SO(n+1,1)$ of $\mathbb{H}^{n+1}$), Tukia's QI Rigidity Theorem for the class uniform lattices in $SO(n+1,1)$ and Schwarz' QI Rigidity Theorem for nonuniform lattices in $SO(n+1,1)$. These theorems will be proven in Chapters 21 and 22.

Historical Remark 20.1. Quasiconformal mappings between open subsets of the complex plane were introduced in the 1920s by Groetch as a generalization of conformal mappings. Quasiconformal mappings in higher dimensions were defined by Lavrentiev in the 1930s as a tool in applied mathematics (hydrodynamics). The discovery of the relation between quasiisometries of hyperbolic spaces and quasiconformal mappings was made independently by Efremovich and Tihomirova [**ET64**] and Mostow [**Mos65**] in the 1960-s.

20.1. Linear algebra and eccentricity of ellipsoids

Suppose that $M \in GL(n, \mathbb{R})$ is an invertible linear transformation. We would like to measure the deviation of M from being a conformal linear transformation, i.e., from being an element of $\mathbb{R}_+ \cdot O(n)$. Geometrically speaking, we are interested in measuring the deviation of the ellipsoid $E = M(\mathbb{D}) \subset \mathbb{R}^n$ from a round ball, where $\mathbb{D} = \mathbb{D}^n$ is the unit ball in $\mathbb{R}^n$.

In the case $n = 2$, there is only one way to define such a measurement, namely, *eccentricity* of the ellipsoid E, which is the ratio of major to minor axes of E. In higher dimensions, there are several invariants which are useful in different situations. This reflects the simple fact that the matrix M has n singular values, while the invariants we are looking for are single real numbers.

Recall that every invertible $n \times n$ matrix M has the *singular value decomposition* (see Theorem 1.115)

$$M = UDV = UDiag(\lambda_1, ..., \lambda_n)V,$$

where the (positive) diagonal entries $\lambda_1 \leqslant ... \leqslant \lambda_n$ of the diagonal matrix $D = Diag(\lambda_1, ..., \lambda_n)$ are the *singular values* of M. Here U, V are orthogonal matrices. Equivalently, if we symmetrize M (i.e., set $A = MM^T$), then the numbers λ_i are square roots of the eigenvalues of A. Geometrically speaking, the singular values λ_i are the half-lengths of the axes of the ellipsoid $E = M(\mathbb{D})$.

We define the following *distortion quantities* for the matrix M:

- Linear dilatation:
$$H(M) := \frac{\lambda_n}{\lambda_1} = \|M\| \cdot \|M^{-1}\|,$$
where $\|M\|$ is the operator norm of a matrix M:
$$\max_{v \in \mathbb{R}^n \setminus 0} \frac{|Mv|}{|v|}.$$
Thus, $H(M) = \epsilon(E)$ is the *eccentricity* of the ellipsoid E, the ratio of lengths of major and minor axes of E. This is the invariant that we will be using most of the time.
- Inner dilatation:
$$H_I(M) := \frac{\lambda_1 .... \lambda_n}{\lambda_1^n} = |\det(M)| \cdot \|M^{-1}\|^n.$$
- Outer dilatation:
$$H_O(M) := \frac{\lambda_n^n}{\lambda_1 .... \lambda_n} = \|M\|^n |\det(M)|^{-1}.$$
- Maximal dilatation:
$$K(M) := \max(H_I(M), H_O(M)).$$

Thus, geometrically speaking, the inner and outer dilatations compute volume ratios of E and inscribed/circumscribed balls, while the linear dilatation compares the radii of inscribed/circumscribed balls. Note that all four dilatations agree for $n = 2$.

EXERCISE 20.2. M is conformal $\iff H(M) = 1 \iff H_I(M) = 1 \iff H_O(M) = 1 \iff K(M) = 1$.

EXERCISE 20.3. Logarithms of linear and maximal dilatations are comparable:

$$(H(M))^{n/2} \leqslant K(M) \leqslant (H(M))^{n-1} .$$

Hint: It suffices to consider the case when M is a diagonal matrix $Diag(\lambda_1, ...\lambda_n)$.

EXERCISE 20.4. 1. $H(M) = H(M^{-1})$ and $H(M_1 \cdot M_2) \leqslant H(M_1) \cdot H(M_2)$.
2. $K(M) = K(M^{-1})$ and $K(M_1 \cdot M_2) \leqslant K(M_1) \cdot K(M_2)$.
Hint: Use geometric interpretation of the four dilatations.

20.2. Quasisymmetric maps

Our next goal is to generalize the dilatation constants of linear maps to nonlinear maps. The linear dilatation is the easiest to generalize, since it deals only with distances. Recall the geometric meaning of the linear dilatation $H(M)$: If E is the image of the round ball $\mathbb{D}$, then $H(M)$ is the ratio of the "outer radius" of E by its "inner radius." Such ratio makes sense not only for ellipsoids but also for arbitrary (closed) topological balls $D \subset \mathbb{R}^n$, where we have chosen a "center" , a point x' in the interior of D: Then we have two real numbers r and R, such that $B(x', r)$ is the largest metric ball (centered at x') contained in D and $\bar{B}(x', R)$ is the smallest metric ball containing D. Then the numbers r and R can be regarded as the inner and outer radii of D. In other words,

$$\frac{R}{r} = \max \frac{|y' - x'|}{|z' - x'|},$$

where the maximum is taken over all points $y', z' \in \partial D$. This ratio is the "eccentricity" of the topological ball $D \subset \mathbb{R}^n$. The idea then is to consider homeomorphisms f which send round balls $B(x, \rho)$ to topological balls of uniformly bounded eccentricity with respect to the "centers" $x' = f(x)$.

This leads to

DEFINITION 20.5. A homeomorphism $f : \Omega \to \Omega'$ between two domains in $\mathbb{R}^n$ is *c-weakly quasisymmetric* if

$$\frac{|f(x) - f(y)|}{|f(x) - f(z)|} \leqslant c \tag{20.1}$$

for all $x, y, z \in \Omega$, such that $|x - y| = |y - z| > 0$.

Note that we do not assume that f preserves orientation. We will be mostly interested in the case $\Omega = \Omega' = \mathbb{R}^n$.

The name *quasisymmetric* comes from the case $n = 1$ (and quasisymmetric maps were originally introduced only for $n = 1$ by Ahlfors and Beurling [**AB56**]). Namely, a homeomorphism $f : \mathbb{R} \to \mathbb{R}$, satisfying $f(0) = 0$, is *symmetric* at the origin if it sends any pair of points symmetric about 0 to points symmetric about 0, i.e., these homeomorphisms are *odd* functions: $f(-y) = -f(y)$. In the case of c-weakly quasisymmetric maps, the exact symmetry is lost, but is replaced by a uniform bound on the ratio of absolute values.

EXERCISE 20.6. Show that 1-weakly quasisymmetric homeomorphisms $f : \mathbb{R} \to \mathbb{R}$ are compositions of dilations and isometries of $\mathbb{R}$.

It turns out that there is a slightly stronger condition, which is a bit easier to work with and which generalizes naturally to metric spaces other than $\mathbb{R}^n$:

DEFINITION 20.7. Fix a homeomorphism $\eta : [0, \infty) \to [0, \infty)$. A homeomorphism $f : \Omega \subset \mathbb{R}^n \to \Omega' \subset \mathbb{R}^n$ is called η-quasisymmetric if for all $x, y, z \in \mathbb{R}^n$ we have

$$\frac{|f(x) - f(y)|}{|f(x) - f(z)|} \leqslant \eta \left(\frac{|x - y|}{|x - z|} \right). \tag{20.2}$$

Thus, if we take $c = \eta(1)$, then every η-quasisymmetric map is also c-weakly quasisymmetric. It is a nontrivial theorem (see e.g. [**Hei01**]) that for $\Omega = \Omega' = \mathbb{R}^n$, the two concepts are equivalent.

EXERCISE 20.8. Show that:

1. Every invertible affine transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ is η-quasisymmetric for $\eta(t) = H(L)t$.
2. L-Lipschitz homeomorphisms are η-quasisymmetric with $\eta(t) = L^2 t$.

As in the case of quasiisometries, we will say that a homeomorphism is (weakly) quasisymmetric if it is η-quasisymmetric (respectively c-weakly quasisymmetric) for some η or $c < \infty$.

The following exercise requires nothing but the definition of quasisymmetry:

EXERCISE 20.9. Show that the composition of quasisymmetric maps is again quasisymmetric. Show that the inverse of a quasisymmetric map is also quasisymmetric.

Recall that we think of $\mathbb{S}^n$ as the 1-point compactification of $\mathbb{R}^n$. Accordingly, we can define quasisymmetric homeomorphisms of $\mathbb{R}^n \cup \{\infty\}$ as extensions of quasisymmetric homeomorphisms $\mathbb{R}^n \to \mathbb{R}^n$. The drawback of this definition of quasisymmetric maps is that we are restricted to maps sending the point ∞ to itself. In particular, we cannot apply this definition to Moebius transformations.

DEFINITION 20.10. A homeomorphism of $\mathbb{S}^n$ is called *quasimoebius* if it is a composition of a Moebius transformation with a quasisymmetric map.

Recall (Theorem 8.4) that Moebius transformations of $\mathbb{S}^n$ can be characterized by the property that they preserve the cross-ratios

$$[x, y, z, w] := \frac{|x - y| \cdot |z - w|}{|y - z| \cdot |w - x|}, x, y, z, w \in \mathbb{S}^n.$$

Similarly, one can prove (see [**Väi85**]) that a homeomorphism f of $\mathbb{S}^n$ is quasimoebius if and only if there exists a homeomorphism $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$[f(x), f(y), f(z), f(w)] \leqslant \kappa([x, y, z, w])$$

for all $x, y, z, w \in \mathbb{S}^n$. While the notion of quasimoebius maps is esthetically appealing, we will be working mostly with quasisymmetric and *quasiconformal maps*, which will be introduced in the next section.

20.3. Quasiconformal maps

The idea of quasiconformality is very natural: We take the definition of weakly quasiconformal maps *via* the ratio (20.1) and then take the limit in this ratio as $\rho = |x - y| = |y - z| \to 0$.

For a homeomorphism $f : \Omega \subset \mathbb{R}^n \to \Omega' \subset \mathbb{R}^n$ between two domains in $\mathbb{R}^n$ and $x \in \Omega$ we define the quantity

$$H_x(f) := \limsup_{\rho \to 0} \left(\max_{y,z} \frac{|f(x) - f(y)|}{|f(x) - f(z)|} \right), \tag{20.3}$$

where, for each $\rho > 0$, the maximum is taken over all $y, z \in \Omega$ with $\rho = |x - y| = |x - z|$. For instance, if f is c-weakly quasisymmetric, then $H_x(f) \leqslant c$ for every $x \in \Omega$.

DEFINITION 20.11. A homeomorphism $f : \Omega \to \Omega'$ is called *quasiconformal* if

$$\sup_{x \in \Omega} H_x(f)$$

finite.

The function $H_x(f)$ is called *the (linear) dilatation function* of f; a quasiconformal map f is said to have *dilatation* $\leqslant H$ if

$$H(f) := ess \sup_{x \in \Omega} H_x(f) \leq H.$$

Note that the essential supremum is the L^∞-norm, thus, it ignores subsets of measure zero. We will see the reason for this discrepancy between the definition of quasiconformality (where $H_x(f)$ is required to be uniformly bounded) and the definition of dilatation $H(f)$, in the next section.

Thus, the intuitive meaning of quasiconformality is that *quasiconformal maps send infinitesimal spheres to infinitesimal ellipsoids of uniformly bounded eccentricity.*

EXERCISE 20.12. Let $f : \mathbb{S}^n \to \mathbb{S}^n$ be a Moebius transformation, $p = f^{-1}(\infty)$. Then the restriction

$$f\big|_{\mathbb{R}^n \setminus \{p\}}$$

is 1-quasiconformal, i.e., conformal. Hint: It suffices to verify conformality only for the inversion in the unit sphere.

Note that here and in what follows *we do not assume that conformal maps preserve orientation.* For instance, in this terminology, complex conjugation is a conformal map $\mathbb{C} \to \mathbb{C}$.

EXERCISE 20.13. 1. Suppose that $f : \Omega \to \Omega'$ is a C^1-diffeomorphism such that $\|D_x(f)\|$ is uniformly bounded above and $|J_x(f)|$ is uniformly bounded below. Show, using the definition of differentiability, that f is quasiconformal. Namely, verify that $H_x(f) = H(D_x(f))$ for every $x \in \Omega$.

2. Show that every C^1-diffeomorphism $\mathbb{S}^n \to \mathbb{S}^n$ is quasiconformal.

20.4. Analytical properties of quasiconformal mappings

In this section we list certain analytical properties of quasiconformal (quasisymmetric) mappings used in the book. We will prove most of them with two notable exceptions: Gehring's version of the Liouville's theorem and Tukia's Strong Convergence Property. (The Measurable Mapping Theorem in the next chapter is another exception.) Proving these theorems would go well beyond the scope of this book.

20.4.1. Some notions and results from real analysis. For a subset $E \subset \mathbb{R}^n$ we let $mes(E)$ denote the n-dimensional Lebesgue measure of E. In what follows, Ω is an open subset in $\mathbb{R}^n$.

20.4.1.A. *Derivatives of measures.* Let μ be a measure on Ω of Lebesgue class, i.e., μ-measurable sets are in the Borel σ-algebra. The *derivative* of μ at $x \in \Omega$, denoted $\mu'(x)$, is defined as

$$\mu'(x) := \limsup \frac{\mu(B)}{mes(B)},$$

where the limit is taken over all balls B containing x whose radii tend to zero. The key fact that we will need is the following theorem (see e.g. [**Fol99**, Theorem 3.22]):

THEOREM 20.14 (Lebesgue–Radon–Nikodym differentiation theorem). *The function $\mu'(x)$ is Lebesgue–measurable and is finite a.e. in Ω. Furthermore, $\mu'(x)$ is the Radon–Nikodym derivative of the component of μ which is absolutely continuous with respect to the Lebesgue measure.*

For a continuous map $f : \Omega \to \mathbb{R}^m$ we define the pull-back measure $\mu = \mu_f$ by

$$\mu(E) := mes(f(E)).$$

20.4.1.B. *Approximate continuity.* A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *approximately continuous* at a point $x \in \mathbb{R}^n$ if for every $\epsilon > 0$

$$\lim_{r\to 0} \frac{mes(\{y \in B(x,r) : |f(x) - f(y)| > \epsilon\})}{mes(B(x,r))} = 0. \tag{20.4}$$

(Here, as before, mes denotes Lebesgue measure.) In other words, as we "zoom into" the point x, "most" points $y \in B(x,r)$, have value $f(y)$ close to $f(x)$, i.e., the rescaled functions $f_r(x) := f(rx)$ converge (with $r \to 0$) in measure to the constant function.

LEMMA 20.15. *Every L_∞-function $f : \mathbb{R}^n \to \mathbb{R}$ is approximately continuous at almost every point.*

PROOF. The proof is an application of *The Lebesgue Density theorem* (see e.g. [**SS05**, p. 106]): For every measurable function h on $\mathbb{R}^n$ and almost every x,

$$\lim_{r\to 0} \frac{1}{mes(B_r)} \int_{B_r} |h(y) - h(x)| dy = 0.$$

Here and below, we set $B_r = B(x,r)$.

Fix $\epsilon > 0$ and let $E_r \subset B_r$ denote the subset consisting of $y \in B_r$ with

$$|f(y) - f(x)| > \epsilon.$$

If the equality (20.4) fails, then

$$\lim_{r\to 0} \frac{mes(E_r)}{mes(B_r)} > 0.$$

By the definition of the subset R_r we have the inequality:

$$\frac{1}{mes(B_r)} \int_{B_r} |f(y) - f(x)| dy \geqslant \epsilon \frac{mes(E_r)}{mes(B_r)}.$$

Since

$$\lim_{r\to 0} \frac{mes(E_r)}{mes(B_r)} > 0,$$

we conclude that

$$\liminf_{r\to 0} \frac{1}{mes(B_r)} \int_{B_r} |f(y) - f(x)| dy \neq 0,$$

contradicting the Lebesgue Density Theorem. □

20.4.1.C. *Rademacher–Stepanov Theorem.* Rademacher–Stepanov theorem is a strengthening of Rademacher's theorem (Theorem 1.75); we will need it in order to prove differentiability a.e. of quasiconformal mappings, among other things.

Recall that a map $f : \Omega \to \mathbb{R}^m$ is called *differentiable* at $x \in \Omega$ with the derivative $D_x f$ at x equal to the matrix A, if

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0.$$

It follows directly from the definition that, for $n = m$, at every point x of differentiability of f, the measure derivative of μ_f equals the absolute value of the Jacobian of f:

$$\mu'_f(x) = |\det(A)| = |J_f(x)|.$$

The other key result that we will use is:

THEOREM 20.16 (Rademacher and Stepanov, see e.g. Theorem 3.4 in [**Hei05**]). *Let $f : \Omega \to \mathbb{R}^m$. For every $x \in \Omega$ define*

$$\|D_x^+(f)\| := \limsup_{h\to 0} \frac{|f(x+h) - f(x)|}{|h|}.$$

Then f is differentiable a.e. in the set $E = \{x \in \Omega : \|D_x^+(f)\| < \infty\}$.

A special case of this theorem is Rademacher's theorem (Theorem 1.75), since for L-Lipschitz maps

$$\|D_x^+(f)\| \leqslant L.$$

20.4.1.D. *Absolutely continuous functions.* Informally, absolutely continuous functions are those which map sets of small measure to sets of small measure. More precisely, suppose that f is a real-valued function defined on an interval I in $\mathbb{R}$. The function f is called *absolutely continuous* (AC) if for every $\epsilon > 0$ there exists $\delta > 0$ such that: For every collection of pairwise disjoint subintervals

$$(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k) \subset I$$

with

$$\sum_{i=1}^{k} |b_i - a_i| < \delta,$$

we have

$$\sum_{i=1}^{k} |f(b_i) - f(a_i)| < \epsilon. \tag{20.5}$$

In particular, AC functions send subsets of zero measure to subsets of zero measure (this is nearly clear from the definition, the reader can find the details in [**Fol99**, Proposition 3.32]).

Clearly, every absolutely continuous function is uniformly continuous, but the converse is false. For instance, the Cantor function, defined using a Cantor set C of zero measure, sends C to the unit interval.

AC functions are characterized by the fact that the Fundamental Theorem of Calculus holds for them:

THEOREM 20.17. *The following properties are equivalent for a function* $f : [a, b] \to \mathbb{R}$*:*

1. f *is absolutely continuous.*

2. There exists a function $h \in L^1([a, b])$ *such that*

$$f(x) = \int_a^x h(t)dt$$

for all $x \in [a, b]$*.*

3. The function f *is differentiable almost everywhere in* $[a, b]$ *with measurable derivative* $f'(x)$*, such that*

$$f(x) = \int_a^x f'(t)dt.$$

for all $x \in [a, b]$*.*

We refer the reader to [**Fol99**, Theorem 3.35] or [**SS05**, Theorem 3.11] for a proof.

The notion of absolutely continuous function generalizes readily to functions of one variable with values in $\mathbb{R}^n$ where in the formula (20.5) instead of the absolute value we use the norm in $\mathbb{R}^n$.

EXERCISE 20.18. Show that a function $f = (f_1, \ldots, f_m) : I \to \mathbb{R}^m$ is absolutely continuous if and only if each component f_i of f is absolutely continuous.

We will need a sufficient condition for absolute continuity of vector-valued functions:

LEMMA 20.19. *Suppose that* $f : I \to \mathbb{R}^m$ *is a continuous function for which there exists a constant* C *such that for every measurable subset* $E \subset I$ *we have*

$$\mu_1(f(E)) \le C\ mes(E).$$

Then f *is absolutely continuous.*

PROOF. It follows immediately from the definition of the 1-dimensional (normalized) Hausdorff measure that each component f_i of the function f also satisfies

$$mes(f_i(E)) \le C\ mes(E).$$

Absolute continuity of f_i follows by taking in the definition of absolute continuity

$$\delta = \epsilon/C. \quad \square$$

We will use the following generalization of the notion of absolute continuity to the case of functions of several variables:

DEFINITION 20.20. A map $f : \Omega \to \mathbb{R}^m$ defined on an open subset $\Omega \subset \mathbb{R}^n$, is called ACL, *absolutely continuous on lines*, if the restriction of f to almost every coordinate line segment in Ω is an absolutely continuous function of one variable. Here and in what follows, a coordinate line segment is a compact straight line segment parallel to one of the coordinate axes in $\mathbb{R}^n$.

20.4.2. Differentiability properties of quasiconformal mappings. We now return to quasiconformal maps. Recall that the dilatation $H_x(f)$ of a homeomorphism f at a point x is defined as

$$H_x(f) := \limsup_{\rho\to 0} \frac{R(x,\rho)}{r(x,\rho)},$$

where

$$R(x,\rho) = \max_{|h|=\rho} |f(x+h) - f(x)|,\ r(x,\rho) = \min_{|h|=\rho} |f(x+h) - f(x)|. \tag{20.6}$$

20.4.2.A. *Differentiability a.e. of quasiconformal homeomorphisms.*

THEOREM 20.21 (F. Gehring, see [**Väi71**]). *Every quasiconformal map $f : \Omega \to \mathbb{R}^n$ is differentiable a.e. in Ω and*

$$\|D_x f\| \leqslant H_x(f)|J_x(f)|^{1/n}$$

for a.e. x in Ω.

PROOF. By the definition of $|D_x^+(f)|$ and $H_x(f)$:

$$\|D_x^+(f)\| = \limsup_{\rho\to 0} \frac{R(x,\rho)}{\rho} = H_x(f) \limsup_{\rho\to 0} \frac{r(x,\rho)}{\rho}.$$

Notice that for $r = r(x,\rho)$, $B(f(x), r) \subset f(B(x,\rho))$, which implies that

$$\omega_n r^n = mes\,(B(f(x), r)) \leqslant mes\,(f(B(x,\rho))),$$

where ω_n is the volume of the unit n-ball. Therefore,

$$\frac{mes(f(B(x,\rho)))}{mes(B(x,\rho))} \geqslant \frac{r^n}{\rho^n}$$

and, thus,

$$\mu_f'(x) = \limsup_{\rho\to 0} \frac{mes(f(B(x,\rho)))}{mes(B(x,\rho))} \geqslant \limsup_{\rho\to 0} \frac{r^n}{\rho^n} = \left(\frac{1}{H_x(f)}\|D_x^+(f)\|\right)^n.$$

It follows that

$$\|D_x^+(f)\| \leqslant H_x(f)(\mu_f'(x))^{1/n}.$$

The right-hand side of this inequality is finite for a.e. x (by Theorem 20.14). Thus, f is differentiable at a.e. x by Rademacher-Stepanov theorem. We also obtain (for a.e. $x \in \Omega$)

$$\|D_x(f)\| = \|D_x^+(f)\| \leqslant H_x(f)(\mu_f'(x))^{1/n} = H_x(f)|J_x(f)|^{1/n} \quad \square$$

20.4.2.B. *ACL property of quasiconformal mappings.* Gehring's differentiability theorem is strengthened as follows:

THEOREM 20.22 (F. Gehring, J. Väisälä, see [**Väi71**] and [**Mos73**]). *For $n \geqslant 2$, quasiconformal maps $f : \Omega \to \mathbb{R}^n$ belong to the Sobolev class $W^{1,n}_{loc}$, i.e., their 1st partial distributional derivatives are locally in $L^n(\Omega)$. This, in particular, implies that quasiconformal maps are ACL.*

We will prove a slightly weaker, but sufficient for our purposes, version of this important theorem:

THEOREM 20.23. *Every weakly quasisymmetric homeomorphism $f : \Omega \to \Omega'$, between domains in $\mathbb{R}^n$, $n \geqslant 2$, is ACL. In particular, the matrix of partial derivatives $D_x(f)$ is a measurable matrix-valued function on Ω.*

PROOF. Our proof closely follows the one given by Mostow in [**Mos73**]. Until the very end of the proof, we will not be using the assumption $n \geqslant 2$. Let κ be the quasisymmetry constant of f. Then:

$$r(p,t) \leqslant R(p,t) \leqslant \kappa r(p,t)$$

for all $t > 0$ and all $p \in \Omega$, where $R(p,t)$ and $r(p,t)$ are defined by the equations (20.6). Let $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be the orthogonal projection, where $\mathbb{R}^{n-1}$ is one of the coordinate hyperplanes in $\mathbb{R}^n$, defined by the equation $x_i = 0$. Fix a bounded open subset $Q \subset \Omega$ and for $y \in \mathbb{R}^{n-1}$ define

$$Q(y) := Q \cap \pi^{-1}(y).$$

For $t > 0$ set

$$Q(y,t) := \mathcal{N}_t(Q(y)) \cap Q.$$

Clearly, the subset $\pi^{-1}(B(y,t)) \cap Q$ contains $Q(y,t)$.

LEMMA 20.24. *For almost all $y \in \mathbb{R}^{n-1}$, the limit*

$$\tau(y) := \limsup_{t\to 0} \frac{mes(f(Q(y,t)))}{t^{n-1}}$$

is finite.

PROOF. For each measurable subset $E \subset \mathbb{R}^{n-1}$, define

$$\phi(E) = mes(f(\pi^{-1}(E) \cap Q)).$$

Then ϕ is a measure of Lebesgue class on $\mathbb{R}^{n-1}$ and by Theorem 20.14 we obtain that for almost all $y \in \mathbb{R}^{n-1}$

$$\limsup_{t\to 0} \frac{\phi(B(y,t))}{mes(B(y,t))} < \infty.$$

Clearly,

$$mes(f(Q(y,t))) \leqslant \phi(B(y,t)).$$

On the other hand, t^{n-1} is (up to a constant factor) equal to $mes(B(y,t))$. The lemma follows. ☐

We now claim that for every y such that $\tau(y) < \infty$, the function f is absolutely continuous on the intersection of the line $\pi^{-1}(y)$ with the set Q. Recall that μ_1 denotes the 1-dimensional Hausdorff measure of subsets in $\mathbb{R}^n$. The following lemma is the key to absolute continuity:

LEMMA 20.25. *There exists a constant C, depending only on n and κ, such that for every y satisfying $\tau(y) < \infty$ and for every compact subset $E \subset Q(y)$, we have*

$$\mu_1(f(E))^n \leqslant C\tau(y)(\mu_1(E))^{n-1}.$$

PROOF. It suffices to prove the lemma in the case when E is a closed interval. Let $L = \mu_1(E)$ denote the length of the interval E. For each compact $K \subset \mathbb{R}^n$ define

$$\Lambda(K,a) = \inf_{\mathcal{U}} \sum_{U_i \in \mathcal{U}} \operatorname{diam}(U_i),$$

where $\mathcal{U}$'s are finite coverings of K by balls U_i satisfying $\operatorname{diam}(U_i) \leqslant 2a$. By the definition of the Hausdorff measure,

$$\liminf_{a\to 0} \Lambda(K,a) = \mu_1(K),$$

the 1-dimensional Hausdorff measure of K. Therefore, in order to prove the lemma, we need to establish a suitable upper bound on $\Lambda(K,a)$ for small values of a.

Now, pick $a > 0$ such that $\mathcal{N}_a(E) \subset Q$. By continuity of f, there exists a constant $c > 0$ such that for all p in E, we have $R(p,c) < a$. Next, pick $t > 0$ such that $t < c$ and $L/t = N \in \mathbb{N}$. Choose points $p_1, \ldots, p_N \in E$ such that

$$|p_i - p_j| = t|i-j|.$$

Then

$$E \subset \bigcup_{i=1}^{N} B(p_i,t).$$

Set

$$s_i = R(p_i,t), i = 1,\ldots,N.$$

Clearly, for all i we have $s_i < a$ and

$$\kappa^{-1}s_i \leqslant r(p_i,t).$$

Furthermore, by the definition of the quasisymmetry constant κ,

$$B(f(p_i),\kappa^{-1}s_i) \subset B(f(p_i),r(p_i,t)) \subset f(B(p_i,t)) \subset B(f(p_i),s_i).$$

Since

$$E \subset \bigcup_{i=1}^{N} B(p_i,t),$$

we obtain

$$K := f(E) \subset \bigcup_{i=1}^{N} f(B(p_i,t)) \subset \bigcup_{i=1}^{N} B(f(p_i),s_i).$$

Therefore, for $B_i := B(f(p_i),s_i)$, the set

$$\mathcal{B} = \{B_i : i = 1,\ldots,N\}$$

is a covering of K and the radius of each ball B_i is less than a. In particular, by the definition of Λ, we have:

$$\Lambda(K,a) \leqslant \sum_i 2s_i.$$

In addition to the balls B_i, we define smaller concentric balls D_i:

$$D_i := \kappa^{-1}B_i = B(f(p_i),\kappa^{-1}s_i) \subset f(B(p_i,t)) \subset B_i, i = 1,\ldots,N.$$

Since the covering $\{B(p_i,t) : i = 1,\ldots,N\}$ has multiplicity 3 and f is 1-1, it follows that the collection of balls $\mathcal{D} = \{D_i : i = 1,\ldots,N\}$ has multiplicity $\leqslant 3$ as well.

Let q be such that

$$\frac{1}{n} + \frac{1}{q} = 1.$$

Then, by the Hølder inequality,

$$\sum_{i=1}^{N} 2s_i \leqslant \left(\sum_{i=1}^{N} 2^q\right)^{1/q} \left(\sum_{i=1}^{N} s_i^n\right)^{1/n} = 2N^{1/q} \cdot \left(\sum_{i=1}^{N} s_i^n\right)^{1/n}.$$

Hence, for ω_n, volume of the unit ball in $\mathbb{R}^n$,

$$\Lambda(K,a) \leqslant 2\omega_n^{1/n} N^{1/q} \left(\sum_{i=1}^{n} mes(B_i)\right)^{1/n}.$$

Since $\mathcal{D}$ has multiplicity $\leqslant 3$, we obtain

$$\sum_m \mu_n(D_i) \leqslant 3\mu_n(\bigcup_i D_i).$$

Recall that we need to estimate the sum of the volumes of the balls B_i from above (this would lead to an upper bound on $\Lambda(K, a)$).

We have:

$$B_i = \kappa D_i$$

and, hence,

$$mes(B_i) = \kappa^n mes(D_i),$$

$$\sum_i mes(B_i) = \kappa^n \sum_i mes(D_i) \leqslant 3\kappa^n mes(\bigcup_i D_i) \leqslant$$

$$3\kappa^n mes\left(f(\bigcup_i B(p_i, t))\right) \leqslant 3\kappa^n mes\left(f(Q(y, t))\right),$$

since

$$B(p_i, t) \subset Q(y, t), i = 1, \ldots, N.$$

By combining the inequalities we obtain

$$\Lambda(f(E), a)^n \leqslant 2\omega_n^{1/n} N^{n/q} \cdot 3\kappa^n mes\left(f(Q(y, t))\right) = C' N^{n-1} mes(f(Q(y, t)),$$

where $C' = 6\omega_n^{1/n}\kappa^n$. Therefore,

$$\Lambda(f(E), a)^n \leqslant C'(Nt)^{n-1} \frac{mes\left(f(Q(y, t)\right)}{t^{n-1}} = C' L^{n-1} \frac{mes\left(f(Q(y, t)\right)}{t^{n-1}},$$

and, by taking the limit as $t \to 0$, we get:

$$\Lambda(f(E), a)^n \leqslant C' L^{n-1} \tau(y)$$

for all sufficiently small t. Since this inequality holds for all sufficiently small $a > 0$, we obtain:

$$\mu_1(f(E))^n \leqslant C' 2^{n-1} (\mu_1(E))^{n-1} \tau(y) = C\tau(y)(\mu_1(E))^{n-1}.$$

The lemma follows. □

We now can finish the proof of Theorem 20.23. Lemma 20.25 implies that for a.e. $y \in \mathbb{R}^{n-1}$,

$$\mu_1(f(E)) \leqslant (C\tau(y))^{1/n} \mu_1(E)^{1-\frac{1}{n}}. \tag{20.7}$$

Since $n \geqslant 2$, this inequality implies that the function

$$f\big|_{\pi^{-1}(y)}$$

satisfies the hypothesis of Lemma 20.19, which, in turn, implies that this restriction is absolutely continuous. We conclude that the map f is ACL. □

Remark 20.26. Note that for $n = 1$, the inequality (20.7) says nothing interesting about the measure of $f(E)$. Furthermore, some quasisymmetric maps fail to be absolutely continuous.

Theorem 20.23 has an important corollary:

Corollary 20.27 (F. Gehring, J. Väisälä, see **[Väi71]**). *For $n \geqslant 2$, every quasiconformal mapping $f : \Omega \to \mathbb{R}^n$ has a.e. nonvanishing Jacobian: $J_x(f) \neq 0$ a.e. in Ω.*

Proof. We will prove a weaker property that will suffice for our purposes, which is that $J_x(f) \neq 0$ on a subset of a positive measure, under the assumption that f is weakly quasisymmetric. Suppose to the contrary that $J_x(f) = 0$ a.e. in Ω. The inequality

$$\|D_x f\| \leqslant H_x(f)|J_x(f)|^{1/n},$$

established in Theorem 20.21, then implies that $D_x f = 0$ a.e. in Ω. Thus, all partial derivatives of f vanish a.e. in Ω. Let $J = [p, q = p + Te_1]$ be a nondegenerate coordinate line segment (parallel to the x_1-axis), connecting p to q, on which f is absolutely continuous. This means that the Fundamental Theorem of Calculus applies to $f\big|_J$:

$$f(q) - f(p) = \int_J \frac{\partial}{\partial x_1} f(x) dx_1 = \int_0^T \frac{d}{dt} f(p_1 + te_1, p_2, ..., p_n) dt = 0.$$

Hence, $f(p) = f(q)$ contradicting injectivity of f. □

Remark 20.28. We refer the reader to **[BK13a]** for a different self-contained proof of this theorem, again, under the quasisymmetry assumption on f.

20.4.2.C. *Analytical definition of quasiconformality.* Since quasiconformal maps are differentiable a.e., it is natural to ask if quasiconformality of a map could be defined analytically, in terms of its derivatives. Theorem below gives two alternative analytical definitions of quasiconformality. Even though we will not need this result, it provides a useful prospective on nature of quasiconformal mappings. We remind the reader that $K(A)$ is the maximal dilatation of the linear transformation A.

Theorem 20.29. *Suppose that $f : \Omega \to \Omega' \subset \mathbb{R}^n$ is a homeomorphism. Then the following are equivalent:*

1. f is a quasiconformal mapping.

2. $D_x(f)$ is in $W^{1,n}_{loc}(\Omega)$ and

$$K(f) := ess \sup_{x \in \Omega} K(D_x(f)) < \infty. \tag{20.8}$$

3. The mapping f is ALC and satisfies (20.8).

Lastly, analytically and geometrically defined dilatations of f are related by:

$$H_x(f) = H(D_x f)$$

for a.e. $x \in \Omega$.

We refer the reader to **[Väi71]** for the proof of this theorem. In view of Theorem 20.29, we arrive to

Definition 20.30. A homeomorphism f is called K-quasiconformal if $K(f) \leqslant K$. The number $K(f)$ is called the *quasiconformality constant* of f.

The reason for defining K-quasiconformality in terms of maximal dilatation $K(D_x(f))$ instead of $H_x(f)$ is that K-quasiconformality is equivalent to yet another, more geometric, definition, in terms of extremal length (modulus) of families of curves. The latter definition, for historic reasons, is the main definition of quasiconformality, see **[Väi71]**.

According to Exercise 20.3, the two key measures of quasiconformality, $H(f)$ and $K(f)$ are log-comparable, therefore, using one or the other is only a matter of convenience. What is most important is that $K(f) = 1$ if and only if $H(f) = 1$. If $n = 2$, then, of course, $K_x(f) = H_x(f)$ and $K(f) = H(f)$.

REMARK 20.31. We can now explain the discrepancy in the definition of maps with bounded dilatation: The condition that $H_x(f)$ is bounded is needed in order to ensure that $f : \Omega \to \mathbb{R}^n$ belongs to $W^{1,n}_{loc}(\Omega)$; on the other hand, the actual bound on dilatation is computed only almost everywhere in Ω. This makes sense since derivatives of f exist only almost everywhere.

20.4.2.D. *Liouville theorem.* Recall that the classical Liouville's theorem which states that smooth conformal maps between domains in $\mathbb{S}^n, n \geqslant 3$, are restrictions of Moebius transformations. Gehring's theorem below shows how one can relax the smoothness assumption in Liouville's theorem:

THEOREM 20.32 (F. Gehring). *Every* 1*-quasiconformal homeomorphism of an open connected domain in* $\mathbb{S}^n$ *(*$n \geqslant 3$*) is the restriction of a Moebius transformation.*

Proofs of this theorem can be found in [**IM01**], [**Res89**] and [**Väi71**].

Liouville's theorem fails, of course, in dimension 2. We will see, however, in §21.5.1 that an orientation-preserving quasiconformal homeomorphism $f : \Omega \to \Omega'$ of two domains in $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$, is 1-quasiconformal if and only if it is conformal. Composing with complex conjugation, we conclude that every 1-quasiconformal map is either holomorphic or antiholomorphic. In particular:

THEOREM 20.33. $f : \mathbb{S}^2 \to \mathbb{S}^2$ *is* 1*-quasiconformal if and only if* f *is a Moebius transformation.*

20.4.2.E. *Quasiconformal and quasisymmetric maps.* So far, we have the implications

$$\text{quasisymmetric} \Rightarrow \text{weakly quasisymmetric} \Rightarrow \text{quasiconformal}$$

for maps between domains in $\mathbb{R}^n$. It turns out that these arrows can be reversed:

THEOREM 20.34 (See e.g. [**Hei01**].). *Every quasiconformal homeomorphism defined on the entire* $\mathbb{R}^n$ *is quasiconformal if and only if it is quasisymmetric.*

20.4.2.F. *Convergence property.* The convergence property of quasiconformal mappings is an analogue of the Arzela–Ascoli theorem for uniformly continuous families of maps between metric spaces. In fact, once Theorems 20.37 and 20.39 are established, one can derive the convergence property from the Coarse Arzela–Ascoli theorem (Proposition 5.32) applied to quasiisometries of the hyperbolic space $\mathbb{H}^{n+1}$. (The three-point normalization for quasiconformal mappings corresponds to a 1-point normalization of quasiisometries.)

Let $z_1, z_2, z_3 \in \mathbb{S}^n$ be three distinct points. A sequence of quasiconformal maps $f_i : \mathbb{S}^n \to \mathbb{S}^n$ is said to be *normalized at* $\{z_1, z_2, z_3\}$ if the limits

$$\lim_{i\to\infty} f_i(z_k), \quad k = 1, 2, 3,$$

exist and are pairwise distinct.

THEOREM 20.35 (See [**Väi71**]). *Let* $\Omega \subset \mathbb{S}^n, n \geqslant 2$, *be a connected open subset and* $f_i : \Omega \to f_i(\Omega) \subset \mathbb{S}^n$ *be a sequence of* K*-quasiconformal homeomorphisms normalized at three points in* Ω*. Then* (f_i) *contains a subsequence which converges to a* K*-quasiconformal map.*

The same theorem holds for $n = 1$, except one replaces *quasiconformal* with *quasimoebius*.

We note that the convergence property is usually stated with our normalization condition replaced by the assumption that the three values $w_{ik} = f_i(z_k), k = 1, 2, 3$, of f_i's are fixed. Recall, however, that Moebius transformations act transitively on three-point subsets of $\mathbb{S}^n$ (see Exercise 8.50). Therefore, there exists a sequence $\gamma_i \in Mob(\mathbb{S}^n)$ satisfying

$$\gamma_i(w_k) = z_k, \quad k = 1, 2, 3.$$

Since the three limits

$$\lim_{i\to\infty} w_{ik} = w'_k, \quad k = 1, 2, 3,$$

are all distinct, the sequence (γ_i) subconverges to a Moebius transformation (Lemma 8.51). Composing the quasiconformal mappings f_i with the conformal mappings γ_i, of course, does not change the quasiconformality constants. Therefore, the normalization used in Theorem 20.35 is equivalent to the standard normalization.

20.4.2.G. *Strong convergence property.* The following strengthening of the convergence property is the key analytical ingredient needed for the proof of Tukia's theorem 21.17 in the next chapter. Fix an open subset $\Omega \subset \mathbb{S}^n$; we will use the notation mes for the Lebesgue measure on $\mathbb{S}^n$ restricted to Ω. For a subset $E \subset \Omega$ we let E^c denote the complement $\Omega \setminus E$.

THEOREM 20.36 (Tukia's Strong Convergence Property, [**Tuk86**]; see also [**IM01**] for a stronger version). *Consider a sequence $f_i : \Omega \to \mathbb{S}^n$ of K-quasiconformal maps defined on an open subset $\Omega \subset \mathbb{S}^n$. Suppose that:*

1. The sequence (f_i) converges to a quasiconformal map f uniformly on compacts in Ω.

2. There exists a sequence of subsets $E_i \subset \Omega$, satisfying

$$\limsup_{i\to\infty} H(f_i|_{E_i^c}) = H,$$

while

$$\lim_{i\to\infty} mes(E_i) = 0.$$

Then $H(f) \leqslant H$. In particular, if $H = 1$, then f is conformal.

This is a nontrivial theorem since its hypothesis is merely C^0 (uniform convergence of mappings), while the conclusion is about infinitesimal quantities (dilatations of quasiconformal mappings).

20.5. Quasisymmetric maps and hyperbolic geometry

The last goal of this chapter is to relate quasisymmetric mappings and quasiisometries of hyperbolic spaces. We will identify the hyperbolic space $\mathbb{H}^{n+1}, n \geqslant 1$, with the upper half-space

$$\mathbb{R}^{n+1}_+\{(x_1, \ldots, x_{n+1}) : x_{n+1} > 0\}.$$

We will be also using the notation

$$e_{n+1} = (0, \ldots, 0, 1) \in \mathbb{H}^{n+1},$$

and δ for the hyperbolicity constant of $\mathbb{H}^{n+1}$.

Let $f : \mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$ be an (L, A)-quasiisometry and let $f_\infty : \mathbb{S}^n = \mathbb{R}^n \cup \{\infty\} \to \mathbb{S}^n$ be its homeomorphic extension to the boundary sphere of the hyperbolic space given by Theorem 9.108. To simplify the notation, we retain the name f for f_∞. After compositing f with an isometry of $\mathbb{H}^{n+1}$, we can assume that $f(\infty) = \infty$.

THEOREM 20.37. *There exists $C = C(L, A)$, such that the restriction $f : \mathbb{R}^n \to \mathbb{R}^n$ is η-quasisymmetric, with*

$$\eta(t) = e^{2C+A} t^L.$$

PROOF. Pick a point $x \in \mathbb{R}^n$ and consider an annulus $\mathbb{A} \subset \mathbb{R}^n$,

$$\mathbb{A} = \{v \in \mathbb{R}^n : R_1 \leqslant |v - x| \leqslant R_2\},$$

where $0 < R_1 \leqslant R_2 < \infty$. We will refer to the ratio $t = \frac{R_2}{R_1}$ as the *eccentricity* of $\mathbb{A}$. In other words, the eccentricity of $\mathbb{A}$ equals the ratio

$$\frac{|y - x|}{|z - x|}$$

for points y, z which belong to the outer and inner boundaries of $\mathbb{A}$ respectively. Consider the smallest annulus $\mathbb{A}'$ centered at $x' = f(x)$, which contains the topological annulus $f(\mathbb{A})$. Let t' denote the eccentricity of $\mathbb{A}'$. Then, by the definition of $\mathbb{A}'$ and t',

$$\frac{|f(y) - f(x)|}{|f(z) - f(x)|} \leqslant t',$$

for all points y and z as above. In order to verify that f is η-quasisymmetric, we need to show that $t' \leqslant \eta(t)$.

After precomposing and postcomposing f with translations of $\mathbb{R}^n$, we can assume that $x = x' = f(x) = 0$. Let $\alpha \subset \mathbb{H}^{n+1}$ denote the vertical geodesic, connecting 0 to ∞, i.e., α is the x_{n+1}-axis in $\mathbb{H}^{n+1}$. Let $\pi_\alpha : \mathbb{H}^{n+1} \to \alpha$ denote the orthogonal projection to α: For every $p \in \mathbb{H}^{n+1}$, $\pi_\alpha(p) = q \in \alpha$, where $q \in \alpha$ is the unique point such that the geodesic pq is orthogonal to α. The map π_α is the nearest-point projection to α. This projection extends continuously to a map

$$\pi_\alpha : \mathbb{H}^{n+1} \cup (\mathbb{R}^n \setminus \{0\}) \to \alpha.$$

Then, $\pi_\alpha(\mathbb{A})$ is the interval

$$\sigma = [R_1 e_{n+1}, R_2 e_{n+1}] \subset \alpha,$$

whose hyperbolic length equals $\ell = \log(R_2/R_1)$, see Exercise 8.14.

By Lemma 9.105, the (L, A)-quasigeodesic $f(\alpha)$ lies within distance $\theta(L, A, \delta)$ from the $\alpha \subset \mathbb{H}^{n+1}$, since we are assuming that $f(0) = 0, f(\infty) = \infty$. According to Proposition 9.107, quasiisometries "almost commute" with nearest-point projections and, thus, we obtain

$$d(f\pi_\alpha(x), \pi_\alpha f(x)) \leqslant C = C(L, A, \delta), \quad \forall x \in \mathbb{H}^{n+1} \cup (\mathbb{R}^n \setminus \{0\}).$$

FIGURE 20.1. Proving quasisymmetry of f_∞.

LEMMA 20.38.

$$\operatorname{diam}(\pi_\alpha(f(\mathbb{A}))) \leqslant 2C + L\ell + A.$$

PROOF. The ideal boundary of the spherical half-shell

$$\tilde{\mathbb{A}} := \pi_\alpha^{-1}(\sigma) \cap \mathbb{H}^{n+1}$$

is the annulus $\mathbb{A}$. Therefore, in view of continuity of

$$f : \mathbb{H}^{n+1} \cup \partial_\infty \mathbb{H}^{n+1} \to \mathbb{H}^{n+1} \cup \partial_\infty \mathbb{H}^{n+1}$$

at ideal boundary points, it suffices to verify the inequality in the lemma with $f(\mathbb{A})$ replaced by $f(\tilde{\mathbb{A}})$.

For any two points $p, q \in \tilde{\mathbb{A}}$ we have

$$d(f\pi_\alpha(p), \pi_\alpha f(p)) \leqslant C, \quad d(f\pi_\alpha(q), \pi_\alpha f(q)) \leqslant C.$$

Since $d(\pi_\alpha(p), \pi_\alpha(q)) \leqslant \ell$,

$$d(f\pi_\alpha(p), f\pi_\alpha(q)) \leqslant L\ell + A,$$

and, by the triangle inequality, we obtain

$$d(\pi_\alpha f(p), \pi_\alpha f(q)) \leqslant 2C + L\ell + A.$$

The lemma follows. □

Now we can finish the proof of the theorem. Lemma 20.38 implies that

$$f(\mathbb{A}) \subset \pi_\alpha^{-1}(\sigma'),$$

where $\sigma' \subset \alpha$ has length $\leqslant \ell' = 2C + L\ell + A$. The eccentricity of the annulus

$$\pi_\alpha^{-1}(\sigma') \cap \mathbb{S}^n$$

is at most $\leqslant e^{\ell'}$. Thus, the eccentricity of the annulus $\mathbb{A}'$ is also at most

$$e^{\ell'} = e^{2C+A} \cdot e^{L\ell} = e^{2C+A} \cdot e^{L\log(t)} = e^{2C+A} t^L,$$

where $t = R_2/R_1$. □

The following converse theorem was first proven by Tukia in the case of hyperbolic spaces and then extended by Paulin to the case of more general Gromov-hyperbolic spaces.

THEOREM 20.39 (P. Tukia [**Tuk94**], F. Paulin [**Pau96**]). *Every quasisymmetric homeomorphism $\mathbb{R}^n \to \mathbb{R}^n$ extends to a quasiisometric map $\mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$. More precisely: For every η-quasisymmetric homeomorphism $f : \mathbb{R}^n \to \mathbb{R}^n$ there exists a an (L, A)-quasiisometric map F of the hyperbolic space $\mathbb{H}^{n+1}$, such that*

$$F_\infty = f,$$

where F_∞ is the boundary extension of the quasiisometry F given by Theorem 9.108. Moreover, the constants L, A depend only on $\eta(1)$ and $\eta(2)$.

PROOF. We let $\Pi : \mathbb{H}^{n+1} = \mathbb{R}^{n+1}_+ \to \mathbb{R}^n$ denote the coordinate projection. We define the extension F as follows. For every $p \in \mathbb{H}^{n+1}$, let $\alpha = \alpha_p$ be the complete vertical geodesic through p. This geodesic is asymptotic to the points ∞ and $x = x_p = \Pi(p) \in \mathbb{R}^n$. Let $y \in \mathbb{R}^n$ be a point such that $\pi_\alpha(y) = p$. (The point y is non-unique, of course: Every point $y \in S(x, R)$ in the sphere centered at x and of the radius $R = |x - p|$ would work.) Let $x' := f(x), y' := f(y)$ and let $\alpha' \subset \mathbb{H}^{n+1}$ denote the vertical geodesic through x' and let $p' := \pi_{\alpha'}(y')$. Lastly, set $F(p) := p'$.

We will prove that F is an (L, A)-coarse Lipschitz map. The quasiinverse to F will be the map $\bar{F}$ defined *via* extension of the map f^{-1} following the same procedure. We will leave it as an exercise to the reader to verify that $\bar{F}$ is indeed a quasiinverse to F and to estimate $d(\bar{F} \circ F, \mathrm{id})$.

FIGURE 20.2. Triangle $\Delta(p_1, p_2, p_3)$.

Suppose that $d(p_1, p_2) \leqslant 1/4$. We are looking for a uniform upper bound on the distance

$$d(p_1', p_2') \leqslant A,$$

with A depending only on $\eta(1)$ and $\eta(2)$: Existence of such upper bound will imply that F is $(4A, A)$-coarse Lipschitz (see Lemma 5.8).

Consider the triangle $\Delta(p_1, p_2, p_3)$ as in Figure 20.2, where the points p_1, p_3 belong to the same vertical geodesic while p_2, p_3 belong to the same horosphere Σ centered at the point $\infty \in \mathbb{S}^n$ (after swapping p_1, p_2 if necessary we may assume that the point p_1 does not belong to the horoball bounded by Σ). Then

$$d(p_1, p_2) \leqslant \frac{1}{4} \Rightarrow d(p_1, p_3) \leqslant \frac{1}{4} \Rightarrow d(p_2, p_3) \leqslant \frac{1}{2}.$$

The uniform upper bounds

$$d(F(p_1), F(p_2)) \leqslant C_1, \quad d(F(p_3), F(p_2)) \leqslant C_3$$

would imply

$$d(F(p_1), F(p_2)) \leqslant A := C_1 + C_2,$$

as required. Therefore, our problem reduces to the two special cases:

Case 1. p_1, p_2 belong to the common vertical geodesic α; in particular, $x_1 = x_2 = x$. We will assume, for concreteness, that $|x - y_1| \leqslant |x - y_2|$. Hence,

$$d(p_1, p_2) = \log\left(\frac{|y_2 - x|}{|y_1 - x|}\right)$$

and the assumption $d(p_1, p_2) \leq 1/2$ implies that

$$\left(\frac{|y_2 - x|}{|y_1 - x|}\right) \leq \sqrt{e} < 2.$$

Since the map f is η-quasisymmetric,

$$\frac{|y_2' - x'|}{|y_1' - x'|} \leqslant \eta\left(\frac{|y_2 - x|}{|y_1 - x|}\right) \leqslant \eta(2).$$

In particular,

$$d(p_1', p_2') \leqslant C_1 := \log(\eta(2)).$$

This estimate shows that different choices of the point $y \in \mathbb{R}^n$ in the definition of F lead to maps which are within distance $\leq C_1$ from each other.

Case 2. Suppose that the points p_1, p_2 have the same last coordinate, i.e., they belong to the same horosphere centered at the point $\infty \in \mathbb{S}^n$. As before, for $i = 1, 2$, we set $x_i := \Pi(p_i)$, let α_i denote the vertical hyperbolic geodesic through p_i and pick arbitrarily $y_i \in \mathbb{R}^n$ with $\pi_{\alpha_i}(y_i) = p_i$. Then

$$R_i := |y_i - x_i| = |p_i - x_i|$$

and $R = R_1 = R_2$ (as p_1, p_2 have the same last coordinate). Set $R_3 := |x_1 - y_2|$.

The reader will verify, using the formula (8.15), that

$$d(p_1, p_2) \leqslant 1/2 \Rightarrow t := |x_1 - x_2| < R \Rightarrow R_3 \leq 2R.$$

In particular, if $n \geqslant 2$, we could have choosen, if we wish, $y_1 = y_2$.

The points $p'_i = F(x_i), i = 1, 2$, belong to the vertical geodesics α'_i, such that

$$\pi_{\alpha'_i}(y'_i) = p'_i,$$

$y'_i = f(y_i)$, $i = 1, 2$. We define the points $x'_i := \Pi(p'_i) \in \mathbb{R}^n$, and set

$$R'_i := |y'_i - x'_i| = |p'_i - x'_i|, i = 1, 2.$$

Lastly, set

$$t' := |x'_1 - x'_2|, \qquad R'_3 := |x'_1 - y'_2|.$$

After switching the roles of p_1 and p_2, we can assume that $R'_1 \leqslant R'_2$. Then

$$d(p'_1, p'_2) \leqslant \frac{t'}{R'_1} + \log\left(\frac{R'_2}{R'_1}\right), \tag{20.9}$$

as we can first travel from p'_1 to the line α'_2 horizontally, along a path of the hyperbolic length

$$\frac{t'}{R'_1} = \frac{|x'_2 - x'_1|}{R'_1},$$

and then vertically, along α'_2, along a path of the length $\log(R'_2/R'_1)$. We now apply the η-quasisymmetry condition (equation (20.2)) to the triple of points x_1, y_1, x_2, (with x_1 playing the role of the center) and get:

$$\frac{t'}{R'_1} \leqslant \eta\left(\frac{t}{R}\right) \leqslant \eta(1). \tag{20.10}$$

Setting $R_3 := |x_1 - y_2|, R'_3 := |x'_1 - y'_2|$ and applying the η-quasisymmetry condition to the triple of points x_1, y_1, y_2 (with x_1 again playing the role of the center), we obtain

$$\frac{R'_3}{R'_1} \leqslant \eta\left(\frac{R_3}{R_1}\right) \leqslant \eta\left(\frac{2R}{R}\right) = \eta(2). \tag{20.11}$$

The inequalities $R'_2 \leqslant t' + R'_3$, (20.10) and (20.11) imply

$$\frac{R'_2}{R'_1} \leqslant \frac{t' + R'_3}{R'_1} \leqslant \eta(1) + \eta(2).$$

Combining this inequality with (20.9), we conclude that

$$d(p'_1, p'_2) \leqslant C_2 := \eta(1) + \log(\eta(1) + \eta(2)).$$

Thus, in general, for $p_1, p_2 \in \mathbb{H}^{n+1}, d(p_1, p_2) \leqslant 1/4$, we have:

$$d(F(p_1), F(p_2)) \leqslant C_1 + C_2 = A.$$

It follows that F is an (A, A)-coarse Lipschitz map, where $A = C_1 + C_2$.

The last thing to observe is that since F, by the construction, sends vertical geodesics to vertical geodesics, its extension $F_\infty : \mathbb{S}^n \to \mathbb{S}^n$, defined by Theorem 9.108, equals f. □

Exercise 20.40. Consider a linear quasiconformal mapping $f : \mathbf{x} \mapsto A\mathbf{x}, \mathbf{x} \in \mathbb{R}^n, A \in GL(n, \mathbb{R})$. Define the linear mapping

$$\tilde{f} : (\mathbf{x}, t) \mapsto (A\mathbf{x}, t), \quad x \in \mathbb{R}^{n-1}, t > 0.$$

Show that $\tilde{f}$ is a quasiisometry of $\mathbb{H}^{n+1}$.

CHAPTER 21

Groups quasi-isometric to $\mathbb{H}^n$

The main result of this chapter is the following theorem, due to P. Tukia, see [**Tuk86**] and [**Tuk94**], as well as the paper by J. Cannon and D. Cooper [**CC92**]:

THEOREM 21.1 (P. Tukia). *If G is a finitely generated group QI to $\mathbb{H}^{n+1}$ (with $n \geqslant 2$), then G acts geometrically on $\mathbb{H}^{n+1}$. In particular, G is virtually isomorphic to a uniform lattice in the Lie group* $\mathrm{Isom}(\mathbb{H}^{n+1})$.

REMARK 21.2. The same result also holds for $n = 1$, but the proof in this case is completely different, see §21.7.

Recall that if a group G is QI to $\mathbb{H}^{n+1}$, then it quasiacts on $\mathbb{H}^{n+1}$, see Lemma 5.61. Furthermore (by Theorem 9.135), every such quasiaction φ determines a topological action

$$\varphi_\infty : G \curvearrowright \mathbb{S}^n$$

on the boundary sphere of $\mathbb{H}^{n+1}$. Since the quasiaction of G is by uniform quasi-isometries, the action of $G \curvearrowright \mathbb{S}^n$ is by uniformly quasiconformal homeomorphisms, see Theorem 20.37. Such group actions are called *uniformly quasiconformal*. According to Lemma 5.61, the quasiaction $G \curvearrowright \mathbb{H}^{n+1}$ is geometric and, by Lemma 9.118, every point $\xi \in \mathbb{S}^n$ is a conical limit point of $G \curvearrowright \mathbb{S}^n$. Lastly, by Theorem 9.135, the fact that the quasiaction $G \curvearrowright \mathbb{H}^{n+1}$ is geometric translates to:

The action $G \curvearrowright \mathrm{Trip}(\mathbb{S}^n)$ is properly discontinuous and cocompact. In particular:

1. The kernel of the homomorphism $\varphi_\infty : G \to Homeo(\mathbb{S}^n)$ is a finite normal subgroup of G.

2. The image $\bar{G} = \varphi_\infty(G)$ is a discrete subgroup of the group of homeomorphisms $Homeo(\mathbb{S}^n)$, where $Homeo(\mathbb{S}^n)$ is equipped with the topology of uniform convergence.

We refer the reader to the Notation 9.88 for the definition of the space $\mathrm{Trip}(\mathbb{S}^n)$.

Our goal, and this is the main result of Sullivan (for $n = 2$) and Tukia (for all $n \geq 2$), is to show that, under the above hypothesis, there exists a quasiconformal homeomorphism $f : \mathbb{S}^n \to \mathbb{S}^n$ which conjugates $\bar{G}$ to a group of Moebius transformations, whose action on $\mathbb{H}^{n+1}$ is geometric. Once the existence of such f is established, Theorem 21.1 would follow. We see that in order to prove Theorem 21.1, one is naturally lead to study *uniformly quasiconformal group actions on* $\mathbb{S}^n$. Our treatment of quasiconformal groups mostly follows the arguments in [**Tuk86**] and in [**IM01**]. A different, but related, proof is given by P. Haissinsky [**Haï09**].

21.1. Uniformly quasiconformal groups

Let $G < Homeo(\mathbb{S}^n)$ be a group of consisting of quasiconformal homeomorphisms. The group G is called *uniformly quasiconformal*, if there exists $K < \infty$

such that $K(g) \leqslant K$ for all $g \in G$. Recall that $K(g)$ is the quasiconformality constant of the homeomorphism $g : \mathbb{S}^n \to \mathbb{S}^n$, see Equation (20.8). Trivial examples of uniformly quasiconformal groups are given by subgroups $\Gamma < Mob(\mathbb{S}^n)$ of Moebius transformations and their quasiconformal conjugates

$$\Gamma^f = f\Gamma f^{-1},$$

where f is k-quasiconformal. Then for every $g \in \Gamma^f$,

$$K(g) = K(f\gamma f^{-1}) \leqslant k^2 = K.$$

We say that a uniformly quasiconformal subgroup $G < Homeo(\mathbb{S}^n)$ is *exotic* if it is not quasiconformally conjugate to a group of Moebius transformations. The following theorem is a fundamental fact of quasiconformal analysis in dimension $n = 2$, observed first by D. Sullivan in [**Sul81**]:

THEOREM 21.3. *There are no exotic uniformly quasiconformal subgroups in* $Homeo(\mathbb{S}^2)$.

This theorem fails rather badly for $n \geqslant 3$. The first examples of exotic uniformly quasiconformal subgroups $G < Homeo(\mathbb{S}^n), n \geqslant 3$, were constructed by P. Tukia [**Tuk81**]. Tukia's subgroups G are nondiscrete, isomorphic to certain connected solvable Lie groups, which do not admit embeddings into $\mathrm{Isom}(\mathbb{H}^m)$ for any m. Algebraically, Tukia's examples are semidirect products $\mathbb{R}^k \rtimes \mathbb{R}^2$, where $(a, b) \in \mathbb{R}^2$ acts on $\mathbb{R}^k$ *via* a diagonal matrix $D(a, b)$ that has (generically) two distinct eigenvalues $\neq \pm 1$. Further examples of*discrete* exotic uniformly quasiconformal subgroups of $Homeo(\mathbb{S}^3)$ were constructed in [**FS88**], [**Mar86**] (these groups have torsion) and in [**Kap92**] (these are certain surface groups acting on $\mathbb{S}^3$). An example of a discrete uniformly quasiconformal subgroup of $Homeo(\mathbb{S}^3)$ which is not *isomorphic* to subgroup of $\mathrm{Isom}(\mathbb{H}^4)$ was constructed in [**Isa90**].

PROBLEM 21.4. Suppose that $G < Homeo(\mathbb{S}^n)$ is a discrete uniformly quasiconformal subgroup. Is it true that G is isomorphic to a subgroup of $\mathrm{Isom}(\mathbb{H}^m)$ for some m?

The answer to this questions is probably negative. One can, nevertheless, ask which algebraic properties of discrete groups of Moebius transformations are shared by discrete uniformly quasiconformal subgroups, e.g.:

PROBLEM 21.5. Suppose that $G < Homeo(\mathbb{S}^n)$ is an infinite discrete uniformly quasiconformal subgroup, $n \geqslant 3$.

1. Is it true that G does not have the Property (T)?
2. Is it true that G has the Haagerup property?
3. Is it true that the action $G \curvearrowright \mathbb{S}^n$ extends to a uniformly quasiconformal action $G \curvearrowright \mathbb{H}^{n+1}$?

Note that, in view of Theorems 21.7 and 5.64, there exists a Gromov-hyperbolic space X quasiisometric to $\mathbb{H}^{n+1}$, such that G acts *isometrically* on X and the actions of G on $\partial_\infty X$ and $\mathbb{S}^n$ are topologically conjugate.

Another problem, open since Tukia's examples of exotic connected solvable uniformly quasiconformal subgroups of $Homeo(\mathbb{S}^n)$ is:

PROBLEM 21.6. Suppose that $N < Homeo(\mathbb{S}^n)$ is a uniformly quasiconformal connected nilpotent subgroup. Is it true that N is abelian?

21.2. Hyperbolic extension of uniformly quasiconformal groups

As we saw, every quasiaction $G \curvearrowright \mathbb{H}^{n+1}$ extends to a uniformly quasiconformal action $G \curvearrowright \mathbb{S}^n$. Our first goal is to prove the converse:

THEOREM 21.7 (P. Tukia, [**Tuk94**]). *Every uniformly quasiconformal action $\rho : G \curvearrowright \mathbb{S}^n$ extends to a quasiaction $\varphi : G \curvearrowright \mathbb{H}^{n+1}$ in the sense that*

$$\varphi(g)_\infty = \rho(g), \quad \forall g \in G,$$

where $h_\infty : \mathbb{S}^n \to \mathbb{S}^n$ is the extension of the quasiisometry $h : \mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$ given by Theorem 20.37.

PROOF. For every $g \in G$ we let $\varphi(g) : \mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$ denote the quasiisometric extension of $\rho(g)$ constructed in Theorem 20.39. In view of the same extension theorem, since every $\rho(g)$ is K-quasiconformal, every $\varphi(g)$ is an (L, A)-quasiisometry, where L and A depend only on K. We need to show that the extension φ defines a quasiaction, i.e., there exists $C = C(L, A)$ such that:

(1) For all $g_1, g_2 \in G$

$$\text{dist}(\varphi(g_1) \circ \varphi(g_2), \varphi(g_1 g_2)) \leqslant C,$$

(2)

$$\text{dist}(\varphi(1_G), id) \leqslant C.$$

It follows immediately from the construction of the quasiisometric extension in the proof of Theorem 20.39 that

$$\varphi(1_G) = id.$$

In order to verify (1), we note that for all $g_1, g_2 \in G$, the composition

$$f' = \varphi(g_1) \circ \varphi(g_2)$$

is an $(L^2, LA + A)$-quasiisometry, while

$$f'' = \varphi(g_1 g_2)$$

is an (L, A)-quasiisometry. Furthermore,

$$f'_\infty = \rho(g_1) \circ \rho(g_2) = \rho(g_1 g_2) = f''_\infty.$$

By homogeneity of $\mathbb{H}^{n+1}$, every point of the hyperbolic space is a centroid of an ideal triangle. Therefore, Lemma 9.112 implies that

$$\text{dist}(f', f'') \leqslant C(L, A) = D(L, A, 0, \delta),$$

where δ is the hyperbolicity constant of $\mathbb{H}^{n+1}$. □

This theorem shows that the study of uniformly quasiconformal groups is equivalent to the study of quasiactions on $\mathbb{H}^{n+1}$. In particular, we can define *conical limit points* for uniformly quasiconformal subgroups $G < Homeo(\mathbb{S}^n)$ as conical limit points of the extended quasiactions.

Our goal, thus, is to prove the following theorem which was first established by D. Sullivan [**Sul81**] for $n = 2$ (without restrictions on conical limit points) and then by P. Tukia in full generality:

THEOREM 21.8 (P. Tukia, [**Tuk86**]). *Suppose that $G < Homeo(\mathbb{S}^n)$ is a countable uniformly quasiconformal subgroup. Assume also that $n \geqslant 2$ and that almost every point of $\mathbb{S}^n$ is a conical limit point of G. Then G is quasiconformally conjugate to a subgroup of the Moebius group $Mob(\mathbb{S}^n)$.*

Before proving this theorem, we will need a few technical tools.

21.3. Least volume ellipsoids

Observe that a closed ellipsoid centered at 0 in $\mathbb{R}^n$ can be described as

$$E = E_A = \{x \in \mathbb{R}^n : \varphi_A(x) = x^T A x \leqslant 1\},$$

where A is some positive-definite symmetric $n \times n$ matrix. The volume of such an ellipsoid is given by the formula

$$Vol(E_A) = \omega_n \left(\det(A)\right)^{-1/2},$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$. A subset $X \subset \mathbb{R}^n$ is called *centrally-symmetric* if $X = -X$.

THEOREM 21.9 (F. John, [**Joh48**]). *For every compact centrally-symmetric subset $X \subset \mathbb{R}^n$ with nonempty interior, there exists a unique ellipsoid $E(X)$ of least volume containing X. The ellipsoid $E(X)$ is called the* John–Loewner *ellipsoid of X.*

PROOF. The existence of $E(X)$ is clear by compactness. We need to prove uniqueness. Consider the function f on the space P_n of positive definite symmetric $n \times n$ matrices, given by

$$f(A) = -\frac{1}{2} \log\left(\det(A)\right).$$

LEMMA 21.10. *The function $f : P_n \to \mathbb{R}$ is strictly convex, in the sense that for every family of matrices $C_t \in P_n$,*

$$C_t = tA + (1-t)B \in P_n, \quad 0 \leqslant t \leqslant 1,$$

the function $g(t) = f(C_t)$ is strictly convex.

PROOF. The matrices A and B in P_n can be simultaneously diagonalized by a matrix $M \in GL(n, \mathbb{R})$:

$$MAM^T = D_A, \quad MBM^T = D_B,$$

where D_A, D_B are diagonal matrices. The matrices

$$D_t = tD_A + (1-t)D_B = M(tA + (1-t)B)M^T$$

are, of course, also diagonal and

$$f(D_t) = f(MC_tM^T) = -\log\det(M) - \frac{1}{2}\log\det(C_t) = -\log\det(M) + f(C_t).$$

Therefore, it suffices to prove strict convexity of f on the space $Diag_n^+$ of positive-definite diagonal $n \times n$ matrices. For each diagonal matrix $D = Diag(x_1, ..., x_n)$ with the diagonal entries $x_1, \ldots, x_n$,

$$f(D) = -\frac{1}{2}\sum_{i=1}^{n} \log(x_i).$$

Lastly, the function $f : Diag_n^+ \to \mathbb{R}$ is strictly convex since log is strictly concave. □

In particular, whenever $V \subset P_n$ is a convex subset and $f\big|_V$ is proper, f attains a unique minimum on V. Since log is a strictly increasing function, the same uniqueness assertion holds for the function $\det^{-1/2}$ on P_n. Let $V = V_X$ denote the set of matrices $C \in P_n$ such that $X \subset E_C$. Since $\varphi_A(x)$ is linear as a function of

A for any fixed $x \in X$, it follows that V convex. Thus, the least volume ellipsoid containing X is unique. $\square$

21.4. Invariant measurable conformal structure

Throughout this section, we assume that n is at least 2 (the discussion becomes meaningless otherwise). Recall (see §2.3) that a measurable Riemannian metric on $\mathbb{S}^n = \mathbb{R}^n \cup \{\infty\}$ is a measurable map g from $\mathbb{S}^n$ to the space P_n of positive definite symmetric $n \times n$ matrices. (Since we are working in the measurable category, we can and will ignore the point ∞.)

A *measurable conformal structure* on $\mathbb{S}^n$ is a measurable Riemannian metric defined up to multiplication by a positive measurable function. In order to avoid the ambiguity with the choice of the conformal factor, one can normalize the measurable metric g so that $\det(g(x)) = 1$ for all $x \in \mathbb{S}^n$. We will refer to such g as a *normalized* measurable Riemannian metric.

Every quasiconformal mapping $f : \mathbb{S}^n \to \mathbb{S}^n$ acts on measurable Riemannian metrics *via* the pull-back by the usual formula:

$$f^*(g) = h, \quad h(x) = (D_x f)\ g(f(x))\ (D_x f)^T.$$

Here we are using the fact that f is differentiable almost everywhere in $\mathbb{S}^n$ and its derivative is a measurable matrix-valued function on $\mathbb{R}^n$, see Theorem 20.23. Measurability of the function $x \mapsto D_x f$ explains why considering measurable Riemannian metrics is the right thing to do in the context of quasiconformal mappings.

Remark 21.11. The reader might have noticed that in the book we proved the ACL property only for quasisymmetric rather than quasiconformal mappings. For the purposes of quasiisometric rigidity this does not matter, since extensions of quasiisometries are quasimoebius mappings and, hence, we can use the analytical properties of quasisymmetric mappings proven in Chapter 20. Furthermore, every quasiconformal mapping of $\mathbb{R}^n$ is also quasisymmetric, §20.4.2.E.

If we consider normalized Riemannian metrics, then the appropriate action is given by the formula:

$$f^\bullet(g) = h, \quad h(x) = (J_x)^{-2n} D_x f g(f(x))(D_x f)^T$$

in order for h to be normalized as well. Here J_x is the Jacobian determinant of f at x. We will think of normalized measurable Riemannian metrics as measurable conformal structures.

A measurable conformal structure μ on $\mathbb{S}^n$ is called *bounded* if it is represented by a bounded normalized measurable Riemannian metric, i.e., a bounded map

$$\mathbb{S}^n \to P_n \cap \{det = 1\}.$$

Below, we interpret boundedness of μ in terms of eigenvalues.

Given a measurable Riemannian metric $\mu(x) = A_x$, we define its *linear dilatation* $H(\mu)$ as the essential supremum of the ratios

$$H(x) := \frac{\sqrt{\lambda_n(x)}}{\sqrt{\lambda_1(x)}},$$

where $\lambda_1(x) \leqslant \ldots \leqslant \lambda_n(x)$ are the eigenvalues of A_x. Geometrically speaking, if $E_x \subset T_x\mathbb{R}^n$ is the unit ball with respect to A_x, then $H(x)$ is the eccentricity of the ellipsoid E_x, i.e., the ratio of the largest to the smallest axis of E_x. In particular, $H(x)$ and $H(\mu)$ are conformal invariants of μ.

EXERCISE 21.12. 1. A measurable conformal structure μ is bounded if and only if $H(\mu) < \infty$.

2. A subset $\mathcal{M}$ in the space normalized measurable conformal structures is bounded if and only if

$$\sup_{\mu \in \mathcal{M}} H(\mu) < \infty.$$

We say that a *measurable conformal structure* $\mu(x) = A_x$ on $\mathbb{R}^n$ is *invariant under a quasiconformal subgroup* $G < Homeo(\mathbb{S}^n)$ if

$$g^{\bullet}\mu = \mu, \forall g \in G.$$

In detail:

$$\forall g \in G, \quad (J_{g,x})^{-\frac{1}{2n}} (D_x g)^T \cdot A_{gx} \cdot D_x g = A_x$$

a.e. in $\mathbb{R}^n$.

The following was first proven by Sullivan in [**Sul81**] for $n = 2$ and, then, by Tukia [**Tuk86**] for arbitrary n:

PROPOSITION 21.13. *Every countable uniformly quasiconformal subgroup $G <$ $Homeo(\mathbb{S}^n)$ admits an invariant measurable conformal structure λ on $\mathbb{S}^n$.*

PROOF. Let μ_0 be the Euclidean metric on $\mathbb{R}^n$, it is given by the constant matrix function $x \mapsto I$. Consider the orbit $G \cdot \mu_0$ in the space of normalized measurable Riemannian metrics. The ideal is to take the "average" of all the measurable conformal structures in this orbit.

Since G is countable, there exists a subset of full measure in $\mathbb{S}^n$ on which we have matrix-valued functions

$$A_{g,x} = g^{\bullet}\mu_0 = (J_{g,x})^{-\frac{1}{2n}} (D_x g)^T \cdot D_x g, \quad g \in G.$$

With this definition, $H(A_{g,x}) = H_g(x)$, is the linear dilatation of g at x, see Definition 20.11. Therefore, the assumption that G is uniformly quasiconformal is equivalent to the assumption that the family of measurable conformal structures $G \cdot \mu_0$ is uniformly bounded:

$$H := \sup_{g \in G} H(g^{\bullet}\mu_0) < \infty.$$

Geometrically, one can think of this as follows. For a.e. x we let $E_{g,x}$ denote the unit ball in $T_x\mathbb{R}^n$ with respect to $g^{\bullet}(\mu_0)$. From the Euclidean viewpoint, $E_{g,x}$ is just an ellipsoid of the volume ω_n (since $g^{\bullet}(\mu_0)$ is normalized). This ellipsoid (up to scaling) is the image of the unit ball under the inverse of the derivative $D_x g$. Then uniform boundedness of the conformal structures $g^{\bullet}(\mu_0)$ simply means the that the eccentricities of the ellipsoids $E_{g,x}$ are bounded by the number H, which is independent of g and x. Since the volume of each $E_{g,x}$ is fixed, it follows that the diameters of these ellipsoids are uniformly bounded above and below: There exists $0 < R < \infty$ such that

$$B(0, R^{-1}) \subset E_{g,x} \subset B(0, R), \forall g \in G$$

for a.e. $x \in \mathbb{R}^n$.

Let U_x denote the union of the ellipsoids

$$\bigcup_{g \in G} E_{g,x}.$$

Since each ellipsoid $E_{g,x}$ is centrally-symmetric, so is U_x. By the construction, the family of sets $\{U_x, x \in \mathbb{R}^n\}$ is invariant under the group G:

$$(J_{g,x})^{-1/n} D_x g(U_x) = U_{g(x)}, \quad \forall g \in G.$$

For each x (where U_x is defined, which is a subset of full measure), we define an ellipsoid E_x, the John-Loewner ellipsoid of the set U_x. Since the group G preserves the family of sets U_x and since, after normalization, the action of $D_x g$ on the tangent space is volume-preserving, it follows (by uniqueness of the John-Loewner ellipsoid, Theorem 21.9) that G also preserves the family of ellipsoids E_x.

Clearly,

$$B(0, R^{-1}) \subset E_x \subset B(0, R)$$

a.e. in $\mathbb{R}^n$, and, hence, the eccentricities of the ellipsoids E_x are uniformly bounded above and below. Let $\mu(x)$ denote the (a.e. defined) function $\mathbb{R}^n \to P_n$ which sends x to the matrix A_x such that E_x is the unit ball with respect to the quadratic form defined by A_x. Then, $H(A_x) \leqslant R^2$ a.e..

Lemma 21.14. *The function $\mu : x \to A_x$ is measurable.*

Proof. Since G is countable, we can represent G as an increasing union of finite subsets $G_i \subset G$. For each i we define the sets

$$U_{x,i} = \bigcup_{g \in G_i} E_{g,x}$$

and the corresponding John-Loewner ellipsoids $E_{x,i}$. We leave it to the reader to check that since each ellipsoid $E_{g,x}$ is measurable as a function of y, then $E_{x,i}$ is also measurable. Note also that

$$E_x = \bigcup_{i \in \mathbb{N}} E_{x,i}.$$

Let $\mu_i : \mathbb{R}^n \to P_n$ denote the measurable functions defined by the ellipsoids $E_{x,i}$. We will think of these functions as functions $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$,

$$(x, v) \mapsto v^T \mu_i(x) v \in \mathbb{R}_+.$$

Then the fact that $E_i \subset E_{i+1}$ means that

$$\mu_i(x, v) \geqslant \mu_{i+1}(x, v).$$

Furthermore,

$$\mu = \lim_i \mu_i.$$

Now, the lemma follows from the Lebesgue monotone convergence theorem (Beppo Levi's theorem), see e.g. [**SS05**]. □

This also concludes the proof of the proposition. □

The above proposition also holds for uncountable uniformly quasiconformal groups, see [**Tuk86**], but we will not need this fact.

21.5. Quasiconformality in dimension 2

In this section we reformulate quasiconformality of a map in the 2-dimensional case in terms of the *Beltrami equation* and explain the relation between measurable conformal structures on domains in $\mathbb{C} = \mathbb{R}^2$ and *Beltrami differentials*. We refer to **[Ahl06]** and **[Leh87]** for further details.

21.5.1. Beltrami equation. For computational purposes, we will use the complex differentials $dz = dx + idy$ and $d\bar{z} = dx - idy$. These differentials define coordinates on the complexification of the real tangent space $T_z\Omega$ of open subsets $\Omega \subset \mathbb{R}^2$. Accordingly,

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right),$$
$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

To simplify the notation, we let ∂f denote $\frac{\partial f}{\partial z} = f_z$ and let $\bar{\partial} f$ denote $\frac{\partial f}{\partial \bar{z}} = f_{\bar{z}}$, the *holomorphic* and *antiholomorphic* derivatives respectively.

Consider a function $f(z)$ which is differentiable at a point $z \in \mathbb{C}$. Writing $f = u + iv$, we obtain a formula for the (real) Jacobian determinant of f:

$$J_f = u_x v_y - u_y v_x = |\partial f|^2 - |\bar{\partial} f|^2.$$

We will assume from now on that f is orientation-preserving at z, i.e., $|\partial f(z)| > |\bar{\partial} f(z)|$.

For $\alpha \in [0, 2\pi]$, the directional derivative of f at z in the direction $e^{i\alpha}$ equals

$$\partial_\alpha f = \partial f + e^{-2i\alpha}\bar{\partial} f.$$

We now can compute lengths of the major and minor semi-axes of the ellipse, which is the image of the unit tangent circle under $D_z f$:

$$\max_\alpha |\partial_\alpha f| = |\partial f| + |\bar{\partial} f|,$$
$$\min_\alpha |\partial_\alpha f| = |\partial f| - |\bar{\partial} f|.$$

Thus,

$$H_z(f) = \max_{\alpha,\beta} \frac{|\partial_\alpha f|}{|\partial_\beta f|} = \frac{|\partial f| + |\bar{\partial} f|}{|\partial f| - |\bar{\partial} f|}$$

is the linear dilatation of f at z. Setting $\mu(z) = \frac{\bar{\partial} f}{\partial f}$, we obtain

$$H_z(f) = \frac{1 + |\mu(z)|}{1 - |\mu(z)|}.$$

Suppose now that $f : \Omega \to \mathbb{C}$ and $f \in W^{1,2}_{loc}(\Omega)$; in particular, f is differentiable a.e. in Ω, its derivatives are locally square-integrable in Ω and $J_z(f) > 0$ in Ω, i.e., f is orientation-preserving. Then, we have a measurable function

$$\mu = \mu(z) = \frac{f_{\bar{z}}}{f_z}, \tag{21.1}$$

called the *Beltrami differential* of f; the equation (21.1) is called the *Beltrami equation.* Let $k = k_f = \|\mu\|$ be the L^∞-norm of μ in Ω. Then,

$$K(f) = \sup_{z \in \Omega} H_z(f) = \frac{1+k}{1-k}$$

is the coefficient of quasiconformality of f.

We conclude that the following are equivalent for a function f:

1. f is K-quasiconformal, where $K = \frac{1+k}{1-k}$.
2. f satisfies the Beltrami equation and $k = \|\mu\| < 1$.

In particular, an (orientation-preserving) quasiconformal map is 1-quasiconformal if and only if $k_f = 0$, i.e., $\mu = 0$, equivalently, $\bar{\partial} f = 0$ (almost everywhere). A theorem of Weyl (see e.g. [**Ahl06**]) then states that such maps are holomorphic.

21.5.2. Measurable Riemannian metrics. Let $f : \Omega \to \Omega' \subset \mathbb{C}$ be an orientation-preserving quasiconformal homeomorphism, $w = f(z)$, with the Beltrami differential μ. For $w = u + iv$ it is useful to compute the pull-back of the Euclidean metric $du^2 + dv^2 = |dw|^2$ by the map f:

$$|dw|^2 = |\partial f dz + \bar{\partial} f d\bar{z}|^2 =$$

$$|\partial f|^2 \cdot |dz + \frac{\bar{\partial} f}{\partial f} d\bar{z}|^2 = |\partial f|^2 \cdot |dz + \mu(z) d\bar{z}|^2.$$

Therefore, up to the conformal multiple $|\partial f|^2$, the pull-back metric $f^*(|dw|^2)$ equals the measurable Riemannian metric

$$ds^2_\mu := |dz + \mu(z) d\bar{z}|^2.$$

Our next goal is to show that an arbitrary measurable Riemannian metric ds^2 on a domain (an open connected subset) $\Omega \subset \mathbb{C}$ is conformal to a metric of the form ds^2_μ for some μ. Consider a measurable Riemannian metric

$$ds^2 = Edx^2 + 2Fdxdy + Gdy^2.$$

We will do the computation in the tangent space at each point $z \in \Omega$. Then, by a change of variables $z = e^{i\theta} w$, we convert a general form ds^2 to the one with $F(z) = 0$; the same change of variables converts $|dz + \mu(z) d\bar{z}|^2$ to $|dw + \mu(z) e^{-2i\theta} d\bar{w}|^2$. Therefore, below we assume that $F = 0$. The condition that ds^2_μ is a multiple of ds^2 translates to

$$1 + \mu = t\sqrt{E}, \quad 1 - \mu = t\sqrt{G},$$

for some $t = t(z) \in (0, \infty)$. Solving this system of equations, we obtain that $\mu(z)$ is real,

$$\mu = \frac{\sqrt{E} - \sqrt{G}}{\sqrt{E} + \sqrt{G}} \ .$$

Clearly, $|\mu| < 1$. Furthermore, $\lim_{z \to z_0} |\mu(z)| = 1$ if and only if

$$\lim_{z \to z_0} \frac{E(z)}{G(z)} \in \{0, \infty\}.$$

Thus, the condition that the measurable conformal structure $[ds^2]$ defined by ds^2_μ is bounded is equivalent to the inequality

$$\|\mu\|_\infty < 1.$$

To summarize these computations: The correspondence $\mu \mapsto ds^2_\mu$, establishes an equivalence of Beltrami differentials μ with norm < 1 and bounded measurable conformal structures. Furthermore, if f is a quasiconformal map solving the *Beltrami equation* (21.1), then the measurable Riemannian metric $f^*(|dz|^2)$ is conformal to the metric ds^2_μ.

The following fundamental theorem goes one step further; it will be used for the proof of nonexistence of exotic uniformly quasiconformal groups acting on $\mathbb{S}^2$.

THEOREM 21.15 (Measurable Riemann Mapping Theorem). *For every measurable function $\mu(z)$ on a domain $\Omega \subset \mathbb{S}^2$ satisfying $\|\mu\|_\infty < 1$, there exists a quasiconformal homeomorphism $f : \Omega \to \Omega' \subset \mathbb{S}^2$ with the Beltrami differential μ. Equivalently, every bounded measurable conformal structure $[ds^2]$ on Ω is equivalent to the standard conformal structure on a domain $\Omega' \subset \mathbb{S}^2$ via a quasiconformal map $f : \Omega' \to \Omega$.*

HISTORICAL REMARK 21.16. In the case of smooth Riemannian metric ds^2, a local version of this theorem was proven by Gauss, it is called *Gauss' theorem on isothermal coordinates*. In full generality it was established by Morrey [**Mor38**]. Modern proofs can be found, for instance, in [**Ahl06**] and [**Leh87**].

21.6. Proof of Tukia's theorem on uniformly quasiconformal groups

We are now ready to prove Tukia's theorem. Recall that the notion of *approximate continuity* was defined in §20.4.1.B.

THEOREM 21.17 (P. Tukia, [**Tuk86**]). *Let $G < Homeo(\mathbb{S}^n)$ be a uniformly quasiconformal group and $n \geqslant 2$. Assume also that μ is a G-invariant bounded measurable conformal structure on $\mathbb{S}^n$, which is approximately continuous at a conical limit point ξ of G. Then there exists a quasiconformal homeomorphism $f : \mathbb{S}^n \to \mathbb{S}^n$ which sends μ to the standard conformal structure on $\mathbb{S}^n$ and conjugates G to a group of Moebius transformations.*

PROOF. As before, we will identify $\mathbb{S}^n$ with $\widehat{\mathbb{R}}^n = \mathbb{R}^n \cup \infty$. We first explain Sullivan's proof of this theorem in the case $n = 2$ since it is easier and does not use the conical limit points assumption.

In view of the Measurable Riemann Mapping Theorem for $\mathbb{S}^2$, the bounded measurable conformal structure μ on $\mathbb{S}^2$ is equivalent to the standard conformal structure μ_0 on $\mathbb{S}^2$, i.e., there exists a quasiconformal map $f : \mathbb{S}^2 \to \mathbb{S}^2$ which sends μ to μ_0:

$$f^\bullet \mu_0 = \mu.$$

Since the quasiconformal group G preserves the conformal structure μ on $\mathbb{S}^2$, it follows that the conjugate group $G^f = fGf^{-1}$ preserves the conformal structure μ_0. Therefore, each $h \in G^f$ is a 1-quasiconformal homeomorphism of $\mathbb{S}^2$, hence, a Moebius transformation, see §20.4. Thus, G^f acts as a group of Moebius automorphisms of the round sphere. This proves theorem for $n = 2$.

We now consider the general case. Without loss of generality, we may assume that the conical limit point ξ in the statement of the theorem is the origin in $\mathbb{R}^n$ and (by conjugating G *via* an affine transformation if necessary) that $\mu(0) = \mu_0(0)$ is the standard conformal structure on $\mathbb{R}^n$. We will identify $\mathbb{H}^{n+1}$ with the upper half-space $\mathbb{R}^{n+1}_+$. Define the point

$$e = e_{n+1} = (0, ..., 0, 1) \in \mathbb{H}^{n+1}.$$

Let $\varphi : G \curvearrowright \mathbb{H}^{n+1}$ be the quasiaction, extending the action $G \curvearrowright \mathbb{S}^n$, see Theorem 20.39. (In the context of the proof of Theorem 21.1, which is our main goal, we can use the original quasiaction, of course.) Let (L, A) be the quasiisometry constants for this quasiaction. Every element $g \in G$ is a K-quasiconformal homeomorphism of $\mathbb{S}^n$ for some $K < \infty$.

By the definition of a conical limit point, there exists a sequence $g_i \in G$ and a number $c \in \mathbb{R}$, such that

$$\lim_{i\to\infty} \varphi(g_i)(e) = 0$$

and

$$d(\varphi(g_i)(e), t_i e) \leqslant c, \tag{21.2}$$

where d is the hyperbolic metric on $\mathbb{H}^{n+1}$ and (t_i) is a sequence of positive numbers converging to zero.

Let γ_i denote the hyperbolic isometry given by

$$\mathbf{x} \mapsto t_i \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^{n+1}.$$

The composition

$$f_i := g_i^{-1} \circ \gamma_i$$

is a K-quasiconformal homeomorphism of $\mathbb{S}^n$. This homeomorphism has the (L, A)-quasiisometric extension

$$\varphi(g_i^{-1}) \circ \gamma_i$$

to the hyperbolic space $\mathbb{H}^{n+1}$. Using the inequality (21.2), we obtain the estimate

$$d(\varphi(g_i^{-1})\gamma_i(e), e) \leqslant Lc + A.$$

We claim that the sequence (f_i) contains a subsequence which converges to a K-quasiconformal mapping. One way to prove it is to appeal to the Coarse Arzela-Ascoli theorem (Proposition 5.32). We will use the Convergence Property of quasiconformal mappings instead.

Let T be an ideal hyperbolic triangle with the centroid e and the set of ideal vertices $Z = \{\zeta_1, \zeta_2, \zeta_3\}$. By the Extended Morse Lemma (Lemma 9.105), the quasigeodesic triangles $\phi(g_i^{-1})\gamma_i(T)$ are uniformly close to ideal geodesic triangles T_i in $\mathbb{H}^{n+1}$, such that the distances from centroids of T_i's to the point e are uniformly bounded (cf. the proof of Proposition 9.107). After passing to a subsequence (which we suppress) vertex sets Z_i of ideal triangles T_i converge to a three-point set $Z' \subset \mathbb{S}^n$. In particular, the K-quasiconformal maps f_i restricted to the set Z, subconverge to a bijection

$$Z \to Z' \subset \mathbb{S}^n.$$

Therefore, by Theorem 20.35, the sequence (f_i) subconverges to a quasiconformal mapping $f : \mathbb{S}^n \to \mathbb{S}^n$.

We now record the transformations of measurable conformal structures:

$$\mu_i := f_i^\bullet(\mu) = (\gamma_i)^\bullet (g_i)^{-1\bullet}(\mu) = (\gamma_i)^\bullet \mu,$$

since $g^\bullet(\mu) = \mu$. Putting it all together:

$$\mu_i(x) = \mu(\gamma_i x) = \mu(t_i x).$$

In other words, the measurable conformal structure μ_i is obtained by "zooming into" the point 0. Since, by the hypothesis of Theorem 21.17, $\xi = 0$ is an approximate continuity point for μ, the sequence of functions (μ_i) converges (in measure) to the constant function $\mu_0 = \mu(0)$. This leads to the diagram:

$$\begin{array}{ccc} \mu & \xrightarrow{f_i} & \mu_i \\ & & \downarrow \\ \mu & \xrightarrow{f} & \mu_0 \end{array}$$

If we knew that the derivatives Df_i subconverge (in measure) to the derivative Df, then we would conclude that

$$f^{\bullet}\mu = \mu_0.$$

Then f would conjugate the group G (preserving μ) to a group G^f preserving μ_0 and, hence, acting conformally on $\mathbb{S}^n$.

However, derivatives of quasiconformal maps (in general), converge only in the "biting" sense (see [**IM01**]), which does not suffice for our purposes. Thus, we have to use a less direct argument below.

We claim that every element of G^f is 1-quasiconformal. Since it suffices to verify 1-quasiconformality locally, we restrict to a certain round ball $B = B(0, R)$ in $\mathbb{R}^n$. Since μ is approximately continuous at 0, for every $\epsilon \in (0, \frac{1}{2})$,

$$\|\mu_i(x) - \mu(0)\| < \epsilon,$$

away from a subset $W_i \subset B$ of measure $< \epsilon_i$, where $\lim_{i\to\infty} \epsilon_i = 0$. Thus, for $x \in B \setminus W_i$,

$$1 - \epsilon < \lambda_1(x) \leqslant ... \leqslant \lambda_n(x) < 1 + \epsilon,$$

where $\lambda_k(x)$'s are the eigenvalues of the matrix $A_{i,x}$ of the normalized metric $\mu_i(x)$. It follows that

$$H_x(\mu_i) < \frac{\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \leqslant \sqrt{1+4\epsilon} \leqslant 1 + 2\epsilon,$$

away from the subsets W_i. For every $g \in G$, each map $h_i := f_i g f_i^{-1}$ is conformal with respect to the structure μ_i and, hence, $(1+2\epsilon)$-quasiconformal away from the set W_i. Since measures of the subsets W_i converge to zero, we conclude, by the Strong Convergence Property (Theorem 20.36), that each $h := \lim h_i$ is $(1+2\epsilon)$-quasiconformal. As this holds for arbitrary $\epsilon > 0$ and arbitrary $R > 0$, we conclude that each h is 1-quasiconformal (with respect to the standard conformal structure on $\mathbb{S}^n$). By the Liouville's Theorem for quasiconformal mappings (Theorem 20.32), it follows that h is Moebius.

This proves that the group $G^f = fGf^{-1}$ consists of Moebius transformations and concludes the proof of Theorem 21.17. □

REMARK 21.18. The key idea of the above proof is the *zooming* argument : The fraction appearing in the definition of the derivative of a function f of several real variables is nothing but a pre- and post-composition of f with some Moebius transformations. This argument will be used again in the proofs of Mostow and Schwarz Rigidity Theorems (sections 22.3) and 22.4).

Proof of Theorem 21.8. According to Proposition 21.13, there exists a G-invariant measurable conformal structure μ on the sphere $\mathbb{S}^n$. By Lemma 20.15, almost every point of $\mathbb{S}^n$ is a point of approximate continuity of μ. Therefore, Theorem 21.17 applies and the action of G is conjugate to a Moebius action. □

Historical Remark 21.19. Theorem 21.17 was first stated by Gromov in [**Gro81b**] in the same volume where Sullivan proved it for $n = 2$, [**Sul81**]. Gromov's sketch of the proof includes the zooming argument; this seems to be the first time when this argument appeared in the literature. However, Gromov did not have Theorem 20.36, which is the key analytical ingredient in the proof.

Proof of QI rigidity of groups acting geometrically on $\mathbb{H}^{n+1}$. We now can conclude the proof of Theorem 21.1. Let G be a finitely generated group quasi-isometric to $\mathbb{H}^{n+1}, n \geqslant 2$. Then there exists a quasi-action $\varphi : G \curvearrowright \mathbb{H}^{n+1}$ and this quasi-action extends to a uniformly quasiconformal action $\varphi_\infty : G \curvearrowright \mathbb{S}^n$. By Lemma 9.118, every point of $\mathbb{S}^n$ is a conical limit point for this action. Since the quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ is geometric, the action $G \curvearrowright \mathbb{S}^n$ is a *uniform convergence action*, see Theorem 9.135. Note that the action $G \curvearrowright \mathbb{S}^n$ is not necessarily faithful, but, by the same theorem, it has to have finite kernel. We will ignore the kernel and identity G with its image in the group of homeomorphisms of $\mathbb{S}^n$. By Proposition 21.13, there exists a G-invariant bounded measurable conformal structure μ on $\mathbb{S}^n$. By Lemma 20.15, almost every point of $\mathbb{S}^n$ is a point of approximate continuity of μ. Lastly, by Theorem 21.8, the action $G \curvearrowright \mathbb{S}^n$ is quasiconformally conjugate to a Moebius action $G^f \curvearrowright \mathbb{S}^n$.

Being a uniform convergence group is a purely topological concept invariant under homeomorphic conjugation. Thus, the group G^f also acts on $\mathbb{S}^n$ as a uniform convergence group. Recall that the Moebius group $Mob(\mathbb{S}^n)$ is isomorphic to the isometry group $\mathrm{Isom}(\mathbb{H}^{n+1})$ *via* the extension map from hyperbolic space to the boundary sphere, see Corollary 8.21. Therefore, by applying Theorem 9.132, we conclude that the isometric action $G^f \curvearrowright \mathbb{H}^{n+1}$ is again geometric. It follows that the group G admits a geometric action on $\mathbb{H}^{n+1}$, which finishes the proof of Theorem 21.1. □

21.7. QI rigidity for surface groups

Note that the proof of Tukia's theorem given above fails in the case $n = 1$, i.e., for groups G quasi-isometric to the hyperbolic plane. However, Theorem 9.135 still implies that G acts on $\mathbb{S}^1$ as a uniform convergence group. It was proven as a result of the combined efforts of Tukia, Gabai, Casson and Jungreis in 1988—1994 (see [**Tuk88, Gab92, CJ94**]) that every uniform convergence group acting on $\mathbb{S}^1$ is isomorphic to a Fuchsian group, i.e., a discrete cocompact subgroup of $\mathrm{Isom}(\mathbb{H}^2)$. Below we outline an alternative argument, which relies, however, on Thurston's Geometrization Conjecture for 3-dimensional manifolds/Perelman's Theorem (see [**KL08**] and [**BBB$^+$10**] for the detailed proofs). For the required background (related to the statement of Thurston's Geometrization Conjecture/Perelman's Theorem) we refer the reader to [**Sco83**], [**Kap01**] and [**Thu97**].

Theorem 21.20. *If a group G is QI to the hyperbolic plane then G admits a geometric action on $\mathbb{H}^2$.*

Proof. Let $\tilde{M} \subset \mathrm{Trip}(\mathbb{S}^1)$ denote the set of positively oriented ordered triples of distinct points on $\mathbb{S}^1$, i.e., points ξ_1, ξ_2, ξ_3 in $\mathbb{S}^1$ which appear in the counter-clockwise order on the circle. Thus, $\tilde{M}$ is a connected 3-dimensional manifold, an open subset of $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$. The group G acts as a uniform convergence group on $\mathbb{S}^1$; i.e., it acts properly discontinuously and cocompactly on $\mathrm{Trip}(\mathbb{S}^1)$; in particular, the restricted action $G \curvearrowright \tilde{M}$ is also properly discontinuous and cocompact. (See Theorem 9.135.)

Lemma 21.21. *If $g \in G$ fixes a point in $\tilde{M}$ then it fixes the entire $\tilde{M}$.*

Proof. Assume that $g \in G$ fixes three distinct points ξ_1, ξ_2, ξ_3 in $\mathbb{S}^1$. In particular, g preserves each component of $\mathbb{S}^1 \setminus \{\xi_1, \xi_2, \xi_3\}$. These components are arcs $\alpha_i, i = 1, 2, 3$. Since g fixes points ξ_i, it also preserves orientation on each α_i. Proper discontinuity of the action $G \curvearrowright \tilde{M}$ implies that the element g has finite order. We claim that g fixes each arc α_i pointwise. We identify each α_i with $\mathbb{R}$; then $g : \mathbb{R} \to \mathbb{R}$ is an orientation-preserving homeomorphism of finite order. Pick a point $x \in \mathbb{R}$ not fixed by g and suppose that $y = g(x) > x$. Then, since g preserves orientation, $g(y) > y$; similarly, $g^i(x) > g^{i-1}(x)$ for every $i \in \mathbb{Z}$. Thus, g cannot have finite order. Contradiction. The same argument applies if $y < x$. $\square$

Let, therefore, $\bar{G}$ denote the quotient of G by the (finite) kernel of the action $G \curvearrowright \mathbb{S}^1$. According to Lemma 21.21, the group $\bar{G}$ acts freely on $\tilde{M}$.

Lemma 21.22. *$\tilde{M}$ is homeomorphic to $\mathbb{H}^2 \times \mathbb{S}^1$.*

Proof. Given a triple $\xi = (\xi_1, \xi_2, \xi_3) \in \tilde{M}$ of distinct points in $\mathbb{S}^1$; we let T_ξ denote the a unique ideal hyperbolic triangle with ideal vertices $\xi_i, i = 1, 2, 3$. Let p_ξ denote the center of this triangle, i.e. the center of the inscribed circle.

Clearly, the map $\xi \to p_\xi$ is continuous as a map $\tilde{M} \to \mathbb{H}^2$. Furthermore, let ρ_i denote the geodesic rays emanating from p_ξ and asymptotic to $\xi_i, i = 1, 2, 3$. These rays meet at the angles equal to $2\pi/3$ at the points p_ξ. Thus, the ray ρ_1 uniquely determines the rays ρ_2, ρ_3 (since the triple ξ is positively oriented). Let v_ξ be the derivative of ρ_1 at p_ξ. Thus, we obtain a continuous map

$$c : \tilde{M} \to U\mathbb{H}^2, \quad c : \xi = (\xi_1, \xi_2, \xi_3) \mapsto v_\xi \in T_p\mathbb{H}^2,$$

where $U\mathbb{H}^2$ is the unit tangent bundle of $\mathbb{H}^2$. The map c also has continuous inverse: Given $(p, v) \in U\mathbb{H}^2, v \in T_p\mathbb{H}^2$, we let ρ_1 be the geodesic ray emanating from p with the derivative v. From this ray ρ_1 we construct rays ρ_2, ρ_3 (meeting ρ_1 at angles $2\pi/3$) and, therefore, the points $\xi_i, i = 1, 2, 3 \in \mathbb{S}^1$. Since $\mathbb{H}^2$ is contractible, the unit tangent bundle $U\mathbb{H}^2$ is trivial and, hence, $\tilde{M}$ is homeomorphic to $U\mathbb{H}^2 \cong \mathbb{H}^2 \times \mathbb{S}^1$. $\square$

In particular, $\pi_i(\tilde{M}) = 0, i \geqslant 2$, and $\pi_1(\tilde{M}) \cong \mathbb{Z}$. We now consider the quotient $M = \tilde{M}/\bar{G}$. Since the action $\bar{G} \curvearrowright \mathbb{S}^1$ is free, properly discontinuous and cocompact, M is a compact 3-dimensional manifold. Furthermore, $C = \pi_1(\tilde{M}) < \pi_1(M)$ is a normal infinite cyclic subgroup and

$$\bar{G} \cong \pi_1(M)/C.$$

Hence, the normal subgroup C has infinite index in $\pi_1(M)$. Since $\pi_i(\tilde{M}) = 0, i \geqslant 2$, the manifold M also has trivial homotopy groups $\pi_i(M), i \geqslant 2$, i.e., the manifold M is aspherical.

We next review, briefly, the classification of closed 3-dimensional manifolds given by Perelman's Geometrization Theorem (Thurston's Conjecture). The description of closed connected oriented 3-dimensional manifolds starts with the connected sum decomposition of a closed 3-manifold into *prime 3-manifolds*:

$$M = M_1 \# \ldots \# M_k,$$

where each manifold M_i does not admit a nontrivial connected sum decomposition. If $M \neq M_1$, then each 2-sphere, along which M splits as a connected sum, defines a nontrivial element of $\pi_2(M)$. Thus, in our case, $M = M_1$.

We now consider the case of (closed, connected and oriented) prime 3-manifolds. Every prime manifold M is either *geometric* or is obtained by gluing geometric manifolds along boundary tori and Klein bottles, which are *incompressible surfaces*.

1. If this decomposition is not void, then the manifold M is *Haken* and classification of Haken manifolds was known before Perelman, primarily, due to work of Waldhausen, Jaco, Shalen, Johannson and Thurston.

2. Otherwise, if this secondary decomposition of M is void, then M is geometric and we explain below what this means.

There are eight types of closed 3-dimensional geometric manifolds, they are homeomorphic to quotients X/Γ of certain simply-connected homogeneous 3-dimensional Riemannian manifolds X. The groups Γ are discrete subgroups of $\mathrm{Isom}(X)$, acting on X freely and cocompactly.

The list of homogeneous manifolds X is:

$$\mathbb{H}^3, \mathbb{E}^3, \mathbb{S}^3, \mathbb{H}^2 \times \mathbb{R}, \mathbb{S}^2 \times \mathbb{R}, \widetilde{SL}(2,\mathbb{R}), Nil_3, Sol_3.$$

Note that the first five of these homogeneous manifolds are symmetric spaces (three of which have nonpositive curvature); the remaining three are Lie groups equipped with left-invariant Riemannian metrics.

In case when $\pi_2(M) = \pi_3(M) = 0$, as in our situation, the manifold X cannot be $\mathbb{S}^3$ and $\mathbb{S}^2 \times \mathbb{R}$. This leaves only the spaces X isometric to:

In the case $X \cong \mathbb{H}^3$, the quotient manifold $M = X/\Gamma$ is hyperbolic, and, hence, its fundamental group $\pi_1(M)$ is Gromov-hyperbolic. In particular, $\pi_1(M)$ cannot contain a normal infinite cyclic subgroup, see §9.14. This excludes hyperbolic manifolds ($X \cong \mathbb{H}^3$). Similarly, every cocompact lattice Γ acting on Sol_3 is isomorphic to the semidirect product

$$\Gamma \cong \mathbb{Z}^2 \rtimes_A \mathbb{Z},$$

where the matrix $A \in GL(2,\mathbb{Z})$ has both eigenvalues different from ± 1. This shows that Γ cannot contain a normal infinite cyclic subgroup. Thus, *Sol*-manifolds are also excluded.

Closed manifolds M homeomorphic to quotients of the remaining homogeneous spaces ($\mathbb{H}^2 \times \mathbb{R}, \widetilde{SL}(2,\mathbb{R}), \mathbb{E}^3, Nil_3$) have an important common feature, they are *Seifert manifolds*. Fundamental groups of all aspherical closed Seifert manifolds M admit short exact sequences:

$$1 \to \mathbb{Z} \to \pi_1(M) \xrightarrow{\psi} F \to 1, \tag{21.3}$$

where F's are groups acting faithfully and geometrically either on the Euclidean plane or on the hyperbolic plane. The former occurs in the case of the geometries $X = \mathbb{E}^3$ and $X = Nil$. In both cases, the group $\pi_1(M)$ is amenable; thus, all its quotients are amenable as well. However, a group G quasi-isometric to the hyperbolic plane cannot be amenable. This leaves only the cases $X \cong \mathbb{H}^2 \times \mathbb{R}$ and $X \cong \widetilde{SL}(2,\mathbb{R})$. The infinite cyclic normal subgroup $C \lhd \pi_1(M)$ described above, projects to a normal subgroup of the group F. Since the latter cannot have any nontrivial normal cyclic subgroups (Corollary 10.21), the group C has to be contained in the kernel of the homomorphism ψ in the sequence (21.3). We conclude, therefore, that

$$\bar{G}/\Phi \cong F,$$

where $\Phi \cong \mathrm{Ker}(\psi)/C$ is a finite cyclic group. Thus, the group $\bar{G}$ (and, hence G) admits a geometric action on the hyperbolic plane, as required.

REMARK 21.23. With a bit more work, one shows that $C = \mathrm{Ker}(\psi)$, and, hence, $F \cong \bar{G}$. Furthermore, one verifies that $X \cong \widetilde{SL}(2,\mathbb{R})$.

We are now done with the case when the manifold M itself is geometric. It remains to rule out the case when M is obtained by gluing geometric 3-manifolds along their boundary tori and Klein bottles. Such a manifold M is necessarily Haken and, hence, Seifert (since $\pi_1(M)$ contains an infinite cyclic normal subgroup): A proof of this theorem can be found for instance in Hempel's book [**Hem78**]. An alternative to this reference is to argue that existence of a nontrivial infinite cyclic normal subgroup of $\pi_1(M)$ implies that the manifold M is obtained by gluing only Seifert manifolds (as hyperbolic ones are excluded by the same argument we used to rule out the entire M from being hyperbolic). Then, similarly to the proof in Hempel's book, one argues that the gluing has to preserve (up to isotopy) *Seifert fibrations* of the geometric pieces and, hence, the manifold M itself is Seifert. □

COROLLARY 21.24. *The class of fundamental groups of closed surfaces is QI rigid.*

PROOF. Suppose that S is a closed connected surface. Since we are interested in VI invariance, we can assume that S is oriented. If $S \cong \mathbb{S}^2$, then its fundamental group is obviously QI rigid. For surfaces of genus $\geqslant 2$, QI rigidity follows from Theorem 21.20. Lastly, suppose that $S = T^2$, is the torus. Then any group G which is QI to $\pi_1(S)$ is virtually abelian of rank 2, see Theorem 14.26. □

CHAPTER 22

Quasiisometries of nonuniform lattices in $\mathbb{H}^n$

Suppose that G is either a Lie group or a finitely generated group and $\Gamma \leqslant G$ is a finitely-generated subgroup. For each element of the commensurator $g \in Comm_G(\Gamma)$, the Hausdorff distance between Γ and $g\Gamma g^{-1}$ is finite. Therefore, g defines a quasiisometry $q = q_g : \Gamma \to \Gamma$, which sends $\gamma \in \Gamma$ to an element $\gamma' \in \Gamma$ nearest to $g\gamma g^{-1}$.

The main goal of this chapter is to prove a converse to this elementary observation, as well as QI rigidity for nonuniform lattices in $PO(n,1), n \geqslant 3$. Along the way, we give a proof of the Mostow Rigidity Theorem.

THEOREM 22.1 (R. Schwartz [**Sch96b**]). *Let $\Gamma < G = PO(n,1)$ be a nonuniform lattice, $n \geqslant 3$. Then:*

(a) For each quasiisometry $f : \Gamma \to \Gamma$ there exists $g \in Comm_G(\Gamma)$, defining quasiisometry q_g, which is a within finite distance from f. The distance between these quasiisometries depends only on Γ and on the quasiisometry constants of f.

(b) Suppose that $\Gamma, \Gamma' < G$ are nonuniform lattices quasiisometric to each other. Then there exists an isometry $g \in \mathrm{Isom}(\mathbb{H}^n)$ such that the groups Γ' and $g\Gamma g^{-1}$ are commensurable.

(c) Suppose that Γ' is a finitely generated group which is quasiisometric to a nonuniform lattice Γ above. Then the groups Γ, Γ' are virtually isomorphic.

Note that the above theorem fails in the case of the hyperbolic plane (except for the last part). Indeed, every free group F_r of rank $\geqslant 2$ can be realized as a nonuniform lattice Γ acting on $\mathbb{H}^2$. In view of the thick-thin decomposition of the hyperbolic surface $M = \mathbb{H}^2/\Gamma$, Γ contains only finitely many Γ-conjugacy classes of maximal parabolic subgroups: Every such class corresponds to a component of $M \setminus M_c$. Suppose now that $r \geqslant 3$. Then there are *atoroidal* automorphisms ϕ of F_r, such that for every nontrivial cyclic subgroup $C < F_n$ and every m, $\phi^m(C)$ is not conjugate to C, see e.g. [**BFH97**]. Therefore, such ϕ cannot send parabolic subgroups of Γ to parabolic subgroups of Γ. Hence, the quasiisometry of F_n given by ϕ cannot extend to a quasiisometry $\mathbb{H}^2 \to \mathbb{H}^2$. It follows that Part (a) fails for $n = 2$. Similarly, one can show that Part (b) also fails, since commensurability preserves arithmeticity and there are both arithmetic and non-arithmetic lattices in $\mathrm{Isom}(\mathbb{H}^2)$. All these lattices are virtually free, hence, virtually isomorphic.

Our proof of Theorem 22.1 mostly follows [**Sch96b**].

22.1. Coarse topology of truncated hyperbolic spaces

Suppose that $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ is a nonuniform lattice. In §10.6.3 we defined the *truncated hyperbolic space* $\Omega \subset \mathbb{H}^n$ associated with Γ. The space Ω is a certain manifold with boundary; its boundary components are *peripheral horospheres* Σ_j.

We equip the truncated hyperbolic space Ω with the path-metric $d = d_\Omega$, induced by the restriction of the Riemannian metric of $\mathbb{H}^n$ to Ω:

$$d(x, y) = \inf_p \operatorname{length}(p)$$

where the infimum is taken over all paths p in Ω connecting x to y. The metric d is invariant under Γ and, since the quotient Ω/Γ is compact, (Ω, d) is quasiisometric to the group Γ. We will use the notation dist for the hyperbolic distance function in $\mathbb{H}^n$.

LEMMA 22.2. *The identity map $\iota : (\Omega, d) \to (\Omega, \operatorname{dist})$ is 1-Lipschitz and uniformly proper.*

PROOF. If p is a path in Ω, then p has the same length with respect to the metrics d and dist. This immediately implies that ι is 1-Lipschitz. Uniform properness follows from the fact that the group Γ acts geometrically on both $(\Omega, d), (\Omega, \operatorname{dist})$ and that the map ι is Γ-equivariant, see Lemma 5.41. □

LEMMA 22.3. *The restriction of d to each peripheral horosphere $\Sigma \subset \partial\Omega$ equals the Riemannian distance function defined by the restriction of the hyperbolic Riemannian metric to Σ. In particular, (Σ, d_Σ) is isometric to the Euclidean space $\mathbb{E}^{n-1}$.*

PROOF. Without loss of generality, we may assume that (in the upper half-space model of $\mathbb{H}^n$), $\Sigma = \{(x_1, \ldots, x_n) : x_n = 1\}$. Hence,

$$\Omega \subset \{(x_1, \ldots, x_n) : 0 < x_n \leqslant 1\}.$$

The hyperbolic Riemannian metric restricted to Σ equals the flat metric on Σ. Therefore, it is enough to show that for every path p in Ω, connecting points $x, y \in \Sigma$, there exists a path q in Σ (still, connecting x to y), such that $\operatorname{length}(q) \leqslant \operatorname{length}(p)$. Consider the vertical projection

$$\pi : \Omega \to \Sigma, \quad \pi(x_1, \ldots, x_{n-1}, x_n) = (x_1, \ldots, x_{n-1}, 1).$$

According to Exercise 8.63, $\|d\pi\| \leqslant 1$ (with respect to the hyperbolic metric). Therefore, setting $q := \pi \circ p$, we obtain

$$\operatorname{length}(q) \leqslant \operatorname{length}(p). \quad \square$$

LEMMA 22.4. *For every horoball $B \subset \mathbb{H}^n$, the R-neighborhood $\mathcal{N}_R(B)$ of B in $\mathbb{H}^n$ is also an open horoball $B' \subset \mathbb{H}^n$.*

PROOF. We again work in the upper half-space model where

$$B = \{(x_1, \ldots, x_n) : x_n > 1\}.$$

We let Σ denote the boundary of B and let $\pi : \mathbb{H}^n \setminus B \to \Sigma$ denote the vertical projection as in the proof of the previous lemma. We leave it to the reader to check that

$$\operatorname{dist}(x, \Sigma) = \operatorname{dist}(x, \pi(x)).$$

It follows, in view of Exercise 8.14, that $\mathcal{N}_R(B)$ is the horoball

$$\{(x_1, \ldots, x_n) : x_n > e^{-R}\}. \quad \square$$

We refer the reader to §6.6 for the notion of *coarse separation*, *deep/shallow components* and *inradii* of shallow components, used below. The following lemma is the key for distinguishing the case of the hyperbolic plane from the higher-dimensional hyperbolic spaces (of dimension $\geqslant 3$):

Lemma 22.5. *Let Ω is a truncated hyperbolic space of dimension $\geqslant 3$. Then each peripheral horosphere $\Sigma \subset \Omega$ does not coarsely separate Ω.*

Proof. Let $B \subset \mathbb{H}^n$ denote the horoball bounded by Σ. For each R, the R-neighborhood $\mathcal{N}_R(B)$ of B in $\mathbb{H}^n$ is again a horoball. We claim that B' does not separate Ω. Indeed, the horoball B' does not separate $\mathbb{H}^n$. Therefore, for each pair of points $x, y \in \Omega \setminus B'$, there exists a piecewise-geodesic path p connecting them within $\mathbb{H}^n \setminus B'$. If the path p is entirely contained in Ω, we are done. Otherwise, we subdivide p into finitely many subpaths, each of which is either contained in Ω or connects a pair of points in the boundary of one of the complementary horoballs $B_j \subset \mathbb{H}^n \setminus \Omega$.

Exercise 22.6. The intersection of B' with $\Sigma_j = \partial B_j$ is isometric to a metric ball in the Euclidean space (Σ_j, d). Hint: Use the upper half-space model, where that B_j is given by the inequality $x_n > 1$. Then $B' \cap \Sigma_j$ is the intersection of a Euclidean hyperplane with a Euclidean metric ball.

Note that a round ball cannot separate $\mathbb{R}^{n-1}$, provided that $n-1 \geqslant 2$. Thus, we can replace $p_j = p \cap B_j$ with a new path p_j' which connects the end-points of p_j within the complement $\Sigma_j \setminus B'$. By making these replacements for each j, we get a new path q connecting x to y within $\Omega \setminus B'$. Therefore, B' does not separate Ω.

We are now ready to show that Σ cannot coarsely separate (Ω, d). We will use the notation $\mathcal{N}_{R,d}$ for the R-neighborhood with respect to the metric d. Suppose that for some R, $Y := \Omega \setminus \mathcal{N}_{R,d}(B)$ contains at least two deep components C_1, C_2. Let $x_i \in C_i, i = 1, 2$. By the definition of a deep component of Y, there are continuous proper paths $\alpha_i : \mathbb{R}_+ \to C_i, \alpha_i(0) = x_i$, $i = 1, 2$. Thus,

$$\lim_{t\to\infty} d(\alpha_i(t), \Sigma) = \infty.$$

By Lemma 22.2, there exists T, such that $y_i := \alpha_i(T) \notin B'$, $i = 1, 2$. Therefore, as we proved in Lemma 22.5, we can connect y_1 to y_2 by a path in $\Omega \setminus B' \subset Y$. We conclude that $C_1 = C_2$, which is a contradiction. $\square$

Exercise 22.7. Show that Lemma 22.5 fails for $n = 2$. Hint: Use the fact that each Cayley graph of a free nonabelian group of finite rank has infinitely many ends.

In order to appreciate the difficulty of the proof of Proposition 22.9 below, we encourage the reader to do first the following exercise:

Exercise 22.8. Suppose that α is an isometry of $\mathbb{H}^n, n \geqslant 2$, such that

$$\mathrm{dist}_{Haus}(\Omega, \alpha(\Omega)) \leqslant C.$$

Show that for each peripheral horosphere $\Sigma \subset \partial\Omega$, there exists a peripheral horosphere $\Sigma' \subset \partial\Omega$ satisfying

$$\mathrm{dist}_{Haus}(\Sigma', \alpha(\Sigma)) \leqslant R,$$

where R depends only on C and not on Σ.

Now, suppose that Ω, Ω' are truncated hyperbolic spaces for lattices $\Gamma, \Gamma' < \mathrm{Isom}(\mathbb{H}^n)$, and $f : \Omega \to \Omega'$ is an (L, A)–quasiisometry. Let Σ be a peripheral horosphere of Ω. Recall that we are assuming that $n \geqslant 3$.

Proposition 22.9. *There exists a peripheral horosphere $\Sigma' \subset \partial\Omega'$ which is within a finite Hausdorff distance from $f(\Sigma)$.*

Proof. We start with the idea of the proof. Suppose that $h : M \to M'$ is a homeomorphism of compact connected n-dimensional manifolds with boundary, satisfying $H_{n-1}(M) = 0, H_{n-1}(M') = 0$. Then $h(\partial M) = \partial M'$. Of course, one can prove it in many ways (and without using our homological assumption), but the following, admittedly, somewhat silly, proof is a model of the proof of the proposition. We first note that no boundary component of M separates M, while a connected hypersurface which is not contained in the interior of M' has to separate M' (due to our homological assumptions). The proof below is a coarse version of this argument, where we use coarse separation arguments. The case 1 in this proof corresponds to the possibility that the entire boundary of M' is contained in one component of $M' \setminus f(\partial M)$, while the case 2 corresponds to the possibility that $f(\partial M)$ separates two boundary components of M'.

We now proceed with the actual proof. Since Ω'/Γ' is compact, there exists $D < \infty$, such that for every $x \in \Omega'$,

$$\operatorname{dist}(x, \partial\Omega') \leqslant D. \tag{22.1}$$

The horosphere Σ, being isometric to $\mathbb{R}^{n-1}$ (with respect to the metric d), has bounded geometry and is uniformly contractible. Therefore, according to Theorem 6.71, $f(\Sigma)$ coarsely separates $\mathbb{H}^n$. However $f(\Sigma)$ cannot coarsely separate Ω', since f is a quasiisometry and Σ does not coarsely separate Ω (Lemma 22.5).

Exercise 22.10. Suppose that $Y \subset X$ coarsely separates subsets $X_1, X_2 \subset X$. Then, under any quasiisometry $f : X \to X'$, the set $f(Y)$ coarsely separates $f(X_1)$ from $f(X_2)$.

Let $r < \infty$ be such that $\mathcal{N}_r(f(\Sigma))$ separates $\mathbb{H}^n$ into (two) deep components X_1, X_2. We define a new truncated hyperbolic space

$$\Omega'' := \mathcal{N}_r(\Omega').$$

We will use the notation $B_j'' := B_j' \setminus \Omega''$ for the complementary horoballs of Ω''.

Case 1. Suppose first that for each B_j', the smaller horoball B_j'' is contained in the complementary region X_1. In view of (22.1), for every $x \in \Omega'$, we obtain:

$$\operatorname{dist}(x, X_1) \leqslant r + D,$$

since the hyperbolic distance from x to some point of Ω'' is at most $D + r$. Furthermore, every $x \in \mathbb{H}^n \setminus \Omega'$ is within a distance $\leqslant r$ from Ω''. Therefore, for every $x \in \mathbb{H}^n$,

$$\operatorname{dist}(x, X_1) \leqslant 2r + D.$$

In particular, if we take any point $x \in X_2$, there exists a path p of length $\leqslant 2r + D$ connecting it to X_1. This path has to cross the neighborhood $\mathcal{N}_r(f(\Sigma))$ separating X_1 from X_2. Therefore, the entire set X_2 is *shallow*: It is contained within distance $\leqslant 2r + D$ from $\mathcal{N}_r(f(\Sigma))$. This contradicts the property that the set X_2 is deep.

Similarly (renaming X_1 and X_2), we rule out the possibility that all horoballs B_j'' are contained in X_2.

Case 2. There are two complementary horoballs B_1', B_2' of Ω' such that

$$B_1'' \subset X_1, \quad B_2'' \subset X_2.$$

Set $\Sigma_i' := \partial B_i', i = 1, 2$. If both intersections

$$T_i' := \Sigma_i' \cap X_i, i = 1, 2$$

contain points which are arbitrarily far from $f(\Sigma)$, then $f(\Sigma)$ coarsely separates Ω', which is again a contradiction. Therefore, say, for $i = 1$, there exists $r' < \infty$ such that $\Sigma' := \Sigma_1'$ satisfies

$$\Sigma' \subset \mathcal{N}_{r'}(f(\Sigma)). \tag{22.2}$$

Our goal is to show that $f(\Sigma) \subset \mathcal{N}_R(\Sigma')$ for some $R < \infty$.

The inclusion (22.2) implies that the nearest-point projection $\Sigma' \to f(\Sigma)$ defines a quasiisometric embedding $h : \Sigma' \to \Sigma$, see Exercise 5.12. However, Lemma 7.83 proves that a quasiisometric embedding between two Euclidean spaces of the same dimension is a quasiisometry. Thus, there exists $R' < \infty$ such that $f(\Sigma) \subset \mathcal{N}_{R'}(\Sigma')$. Hence, $f(\Sigma)$ is Hausdorff-close to Σ'.

□

EXERCISE 22.11. Show that the horosphere Σ' in Proposition 22.9 is unique.

We note that there are alternative proofs of Proposition 22.9, which use asymptotic cones instead of coarse topology, see for instance, [**KL97**] or [**BDM09**] (Theorem 23.40 in the next chapter).

We now improve Proposition 22.9 and establish uniform control on the distance from $f(\Sigma)$ to the boundary horosphere $\Sigma' \subset \Omega'$ in this proposition.

LEMMA 22.12. *In Proposition 22.9, for all peripheral horospheres $\Sigma \subset \partial\Omega$,*

$$\mathrm{dist}_{Haus}(f(\Sigma), \Sigma')) \leqslant c(L, A),$$

where $c(L, A)$ is independent of Σ.

PROOF. The proof is by inspection of the arguments in the proof of Proposition 22.9. First of all, the constant r depends only on the quasiisometry constants of the mapping f and the uniform geometry/uniform contractibility bounds for $\mathbb{R}^{n-1}$ and $\mathbb{H}^n$. The *inradii* of the *shallow* components of

$$\Omega' \setminus \mathcal{N}_r(f(\Sigma))$$

again depend only on the above data. Therefore, there exists a uniform constant r' such that one of the horospheres Σ_1' or Σ_2' in the proof of Proposition 22.9 is contained in $\mathcal{N}_{r'}(f(\Sigma))$. Finally, an upper bound on R' such that $\mathcal{N}_{R'}(Image(h)) = \Sigma'$ (coming from Lemma 7.83) again depends only on the quasiisometry constants of the projection $h : \Sigma' \to \Sigma$. □

REMARK 22.13. Proposition 22.9 and Lemma 22.9 combine into the *Quasiflat Lemma* from [**Sch96b**], §3.2. This lemma can be seen as a version of the Morse Lemma 9.40 for truncated hyperbolic spaces. The spaces Ω, Ω' are, in fact, *relatively hyperbolic* in the strong sense. See Chapter 9.26 for further details.

22.2. Hyperbolic extension

Let $\Omega, \Omega' \subset \mathbb{H}^n$ be truncated hyperbolic spaces ($n \geqslant 3$) of lattices $\Gamma, \Gamma' < \mathrm{Isom}(\mathbb{H}^n)$ and let C, C' denote the sets whose elements are peripheral horospheres of Ω, Ω' respectively. The main result of this section is

THEOREM 22.14 (Horoball QI extension theorem). *Each quasiisometry $f : \Omega \to \Omega'$ admits a quasiisometric extension $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$. Moreover, the extension $\tilde{f}$ satisfies the following equivariance property:*

Suppose that $f : X \to X'$ is quasiequivariant with respect to an isomorphism

$$\rho : \Gamma \to \Gamma'.$$

Then the extension $\tilde{f}$ is also ρ–quasiequivariant.

PROOF. By Lemma 22.12, for every peripheral horosphere $\Sigma \subset \Omega$ there exists a peripheral horosphere Σ' of Ω' such that $\text{dist}_{Haus}(f(\Sigma), \Sigma') \leqslant c < \infty$, where c depends only on the quasiisometry constants of f. By uniqueness of the horosphere Σ', we obtain a map

$$\theta : C \to C', \theta : \Sigma \mapsto \Sigma',$$

which is ρ–equivariant, provided that f was ρ–quasiequivariant. We will use the notation B' for the horoball bounded by Σ'.

We first alter f on $\partial\Omega$ by postcomposing $f\big|_{\Sigma}$ with the nearest-point projection to Σ' for every $\Sigma \in C$. The new map is again a quasiisometry, since its distance from f is finite. The modification clearly preserves ρ–quasiequivariance. We retain the notation f for the new quasiisometry, which now satisfies

$$f(\Sigma) \subset \Sigma', \forall \Sigma \in C.$$

We construct an extension $\tilde{f} : B \to B'$ of $f\big|_{\Sigma}$ for each complementary horoball $B \subset \mathbb{H}^n \setminus \Omega$ as follows.

For points $x \in \Sigma, x' \in \Sigma'$ and $t \in \mathbb{R}_+$ we define $x_t \in B, x'_t \in B'$, so that the maps

$$t \mapsto x_t, q \mapsto x'_t, t \in \mathbb{R}_+$$

are geodesic rays asymptotic to the centers of the horoballs B, B' respectively. Of course, every point $y \in B$ has the form $y = x_t$ for unique $x \in \Sigma, t > 0$. Then we define the extension $\tilde{f} : B \to B'$ by the formula:

$$x_t \mapsto x'_t, \quad x' = f(x), x \in \Sigma.$$

By the construction, this extension is quasiequivariant if f is.

We will now verify that this extension is coarsely Lipschitz. Since being coarse Lipschitz is a local property, it suffices to show that for each horoball B, the map $\tilde{f} : B \to B'$ is coarse Lipschitz. By composing $\tilde{f}$ with isometries of $\mathbb{H}^n$, the problem reduces to the case when $B = B'$ is given by the inequality $x_n \geqslant 1$ (in the upper half-space model of $\mathbb{H}^n$). It suffices to consider points $y, z \in B$ within unit distance from each other. We need to show that

$$\text{dist}(\tilde{f}(z), \tilde{f}(w)) \leqslant Const.$$

If $z, w \in B$ have the form $z = x_t, w = x_s$ for some $x \in \Sigma, s$ and t, then, by the construction,

$$\text{dist}(\tilde{f}(z), \tilde{f}(w)) = \text{dist}(z, w) = |t - s|.$$

Therefore, by the triangle inequality, the problem reduces to getting a uniform upper bound on the distances $\text{dist}(\tilde{f}(z), \tilde{f}(w))$ for points z and w belonging to the same horosphere $\Sigma_t \subset B$:

$$z = x_t, y = z_t, \quad x \in \Sigma, y \in \Sigma.$$

We will use the notation

$$\text{dist}_{\Sigma_t}(z, w)$$

for the distance between z and w computed with respect to the Riemannian distance function on Σ_t, where the Riemannian metric on Σ_t is the restriction of the hyperbolic Riemannian metric. In other words,

$$\mathrm{dist}_{\Sigma_t}(z, w) = e^{-t}|z - w|.$$

It follows from Exercise 8.56 that

$$\mathrm{dist}_{\Sigma_t}(z, w) \leqslant \epsilon := \sqrt{2(e-1)},$$

since we are assuming that $\mathrm{dist}(z, w) \leqslant 1$. Accordingly,

$$\mathrm{dist}_{\Sigma}(x, y) \leqslant \epsilon e^t.$$

Since $f : (\Omega, d_\Omega) \to (\Omega', d_{\Omega'})$ is (L, A)–coarse Lipschitz,

$$\mathrm{dist}_{\Sigma}(f(x), f(y)) \leqslant e^t L\epsilon + A.$$

It follows that

$$d(\tilde{f}(z), \tilde{f}(w)) \leqslant \mathrm{dist}_{\Sigma_t}(\tilde{f}(z), \tilde{f}(w)) \leqslant L\epsilon + Ae^{-t} \leqslant L\epsilon + A.$$

This proves that the extension $\tilde{f}$ is coarse Lipschitz in the horoball B and, hence, in the entire $\mathbb{H}^n$. The same argument applies to the extension $\tilde{f}'$ of the coarse inverse f' to the mapping f. We leave it to the reader to verify that the inequality

$$d_\Omega(f' \circ f, id) \leqslant A$$

implies

$$\mathrm{dist}(\tilde{f}' \circ \tilde{f}, id) \leqslant A.$$

Thus, $\tilde{f}$ is a quasiisometry. □

Since $\tilde{f}$ is a quasiisometry of $\mathbb{H}^n$, it admits a quasiconformal extension $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$ (see Theorems 9.108 and 20.37). By Corollary 9.111, the homeomorphism h is ρ-equivariant, provided that f is quasiequivariant.

Let Λ, Λ' denote the sets of the centers of the peripheral horospheres of Ω, Ω' respectively. Since $f(\Sigma) = \Sigma'$ for every peripheral horosphere of Ω, continuity of the extension also implies that $h(\Lambda) = \Lambda'$.

22.3. Mostow Rigidity Theorem

The Mostow Rigidity Theorem that we will prove in this section was a precursor and inspiration for the Schwartz Rigidity Theorem. We prove this theorem first, since its proof is technically simpler and also illustrates some of the ideas behind Schwartz' proof. Our arguments are inspired by the ones of P. Tukia [**Tuk85**] and N. Ivanov [**Iva96**].

THEOREM 22.15 (Mostow Rigidity Theorem). *Suppose that $n \geqslant 3$ and $\Gamma, \Gamma' <$ $\mathrm{Isom}(\mathbb{H}^n)$ are lattices and $\rho : \Gamma \to \Gamma'$ is an isomorphism. Then ρ is induced by an isometry, i.e. there exists an isometry $\alpha \in \mathrm{Isom}(\mathbb{H}^n)$ such that*

$$\alpha \circ \gamma = \rho(\gamma) \circ \alpha$$

for all $\gamma \in \Gamma$.

PROOF. **Step 1.** We first observe that Γ is uniform if and only if Γ' is uniform. Indeed, if Γ is uniform, it is Gromov-hyperbolic and, hence, cannot contain a noncyclic free abelian group. On the other hand, if Γ' is nonuniform then Corollary 10.28 implies that Γ' contain free abelian subgroups of rank $n - 1 > 1$.

Lemma 22.16. *There exists a ρ–equivariant quasiisometry $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$.*

Proof. As in the proof of Theorem 22.1, we choose truncated hyperbolic spaces $\Omega \subset \mathbb{H}^n, \Omega' \subset \mathbb{H}^n$ for the lattices Γ and Γ' respectively. (If Γ is a uniform lattice, we take, of course, $\Omega = \Omega' = \mathbb{H}^n$.) Lemma 5.43 implies that there exists a ρ–quasiequivariant quasiisometry

$$f : \Omega \to \Omega'.$$

Therefore, according to Theorem 22.14, f extends to a ρ–equivariant quasiisometry $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$. □

Remark 22.17. The most difficult part of the proof of Theorem 22.14 was to show that f sends peripheral horospheres uniformly close to peripheral horospheres. In the equivariant setting the proof is much easier: The homomorphism ρ sends maximal abelian subgroups of Γ to maximal abelian subgroups of Γ'. The stabilizers of peripheral horospheres are virtually $\mathbb{Z}^{n-1}$. Therefore, ρ sends stabilizers of peripheral horospheres to stabilizers of peripheral horospheres. From this, it is immediate that peripheral horospheres map uniformly close to peripheral horospheres.

Step 2. Let h denote the ρ–equivariant quasiconformal homeomorphism

$$h : \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$$

extending the quasiequivariant quasiisometry $\tilde{f}$. Our goal is to show that h is Moebius. We argue as in the proof of Theorem 22.31. We will identify $\mathbb{S}^{n-1}$ with the extended Euclidean space $\mathbb{R}^{n-1} \cup \{\infty\}$. Accordingly, we will identify $\mathbb{H}^n$ with the upper half-space. The key to the proof is the fact that h is differentiable almost everywhere on $\mathbb{R}^{n-1}$ and that its Jacobian determinant is nonzero for almost every $z \in \mathbb{R}^{n-1}$. (In fact, we need only uncountably many points of differentiability, where the Jacobian determinant is nonzero.)

In Theorem 10.29 we proved that every point of $\mathbb{S}^{n-1}$ is either a conical limit point of Γ or is a parabolic fixed point. Since Γ has only countably many parabolic elements and each has only one fixed point, almost every of $\mathbb{S}^{n-1}$ is a conical limit point ofΓ. Hence, we find a conical limit point $z \in \mathbb{S}^{n-1} \setminus \{\infty\}$, which is a point of differentiability of h, where $J_z(h) \neq 0$. After applying a Moebius change of coordinates, we can assume that $z = h(z) = 0 \in \mathbb{R}^{n-1}$ and that $h(\infty) = \infty$.

The following proof is yet another version of the *zooming argument*. Let $L \subset \mathbb{H}^n$ be the vertical geodesic emanating from 0; pick a base-point $y_0 \in L$. Since z is a conical limit point, there is a sequence of elements $\gamma_i \in \Gamma$ such that

$$\lim_{i\to\infty} \gamma_i(y_0) = z$$

and

$$\mathrm{dist}(\gamma_i(y_0), L) \leqslant Const$$

for each i. Let y_i denote the nearest-point projection of $\gamma_i(y_0)$ to L. Take the sequence of hyperbolic translations $T_i : y \mapsto t_i y$ with the axis L, such that $T_i(y_0) = y_i$. Then the sequence $k_i := \gamma_i^{-1} T_i$ lies in a compact $C \subset G = \mathrm{Isom}(\mathrm{H}^n)$.

Now, the proof turns from mappings of the hyperbolic n-space to mappings of $\mathbb{R}^{n-1}$. We form a sequence of quasiconformal homeomorphisms

$$h_i(x) := t_i^{-1} h(t_i x) = T_i^{-1} \circ h \circ T_i(x), x \in \mathbb{R}^{n-1},$$

$$\lim_{i\to\infty} t_i = 0.$$

Since the mapping h is assumed to have invertible derivative at the origin, there is a linear transformation $A \in GL(n-1,\mathbb{R})$ such that

$$\lim_{i\to\infty} h_i(x) = Ax$$

for all $x \in \mathbb{R}^{n-1}$. Since $h(\infty) = \infty$, it follows that

$$\lim_{i\to\infty} h_i = A$$

pointwise on $\mathbb{S}^{n-1}$.

By the construction, h_i conjugates the group $\Gamma_i := T_i^{-1}\Gamma T_i \subset Mob(\mathbb{S}^{n-1})$ into the group of Moebius transformations $Mob(\mathbb{S}^{n-1})$. We have

$$\Gamma_i = T_i^{-1}\Gamma T_i = (k_i^{-1}\gamma_i)\Gamma(k_i^{-1}\gamma_i)^{-1} = k_i^{-1}\Gamma k_i.$$

After passing to a subsequence, we can assume that

$$\lim_{i\to\infty} k_i = k \in \mathrm{Isom}(\mathbb{H}^n)\,.$$

Therefore the *sequence of subsets* $\Gamma_i \subset G = Mob(\mathbb{S}^{n-1})$ converges to $\Gamma_\infty := k^{-1}\Gamma k$; here convergence is understood in the Chabauty topology on the set of closed subsets of G, see §1.9. For each sequence $\beta_i \in \Gamma_i$ which converges to some $\beta \in G$ we have

$$\lim_{i\to\infty} h_i\beta_i h_i^{-1} = A\beta A^{-1}.$$

Since the subgroup G is closed in $\mathrm{Homeo}(\mathbb{S}^{n-1})$ (with respect to the topology of pointswise convergence, see Corollary 8.5), it follows that the limit $A\beta A^{-1}$ of the sequence of Moebius transformations $(h_i\beta_i h_i^{-1})$, is again a Moebius transformation. This shows that $A\beta A^{-1} \in G$, for each $\beta \in \Gamma_\infty$. Thus,

$$A\Gamma_\infty A^{-1} \subset G.$$

The subgroup $\Gamma_\infty < G$ is conjugate to the lattice Γ and, hence, it cannot have a finite orbit in $\mathbb{S}^{n-1}$, see Corollary 10.20. In particular, the Γ_∞-orbit of ∞ is infinite, which implies that Γ_∞ contains an element γ such that $\gamma(\infty) \notin \{\infty, 0\}$.

LEMMA 22.18. *Suppose that $\gamma \in G = Mob(\mathbb{S}^{n-1})$ is such that $\gamma(\infty) \neq \infty, 0$, $A \in GL(n-1,\mathbb{R})$ is an element which conjugates γ to $A\gamma A^{-1} \in G$. Then A is a Euclidean similarity, i.e., it belongs to $\mathbb{R}_+ \times O(n-1)$.*

PROOF. Suppose that A is not a similarity. Let P be a hyperplane in $\mathbb{R}^{n-1}$ which contains the origin 0 but does not contain $A\gamma^{-1}(\infty)$. Then $\gamma \circ A^{-1}(P)$ does not contain ∞ and, hence, is a round sphere S in $\mathbb{R}^{n-1}$. Since A is not a similarity, the image $A(S)$ is an ellipsoid, which is not a round sphere. Hence, the composition $A\gamma A^{-1}$ does not send planes to round spheres and, therefore, it is not Moebius. This is a contradiction. □

We conclude that the derivative of h at 0 is a similarity $A \in \mathbb{R}_+ \times O(n-1)$. Thus, h is conformal at a.e. point of $\mathbb{R}^n$. One option now is to use Liouville's theorem for quasiconformal maps (Theorem 20.32). Instead, we give a direct argument.

Step 3. We will be using the notation of the Step 2. Consider the quotient

$$Q = G\backslash \mathrm{Homeo}(\mathbb{S}^{n-1})$$

consisting of the cosets $[f] = \{g \circ f : g \in G\}$. We equip this quotient with the quotient topology, where we endow Homeo($\mathbb{S}^{n-1}$) with the topology of pointwise convergence. Since G is a closed subgroup in Homeo($\mathbb{S}^{n-1}$), it follows that every point in Q is closed. (Actually, Q is Hausdorff, but we will not need this.) The group Homeo($\mathbb{S}^{n-1}$) acts on Q by the formula

$$[f] \mapsto [f \circ g], g \in \text{Homeo}(\mathbb{S}^{n-1}).$$

It is clear from the definition of the quotient topology on Q, that this action is continuous, i.e., the map

$$Q \times \text{Homeo}(\mathbb{S}^{n-1}) \to Q$$

is continuous.

Since h is a ρ–equivariant homeomorphism, we have

$$[h] \circ \gamma = [h], \quad \forall \gamma \in \Gamma.$$

Recall that we have two sequences: $\gamma_i \in \Gamma$, $k_i \in G$, such that

$$\lim_{i \to \infty} k_i = k \in G.$$

We also have a sequence of dilations $T_i = \gamma_i \circ k_i$ (fixing the origin in $\mathbb{R}^{n-1}$). Furthermore,

$$\lim_{i \to \infty} h_i = A \in \mathbb{R}_+ \times O(n-1) \subset G,$$

where

$$h_i = T_i^{-1} \circ h \circ T_i.$$

Therefore

$$[h_i] = [h\gamma_i k_i] = [h] \circ k_i,$$
$$[1] = [A] = \lim_i [h_i] = \lim_i ([h_i] \circ k_i) = [h] \circ \lim_{i \to \infty} k_i = [h] \circ k.$$

(Recall that every point in Q is closed.) Thus, $[h] = [1] \circ k^{-1} = [1]$, which implies that h is in G, i.e., h is a Moebius transformation, which we now denote by α. Regarding α as an isometry of $\mathbb{H}^n$ and taking into account that the map $\alpha : \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$ is ρ-equivariant, we conclude that the isometry

$$\alpha : \mathbb{H}^n \to \mathbb{H}^n$$

is also ρ–equivariant. The Mostow Rigidity Theorem follows. ☐

22.4. Zooming in

We now return to the proof of the Schwartz Rigidity Theorem. In §22.2, given a quasiisometry $f : \Omega \to \Omega'$, we constructed its quasiisometric extension $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$; the latter, in turn, has a quasiconformal extension $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$. Our main goal is to show that h is Moebius. By the Liouville's theorem for quasiconformal mappings (Theorem 20.32), h is Moebius provided that it is 1-quasiconformal, i.e., for a.e. point $\xi \in \mathbb{S}^{n-1}$, the derivative $D_\xi h$ of h at ξ is a Euclidean similarity.

We will continue to work with the upper half-space model of the hyperbolic space $\mathbb{H}^n$.

Proposition 22.19. *Suppose that h is not Moebius. Then, there exist lattices $\Gamma_\infty, \Gamma'_\infty$ conjugate to Γ, Γ' respectively, with truncated hyperbolic spaces $\Omega_\infty, \Omega'_\infty$ and a quasiisometry $\tilde{F} : \mathbb{H}^n \to \mathbb{H}^n$, such that:*

1. $\tilde{F}(\Omega_\infty) = \Omega'_\infty$.

2. The extension A of $\tilde{F}$ to $\partial_\infty \mathbb{H}^n = \mathbb{R}^{n-1} \cup \{\infty\}$ fixes the points $0, \infty$.

3. *The mapping A is a linear map, but not a similarity.*

PROOF. Our arguments follow the one in the proof of the Mostow Rigidity Theorem. Since h is differentiable a.e. and is not Moebius, there is a point $\xi \in \mathbb{S}^{n-1}$ such that $D_\xi h$ exists, is invertible, but is not a similarity. Since the subset $\Lambda \subset \mathbb{S}^{n-1}$ consisting of fixed points of parabolic element of Γ is countable, we can assume that $\xi \notin \Lambda$, i.e., is not the center of a complementary horoball of Ω. By pre- and post-composing with isometries of $\mathbb{H}^n$ we can assume that $\xi = 0 = h(\xi)$. We will use the notation

$$A = D_\xi h \in GL(n-1, \mathbb{R})$$

for the derivative of h at 0.

Let $L \subset \mathbb{H}^n$ denote the vertical geodesic asymptotic to 0. Since 0 is not the center of a complementary horoball of Ω, there exists a sequence

$$y_i \in L \cap \Omega, \lim_{i\to\infty} y_i = 0.$$

We now break the symmetry between the lattices Γ, Γ' and, instead of taking points in $\Omega' \cap L$, we take the images $y_i' = f(y_i) \in \Omega'$.

For each i there exists a hyperbolic isometry

$$T_i(y) = t_i y, \quad y \in \mathbb{H}^n, t_i > 0,$$

which maps y_1 to y_i. The compositions

$$\tilde{f}_i := T_i^{-1} \circ \tilde{f} \circ T_i$$

are uniform quasiisometries of $\mathbb{H}^n$. The quasiconformal extensions of these quasi-isometries to $\mathbb{R}^{n-1}$, are given by

$$h_i(x) = \frac{h(t_i x)}{t_i}, \quad x \in \mathbb{R}^{n-1}.$$

By the definition of the derivative of h at the origin,

$$\lim_{i\to\infty} h_i = A, \tag{22.3}$$

where convergence is uniform on compact subsets of $\mathbb{R}^{n-1}$. We now claim that the sequence of quasiisometries $\tilde{f}_i$ coarsely subconverges to a quasiisometry of $\mathbb{H}^n$. As in the proof of Theorem 21.17, there are two ways to argue. One argument is that the claim follows from the convergence property of quasiconformal mappings h_i (in conjunction with the extension Theorem 20.39). Alternatively, the claim follows from the coarse Arzela-Ascoli theorem (Theorem 5.32), since the limit (22.3) forces the quasiisometries $\tilde{f}_i$ to send a base-point $p \in \mathbb{H}^n$ to points $q_i \in \mathbb{H}^n$ satisfying

$$\sup_i \operatorname{dist}(p, q_i) < \infty.$$

In either case, after passing to a subsequence, which we suppress, the sequence of quasiisometries $\tilde{f}_i$ coarsely subconverges to a quasiisometry of $\mathbb{H}^n$. In view of the limit (22.3), we can take the linear extension $\tilde{A}$ of the linear mapping A (defined in Exercise 20.40) as the (coarse) limit quasiisometry.

At this point, there is no reason for $\tilde{A}$ to send Ω to a subset of $\mathbb{H}^n$ within a finite Hausdorff distance from Ω' (after all, we were composing with the mappings $T_i^{\pm 1}$ which do not preserve Ω and Ω'). The reader, who went through the proofs of Tukia's and Mostow's theorems, probably already knows what to do: We need to compose the hyperbolic isometries $T_i^{\pm 1}$ of $\mathbb{H}^n$ with suitable elements of the groups

Γ and Γ'. Recall that the quotients Ω/Γ and Ω'/Γ' are compact. Therefore, there exist $R < \infty$ and sequences $\gamma_i \in \Gamma, \gamma'_i \in \Gamma'$ such that for all i,

$$\mathrm{dist}(\gamma_i(y_1), y_i) \leqslant R, \quad \mathrm{dist}(\gamma'_i(y_1), y'_i) \leqslant R.$$

(This is where we are using the fact that $y_i \in \Omega$ and $y'_i \in \Omega'$.) Hence, both sequences

$$k_i := T_i^{-1} \circ \gamma_i, \quad k'_i := T_i^{-1} \circ \gamma'_i$$

belong to a compact subset in $\mathrm{Isom}(\mathbb{H}^n)$. After passing to a subsequence (which we suppress), we obtain

$$\lim_{i\to\infty} k_i = k \in \mathrm{Isom}(\mathbb{H}^n), \quad \lim_{i\to\infty} k'_i = k' \in \mathrm{Isom}(\mathbb{H}^n)\,.$$

For $i \in \mathbb{N}$ we define truncated hyperbolic spaces

$$\Omega_i := T_i^{-1}\Omega = k_i \circ \gamma_i^{-1}\Omega = k_i\Omega,$$

and

$$\Omega'_i := T_i^{-1}\Omega' = k'_i\Omega',$$

for the lattices $\Gamma_i = k_i^{-1}\Gamma k_i$ and $\Gamma'_i = {k'_i}^{-1}\Gamma' k'_i$ respectively. By the definition of these truncated hyperbolic spaces, the quasiisometry $\tilde{f}_i$ sends Ω_i to Ω'_i. Since (k_i) converges to k and (k'_i) converges to k', we have limits (in the Chabauty topology on the set of closed subsets of $\mathrm{Isom}(\mathbb{H}^n)$, §1.9)

$$\lim_{i\to\infty} \Gamma_i = \Gamma_\infty, \quad \lim_{i\to\infty} \Gamma'_i = \Gamma'_\infty.$$

Since the groups $\Gamma_\infty, \Gamma'_\infty$ are conjugate to the lattices Γ, Γ' respectively, they are lattices themselves. We leave it to the reader to verify that the sets

$$\Omega_\infty := k(\Omega), \quad \Omega'_\infty := k'(\Omega')$$

are truncated hyperbolic spaces for the lattices Γ_∞ and Γ'_∞ respectively and that

$$\lim_{i\to\infty} \Omega_i = \Omega_\infty, \quad \lim_{i\to\infty} \Omega'_i = \Omega'_\infty,$$

again, in the Chabauty topology. Since the sequence $(\tilde{f}_i)$ coarsely converges to $\tilde{A}$, it follows that the affine map $\tilde{A}$ defines a quasiisometry $\Omega_\infty \to \Omega'_\infty$ in the sense that the sets $\tilde{A}(\Omega_\infty)$ and Ω'_∞ are Hausdorff-close to each other. Since the lattice Γ is conjugate to Γ_∞ and Γ' is conjugate to Γ'_∞, the proposition follows. □

We are aiming for a contradiction, therefore, from now on, we rename Γ_∞ to Γ and Γ'_∞ to Γ', etc. The situation when we have a linear mapping (that is not a similarity!) sending Λ to Λ' seems, at the first glance, impossible. Here, however, is an example:

Example 22.20. Let $\Gamma := PSL(2, \mathbb{Z}[i]), \Gamma' := PSL(2, \mathbb{Z}[\sqrt{-2}])$ be Bianchi subgroups of $PSL(2, \mathbb{C})$. Then

$$\Lambda = \mathbb{Q}(i) \cup \{\infty\}, \quad \Lambda' = \mathbb{Q}(\sqrt{-2}) \cup \{\infty\}.$$

Take the real linear mapping $A : \mathbb{C} \to \mathbb{C}$ sending 1 to 1 and i to $\sqrt{-2}$. Then A is invertible, is not a similarity, but $A(\Lambda) = \Lambda'$.

Thus, in order to get a contradiction, we have to exploit the fact that the linear map A we constructed, is the quasiconformal extension of a quasiisometry $\tilde{F}$, $\tilde{F}(\Omega) \subset \Omega'$. We will show (Theorem 22.30) the following:

For every peripheral horosphere $\Sigma \subset \partial\Omega$ whose center is not ∞, there exists a sequence of peripheral horospheres $\Sigma_k \subset \partial\Omega$ such that:

$$\mathrm{dist}(\Sigma, \Sigma_k) \leqslant Const, \quad \lim_{k\to\infty} \mathrm{dist}(\Sigma', \Sigma'_k) = \infty.$$

(We remind the reader that $\mathrm{dist}(\cdot,\cdot)$ denotes the minimal distance between the horospheres and $\theta(\Sigma) = \Sigma', \theta(\Sigma_k) = \Sigma'_k$, $k \in \mathbb{N}$.)

Of course, this means that $\tilde{F}$ cannot be coarse Lipschitz. We will prove the above statement by conjugating $\tilde{F}$ by an inversion which interchanges a horosphere with the center at ∞ and the horospheres Σ, Σ' above. This will amount to replacing the linear map A (from Proposition 22.19) with an *inverted linear mapping.* Inverted linear mappings are defined and analyzed in the next section.

22.5. Inverted linear mappings

Let $A : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be an (invertible) linear mapping. Recall that the inversion J in the unit sphere $\mathbb{S}^{n-1}$ is given by the formula

$$J(x) = \frac{\mathbf{x}}{|\mathbf{x}|^2}, \quad \mathbf{x} \in \mathbb{R}^{n-1}.$$

Definition 22.21. An *inverted linear map* is the conjugate of an invertible linear map A by the inversion J, i.e., the composition

$$h := J \circ A \circ J, \quad h(\mathbf{x}) = \frac{|\mathbf{x}|^2}{|A\mathbf{x}|^2} A(\mathbf{x}).$$

We will introduce the notation

$$\phi = \phi_h = \frac{|\mathbf{x}|^2}{|A\mathbf{x}|^2}$$

and will refer to this function as the *nonlinear factor* of the inverted linear map h.

Lemma 22.22. *The function* $\phi(\mathbf{x}) = \frac{|\mathbf{x}|^2}{|A\mathbf{x}|^2}$ *is* asymptotically constant*, in the sense that*

$$|\nabla\phi(\mathbf{x})| = O(|\mathbf{x}|^{-1}), \quad \|Hess\,(\phi(\mathbf{x}))\| = O(|\mathbf{x}|^{-2})$$

as $|\mathbf{x}| \to \infty$. *Here* $Hess(\phi(\mathbf{x}))$ *stands for the Hessian of the function* $\phi(\mathbf{x})$.

Proof. The function ϕ is a rational vector-function of zero degree, hence, its gradient is a rational vector-function of the degree -1, while every component of its Hessian is a rational function of the degree -2. □

Exercise 22.23. The following are equivalent:
1. The function ϕ is constant.
2. The mapping h is linear.
3. The linear transformation A is a similarity.

Exercise 22.24. The mapping h is differentiable at 0 if and only if A is a similarity.

Since each inverted linear mapping h is, clearly, differentiable everywhere on $\mathbb{R}^{n-1}\setminus\{0\}$, it follows that h determines the origin 0 in the Euclidean space. Hence, h also determines its nonlinear factor ϕ_h (up to a scalar multiple). The next exercise also shows that h determines the origin in $\mathbb{R}^n$:

EXERCISE 22.25. Suppose that h is an inverted nonlinear map with nonconstant factor ϕ_h. Show that for each Euclidean hyperlane $P\subset\mathbb{E}^{n-1}$ not passing through the origin, $h(P)$ is not a Euclidean hyperplane. In contrast, show that h sends each linear subspace in $\mathbb{R}^{n-1}$ to a linear subspace in $\mathbb{R}^{n-1}$.

COROLLARY 22.26. *Fix a positive real number R, and let $(\mathbf{v}_k)$ be a sequence diverging to infinity in $\mathbb{R}^{n-1}$. Then the sequence of maps*

$$h_k(\mathbf{x}) := h(\mathbf{x}+\mathbf{v}_k) - h(\mathbf{v}_k)$$

subconverges (uniformly on the R-ball $B = B(0,R)\subset\mathbb{R}^{n-1}$) to an affine map, as $k\to\infty$.

PROOF. We have:

$$h(\mathbf{x}+\mathbf{v}_k) - h(\mathbf{v}_k) = \phi(\mathbf{x}+\mathbf{v}_k)A(\mathbf{x}+\mathbf{v}_k) - \phi(\mathbf{v}_k)A(\mathbf{v}_k) =$$
$$\phi(\mathbf{x}+\mathbf{v}_k)A(\mathbf{x}) + (\phi(\mathbf{x}+\mathbf{v}_k) - \phi(\mathbf{v}_k))\,A(\mathbf{v}_k).$$

Since $\phi(y)$ is asymptotically constant, $\lim_{k\to\infty}\phi(\mathbf{x}+\mathbf{v}_k)A(\mathbf{x}) = c\cdot A(\mathbf{x})$ for some constant c (uniformly on $B(0,R)$). Since $(\phi(\mathbf{x}+\mathbf{v}_k)-\phi(\mathbf{v}_k)) = O(|\mathbf{v}_k|^{-1})$ (as $k\to\infty$), the sequence of vectors

$$(\phi(\mathbf{x}+\mathbf{v}_k)-\phi(\mathbf{x}_k))A(\mathbf{v}_k)$$

is uniformly bounded for $x\in B(0,R)$. Furthermore, for every pair of indices $1\leqslant i,j\leqslant n-1$,

$$\frac{\partial^2}{\partial x_i\partial x_j}\left(\phi(\mathbf{x}+\mathbf{v}_k)-\phi(\mathbf{v}_k)\right)A(\mathbf{v}_k) = \frac{\partial^2}{\partial x_i\partial x_j}\phi(\mathbf{x}+\mathbf{v}_k)\cdot A(\mathbf{v}_k) =$$
$$O(|\mathbf{v}_k|^{-2})A(\mathbf{v}_k) = O(|\mathbf{v}_k|^{-1}).$$

Therefore, the Hessians of $h_k\big|_B$ uniformly converge to zero as $k\to\infty$. $\square$

We would like to strengthen the assertion that ϕ is not constant, unless A is a similarity, which we assume is not the case. Let Φ be a group of Euclidean isometries acting cocompactly on $\mathbb{E}^{n-1}$. Fix a point $\mathbf{v}\in\mathbb{E}^{n-1}$. We say that a function ψ defined on a subset E of $\mathbb{R}^n$ is linear if extends from E to a linear function on $\mathbb{R}^n$.

LEMMA 22.27. *Suppose that the linear transformation A is not a similarity. Then there exists a number R and a sequence of points $\mathbf{v}_k\in\Phi\cdot\mathbf{v}$ diverging to infinity, such that the restrictions of h to $B(\mathbf{v}_k,R)\cap\Phi\cdot\mathbf{v}$ are nonlinear for all k.*

PROOF. Let R be such that

$$\bigcup_{g\in\Phi} B(gv,R) = \mathbb{E}^{n-1}$$

and that the intersection $B(\mathbf{v},R)\cap\Phi\cdot\mathbf{v}$ contains a subset $\{\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}\}$ such that the vectors

$$\mathbf{u}_i = \mathbf{v}_i - \mathbf{v}, i = 1,\ldots,n-1,$$

span $\mathbb{R}^{n-1}$.

Suppose that the sequence $\mathbf{v}_k$ as required does not exist. This means that there exists $r < \infty$ such that the restriction of h to $B(\mathbf{v}_k, 4R) \cap \Phi\mathbf{v}$ is linear for each $\mathbf{v}_k \in \Phi\mathbf{v} \setminus B(\mathbf{v}, r)$. In particular, if $|\mathbf{v}_i - \mathbf{v}_j| \leqslant R$ and

$$|\mathbf{v}_i - \mathbf{v}| > 4R + r, \quad |\mathbf{v}_i - \mathbf{v}| > 4R + r,$$

then $B(\mathbf{v}_i, R)$ is contained in $B(\mathbf{v}_j, 4R) \cap \Phi\mathbf{v}$ and vice-versa. In view of the 'spanning' assumption on the number R, it follows that

$$\phi\big|_{B(\mathbf{v}_i,4R)} = \phi\big|_{B(\mathbf{v}_j,4R)}.$$

Define the set $Y = \{\mathbf{v}_k : |\mathbf{v}_k - \mathbf{v}| > 4R + r\}$. The collection $\mathcal{U}$ of the balls

$$B(\mathbf{v}_k, R), \quad \mathbf{v}_k \in Y,$$

is an open covering of the complement $\mathbb{E}^{n-1} \setminus B(\mathbf{v}, 4R+r)$. The latter is connected since $\mathbb{E}^{n-1}$ has dimension $\geqslant 2$. It follows that the nerve of $\mathcal{U}$ is also connected. Since whenever $B(\mathbf{v}_i, R) \cap B(\mathbf{v}_j, R)$ is nonempty, h is linear on

$$\bar{B}(\mathbf{v}_i, 4R) \cap \bar{B}(\mathbf{v}_j, 4R) \cap \Phi \cdot \mathbf{v},$$

we conclude that h is linear on the union

$$\bigcup_{\mathbf{v}_k \in Y} B(\mathbf{v}_k, R) \cap \Phi \cdot \mathbf{v}.$$

Hence, ϕ is constant on the set $\Phi \cdot \mathbf{v} \setminus B(0, 4R + r)$.

According to the Bieberbach Theorem (see e.g. [**Rat06**, Theorem 7.5.2]), the group Φ contains a free abelian subgroup of rank $n - 1$ acting on $\mathbb{E}^{n-1}$ by translations. Up to an affine conjugation, this subgroup is $\mathbb{Z}^{n-1} \subset \mathbb{R}^{n-1}$. The projection of $\mathbb{Z}^{n-1}$ to $\mathbb{R}P^{n-1}$ is dense in the latter. Therefore, the set

$$\left\{ \frac{\mathbf{y}}{|\mathbf{y}|} : \mathbf{y} \in Y \right\}$$

is also dense in the unit sphere. Since $\phi(\mathbf{u}/|\mathbf{u}|) = \phi(\mathbf{u})$ for all nonzero vectors $\mathbf{u} \in \mathbb{R}^{n-1}$, it follows that the function $\phi : \mathbb{R}^n \to \mathbb{R}$ is constant. This is a contradiction. □

22.6. Scattering

We now return to the discussion of quasiisometries. We continue with the notation of §22.4. In particular, we have a linear transformation (that is not a similarity) $A \in GL(n - 1, \mathbb{R})$, $A(\Lambda) = \Lambda'$, where $\Lambda, \Lambda' \subset \mathbb{E}^{n-1}$ are the sets of centers of peripheral horospheres of truncated hyperbolic spaces Ω, Ω'. Moreover, after composing A with hyperbolic isometries, we can assume the origin in $\mathbb{R}^n$ belongs to $\Lambda \cap \Lambda'$. (Recall that earlier, in §22.4 we were carefully choosing the point of differentiability, the origin, not to be in Λ.) Lastly, A extends to a quasiisometry $\tilde{f}$ of $\mathbb{H}^n$, which restricts to a quasiisometry $\Omega \to \Omega'$.

Let $J : \mathbb{E}^{n-1} \cup \{\infty\} \to \mathbb{E}^{n-1} \cup \{\infty\}$ be the inversion in the unit sphere $\mathbb{S}^{n-1}$. Then $\infty = J(0)$ belongs to both $J(\Lambda)$ and $J(\Lambda')$. The conjugate quasiisometry

$$J \circ \tilde{f} \circ J : \mathbb{H}^n \to \mathbb{H}^n$$

sends $J(\Omega)$ to $J(\Omega')$ and $J \circ A \circ J$ is the boundary extension of this quasiisometry. Since ∞ now belongs to $J(\Lambda) \cap J(\Lambda')$, we have two horoballs B_∞, B'_∞ (with centers at ∞) in the complements of $J(\Omega), J(\Omega')$. The latter are the truncated hyperbolic spaces of the lattices $J\Gamma J, J\Gamma' J$ respectively.

In order to simplify the notation, we now set

$$\Gamma := J\Gamma J, \quad \Gamma' := J\Gamma' J, \quad \Omega := J(\Omega), \quad \Omega' := J(\Omega'), \quad \Lambda := J(\Lambda), \quad \Lambda' := J(\Lambda'),$$

use h for the inverted linear map $J \circ A \circ J$ and $\tilde{h}$ for its quasiisometric extension to $\mathbb{H}^n$ that sends Ω to Ω'. Further, let Γ_∞, Γ'_∞ denote the stabilizers of ∞ in Γ, Γ' respectively. Then the groups Γ_∞, Γ'_∞ act cocompactly on the boundaries of the horoballs B_∞, B'_∞, since the quotients $\Omega/\Gamma, \Omega'/\Gamma'$ are both compact. Therefore, the groups Γ_∞, Γ'_∞ also act cocompactly on the Euclidean space $\mathbb{E}^{n-1}$.

Remark 22.28. We realize that all this is very inconsistent with the notation from §22.4, but, we no longer need the notation used there.

Lastly, given a point $\mathbf{x} \in \mathbb{E}^{n-1}$, we define the subset $h_*(\mathbf{x}) := h(\Gamma_\infty \mathbf{x}) \subset \mathbb{E}^{n-1}$.

Lemma 22.29 (Scattering lemma). *Suppose that the nonlinear factor $\phi = \phi_h$ of h is nonconstant. Then for each $\mathbf{x} \in \mathbb{E}^{n-1}$, the set $h_*(\mathbf{x})$ is not contained in the union of finitely many Γ'_∞–orbits.*

Proof. Suppose that $h_*(\mathbf{x})$ is contained in the union of finitely many Γ'_∞–orbits, then. Each Γ'_∞–orbit is a discrete subset of $\mathbb{E}^{n-1}$; the same is true for a finite union of such orbits. Therefore, for every Euclidean ball $B(\mathbf{x}, R) \subset \mathbb{E}^{n-1}$, the intersection

$$(\Gamma'_\infty \cdot h_*(\mathbf{x})) \cap B(\mathbf{x}, R)$$

is finite. We will show that this cannot be the case.

We apply Lemma 22.27 to the discrete group $\Phi := \Gamma_\infty$ and the point $\mathbf{x} = \mathbf{v}$. The lemma gives us a positive number R and an infinite sequence $\mathbf{x}_k$, $\mathbf{x}_k = \gamma_k(\mathbf{x})$, $\gamma_k \in \Phi$, satisfying

$$\lim_{k \to \infty} |\mathbf{x}_k| = \infty.$$

Since the group Γ'_∞ also acts cocompactly on $\mathbb{E}^{n-1}$, there exists a sequence $\gamma'_k \in \Gamma'_\infty$, such that the set

$$\{\gamma'_k h(\mathbf{x}_k), k \in \mathbb{N}\}$$

is relatively compact in $\mathbb{E}^{n-1}$.

According to Lemma 22.27, the restriction

$$h\big|_{B(\mathbf{x}_k, R) \cap \Phi \mathbf{x}}$$

is nonlinear for each k. Therefore, the maps

$$h_k := \gamma'_k \circ h \circ \gamma_k$$

cannot be affine on $B(\mathbf{x}, R) \cap \Phi\mathbf{x}$. On the other hand, Corollary 22.26 implies that the sequence of maps

$$h_k\big|_{B(\mathbf{x}, R)}$$

subconverges to an affine mapping h_∞. Since each h_k is not affine on $B(\mathbf{x}, R) \cap \Phi\mathbf{x}$, this subconvergence cannot be eventually constant. In other words, there exists $\mathbf{y} \in B(\mathbf{x}, R) \cap \Phi\mathbf{x}$ such that the set $\{h_k(\mathbf{y}) : k \in \mathbb{N}\}$ is infinite. We conclude that the union

$$\bigcup_{k \in \mathbb{N}} h_k \left(\Phi\mathbf{x} \cap B(\mathbf{x}, R)\right) \subset (\Gamma'_\infty \cdot h_*(\mathbf{x})) \cap B(\mathbf{x}, R)$$

is an infinite set. This is a contradiction. The lemma follows. □

THEOREM 22.30. *Suppose that h is an inverted linear map that admits a quasi-isometric extension $\tilde{h} : \mathbb{H}^n \to \mathbb{H}^n$ sending Ω to Ω'. Then ϕ_h is constant, i.e., h is a similarity map.*

PROOF. Let $\mathbf{x}$ be the center of a complementary horoball B of Ω, $B \neq B_\infty$. Suppose that ϕ_h is nonconstant.

According to the scattering lemma, $h_*(\mathbf{x})$ is *not contained* in a finite union of Γ'_∞–orbits. Let $\gamma_k \in \Gamma_\infty$ be a sequence such that the Γ'_∞–orbits of the points $h\gamma_k(\mathbf{x})$ are all distinct. Since Γ'_∞ acts on $\mathbb{E}^{n-1}$ cocompactly, there exists an infinite sequence (k_i) and elements

$$\gamma'_{k_i} \in \Gamma'_\infty,$$

such that the sequence

$$\mathbf{x}'_{k_i} := \gamma'_{k_i} h \gamma_{k_i}(\mathbf{x})$$

converges to a point $\mathbf{x}' \in \mathbb{E}^{n-1}$. According to our assumption, all the points $\mathbf{x}'_{k_i}$ are distinct. Let B'_{k_i} denote the complementary horoball to Ω' whose center is $\mathbf{x}'_{k_i}$. All these horoballs are distinct since their centers are. As the horoballs B'_{k_i} are also pairwise disjoint, we obtain

$$\lim_{i\to\infty} \operatorname{diam}_{\mathbb{E}^n}(B'_{k_i}) = 0.$$

Let B_k be the complementary horoball to Ω whose center is $\gamma_k \mathbf{x}$. Then

$$D := \operatorname{dist}(B_k, B_\infty) = \operatorname{dist}(B_1, B_\infty) = -\log(\operatorname{diam}_{\mathbb{E}^n}(B_1)).$$

At the same time,

$$\lim_{i\to\infty} \operatorname{dist}(B'_{k_i}, B'_\infty) = -\lim_{i\to\infty} \log(\operatorname{diam}(B'_{k_i})) = \infty.$$

Recall that we are assuming that there exists an (L, A) quasiisometric extension $\tilde{h}$ of h such that $\tilde{h} : \Omega \to \Omega'$. According to Lemma 22.12,

$$\operatorname{dist}(B'_j, B'_\infty) \leqslant R(L, A) + LD + A.$$

This is a contradiction. ☐

By combining all these results, we conclude:

THEOREM 22.31. *Suppose that $f : \Omega \to \Omega'$ is a quasiisometry of truncated hyperbolic spaces, $n \geqslant 3$. Then the extension of f to $\partial_\infty \mathbb{H}^n$ is a Moebius transformation.*

22.7. Schwartz Rigidity Theorem

Before proving Theorem 22.1 we will need two technical assertions concerning isometries of $\mathbb{H}^n$ which "almost preserve" truncated hyperbolic spaces.

Let Ω be the truncated hyperbolic space of a nonuniform lattice $\Gamma < G = \operatorname{Isom}(\mathbb{H}^n)$. We will say that a subset $A \subset G$ *almost preserves* Ω if there exists $C < \infty$ such that

$$\operatorname{dist}_{Haus}(\Omega, \alpha\Omega) \leqslant C, \forall \alpha \in A.$$

Note that each $\alpha \in A$ determines an (L, A)-quasiisometry $\Omega \to \Omega$, defined by composing α with the nearest-point projection $\pi_\Omega : \alpha(\Omega) \to \Omega$, see Exercise 5.12, Part 1.

LEMMA 22.32. *Suppose that $\beta_k \in G$ is a sequence almost preserving Ω and $\lim_k \beta_k = \beta \in G$. Then the sequence (β_k) consists of finitely many elements of G.*

Proof. Assume to the contrary that the sequence (β_k) consists of distinct elements. The lattice Γ cannot preserve a proper round sphere in $\partial_\infty \mathbb{H}^n$. Therefore, there exists a finite subset $\Lambda_o \subset \Lambda$ not contained in a proper round sphere in $\mathbb{S}^{n-1}$. It follows that (after passing to a subsequence, which we will suppress), there exists a peripheral horosphere Σ centered at a point $\xi \in \Lambda_o$, such that all elements of the sequence $\xi_k := \beta_k(\xi)$ are distinct. Since $\lim_{k\to\infty} \beta_k = \beta$, the horospheres $\beta_k(\Sigma)$ converge to the horosphere $\beta(\Sigma)$. For each k we have a unique horosphere $\widehat{\Sigma}_k \subset \partial\Omega$ whose Hausdorff distance from $\beta_k(\Sigma)$ is uniformly bounded. It follows that the horospheres $\widehat{\Sigma}_k$ have to have nonempty intersections for all large k. This forces the equality

$$\widehat{\Sigma}_k = \widehat{\Sigma}_{k+1}, \forall k \geqslant k_o.$$

Hence, the centers ξ_k of these horospheres are also equal, which is a contradiction. □

Proposition 22.33. *Let Γ, Γ' be nonuniform lattices in $G = \mathrm{Isom}(\mathbb{H}^n)$ such that Γ' almost preserves Ω, the truncated hyperbolic space of Γ. Then the groups Γ, Γ' are commensurable.*

Proof. Suppose the assertion fails. Then the projection of Γ' to $\Gamma \setminus G$ is infinite. Therefore, there exists an infinite sequence (ψ_k) of elements of Γ' whose projections to $\Gamma \setminus G$ are all distinct. Since $\partial\Omega/\Gamma$ is compact, there are only finitely many Γ-orbits of peripheral horospheres of Ω. The set Λ/Γ, consisting of the Γ-orbits of their centers, is also finite. Therefore, after passing to a subsequence in (ψ_k), we can assume that for some horosphere $\Sigma \subset \partial\Omega$, for every k, the centers of all the horospheres $\psi_k(\Sigma)$ lie in the same Γ-orbit. In other words, there are elements $\gamma_k \in \Gamma$ such that every $\alpha_k := \gamma_k\psi_k$ fixes the center ξ of Σ.

Since all γ_k's preserve Ω and all ψ_k's almost preserve Ω, the infinite set

$$A = \{\alpha_k : k \in \mathbb{N}\} \subset G$$

also almost preserves Ω. Projections of all ψ_k's to $\Gamma \setminus G$ are all distinct, thus, A projects injectively into $\Gamma \setminus G$.

Without loss of generality, we may assume that $\xi = \infty$ in the upper half-space model of $\mathbb{H}^n$ and Σ is given by the equation $\{x_n = 1\}$. Then the elements of A are Euclidean similarities (they all fix the point ξ). Since the stabilizer Γ_∞ of ∞ in Γ acts cocompactly on the Euclidean space $\mathbb{E}^{n-1}$, there exists a constant C' and a sequence $\tau_k \in \Gamma_\infty$ such that the compositions $\alpha'_k := \tau_k\alpha_k$ satisfy,

$$|\beta_k(0)| \leqslant C'.$$

Set $A' := \{\alpha'_k : k \in \mathbb{N}\}$. As before, the subset $A' \subset G$ is infinite, almost preserves Ω and projects injectively to $\Gamma \setminus G$. Since every $\alpha' \in A'$ determines a uniform quasiisometry $\Omega \to \Omega$, there exists $C < \infty$ such that for every $\alpha' \in A'$,

$$\mathrm{dist}_{Haus}(\Sigma, \alpha'\Sigma) \leqslant C.$$

(This is a special case of Proposition 22.9. Cf. Exercise 22.8.) In other words, the value of the coordinate x_n on $\alpha'\Sigma$ satisfies the inequality

$$e^{-C} \leqslant x_n \leqslant e^C.$$

Thus, the subset A' is contained in the compact set of similarities

$$\{\beta : \mathbf{x} \mapsto tU\mathbf{x} + \mathbf{v} : \quad e^{-C} \leqslant t \leqslant e^C, U \in O(n-1), |\mathbf{v}| \leqslant C'\}.$$

Therefore, the set A' is infinite and has compact closure in G. This contradicts Lemma 22.32. $\square$

Proof of Theorem 22.1. Suppose that $\Gamma < G = \mathrm{Isom}(\mathbb{H}^n), n \geqslant 3$, is a nonuniform lattice.

(a) For each (L, A)–quasiisometry $f : \Gamma \to \Gamma$, there exists $\alpha \in Comm_G(\Gamma)$, satisfying

$$\mathrm{dist}(f, \alpha) < \infty.$$

PROOF. The quasiisometry f extends to a quasiisometry of the hyperbolic space $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ (Theorem 22.14). This quasiisometry extends to a quasiconformal mapping $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$. The quasiconformal mapping h has to be Moebius according to Theorem 22.31. Therefore, $\tilde{f}$ is within a finite distance from an isometry α of $\mathbb{H}^n$ (which is the unique isometric extension of h to $\mathbb{H}^n$), see Lemma 9.112.

EXERCISE 22.34. Verify that $\mathrm{dist}(\tilde{f}, \alpha)$ depends only on Γ and quasiisometry constants L, A.

It remains to show that α belongs to $Comm_G(\Gamma)$. We note that f (and, hence, α) almost preserves Ω, the truncated hyperbolic space of Γ. Since Γ preserves Ω, the entire group

$$\Gamma' := \alpha \Gamma \alpha^{-1}$$

almost preserves Ω. By Proposition 22.33, the groups Γ, Γ' are commensurable. Thus, α belongs to the commensurator $Comm_G(\Gamma)$. $\square$

(b) Suppose that $\Gamma, \Gamma' < G = \mathrm{Isom}(\mathbb{H}^n)$ are nonuniform lattices quasiisometric to each other. Then there exists $\alpha \in \mathrm{Isom}(\mathbb{H}^\mathrm{n})$ such that the groups Γ and $\alpha \Gamma' \alpha^{-1}$ are commensurable.

PROOF. The proof is analogous to (a): The quasiisometry $f : \Omega' \to \Omega$ of the truncated hyperbolic spaces of the lattices Γ', Γ is within a finite distance from an isometry α. The group $\Gamma'' := \alpha \Gamma' \alpha^{-1}$ again almost preserves Ω. By Proposition 22.33, the groups Γ, Γ'' are commensurable. $\square$

(c) Suppose that Γ' is a finitely generated group quasiisometric to a nonuniform lattice $\Gamma < \mathrm{Isom}(\mathbb{H}^\mathrm{n})$. Then the groups Γ, Γ' are virtually isomorphic; more precisely, there exists a finite normal subgroup $K \subset \Gamma'$ such that the groups $\Gamma, \Gamma'/K$ contain isomorphic subgroups of finite index.

PROOF. Let $f : \Gamma \to \Gamma'$ be a quasiisometry and let $f' : \Gamma' \to \Gamma$ be its quasiinverse. Then, by Lemma 5.61, we have a quasiaction $\Gamma' \curvearrowright \Omega$ via

$$\gamma' \mapsto \rho(\gamma') := f' \circ \gamma' \circ f \in QI(\Omega).$$

According to Part (a), each quasiisometry $g = \rho(\gamma')$ is within a (uniformly) bounded distance from a quasiisometry of Ω induced by an element g^* of $Comm_G(\Gamma)$. We obtain a map

$$\psi : \gamma' \mapsto \rho(\gamma') = g \mapsto g^* \in Comm_G(\Gamma).$$

We claim that this map is a homomorphism with finite kernel. For each quasiisometries $h : \mathbb{H}^n \to \mathbb{H}^n$ we let h_∞ denote its extension to $\partial_\infty \mathbb{H}^n$. Then, since ρ is a quasiaction, ψ induces a homomorphism

$$\psi_\infty : \gamma' \mapsto g_\infty = g^*_\infty, \quad \psi_\infty : \Gamma' \to Comm_G(\Gamma),$$

see Theorem 9.135. Since the quasiaction $\rho : \Gamma' \curvearrowright \Omega'$ is geometric (see Lemma 5.61), by Lemma 9.117 the kernel K of the quasiaction $\Gamma' \curvearrowright \Omega'$ is quasifinite. The subgroup $K \lhd \Gamma'$ is also the kernel of the homomorphism ψ_∞; by Lemma 9.112, the subgroup K is finite.

The rest of the proof is the same as for (a) and (b): The group $\Gamma'' := \psi(\Gamma')$ almost preserves Ω, hence, it is commensurable to Γ. $\square$

CHAPTER 23

A survey of quasiisometric rigidity

In this chapter we review results and open problems on quasiisometric rigidity of groups and metric spaces. We refer the reader to §5.6 for the basic terminology that we will be using. Our survey covers three types of problems within the theme of quasiisometric rigidity:

(1) **The description of the group of quasiisometries $QI(X)$ of specific metric spaces X and $QI(G)$ for finitely generated groups G.** For instance, in some cases, $QI(X)$ coincides with the subgroup of isometries of X, or with the subgroup of virtual automorphisms of G, or with the commensurator of G, either abstract or considered in a larger group.

(2) **The identification of the classes of groups $\mathcal{G}$ that are QI rigid.** This problem was formulated for the first time (with slightly different terminology) by M. Gromov in [**Gro83**]. It is sometimes related to the first problem. Indeed, if a group G' is quasiisometric to G, then there exists a homomorphism $G' \to QI(G)$, which, in many cases, has finite kernel (see Lemma 5.62). If $QI(G)$ is either very close to G, or very close to an ambient group in which all groups in the class $\mathcal{G}$ lie, then one is halfway through a proof of QI rigidity of the class $\mathcal{G}$.

(3) **The quasiisometric classification within a given class of groups.** This can be achieved either by a complete description of the equivalence classes or by using QI invariants. An extreme case is when the QI class of a group G contains only finite index subgroups of G, their quotients by finite normal subgroups and finite extensions of these quotients, i.e., the group G is QI rigid.

23.1. Rigidity of symmetric spaces, lattices and hyperbolic groups

23.1.1. Uniform lattices. The oldest QI rigidity theorem in the context of symmetric spaces was proven by P. Pansu:

<u>Theorem</u> 23.1 (P. Pansu, [**Pan89**]). *Let X be a quaternionic hyperbolic space $\mathbf{H}\mathbb{H}^n$ $(n \geqslant 2)$ or the octonionic hyperbolic plane $\mathbf{O}\mathbb{H}^2$. Then X is strongly QI rigid.*

Even though real and complex hyperbolic spaces are not strongly QI rigid, nevertheless, the classes of uniform lattices in their isometry groups are QI rigid. We saw that the class of uniform lattices in the group $PO(n, 1)$ is QI rigid (see Chapter 21), with rigidity results primarily due to P. Tukia, D. Gabai, A. Casson and D. Jungreis. An analogous QI rigidity theorem was proven by R. Chow [**Cho96**] for complex-hyperbolic spaces $\mathbf{C}\mathbb{H}^n, n \geqslant 2$, by methods similar to the proof of Tukia's theorem. We summarize these results as follows:

Theorem 23.2. *Let X be a symmetric space of negative curvature. Then the class of uniform lattices in X is QI rigid.*

For higher rank symmetric spaces, strong QI rigidity follows from a series of results of B. Kleiner and B. Leeb, which were independently (although, a bit later) obtained by A. Eskin and B. Farb in [**EF97b**].

Theorem 23.3 (B. Kleiner, B. Leeb, [**KL98b**]). *Let X be a symmetric space of of noncompact type, without rank one de Rham factors. Assume that if two irreducible factors of X are homothetic to each other then they are isometric. Then X is strongly QI rigid.*

As an application of this rigidity theorem, Kleiner and Leeb, as well as Eskin and Farb, obtained:

Theorem 23.4 (B. Kleiner, B. Leeb, [**KL98b**]). *Let X be a symmetric space of noncompact type, without rank one de Rham factors. Then the class of uniform lattices in* $\mathrm{Isom}(X)$ *is QI rigid.*

Furthermore, Kleiner and Leeb proved that even if the de Rham decomposition of X

$$X = \prod_{i=1}^{n} X_i$$

does have rank one de Rham factors, the associated quasi-action of Γ on X is within finite distance from another quasi-action, which preserves the de Rham factors X_i (except, it might permute them).

Kleiner and Leeb also established strong QI rigidity for Euclidean buildings:

Theorem 23.5 (B. Kleiner, B. Leeb, [**KL98b**]). *Let X be a Euclidean building such that each de Rham factor of X is a Euclidean building of rank $\geqslant 2$. Assume that if two irreducible factors of X are homothetic to each other then they are isometric. Then X is strongly QI rigid.*

The overall QI rigidity result for uniform lattices reads as follows:

Theorem 23.6. *Suppose that X is a symmetric space of noncompact type. Then the class of uniform lattices in* $\mathrm{Isom}(X)$ *is rigid.*

23.1.2. Nonuniform lattices. Turning to nonuniform lattices, one should first note that Theorem 22.1 of R. Schwartz (see Chapter 22) in its most general form holds even when the space $\mathbb{H}^n, n \geqslant 3$, is replaced by an arbitrary negatively curved symmetric space of dimension > 2. This theorem answers the three types of problems described in the beginning of this chapter, and can be stated as follows.

Theorem 23.7 (R. Schwartz [**Sch96a**]). *Suppose that X is a negatively curved symmetric space of dimension > 2; we let G denote the isometry group of X. Then:*

(1) *Each nonuniform lattice in $\Gamma < G$ is strongly QI rigid: The natural homomorphism $Comm_G(\Gamma) \to QI(\Gamma)$ is an isomorphism.*
(2) *The class of nonuniform lattices in G is QI rigid.*
(3) *If Γ and Γ' are quasiisometric nonuniform lattices of isometry groups of negatively curved symmetric spaces X and respectively X', then $X \cong X'$ (up to rescaling the metric) and the lattices Γ and Γ' are commensurable in G.*

A nonuniform lattice of $\mathbb{H}^2$ contains a finite index subgroup which is free nonabelian. In this case we may therefore apply QI rigidity of virtually free groups (Theorem 18.44) and conclude:

THEOREM 23.8. *Each nonuniform lattice in* $\mathrm{Isom}(\mathbb{H}^2)$ *is QI rigid.*

In the special case when X is the 3-dimensional hyperbolic space, the group of orientation-preserving isometries of X, $\mathrm{Isom}_+(X)$, is isomorphic to $PSL(2, \mathbb{C})$. Schwartz's result has the following arithmetic version. Let F_1 and F_2 be imaginary quadratic extensions of $\mathbb{Q}$ and let $\mathcal{O}_1$ and $\mathcal{O}_2$ be their respective rings of integers. Then the arithmetic lattices $PSL(2, \mathcal{O}_1)$ and $PSL(2, \mathcal{O}_2)$ (Bianchi groups) are commensurable (quasiisometric) if and only if the fields F_1 and F_2 are isomorphic.

When instead of imaginary quadratic extensions, one takes totally real quadratic extensions, the corresponding groups $PSL(2, \mathcal{O}_i)$ become nonuniform $\mathbb{Q}$-rank one lattices in $\mathrm{Isom}(\mathbb{H}^2 \times \mathbb{H}^2)$, a rank two semisimple Lie group. In general, when F_i's are algebraic extensions of $\mathbb{Q}$, the groups $PSL(2, \mathcal{O}_i)$ are isomorphic to nonuniform $\mathbb{Q}$-rank one lattices in $\mathrm{Isom}(X)$, where the symmetric space X is isometric to a product of several copies of $\mathbb{H}^2$ and $\mathbb{H}^3$. The QI rigidity theorem in this context was first proven by B. Farb and R. Schwartz **[FS96]** in the case of fields F_i of degree 2 and by R. Schwartz **[Sch96a]** in full generality:

THEOREM 23.9.
(1) *Let F be an algebraic extension of $\mathbb{Q}$, let $\mathcal{O}$ be the ring of integers of F, let $\Gamma = PSL(2, \mathcal{O})$ and let X be the product of hyperbolic spaces on which Γ acts as a lattice. Then the group Γ is strongly QI rigid.*
(2) *Let F_1, F_2 be two algebraic extensions of $\mathbb{Q}$, and let $\mathcal{O}_i$ be their corresponding rings if integers. Then the lattices $PSL(2, \mathcal{O}_1)$ and $PSL(2, \mathcal{O}_2)$ are quasiisometric if and only if the fields F_1 and F_2 are isomorphic.*

It is known that every irreducible lattice Γ in a semisimple group of $\mathbb{R}$-rank at least 2 and at least one factor of $\mathbb{R}$-rank one, is an arithmetic $\mathbb{Q}$-rank one lattice **[Pra73**, Lemma 1.1]. However, such lattices Γ are, in general, quite different from the arithmetic group $PSL(2, \mathcal{O})$.

The case of *higher $\mathbb{Q}$-rank groups* was settled by A. Eskin:

THEOREM 23.10 (A. Eskin **[Esk98]**). *Let X, X_1, X_2 be a symmetric space of noncompact type with all de Rham factors of rank at least 2. Then:*
(1) *Every nonuniform lattice $\Gamma < \mathrm{Isom}(X)$ is strongly QI rigid.*
(2) *The class of nonuniform irreducible lattices in $\mathrm{Isom}(X)$ is QI rigid.*
(3) *If Γ_1 and Γ_2 are quasiisometric nonuniform irreducible lattices in isometry groups of symmetric spaces X_1 and X_2, then X_1 is isometric to X' (up to rescaling the metrics on de Rham factors). Moreover, the lattices Γ_1 and Γ_2' are commensurable in $\mathrm{Isom}(X)$.*

As an application of various QI rigidity results for nonuniform lattices B. Farb proves in **[Far97]** the following rigidity result:

THEOREM 23.11. *The class of irreducible nonuniform lattices in each connected semisimple Lie group G is QI rigid.*

We refer the reader to **[Far97]** for more details.

The class of groups which are left out from this classification include, for instance, the group
$$PSL(2, \mathbb{Z}[\sqrt{-1}]) \times SL(3, \mathbb{Z}).$$

Problem 23.12 (I. Belegradek). Prove QI rigidity of the class of non-uniform, possibly reducible, lattices, for a general symmetric space X of noncompact type.

The QI rigidity results for nonuniform lattices in semisimple Lie groups were extended by K. Wortman [**Wor07**] to *S-arithmetic lattices*, which are lattices in products of some semisimple Lie groups and some p-adic Lie groups.

Comparison of QI rigidity properties of uniform and nonuniform lattices. Assume, for simplicity, that X is an irreducible symmetric space of nonpositive curvature, not isometric to the hyperbolic plane. The results we described above, show that two nonuniform lattices in $G = \mathrm{Isom}(X)$ are commensurable if and only if they are quasiisometric. This fails in the case of uniform lattices, as all such lattices are quasiisometric to each other and it is known that G contains infinitely many virtual isomorphism classes of uniform arithmetic lattices. Restricting to non-arithmetic lattices in $PO(n, 1)$ still leads to infinitely many QI equivalence classes. (We refer the reader to §10.3 for the references.)

23.1.3. Symmetric spaces with Euclidean de Rham factors and Lie groups with nilpotent normal subgroups. So far, we considered only nonpositively symmetric spaces X of *noncompact type*. The naive QI rigidity fails for uniform lattices in isometry groups of nonpositively curved symmetric spaces which are not of noncompact type:

Theorem 23.13. *Suppose that $G = PO(n, 1) \times \mathbb{R}, G = PU(n, 1) \times \mathbb{R}$ or $G = SO(n, 2) \times \mathbb{R}$. Then there are uniform lattices $\Lambda < G = \mathrm{Isom}(X)$, quasiisometric to groups which are not VI to lattices in G. In other words, the class of uniform lattices in G is not QI rigid.*

Proof. Each of the Lie groups in the theorem has the form $G = G_1 \times \mathbb{R}$. According to Corollary 10.33, there exists a lattice $\Gamma < G_1$ which admits a central coextension

$$1 \to \mathbb{Z} \to \tilde{\Gamma} \to \Gamma \to 1$$

such that $\tilde{\Gamma}$ is not VI to any product group $\mathbb{Z} \times \Gamma'$ and, at the same time, $\tilde{\Gamma}$ is QI to the product lattice

$$\Lambda = \Gamma \times \mathbb{Z} < G_1 \times \mathbb{R}.$$

Furthermore, every lattice $\Lambda' < G_1 \times \mathbb{R}$ is isomorphic to the direct product $\Gamma' \times \mathbb{Z}$, where Γ' is the projection of Λ' to G_1. □

On the other hand, the following theorem shows that the central coextension construction is the only source of failure of QI rigidity for uniform lattices in symmetric spaces of nonpositive curvature.

Theorem 23.14 (B. Kleiner, B. Leeb, [**KL01**]). *Suppose that G is a connected Lie group, which fits into a short exact sequence*

$$1 \to N \to G \to \bar{G} \to 1,$$

where the group N is connected and nilpotent, the group $\bar{G}$ is semisimple and acts via the trivial representation on N. Equip G with a left-invariant Riemannian metric and the associates Riemannian distance function. Then every finitely generated group Γ quasiisometric to G fits into a short exact sequence

$$1 \to K \to \Gamma \to \bar{\Gamma} \to 1,$$

where K is quasiisometric to N and $\bar{\Gamma}$ is virtually isomorphic to a uniform lattice in $\bar{\Gamma}$.

An example of the situation covered by this theorem is a symmetric space $X = Y \times \mathbb{E}^k$, where Y is a symmetric space of noncompact type. The group $G = \mathrm{Isom}_o(Y) \times \mathbb{R}^k = \bar{G} \times N$ acts transitively and isometrically on X.

COROLLARY 23.15. *Each finitely generated group Γ quasiisometric to X fits into a short exact sequence*

$$1 \to K \to \Gamma \to \bar{\Gamma} \to 1,$$

where the group K is virtually abelian and $\bar{\Gamma}$ acts as a cocompact lattice on Y.

These results leave out the case of general connected Lie groups G; these groups admit the Levi-Mal'cev decomposition

$$1 \to S \to G \to \bar{G} \to 1,$$

where S is a solvable Lie group and $\bar{G}$ s semisimple.

The following problem is known to be quite difficult. Note, however, that it bypasses the notoriously difficult problem of QI rigidity for polycyclic groups.

PROBLEM 23.16. Prove an analogue of Theorem 23.14 for all connected Lie groups G.

23.1.4. QI rigidity for hyperbolic spaces and groups. We turn now to QI rigidity in the context of Gromov-hyperbolic spaces and groups. As we saw before, proofs of QI rigidity theorems for lattices in rank one Lie groups, to large degree, are based on a well-developed theory of quasiconformal mappings of the ideal boundaries of rank one symmetric spaces. Such theory was developed by M. Bourdon and H. Pajot in [**BP00**], and extended further by X. Xie [**Xie06**], for *2-dimensional hyperbolic buildings* and resulted in strong QI rigidity these buildings. Instead of giving precise definitions of hyperbolic buildings, we note here only that n-dimensional hyperbolic buildings are certain CAT(-1) spaces X, covered by isometric copies of $\mathbb{H}^n$ (called "apartments"), which, in turn, are tiled by some compact convex hyperbolic polyhedra, called *fundamental domains.*

<u>THEOREM</u> 23.17 (M. Bourdon, H. Pajot [**BP00**]). *Suppose that X is a thick 2-dimensional hyperbolic building, whose links are complete bipartite graphs and whose fundamental domains are right-angled fundamental polygons. Then X is strongly QI rigid.*

Some of the restrictions in this theorem were removed later on by X. Xie, using techniques similar to the ones of Bourdon and Pajot:

<u>THEOREM</u> 23.18 (X. Xie [**Xie06**]). *Each thick 2-dimensional hyperbolic building is strongly QI rigid.*

COROLLARY 23.19. *The class $\mathcal{C}_X$ of groups acting geometrically on the thick 2-dimensional hyperbolic building X, is QI rigid.*

Note that there are many examples of hyperbolic groups acting geometrically on thick 2-dimensional hyperbolic buildings. However, such groups need not be commensurable to each other, which leads to:

Problem 23.20. Construct examples of QI rigid hyperbolic groups whose boundaries are homeomorphic to the Menger curve. (The ideal boundary of each thick 2-dimensional hyperbolic building is a Menger curve.)

The restriction to 2-dimensional hyperbolic buildings also appears to be unnatural, since higher-dimensional hyperbolic spaces tend to be more rigid than low-dimensional ones.

Conjecture 23.21. *Each thick hyperbolic building is strongly QI rigid.*

The QI rigidity problem is wide-open for Kleinian groups, i.e., discrete groups of isometries of real-hyperbolic spaces of dimensions $n \geqslant 3$.

Conjecture 23.22. *The class of finitely generated discrete subgroups of* $PO(3,1)$ *is QI rigid.*

Recently, this conjecture was proven by P. Haissinsky in the case of the class of Gromov-hyperbolic discrete subgroups of $PO(3,1)$:

Theorem 23.23 (P. Haissinsky [**Haï15**]). *1. The class of Gromov-hyperbolic discrete subgroups of* $PO(3,1)$ *is QI rigid.*

2. Moreover, if Γ *is a finitely generated group which admits a QI embedding into* $\mathbb{H}^3$*, then* Γ *is virtually isomorphic to a Gromov-hyperbolic discrete subgroup of* $PO(3,1)$.

One can ask for a stronger *topological* rigidity property in this regard, namely:

Conjecture 23.24. *Let* Γ *be a hyperbolic group whose ideal boundary is* planar*, i.e., topologically embeds in the 2-sphere. Then* Γ *is virtually isomorphic to a discrete subgroup of* $PO(3,1)$.

Note that this conjecture includes two well-known problems as special cases:

1. *Cannon's conjecture*: A hyperbolic group whose ideal boundary is homeomorphic to $\mathbb{S}^2$, acts geometrically on $\mathbb{H}^3$.

2. *Sierpinsky carpet conjecture* [**KK00**]: A hyperbolic group whose ideal boundary is homeomorphic to the Sierpinsky carpet, is virtually isomorphic to a uniform lattice in $PO(3,1)$.

We refer the reader to Haissinsky's paper [**Haï15**] for the most recent results in this direction.

Very little is known about strong QI rigidity of hyperbolic groups with higher-dimensional boundary. Below we list what is known and some open problems.

Theorem 23.25 (M. Kapovich, B. Kleiner, [**KK00**]). *There are hyperbolic groups* G *with 2-dimensional boundaries, which are strongly QI rigid. Furthermore, in these examples, all homeomorphisms of* $\partial_\infty G$ *are restrictions of elements of* G.

Suppose that M is a compact n-dimensional Riemannian manifold of constant curvature -1 and totally-geodesic boundary. We will refer to the fundamental groups of boundary components of M as *peripheral subgroups* of $G = \pi_1(M)$. The universal cover $\tilde{M}$ is isometric to a certain convex subset $C \subset \mathbb{H}^n$ bounded by pairwise disjoint hyperbolic subspaces of codimension 1. The group G acts geometrically on C and, hence, embeds as a discrete subgroup of $PO(n,1)$. The ideal boundary of C in $\mathbb{S}^{n-1}$ is a *round Sierpinsky carpet*: It is nowhere dense in $\mathbb{S}^{n-1}$ and its complement is a union of open round balls with pairwise disjoint closures. Such groups G are called *round Sierpinsky carpet groups*. The following theorem was known to various people, its proof is sketched by J.-F. Lafont in [**Laf04**]:

THEOREM 23.26. *1. The convex sets C above are strongly QI rigid. 2. In particular, each round Sierpinsky carpet group is QI rigid.*

It is a corollary of Thurston's Geometrization Theorem (see [**Kap01**]) that each Gromov-hyperbolic discrete subgroup of $PO(3,1)$ whose ideal boundary is homeomorphic to a Sierpinsky carpet is actually a round Sierpinsky carpet group.

PROBLEM 23.27. Suppose that $G < PO(n,1)$ is a discrete subgroup, which is Gromov-hyperbolic and $\partial_\infty G$ is homeomorphic to a Sierpinsky carpet. Is it true that G is QI rigid?

Round Sierpinsky carpet groups can be amalgamated along their peripheral subgroups. More specifically, consider an n-dimensional complex C obtained by gluing finite many n-dimensional compact hyperbolic manifolds with geodesic boundary in such a way that along each boundary component we glue at least three n-dimensional manifolds. The fundamental groups of the complexes C are hyperbolic, e.g. by Bestvina-Feighn Combination Theorem [**BF92**]. J.-F. Lafont in [**Laf04**] proved that the groups $\pi_1(C)$ are QI rigid.

Fundamental groups of complexes of simplicial negative curvature (see [**JŚ03, JŚ06**]) provide a major new class of higher dimensional hyperbolic groups.

PROBLEM 23.28. Are any of these higher-dimensional hyperbolic groups QI rigid?

Recall that, in view of Thurston's Hyperbolization Theorem, for all closed hyperbolic surfaces S and pseudo-Anosov homeomorphisms $f : S \to S$, the mapping tori M_f of $f : S \to S$ are hyperbolic. The fundamental groups of every such M, is a semidirect product

$$\pi_1(S) \rtimes_\phi \mathbb{Z},$$

where $\phi \in Aut(\pi_1(S))$ is the automorphism induced by f (it is defined up to an inner automorphism of the fundamental group). Then all of the groups $\pi_1(M)$ are quasiisometric to each other, since they embed as uniform lattices in $\mathrm{Isom}(\mathbb{H}^3)$. The question is what happens if we replace closed surface groups in such extensions by finitely generated free groups F of finite rank $r \geqslant 3$. The natural generalization of the notion of a pseudo-Anosov homeomorphism in this setting is the one of an *atoroidal fully irreducible (iwip) automorphism ϕ of F*: These are automorphisms so that no power of ϕ preserves (up to conjugation) a free factor of F and so that ϕ has no periodic conjugacy classes. The corresponding semidirect products

$$F_\phi := F \rtimes_\phi \mathbb{Z}$$

are hyperbolic groups and their boundaries are Menger curves, see [**Bri00**]. One can imagine several possibilities of the quasiisometric behavior of such semidirect products, the question is which (if any) of them occur:

QUESTION 23.29 (P. Sardar). Which (if any) of the following hold:

1. All groups F_ϕ are quasiisometric to each other.
2. The groups F_ϕ are quasiisometric if and only if they are virtually isomorphic.
3. If a group G is quasiisometric to some F_ϕ, then G is VI to F_ϕ.
4. Every hyperbolic group G with the Menger curve boundary is quasiisometric to one of the groups F_ϕ.

Note that (3) and (4) are mutually exclusive, since there are hyperbolic groups with Menger curve boundary which have the Property (T), while none of the free-by-cyclic groups has this property.

Lastly, we note that numerous results about structure of quasiisometries of solvable Lie groups with negatively curved left-invariant Riemannian metrics, were obtained by X. Xie, [**SX12, Xie12, Xie13, Xie14a, Xie14b**].

23.1.5. Failure of QI rigidity. So far we discussed QI rigidity in various forms. Below are examples of failure of QI rigidity.

Central coextensions. As we saw in Theorem 23.13, there are numerous example of groups Γ which admit central extensions

$$1 \to \mathbb{Z} \to \tilde{\Gamma} \to \Gamma \to 1,$$

such that $\tilde{\Gamma}$ is QI to the product $\Gamma \times \mathbb{Z}$, but is not VI to it. For instance:

EXAMPLE 23.30. Let S be a closed hyperbolic surface and let M be the unit tangent bundle of S. Then we have an exact sequence

$$1 \to \mathbb{Z} \to G = \pi_1(M) \to Q := \pi_1(S) \to 1.$$

This sequence does not split even after passage to a finite index subgroup in G, hence, G is not virtually isomorphic to $Q \times \mathbb{Z}$. However, since Q is hyperbolic, the group G is quasiisometric to $Q \times \mathbb{Z}$, Theorem 9.150. Note that the group $Q \times \mathbb{Z}$ is CAT(0), while the group G is not (see e.g. [**BH99**] or [**KL98a**]. In particular, the class of $CAT(0)$ groups is not QI rigid.

Groups quasiisometric to products of trees. In [**BM00**], M. Burger and S. Mozes constructed examples of *simple* groups G acting geometrically on products of locally finite simplicial trees $T_1 \times T_2$. In their examples, each tree T_i was has infinitely many ends and large group of automorphisms (it is transitive on the set of all embedded edge-paths of length n for each n). In particular, the trees T_i have constant valence $\geqslant 3$ and, hence, are quasiisometric to the free group F_2. Therefore, in these examples, the group G is quasiisometric to the product group $F_2 \times F_2$.

COROLLARY 23.31. *The product of free groups $G = F_n \times F_m$, $(n, m \geqslant 2)$ is not QI rigid: It is quasiisometric to a group G above, but is not virtually isomorphic to it.*

The group $F_2 \times F_2$ is co-large and not simple; therefore:

COROLLARY 23.32 (M. Burger, S. Mozes). *Virtual simplicity and co-largeness are not QI invariant.*

Here a group G is *virtually simple* if it is VI to a simple group.

These examples, of course, have the same *geometric model*, Definition 5.53. According to Theorem 5.54, there are commensurable groups without a common geometric model.

PROBLEM 23.33. Find an example of a pair of groups G_1, G_2 which are QI to each other, but they are not VI to groups with a common geometric model.

Virtual torsion-freeness. There are example of virtually isomorphic groups G_1, G_2 such that G_1 is virtually torsion free (even linear), while G_2 is not: Millson [**Mil79**] constructed examples of lattices G_1 in a linear Lie group and central coextensions

$$0 \to \mathbb{Z}_2 \to G_2 \to G_1 \to 1$$

such that the groups G_2 are not virtually torsion free. Being linear, the lattice G_1 is, of course, virtually torsion-free by Selberg's lemma. Further examples like this were constructed by Raghunathan [**Rag84**].

PROBLEM 23.34. 1. Is it true that every finitely generated group is QI to a torsion-free group?

2. Construct examples of groups G which are QI to torsion-free groups but not VI to torsion-free groups.

Note that a positive answer to the 1st question (which seems unlikely) would be an ultimat form of Selberg's lemma. As for the question 2, natiral candidates would be simple groups acting geometrically on products of trees. However, it appears that all currently known examples of simple groups acting geometrically on products of trees are torsion-free.

Hopfian and cohopfian properties. Both properties are not preserved by virtual isomorphisms, see §4.12.

Unbounded group actions on trees. We will say that a group Γ *virtually splits* if it is virtually isomorphic to a group which admits a nontrivial graph of groups decomposition. Recall that, in view of the Bass-Serre theory, a group Γ admits a nontrivial decomposition as a graph of groups if and only if Γ acts on a simplicial tree (without inversions) without a fixed vertex. The Lie group $G = SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$ contains uniform irreducible lattices Γ, for instance $\Gamma \cong SL(2,\mathbb{Z}[\sqrt{2}])$. According to Theorem 10.13 of Margulis, every action of such irreducible lattice on a simplicial tree has a fixed point. Therefore, irreducible lattices in G do not virtually split. On the other hand, G also contains reducible uniform lattices Λ, e.g., discrete subgroups isomorphic to $\pi_1(S) \times \pi_1(S)$, where S is a compact hyperbolic surface. Such lattices do split nontrivially as graphs of groups, since $\pi_1(S)$ does. Therefore, we obtain examples of quasiisometric groups Γ and Λ such that Γ does not virtually split, while Λ does.

The examples of failure of quasiisometric invariance of Property T, see Theorem 17.62, also show that the property FA is not a quasiisometric invariant either.

QUESTION 23.35. 1. Are Properties T and Haagerup quasiisometrically invariant in the class of hyperbolic groups?

2. Suppose that G_1, G_2 are quasiisometric hyperbolic groups and G_1 virtually splits. Is it true that G_2 virtually splits as well?

The second part of the problem is open even for arithemetic lattices in $PU(2,1)$. On the other hand, both parts have positive answers in some notable examples, like lattices in $\mathrm{Isom}(\mathbf{H}\mathbb{H}^n)$ $(n \geqslant 2)$ and in isometry groups of 2-dimensional thick hyperbolic buildings.

23.1.6. Rigidity of random groups. At this time, it is far from clear how common is the phenomenon of QI rigidity for finitely presented groups. The reader might have noticed that all the examples of QI rigid groups and classes of groups are quite special: They come from groups acting discretely and isometrically on some highly homogeneous metric spaces (symmetric spaces and buildings). On the other hand, all the examples of groups failing QI rigidity are rather special as well. This leads to:

PROBLEM 23.36. Let G be an infinite *random*[1] finitely presented group. Is G QI rigid?

We refer the reader to §9.25 for the description of some models of *randomness* among finitely presented groups. Here we recall only that, according to all these models, *random groups* are Gromov-hyperbolic. Generic QI rigidity is an open problem for all these models. (Except for the regime in Gromov's model of randomness, in which random groups are finite.) One reason to expect random groups to be QI rigid is that I. Kapovich and P. Schupp proved in [**KS08**] that, in one of the models, random groups are "algebraically rigid", i.e., their isomorphisms are induced by Nielsen transformations.

23.2. Rigidity of relatively hyperbolic groups

As we know, see Corollary 9.43, the class of hyperbolic groups is QI rigid. In this section we discuss QI rigidity properties of relatively hyperbolic groups. In the following discussion, by a relatively hyperbolic group we always mean a group G which admits a relatively hyperbolic structure where each peripheral subgroup has infinite index in G. Thus, a group G is *not relatively hyperbolic (NRH)* if G contains no finite collection of infinite index subgroups with respect to which it is relatively hyperbolic.

THEOREM 23.37 (C. Druţu, [**Dru09**]). *The class of relatively hyperbolic groups is QI rigid. More precisely, if a group G_1 is relatively hyperbolic and a group G_2 is quasiisometric to G_1, then G_2 is also relatively hyperbolic.*

Other theorems appearing in this chapter emphasize that various subclasses of relatively hyperbolic groups are, likewise, QI rigid: Nonuniform lattices in rank one symmetric spaces (Theorem 23.7), fundamental groups of non-geometric Haken manifolds (Theorem 23.69), fundamental groups of graphs of groups with finite edge groups [**PW02**].

Theorem 23.37 suggests the following natural question.

PROBLEM 23.38 (P. Papasoglou). Is there a geometric criterion allowing to recognize whether a finitely generated group is relatively hyperbolic (without any reference to peripheral subgroups or subsets)?

Concerning the proof of Theorem 23.37, it is not difficult to see that if $f : X \to Y$ is a quasiisometry between two metric spaces and X is hyperbolic relative to $\mathcal{A}$ then Y is hyperbolic relative to $\{f(A) \;:\; A \in \mathcal{A}\}$. Thus, the main step towards proving Theorem 23.37 is to show that if a group G is hyperbolic relative to some collection of subsets $\mathcal{A}$, then it is also hyperbolic relative to some collection of subgroups $H_1, \ldots, H_n$, such that each H_i is contained in a metric neighborhood of some A_i, [**Dru09**]. A variation of the same argument, appears in [**BDM09**]:

[1] In any currently existing model of randomness.

Theorem 23.39. *Let X be a metric space which is hyperbolic relative to a collection of subsets $\mathcal{A}$. Suppose that $f : G \to X$ is a quasiisometric embedding of a finitely generated group G. Then G is hyperbolic relative to the collection of pre-images $f^{-1}(\mathcal{N}_C(A)), A \in \mathcal{A}$, for some $C < \infty$.*

This result and the above argument imply that either $f(G)$ lies in an M–tubular neighborhood of some set $A \in \mathcal{A}$, or G is hyperbolic relative to finitely many (proper) subgroups H_i with each $f(H_i)$ contained in an M–tubular neighborhood of some set $A_i \in \mathcal{A}$. Thus, we have the following generalization of the Quasi-Flat Lemma of R. Schwartz [**Sch96b**] (see Proposition 22.9, Lemma 22.12 as well as Remark 22.13 in Chapter 22). This generalization was proven by J. Behrstock, C. Druţu, L. Mosher in [**BDM09**], Theorem 4.1:

Theorem 23.40 (NRH subgroups are always peripheral). *Let X be a metric space hyperbolic relative to a collection $\mathcal{A}$ of subsets. For every $L \geqslant 1$ and $C \geqslant 0$ there exists $R = R(L, C, X, \mathcal{A})$ such that the following holds.*

If G is a finitely generated group with a word metric dist *and G is NRH, then the image of any (L, C)–quasiisometric embedding $f : G \to X$ is contained in the R–neighborhood of some set $A \in \mathcal{A}$.*

In this theorem, the constant R *does not* depend on the group G. In [**DS05b**] the same theorem was proved under the stronger hypothesis that the group G has one asymptotic cone without global cut-points.

As in the case of the proof of Theorem 22.1, Theorem 23.40 is a step towards the classification of relatively hyperbolic groups:

Theorem 23.41 (J. Behrstock, C. Druţu, L. Mosher, [**BDM09**], Theorem 4.8). *Let G be a finitely generated group hyperbolic relative to a finite collection of finitely generated subgroups $\mathcal{H}$, such that each $H \in \mathcal{H}$ is NRH. If G' is a finitely generated group quasiisometric to G, then G' is hyperbolic relative to a finite collection of finitely generated subgroups $\mathcal{H}'$, where each subgroup in $\mathcal{H}'$ is quasiisometric to one of the subgroups in $\mathcal{H}$.*

When working in full generality, it is impossible to establish a relation between peripheral subgroups of QI relatively hyperbolic groups; hence, this is not mentioned in Theorem 23.37. For instance, when $G = G' = A \star B \star C$, the group G is hyperbolic relative to $\{A, B, C\}$, and and it is also hyperbolic relative to $\{A \star B, C\}$.

By the results in [**PW02**], the classification of relatively hyperbolic groups reduces to the classification of one-ended relatively hyperbolic groups. Theorem 23.41 points out a fundamental necessary condition for two one-ended relatively hyperbolic groups (with NRH peripheral subgroups) to be quasiisometric: Their peripheral subgroups have to define the same collection of quasiisometry classes.

Related to this, one may ask whether every relatively hyperbolic group G admits a relatively hyperbolic structure $(G; \mathcal{P})$, such that all peripheral subgroups $P_i \in \mathcal{P}$ are NRH. The answer is negative in general, a counter-example is Dunwoody's inaccessible group [**Dun93**]. Since finitely presented groups are accessible, this raises the following natural question:

Problem 23.42 (J. Behrstock, C. Druţu, L. Mosher, [**BDM09**]). Is there an example of a finitely presented relatively hyperbolic group G such that for every relatively hyperbolic structure $(G; \mathcal{P})$ at least one group $P_i \in \mathcal{P}$ is a relatively hyperbolic group?

23.3. Rigidity of classes of amenable groups

The class of amenable groups is QI rigid, see Theorem 16.12. Recall that, by Corollary 16.50, the set of finitely generated groups splits into amenable groups and paradoxical groups. This implies that the class of paradoxical groups is also QI rigid. Since the latter class is characterized by the fact that the Cheeger constant is positive (Theorem 16.4), it follows that having a positive Cheeger constant is a QI invariant property. As noted, the property of having positive Cheeger constant is QI invariant not only among groups, but also among graphs or manifolds of locally bounded geometry.

Various subclasses of amenable groups behave quite differently with respect to QI rigidity, and relatively little is known about their QI classification and the description of groups of quasiisometries.

The class of virtually nilpotent groups is QI rigid by Theorem 14.25. Concerning the QI classification of nilpotent groups, the following is known:

<u>Theorem</u> 23.43 (P. Pansu [**Pan89**]). *If G and H are finitely generated quasi-isometric nilpotent groups, then the graded Lie groups associated with $G/\mathrm{Tor}\,(G)$ and $H/\mathrm{Tor}\,(H)$ are isomorphic.*

One of the steps in the proof of Theorem 23.43 is that all asymptotic cones of a finitely generated nilpotent group G with a canonically chosen word metric are isometric to the graded Lie group associated to $G/\mathrm{Tor}\,(G)$, endowed with a Carnot-Caratheodory metric, see Theorem 14.28.

Theorem 23.43 establishes new quasiisometry invariants in the class of nilpotent groups: The nilpotency class of $\bar{G} = G/\mathrm{Tor}\,(G)$ and ranks of the abelian groups $C^i\bar{G}/C^{i+1}\bar{G}$, where $C^i\bar{G}$ is the i-th group in the lower central series of $\bar{G}$.

Other QI invariants in the class of nilpotent group that help to distinguish nilpotent groups with the same associated nilpotent graded Lie groups are the virtual Betti numbers [**Sha04**, Theorem 1.2]. Recall that Pansu also proves QI rigidity of abelian groups, see Theorem 14.26.

Unlike abelian groups, nilpotent groups are not completely classified up to QI. In particular, the following remains an open problem:

Problem 23.44. Is it true that two nilpotent simply-connected Lie groups (endowed with left-invariant Riemannian metrics) are quasiisometric if and only if they are isomorphic?

This answer to this problem is positive for graded nilpotent Lie groups, according to Theorem 23.43.

The group of quasiisometries is very large already for abelian groups, see Examples 5.15 and 5.16. However, it is a reasonable and not well-understood problem to classify uniformly quasiisometric discrete subgroups of quasiisometries of Euclidean spaces and of nilpotent groups. For instance:

Problem 23.45. Is there a discrete quasi-action $G \curvearrowright \mathbb{E}^n$ of a finitely generated nilpotent group G, which is not virtually abelian?

In view of Erschler's Theorem 12.40, the class of (virtually) solvable groups is not QI rigid. Note, however, that:

1. Groups constructed in the proof are not finitely presented.
2. Both groups G_A and G_B in the proof are elementary amenable.

This lead to:

PROBLEM 23.46. 1. Is the class of finitely presented solvable groups QI rigid?

2. Is the class of finitely presented metabelian groups (i. e., solvable groups of derived length 2) QI rigid?

3. Is the class of elementary amenable groups QI rigid?

Below we review partial results and open problems in this direction.

THEOREM 23.47 (B. Farb and L. Mosher [**FM00**]). *The class of finitely presented non-polycyclic abelian-by-cyclic groups is QI rigid.*

The starting point in the proof of this theorem is to consider torsion-free finite index subgroups and apply a result of R. Bieri and R. Strebel [**BS78**]. The latter states that for every torsion-free finitely presented abelian-by-cyclic group G, there exists $n \in \mathbb{N}$ and a matrix $M = (m_{ij}) \in M(n, \mathbb{Z})$ with non-zero determinant, such that the group G has the presentation

$$\left\langle a_1, a_2, \ldots, a_n, t \mid [a_i, a_j], t a_i t^{-1} a_1^{m_{1i}} a_2^{m_{2i}} \cdots a_n^{m_{ni}} \right\rangle . \tag{23.1}$$

Let Γ_M be the group with the presentation in (23.1) for the integer matrix M. The group Γ_M is polycyclic if and only if $|\det(M)| = 1$, see [**BS80**].

In [**FM00**], Farb and Mosher prove that if a finitely generated group G is quasiisometric to the group Γ_M, for an integer matrix M with $|\det\ M| > 1$, then a quotient G/F of G by a finite normal subgroup F, is virtually isomorphic to a group Γ_N defined by an integer matrix N with $|\det(N)| > 1$.

THEOREM 23.48 (B. Farb, L. Mosher, [**FM00**]). *Let M_1 and M_2 be integer matrices with $|\det(M_i)| > 1,\ i = 1, 2$. The groups Γ_{M_1} and Γ_{M_2} are quasiisometric if and only if there exist two positive integers k_1 and k_2 such that $M_1^{k_1}$ and $M_2^{k_2}$ have the same absolute Jordan form.*

The absolute Jordan form of a matrix is obtained from the Jordan form over $\mathbb{C}$ by replacing the diagonal entries with their absolute values and arranging the Jordan blocks in a canonical way.

In the case of solvable Baumslag–Solitar groups, which form a subclass in the class of groups in Theorem 23.48, more can be said:

THEOREM 23.49 (B. Farb, L. Mosher, [**FM98**], [**FM99**]). *Each solvable Baumslag–Solitar group*

$$BS(1, m) = \left\langle x, y : xyx^{-1} = y^m \right\rangle$$

is QI rigid.

This theorem is complemented by:

THEOREM 23.50 (K. Whyte, [**Why01**]). *All non-solvable Baumslag–Solitar groups*

$$BS(n, m) = \left\langle x, y | xy^n x^{-1} = y^m \right\rangle,$$

$|n| \neq 1, |m| \neq 1$ *are QI to each other.*

Since non-solvable Baumslag–Solitar groups are all nonamenable, these results complete the QI classification of Baumslag–Solitar groups.

The fact that polycyclic groups were excluded from theorems of Farb and Mosher is not an accident: These groups are much harder to handle since the tools of coarse topology do not apply to them.

PROBLEM 23.51. (1) Is the class of finitely generated polycyclic groups QI rigid?
(2) What is the QI classification of finitely generated polycyclic groups?

Every virtually polycyclic group of course has a finite index subgroup with infinite abelianization. Shalom in [**Sha04**] proved (among other QI rigidity properties for various classes of amenable groups) the following:

THEOREM 23.52. *Suppose that G is a group QI to a polycyclic group. Then G contains a finite index subgroup with infinite abelianization.*

Even the problem of QI rigidity for finitely generated polycyclic abelian-by-cyclic groups has remained open for some time. The papers of Eskin, Fisher and Whyte [**EFW13, EFW12**] made a major progress in this direction. In particular, they prove:

THEOREM 23.53. *Consider the class $Poly_3$ of groups G which are not virtually nilpotent and admit short exact sequences*

$$1 \to \mathbb{Z}^2 \to G \to \mathbb{Z} \to 1.$$

Then the class $Poly_3$ is QI rigid.

Note that the groups in $Poly_3$ play an important role in 3-dimensional topology. Namely, there exists a 3-dimensional simply-connected solvable Lie group Sol_3, such that each $\Gamma \in Poly_3$ is isomorphic to a uniform lattice in Sol_3. Accordingly, Γ is isomorphic to the fundamental group of a closed 3-dimensional manifold $M = \Gamma \setminus Sol_3$. The manifolds M of this form (and manifolds which are covered by such M's) are called *Sol_3-manifolds*, they appear in the classification theory of 3-dimensional manifolds.

In view of the Problem 23.51, Part (1), one may ask about the QI classification of solvable Lie groups.

Y. Cornulier proved that each connected Lie group is quasiisometric to a closed connected subgroup of the group of real upper triangular matrices [**dC08**, Lemma 6.7]. This has lead him to ask:

PROBLEM 23.54 (Y. Cornulier [**dC09**]). Suppose that G_1, G_2 are closed connected subgroups of the group of real upper triangular matrices, endowed with left-invariant Riemannian metrics. Is it true that G_1, G_2 are quasiisometric if and only if are they isomorphic?

Groups QI to abelian-by-abelian solvable groups. Generalizing the results of [**EFW13, EFW12**], I. Peng in [**Pen11a, Pen11b**] considered quasiisometries of lattices in solvable Lie groups G of the type

$$G = G_\varphi = \mathbb{R}^n \rtimes_\varphi \mathbb{R}^m,$$

where $\varphi : \mathbb{R}^m \to GL(n, \mathbb{R})$ is an action of $\mathbb{R}^m$ on $\mathbb{R}^n$. The number m is called the *rank* of G. The group G is clearly a Lie group and we equip G with a left-invariant Riemannian metric. Then G admits *horizontal* and *vertical* foliations by the left translates of $\mathbb{R}^n$ and $\mathbb{R}^m$.

DEFINITION 23.55. 1. G_φ is *unimodular* if $\varphi(\mathbb{R}^m) \subset SL(n, \mathbb{R})$.

2. G_φ is *nondegenerate* if for each $\mathbf{x} \in \mathbb{R}^m \setminus \{0\}$, $\varphi(\mathbf{x})$ has at least one eigenvalue with the absolute value $\neq 1$.

3. G_φ is *split* if there are no nonzero $\varphi(\mathbb{R}^m)$-invariant subspaces $V \subset \mathbb{R}^n$ such that the image group $\varphi(\mathbb{R}^m) < GL(V)$ is a bounded subgroup.

Note that this definition is given in [**Pen11b**] in terms of root systems, but it is easy to see that our definitions are equivalent to hers.

The main result of [**Pen11a, Pen11b**] is to establish control of quasiisometries $G_\varphi \to G_\psi$ between the solvable groups, the key being that quasiisometries send leaves of horizontal and vertical foliations uniformly close to leaves of horizontal and vertical foliations. The precise result is too technical to be stated here (see [**Pen11b**, Theorem 5.2]); below is its main corollary, which is a combination of the work of I. Peng and T. Dymarz (Corollary 1.0.2 in [**Pen11b**]):

THEOREM 23.56 (I. Peng, T. Dymarz). *Suppose that G_φ is non-degenerate and unimodular, and G' is a finitely generated group quasiisometric to G_φ. Then the group G' is virtually polycyclic.*

As a special case, consider a group $G = G_\varphi$ of rank 1, i.e., $m = 1$. T. Dymarz in [**Dym10**] proved the following:

THEOREM 23.57 (Dymarz rigidity theorem). *Every finitely generated group QI to G is virtually isomorphic to a lattice in G.*

23.4. Bilipschitz vs. quasiisometric

The question about the difference between quasiisometries and bilipschitz maps between finitely generated groups is both very basic and interesting. At the first glance, there should not be any need for passing to a subnet in order to go from quasiisometries to bilipschitz maps of finitely generated groups. Gromov asked in [**Gro93**, § 1.A0] if this is really the case, as the situation was unclear even for separated nets in Euclidean spaces and for free groups of different (finite) ranks:

QUESTION 23.58 (M. Gromov). 1. Suppose that $X_1, X_2 \subset \mathbb{E}^n$ are separated nets. Is there a bilipschitz homeomorphism $X_1 \to X_2$?

2. Suppose that F_m, F_n, $2 \le m, n < \infty$, are free groups of ranks m and n respectively, equipped with the word metrics associated with their free generating sets. Is F_m is bilipschitz to F_n?

The case of free groups was settled quickly by P. Papasoglu [**Pap95a**]:

THEOREM 23.59. *Any two nonabelian free groups of finite ranks are bilipschitz to each other.*

In view of his theorem it was reasonable to expect that any two separated nets in $\mathbb{E}^n$ are bilipschitz to each other and that any two finitely generated quasiisometric groups are also bilipschitz equivalent. In a surprising development, D. Burago and B. Kleiner [**BK02b**] and C. McMullen [**McM98**], independently constructed examples of separated nets in $\mathbb{R}^2$ which are not bi-Lipschitz equivalent.

QUESTION 23.60 (D. Burago, B. Kleiner, [**BK02b**]). 1. When placing a point in the barycenter of each tile of a *Penrose tiling* in $\mathbb{E}^2$, is the resulting separated net bi-Lipschitz equivalent to $\mathbb{Z}^2$?

2. More generally: Embed $\mathbb{R}^2$ into $\mathbb{R}^n$ as a plane P with irrational slope and take B, a bounded subset of $\mathbb{R}^n$ with non-empty interior. Consider the subset $Z \subset \mathbb{Z}^n$ of all z's such that $z + B$ intersects P. The orthogonal projection Z to P composes a separated net in $\mathbb{R}^2$. Is such a net bilipschitz equivalent to $\mathbb{Z}^2$?

Part 1 of this question was answered by Y. Solomon in [**Sol11**], see also [**APCG13**] for an improvement of his results.

This has left open the case of finitely generated groups. The case of non-amenable groups was settled by K. Whyte:

THEOREM 23.61 (K. Whyte, [**Why99**]; see Theorem 16.8 of this book). *Suppose that G_1, G_2 are finitely generated quasiisometric non-amenable groups. Then G_1, G_2 are bilipschitz equivalent. More generally, for every quasiisometry $f : \Gamma_1 \to \Gamma_2$ between nonamenable graphs of bounded geometry, the restriction*

$$f\big|_{V(\Gamma_1)} : V(\Gamma_1) \to V(\Gamma_2)$$

is at a bounded distance from a bilipschitz bijection.

We also note that the theorem about graphs is implicitly contained in the earlier paper [**DSS95**] of W. A. Deuber, M. Simonovits and V. T. Sós.

The case of amenable groups was settled (in the negative) by T. Dymarz in [**Dym10**]. She constructed certain *lamplighter groups* which are quasi-isometric but not bilipschitz equivalent. Her examples, however, are commensurable. This leads to:

PROBLEM 23.62. Generate an equivalence relation CLIP on finitely generated groups by combining commensurability and bilipschitz equivalence. Is CLIP equal to the quasiisometry equivalence relation?

We note that Dymarz' examples are merely finitely generated; finitely presented examples were constructed by Dymarz, Peng and Taback [**DPT15**].

23.5. Various other QI rigidity results and problems

The following theorem was first proven by R. Grigorchuk in [**Gri84a**], who proved in [**Gri84a**] that there are uncountable many equivalence classes of growth functions of groups of intermediate growth. B. Bowditch in [**Bow98a**] gave a different argument, not based on the growth of groups.

THEOREM 23.63 (R. Grigorchuk [**Gri84a**]). *There are uncountable many QI classes of finitely generated groups.*

We note that the most progress in establishing QI rigidity was achieved in the context of lattices in Lie groups or certain solvable groups. Below we review some QI rigidity results for groups which do not belong to these classes.

The following rigidity theorem was proven by J. Behrstock, B. Kleiner, Y. Minsky and L. Mosher in [**BKMM12**]:

THEOREM 23.64. *Let S be a closed surface of genus g with n punctures, so that $3g - 3 + n \geqslant 2$ and $(g, n) \neq (1, 2)$. Then the Mapping Class group $\Gamma = Map(S)$ of S is strongly QI rigid. Moreover, quasiisometries of Γ are uniformly close to automorphisms of Γ.*

Note that for a closed surface S, the group $Map(S)$ is isomorphic to the group of outer automorphisms $Out(\pi)$, where $\pi = \pi_1(S)$, see [**FM11**]. Furthermore, N. Ivanov [**Iva88**] proved that $Out(Map(S))$ is trivial if $3g - 3 + n \geqslant 2, (g, n) \neq (2, 0)$

and $Out(Map(S)) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ if $(g, n) = (2, 0)$. Recall that for a group π, the group of outer automorphisms $Out(\pi)$ is the quotient

$$Out(\pi) = Aut(\pi)/Inn(\pi)$$

where $Inn(\pi)$ consists of automorphisms of π given by conjugations *via* elements of π.

PROBLEM 23.65. Is the group $Out(F_n)$ QI rigid?

Artin groups and Coxeter groups are prominent classes of groups which appear frequently in geometric group theory. Note that some of these groups are not QI rigid, e.g., the group $F_2 \times F_2$, see the above-mentioned examples of Burger and Mozes. In particular, if G is a Coxeter or Artin group which splits as the fundamental group of graph of groups with finite edge groups, where one of the vertex groups G_v is virtually $F_2 \times F_2$, then G cannot be QI rigid. The same applies if one takes a direct product of such G with a Coxeter/Artin group. Also, there are many Coxeter groups which appear as uniform lattices in $O(n, 1)$ (for relatively small n). Such Coxeter groups are QI to non-Coxeter lattices in $O(n, 1)$. This leads to

PROBLEM 23.66. (a) Suppose that G is an Artin group, which does not contain $F_2 \times F_2$. Is such G QI rigid?

(b) Suppose that G is a non-hyperbolic 1-ended Coxeter group, which does not contain $F_2 \times F_2$. Is G QI rigid?

Note that Theorem 23.64 implies QI rigidity of *Artin Braid groups* B_n: The quotient of B_n by its center is isomorphic to the mapping class group $Map(S)$, where S is the 2-sphere with $n + 1$ punctures. QI rigidity results for other classes of Artin groups were obtained by J. Behrstock, T. Januszkiewicz and W. Neumann [**BJN09, BJN10**].

M. Gromov and W. Thurston [**GT87**] constructed interesting examples of closed negatively curved manifolds. The fundamental group of such a manifold is not isomorphic to a lattice in a Lie group (with finitely many components). We will refer to the manifolds constructed in [**GT87**] as *Gromov-Thurston manifolds*. Some of these manifolds are obtained as ramified covers over closed hyperbolic n-manifolds ($n \geqslant 4$), ramified over totally-geodesic submanifolds.

PROBLEM 23.67. Are the fundamental groups of Gromov-Thurston n-manifolds QI rigid?

The reason to be hopeful that these groups Γ are QI rigid is the following. Each Γ is associated with a uniform lattice $\Gamma' < O(n, 1)$ and a sublattice

$$\Gamma'' = \Gamma' \cap O(n - 2, 1).$$

The sublattice Γ'' yields a Γ'-invariant collection of $n - 2$-dimensional hyperbolic subspaces $X_i \subset \mathbb{H}^n$, where X_1 is Γ''-invariant. (For instance, a Gromov-Thurston manifold can appear as a ramified cover over $\mathbb{H}^n/\Gamma'$ which is ramified over the submanifold X_1/Γ''.) While the entire hyperbolic n-space is highly non-rigid, R. Schwartz proved in [**Sch97**] that the pair $(\mathbb{H}^n, \cup_i X_i)$ is QI rigid.

PROBLEM 23.68. Let S be a closed hyperbolic surface. Let M be the 4-dimensional manifold obtained by taking the 2-fold ramified cover over $S \times S$, which is ramified over the diagonal, see [**BGS85**, Exercise 1]. Is $\pi_1(M)$ QI rigid?

3-manifold groups. Another class of groups whose QI rigidity properties are relative well (but not completely) understood are fundamental groups of compact 3-manifolds.

Theorem 23.69 (M. Kapovich, B. Leeb, [**KL97**]). *The class of fundamental groups G of 3-dimensional closed Haken 3-manifolds, which are not Sol_3-manifolds is QI rigid.*

This rigidity theorem was generalized to higher dimensional graph-manifolds by R. Frigerio, J.-F. Lafont and A. Sisto [**RF**].

The combination of several rigidity results for 3-manifold groups, leads to:

Theorem 23.70 (Eskin–Fisher–Whyte, Kapovich–Leeb, Pansu, Schwartz). *The class of fundamental groups of closed connected 3-manifolds is QI rigid.*

Proof. Suppose that G' is a group QI to $G = \pi_1(M)$, where M is a closed connected 3-dimensional manifold. Without loss of generality, we may assume that M is oriented. Recall that, according to Thurston's Geometrization Conjecture/Perelman's Theorem, the manifold M has the following structure: M splits as a connected sum $M = M_1 \# \ldots \# M_k$, where each manifold M_i is either geometric, or M is Haken, obtained by gluing compact 3-dimensional geometric manifolds with boundary along boundary tori. (See §21.7 for more details.)

1. Suppose that $M = M_1$ and M is non-geometric. Then G is VI to the fundamental group of a closed 3-manifold (Theorem 23.69).

2. Suppose that $M = M_1$ and M is geometric. The case when M is hyperbolic is covered by Sullivan's theorem (Theorem 21.1). If M has spherical geometric structure, then $\pi_1(M)$ and, hence, G, is finite. If M has the $\mathbb{S}^2 \times \mathbb{R}$-structure, then G is infinite cyclic or infinite dihedral, i.e., is 2-ended. Hence, the group G' is also 2-ended and, therefore, is VI to $\mathbb{Z}$, see Proposition 6.23. Of course,

$$\mathbb{Z} \cong \pi_1(\mathbb{S}^2 \times \mathbb{S}^1)$$

is a 3-manifold group.

The cases of manifolds with Euclidean geometry and Nil_3-geometry are covered by Gromov–Pansu's rigidity theorems as follows. Since $\pi_1(M)$ is virtually nilpotent in both cases, by applying Gromov's theorem, we can assume that G' is also torsion-free nilpotent.

a. If M is a Euclidean manifold, its fundamental group is VI to $\mathbb{Z}^3$. Therefore, by Pansu's theorem 11.11, G' is VI to $\mathbb{Z}^3 \cong \pi_1(T^3)$.

b. If M has Nil_3-geometry, then the group G (after passing to a finite-index subgroup) is 2-step nilpotent and $C^1(G)/C^2(G) \cong \mathbb{Z}^2$, $C^2(G) \cong \mathbb{Z}$. By Theorem 23.43, the group G' has the same properties. Since $\mathbb{Z}$-central coextensions of $\mathbb{Z}^2$ are classified by $H^2(\mathbb{Z}^2, \mathbb{Z}) \cong \mathbb{Z}$, every coextension is isomorphic to a group

$$G_k = \left\langle a, b, c | [a, b] = c^k, [a, c] = 1.[b, c] = 1 \right\rangle .$$

All these groups (with $k \neq 1$) are commensurable to each other and, hence, to the integer Heisenberg group $H_3(\mathbb{Z})$, which is a cocomapct lattice in $Nil_3 = H_3(\mathbb{R})$. If $k = 0$, then $G_k \cong \mathbb{Z}^3$ and, thus, has the $\mathbb{E}^3$-geometry.

Finally, QI rigidity of the class of Sol_3-manifold groups is the content of Theorem 23.53.

3. Suppose that $k \geqslant 2$, i.e., $\pi_1(M)$ has infinitely many ends. Then the group G splits as an graph of groups with finite edge-groups and vertex groups G_v which

have at most one end. This decomposition corresponds to the connected sum decomposition of the manifold M:

$$M = M_1 \# \dots \# M_k,$$

where each M_i either has finite fundamental group, or 2-ended fundamental group (which is either infinite cyclic or infinite dihedral) or 1-ended fundamental group.

According to Theorem 18.46, the group G' also splits as a graph of groups, with finite edge groups, where each vertex group G'_w is either finite or is QI to one of the 1-ended vertex groups G_v.

By combining 1, 2 and 3, we conclude that G' splits as a finite graph of groups with finite edge groups and vertex groups G'_w which are VI to the fundamental groups of closed 3-manifolds. Up to passing to a finite index subgroup, each group G_w has the form

$$1 \to K_w \to G'_w \to \bar{G}_w \to 1,$$

where $\bar{G}_w$ is a closed 3-manifold group. It is observed in [**Kap07**] that such groups G'_w are virtually torsion-free and, hence, each contains a finite index subgroup which is the fundamental group of a closed 3-manifold. Lastly, as in the proof of Theorem 4.52, one assembles all these finite-index subgroups in the vertex groups G'_w, into a finite-index subgroup $H < G'$, which is a free product of fundamental groups of closed 3-manifolds M'_u. The connected sum of the 3-manifolds M'_u is a closed 3-manifold M' with $\pi_1(M') \cong H$. □

This leaves open the internal QI classification of fundamental groups of closed 3-manifolds.

PROBLEM 23.71. Classify fundamental groups of closed non-geometric irreducible 3-dimensional manifolds up to quasiisometry.

Partial progress towards this problem is achieved in several papers of J. Behrstock and W. Neumann. In the paper [**BN08**] they proved that fundamental groups of all *nongeometric* 3-dimensional *graph-manifolds* are QI to each other. They obtained further QI rigidity results for manifolds with hyperbolic components in [**BN12**].

Most QI rigidity results for fundamental group of 3-manifolds are obtained under the assumption that the 3-manifolds in question have either empty boundary or the boundary consisting of tori and Klein bottles. However, the problem is also interesting in the case of compact 3-dimensional manifolds with more complex boundary surfaces. In view of the recent results of P. Haissinsky [**Haï15**], it is reasonable to ask:

QUESTION 23.72. Is the class of fundamental groups of compact 3-dimensional manifolds with nonempty boundary QI rigid?

Haissinsky's results settle the problem in the case of 3-manifolds with Gromov-hyperbolic fundamental groups.

Quasiisometric invariance of group decompositions.

Two sets of theorems below show that, under certain conditions, quasiisometries respect certain graph of groups decompositions and direct product decompositions. In order to state the first of these results, we note that each 1-ended finitely presented group G admits a special decomposition as a graph of groups with 2-ended edge groups. This special decomposition is called the *JSJ decomposition* of G.

Decompositions of this type first appeared in the context of 3-dimensional manifolds (and their fundamental groups), in the work of Jaco, Shalen and Johannson, hence, the name JSJ. Each finitely presented group G has a unique JSJ decomposition; this decomposition, furthermore, is the refinement of any splitting of G with 2-ended edge groups. We refer the reader to the works of Rips, Sela, Dunwoody, Sageev, Fujuwara and Papasoglu, [**RS97, DS99, FP06**] for the detailed treatment of group-theoretic JSJ decompositions.

<u>Theorem</u> 23.73 (P. Papasoglu, [**Pap05**]). *The class of one-ended finitely presented groups which split over 2-ended groups G is QI rigid. Moreover, quasiisometries every such group G preserve the JSJ decomposition of G.*

The second part of this theorem means that if X is a simply-connected tree of spaces associated with the JSJ decomposition of G, then each quasiisometry $X \to X$ sends each vertex space uniformly close to a vertex space, sends each edge space uniformly close to an edge space, and induces an automorphism of the corresponding tree.

The next theorem deals with QI invariance of direct products of groups in the context of fundamental groups of closed manifolds M of nonpositive curvature. The de Rham decomposition of the universal cover X of such M is a canonical decomposition of X into a Riemannian direct product of manifolds of nonpositive curvature

$$X = \mathbb{E}^m \times X_1 \times \ldots \times X_k.$$

None of the factors X_i is further decomposable as a Riemannian direct product.

<u>Theorem</u> 23.74 (M. Kapovich, B. Kleiner, B. Leeb, [**KKL98**]). *Quasiisometries $X \to X$ preserve the de Rham decomposition. More precisely:*

1. Each (L, A) quasiisometry $f : X \to X$ sends each Euclidean leaf $\mathbb{E}^m \times x$ uniformly Hausdorff-close to another leaf $\mathbb{E}^m \times x'$, where x, x' belong to

$$\overline{X} = \prod_{i=1}^{k} X_i.$$

In particular, f induces a $(\bar{L}, \bar{A})$ quasiisometry $\bar{f} : \overline{X} \to \overline{X}$.

2. Suppose that $f : X \to X$ is an (L, A) quasiisometry, where X does not have a Euclidean de Rham factor. Then, after composing f with a permutation of the factors X_i if necessary, the map f is uniformly close to a product map

$$f_1 \times \ldots \times f_k,$$

where each $f_i : X_i \to X_i$ is a quasiisometry.

On the group-theoretic side:

Corollary 23.75. *Suppose that X does not have a Euclidean de Rham factor ($m = 0$) and that the manifold M splits as the direct product $M = M_1 \times \ldots \times M_k$, where each M_i has the universal cover X_i. Accordingly, the group G splits as the direct product $G = G_1 \times \ldots \times G_k$ Then each quasiisometry $G \to G$ preserves the direct product decomposition of G in the same sense as a quasiisometry $X \to X$ preserves the de Rham decomposition.*

Note that a group acting geometrically on X (say, $G = \pi_1(M)$) need not contain a finite index subgroup which splits as a direct product. This happens, for instance, in the case of irreducible lattices in semisimple Lie groups.

Coarsification of the co-hopfian property. Recall that a group G is *co-hopfian* if every injective homomorphism $G \to G$ is surjective.

DEFINITION 23.76 (I. Kapovich). A metric space group X is *coarsely co-hopfian* if every quasiisometric embedding $X \to X$ is a quasiisometry.

For instance, one can show that Poincaré Duality groups are coarsely co-hopfian. The same holds for some classes of relative Poincaré Duality groups, see [**KL12**]. Recently, coarse co-hopfian property was verified by S. Merenkov [**Mer10**] for some classes of Gromov-hyperbolic spaces (whose ideal boundaries are homeomorphic to Sierpinsky carpets). One can ask what are other "interesting" examples of coarsely co-hopfian spaces and groups. Here is a concrete open problem:

PROBLEM 23.77. Let X be a thick hyperbolic building. Is it true that X is coarsely co-hopfian?

We conclude with a table of algebraic and geometric properties/invariants/classes of finitely generated groups in relation to their QI invariance. Most of the definitions in this table were introduced earlier in the book. The missing ones are:

1. A group G is *noetherian* if every subgroup of G is finitely generated (sometimes, such groups are called *slender*). Notable examples of noetherian groups are polycyclic groups and Tarski monsters (finitely generated groups where every proper subgroup is cyclic, see [**Ol′91a**]).

2. A group G is *coherent* if every finitely generated subgroup of G is finitely presented (sometimes, such groups are called *artinian*). Notable examples of such groups are the fundamental groups of compact 3-dimensional manifolds and polycyclic groups.

3. A group G is *hyperbolike* if it is a direct limit of a sequence of epimorphisms of hyperbolic groups:

$$G_1 \twoheadrightarrow G_2 \twoheadrightarrow G_3 \twoheadrightarrow \ldots$$

Many examples of such groups are *lacunary hyperbolic*, i.e., finitely generated groups for which one asymptotic cone is a tree, see [**OOS09**]. Hyperbolike groups appear frequently in constructions of group-theoretic *monsters*, which are finitely generated groups satisfying some exotic properties (see e.g. [**Ol′91a**]).

4. A finitely generated group G is said to satisfy *Yu's Property A* if G admits a uniformly proper embedding in a Hilbert space, see [**Yu00**] (there are many other definitions).

5. *Shalom's Property H_{FD}*: A group G is said to satisfy the Property H_{FD} if every unitary representation $\pi : G \to U(\mathcal{H})$ of G with $H^1(G, \mathcal{H}_\pi) \neq 0$, there exists a G-invariant finite-dimensional nonzero subspace $\mathcal{H}' \subset \mathcal{H}$. Shalom proved in [**Sha04**] that Property H_{FD} is a QI invariant among amenable groups.

6. It would take too much space here to define *Poincaré duality groups* and *duality groups*; we refer the reader instead to Brown's book [**Bro82b**]. The same applies to *semihyperbolic* and *automatic groups*; the reader is referred to [**JA95**], [**BH99**] and [**ECH$^+$92**]. Both properties capture some features of nonpositive curvature.

QI invariant	**Not QI invariant**	**Unknown**
hyperbolic, Corollary 9.43		CAT(-1)
semihyperbolic	CAT(0), Example 23.30	automatic
relatively hyperbolic, Theorem 23.37		hyperbolike
virtually nilpotent, Theorem 14.25	virtually solvable, Theorem 12.40	virtually polycyclic
solvable word problem	simple	solvable conjugacy problem
virtually free, Theorem 18.44	colarge	small
finite	residually finite	torsion
	virtually torsion-free	bounded generation property
amenable, Theorem 16.12	virtually metabelian, Theorem 12.40	elementary amenable
	hopfian	noetherian
	co-hopfian	coherent
Yu's Property A	Properties T & Haagerup, Theorem 17.62 and [**CAPV14**]	LT/a-LT
Property H_{FD} for amenable groups		
amalgam/HNN with finite edge groups	amalgam/HNN	contains proper infinite subgroups of infinite index
virtually splits with virtually cyclic edge groups	virtually splits	
Type $\mathbf{F}_n$, Theorem 6.55		
virtually a closed surface group, Corollary 21.24		
VI to a closed 3-manifold group, Theorem 23.70		VI to a compact 3-manifold group
cohomological dimension n over $\mathbb{Q}$, Theorem 6.62		Poincaré duality group over $\mathbb{Q}$
		duality group over $\mathbb{Q}$

Bibliography

[AB56] L. V. Ahlfors and A. Beurling, *The boundary correspondence under quasiconformal mappings*, Acta Math. **96** (1956), 125–142.

[ABDY13] A. Abrams, N. Brady, P. Dani, and R. Young, *Homological and homotopical Dehn functions are different*, Proc. Natl. Acad. Sci. USA **110** (2013), no. 48, 19206–19212.

[Ado36] I. D. Ado, *Note on the representation of finite continuous groups by means of linear substitutions*, Bull. Soc. Phys.-Math. Kazan **7** (1936), 3–43.

[Ady79] S.I. Adyan, *The Burnside problem and identities in groups*, Springer, 1979.

[Ady82] ———, *Random walks on free periodic groups*, Izv. Akad. Nauk SSSR Ser. Mat. **46** (1982), no. 6, 1139–1149.

[Aea91] J. M. Alonso and et al., *Notes on word hyperbolic groups*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, Edited by H. Short, pp. 3–63.

[AG99] D. Allcock and S. Gersten, *A homological characterization of hyperbolic groups*, Invent. Math. **135** (1999), 723–742.

[Ago13] I. Agol, *The virtual Haken conjecture*, Doc. Math. **18** (2013), 1045–1087, With an appendix by Agol, Daniel Groves, and Jason Manning.

[Ahl06] L. V. Ahlfors, *Lectures on quasiconformal mappings*, second ed., University Lecture Series, vol. 38, American Mathematical Society, Providence, RI, 2006, With supplemental chapters by C. J. Earle, I. Kra, M. Shishikura and J. H. Hubbard.

[AK11] Y. Algom-Kfir, *Strongly contracting geodesics in outer space*, Geom. Topol. **15** (2011), no. 4, 2181–2233.

[Alp82] R. Alperin, *Locally compact groups acting on trees and Property T*, Monatsh. Math. **93** (1982), 261–265.

[AŁŚ15] S. Antoniuk, T. Łuczak, and J. Świątkowski, *Random triangular groups at density* 1/3, Compos. Math. **151** (2015), no. 1, 167–178. MR 3305311

[AM15] Y. Antolín and A. Minasyan, *Tits alternatives for graph products*, J. Reine Angew. Math. **704** (2015), 55–83. MR 3365774

[And05] J. W. Anderson, *Hyperbolic geometry*, Springer, 2005.

[APCG13] J. Aliste-Prieto, D. Coronel, and J.-M. Gambaudo, *Linearly repetitive Delone sets are rectifiable*, Ann. Inst. H. Poincaré Anal. Non Linéaire **30** (2013), no. 2, 275–290.

[Att94] O. Attie, *Quasi-isometry classification of some manifolds of bounded geometry*, Math. Z. **216** (1994), 501–527.

[AVS57] G.M. Adelson-Velskii and Yu. A. Sreider, *The banach mean on groups*, Uspehi Mat. Nauk. (N.S.) **126** (1957), 131–136.

[AW81] C.A. Akemann and M.E. Walter, *Unbounded negative definite functions*, Canad. J. Math. **33** (1981), 862–871.

[Bal95] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Seminar, Band 25, Birkhäuser, 1995.

[Bas72] H. Bass, *The degree of polynomial growth of finitely generated nilpotent groups*, Proc. London Math. Soc. **25** (1972), 603–614.

[Bas76] ———, *Euler characteristics and characters of discrete groups*, Invent. Math. **35** (1976), 155–196.

[Bat99] M. Batty, *Groups with a sublinear isoperimetric inequality*, IMS Bulletin **42** (1999), 5–10.

[Bav91] C. Bavard, *Longueur stable des commutateurs*, Enseign. Math. (2) **37** (1991), no. 1-2, 109–150.

[BB95] W. Ballmann and M. Brin, *Orbihedra of nonpositive curvature*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 82, 169–209 (1996).

[BBB+10] L. Bessières, G. Besson, M. Boileau, S. Maillot, and J. Porti, *Geometrisation of 3-manifolds*, EMS Tracts in Mathematics, vol. 13, European Mathematical Society, Zurich, 2010.

[BBI01] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Math., vol. 33, American Mathematical Society, Providence, RI, 2001.

[BC01] R. Bishop and R. Crittenden, *Geometry of manifolds*, AMS Chelsea Publishing, Providence, RI, 2001, Reprint of the 1964 original.

[BC12] J. Behrstock and R. Charney, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. **352** (2012), no. 2, 339–356.

[BCG96] G. Besson, G. Courtois, and S. Gallot, *Minimal entropy and Mostow's rigidity theorems*, Ergodic Theory Dynam. Systems **16** (1996), 623–649.

[BCG98] ———, *A real Schwarz lemma and some applications*, Rend. Mat. Appl. (7) **18** (1998), no. 2, 381–410.

[BCM12] J. Brock, R. Canary, and Y. Minsky, *The classification of Kleinian surface groups, II: The Ending Lamination Conjecture*, Ann. of Math. (2) **176** (2012), no. 1, 1–149.

[BDCK66] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine, *Lois stables et espaces L^p*, Ann. IHP, section B **2** (1966), no. 3, 231–259.

[BdlHV08] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, Cambridge, 2008.

[BDM09] J. Behrstock, C. Druţu, and L. Mosher, *Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity*, Math. Annalen **344** (2009), 543–595.

[BE12] L. Bartholdi and A. Erschler, *Growth of permutational extensions*, Invent. Math. **189** (2012), no. 2, 431–455.

[Bea83] A.F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer, 1983.

[Bel03] I. Belegradek, *On cohopfian nilpotent groups*, Bull. London Math. Soc. **35** (2003), 805–811.

[Ber68] G. Bergman, *On groups acting on locally finite graphs*, Annals of Math. **88** (1968), 335–340.

[Bes14] M. Bestvina, *Geometric group theory and 3-manifolds hand in hand: the fulfillment of Thurston's vision*, Bull. Amer. Math. Soc. (N.S.) **51** (2014), no. 1, 53–70.

[BF92] M. Bestvina and M. Feighn, *A combination theorem for negatively curved groups*, J. Diff. Geom. **35** (1992), no. 1, 85–101.

[BF95] ———, *Stable actions of groups on real trees*, Invent. Math. **121** (1995), no. 2, 287–321.

[BF10] N. Brady and M. Forester, *Density of isoperimetric spectra*, Geom. Topol. **14** (2010), no. 1, 435–472.

[BFGM07] U. Bader, A. Furman, T. Gelander, and N. Monod, *Property (T) and rigidity for actions on Banach spaces*, Acta Math. **198** (2007), 57–105.

[BFH97] M. Bestvina, M. Feighn, and M. Handel, *Laminations, trees, and irreducible automorphisms of free groups*, Geom. Funct. Anal. **7** (1997), 215–244.

[BFH00] ———, *The Tits alternative for* $\mathrm{Out}(F_n)$*. I. Dynamics of exponentially-growing automorphisms.*, Ann. of Math. **151** (2000), no. 2, 517–623.

[BFH04] ———, *Solvable subgroups of* $\mathrm{Out}(F_n)$ *are virtually Abelian*, Geom. Dedicata **104** (2004), 71–96.

[BFH05] ———, *The Tits alternative for* $\mathrm{Out}(F_n)$*. II. A Kolchin type theorem*, Ann. of Math. (2) **161** (2005), no. 1, 1–59.

[BG03] E. Breuillard and T. Gelander, *On dense free subgroups of Lie groups*, J. Algebra **261** (2003), no. 2, 448–467.

[BG08a] J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbb{F}_p)$, Ann. of Math. (2) **167** (2008), no. 2, 625–642.

[BG08b] E. Breuillard and T. Gelander, *Uniform independence in linear groups*, Inventiones Math. **173** (2008), 225–263.

[BGS85] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of non-positive curvature*, Progress in Math., vol. 61, Birkhauser, 1985.

[BGT12] E. Breuillard, B. Green, and T. Tao, *The structure of approximate groups*, Publ. Math. Inst. Hautes Études Sci. **116** (2012), 115–221.

[BH99] M. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Springer-Verlag, Berlin, 1999.

[BH05] M. Bridson and J. Howie, *Conjugacy of finite subsets in hyperbolic groups*, Internat. J. Algebra Comput. **5** (2005), 725–756.

[Bie76a] R. Bieri, *Homological dimension of discrete groups*, Mathematical Notes, Queen Mary College, 1976.

[Bie76b] ———, *Normal subgroups in duality groups and in groups of cohomological dimension* 2, J. Pure Appl. Algebra **1** (1976), 35–51.

[Bie79] ———, *Finitely presented soluble groups*, Séminaire d'Algèbre Paul Dubreil 31ème année (Paris, 1977–1978), Lecture Notes in Math., vol. 740, Springer, Berlin, 1979, pp. 1–8.

[Big01] S. Bigelow, *Braid groups are linear*, J. Amer. Math. Soc. **14** (2001), no. 2, 471–486.

[Bis96] C. Bishop, *On a theorem of Beardon and Maskit*, Ann. Acad. Sci. Fenn. Math. **21** (1996), 383–388.

[BJN09] J. Behrstock, T. Januszkiewicz, and W. Neumann, *Commensurability and QI classification of free products of finitely generated abelian groups*, Proc. Amer. Math. Soc. **137** (2009), no. 3, 811–813.

[BJN10] ———, *Quasi-isometric classification of some high dimensional right-angled Artin groups*, Groups Geom. Dyn. **4** (2010), no. 4, 681–692.

[BK98] D. Burago and B. Kleiner, *Separated nets in Euclidean space and Jacobians of bi-Lipschitz maps*, Geom. Funct. Anal. **8** (1998), no. 2, 273–282.

[BK02a] M. Bonk and B. Kleiner, *Quasisymmetric parametrizations of two-dimensional metric spheres*, Invent. Math. **150** (2002), no. 1, 127–183.

[BK02b] D. Burago and B. Kleiner, *Rectifying separated nets*, Geom. Funct. Anal. **12** (2002), 80–92.

[BK05] M. Bonk and B. Kleiner, *Conformal dimension and Gromov hyperbolic groups with 2-sphere boundary*, Geom. Topol. **9** (2005), 219–246 (electronic).

[BK13a] ———, *Quasisymmetric homeomorphisms and Mostow rigidity*, Preprint, 2013.

[BK13b] M. Bourdon and B. Kleiner, *Combinatorial modulus, the combinatorial Loewner property, and Coxeter groups*, Groups Geom. Dyn. **7** (2013), no. 1, 39–107.

[BKMM12] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher, *Geometry and rigidity of mapping class groups*, Geom. Topol. **16** (2012), no. 2, 781–888.

[BL12] A. Bartels and W. Lück, *The Borel conjecture for hyperbolic and* CAT(0)*-groups*, Ann. of Math. (2) **175** (2012), no. 2, 631–689.

[BLW10] A. Bartels, W. Lück, and S. Weinberger, *On hyperbolic groups with spheres as boundary*, J. Differential Geom. **86** (2010), no. 1, 1–16.

[BM00] M. Burger and S. Mozes, *Lattices in product of trees*, Inst. Hautes Études Sci. Publ. Math. (2000), no. 92, 151–194.

[BN08] J. Behrstock and W. Neumann, *Quasi-isometric classification of graph manifold groups*, Duke Math. J. **141** (2008), 217–240.

[BN12] ———, *Quasi-isometric classification of non-geometric 3-manifold groups*, Crelle Journal **669** (2012), 101–120.

[Bol79] B. Bollobás, *Graph theory, an introductory course*, Graduate Texts in Mathematics, vol. 63, Springer, 1979.

[Bon11] M. Bonk, *Uniformization of Sierpinski carpets in the plane*, Inventiones Math. **186** (2011), 559–665.

[Boo57] W. Boone, *Certain simple, unsolvable problems of group theory. V, VI*, Nederl. Akad. Wetensch. Proc. Ser. A. 60, Indag. Math. **19** (1957), 22–27, 227–232.

[Bor60] A. Borel, *Density properties for certain subgroups of semisimple groups without compact components*, Ann. of Math. **72** (1960), 179–188.

[Bor63] ———, *Compact Clifford-Klein forms of symmetric spaces*, Topology **2** (1963), 111–122.

[Bou63] N. Bourbaki, *Éléments de mathématique. Fascicule XXIX. Livre VI: Intégration. Chapitre 7: Mesure de Haar. Chapitre 8: Convolution et représentations*, Actualités Scientifiques et Industrielles, No. 1306, Hermann, Paris, 1963.

[Bou65] ———, *Topologie générale*, Hermann, Paris, 1965.

[Bou02] ———, *Lie groups and Lie algebras. Chapters 4–6*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2002, Translated from the 1968 French original by Andrew Pressley.

[Bou16] M. Bourdon, *Cohomologie et actions isométriques propres sur les espaces L_p*, To appear in "Geometry, topology and dynamics in negative curvature", Proceedings of the 2010 Bangalore Conference. LMS Lecture Notes, Vol. 425, 2016.

[Bow91a] B. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, pp. 64–167.

[Bow91b] ———, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, "Group theory from a geometrical point of view"; ICTP, Trieste April 1990 (E. Ghys, A. Haefliger, and A. Verjovsky, eds.), 1991, pp. 64–167.

[Bow94] ———, *Some results on the geometry of convex hulls in manifolds of pinched negative curvature*, Comment. Math. Helv. **69** (1994), no. 1, 49–81.

[Bow95] ———, *A short proof that a sub-quadratic isoperimetric inequality implies a linear one*, Mich. J. Math. **42** (1995), 103–107.

[Bow98a] ———, *Continuously many quasi-isometry classes of* 2*-generator groups*, Comment. Math. Helv. **73** (1998), no. 2, 232–236.

[Bow98b] ———, *A topological characterisation of hyperbolic groups*, Jour. Amer. Math. Soc. **11** (1998), no. 3, 643–667.

[Bow06a] ———, *A course on geometric group theory*, MSJ Memoirs, vol. 16, Mathematical Society of Japan, Tokyo, 2006.

[Bow06b] ———, *Intersection numbers and the hyperbolicity of the curve complex*, J. Reine Angew. Math. **598** (2006), 105–129.

[Bow12] ———, *Relatively hyperbolic groups*, Internat. J. Algebra Comput. **22** (2012), no. 3, 1250016, 66.

[Bož80] M. Bożejko, *Uniformly amenable discrete groups*, Math. Ann. **251** (1980), 1–6.

[BP92] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Springer, 1992.

[BP00] M. Bourdon and H. Pajot, *Rigidity of quasi-isometries for some hyperbolic buildings*, Comment. Math. Helv. **75** (2000), no. 4, 701–736.

[BP03] ———, *Cohomologie l_p et espaces de Besov*, J. Reine Angew. Math. **558** (2003), 85–108.

[Bra99] N. Brady, *Branched coverings of cubical complexes and subgroups of hyperbolic groups*, J. London Math. Soc. (2) **60** (1999), no. 2, 461–480.

[Bre92] L. Breiman, *Probability*, Classics in Applied Mathematics, vol. 7, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, Corrected reprint of the 1968 original.

[Bre09] W. Breslin, *Thick triangulations of hyperbolic n-manifolds*, Pacific J. Math. **241** (2009), no. 2, 215–225.

[Bri00] P. Brinkmann, *Hyperbolic automorphisms of free groups*, Geom. Funct. Anal. **10** (2000), no. 5, 1071–1089.

[Bro81a] R. Brooks, *The fundamental group and the spectrum of the Laplacian*, Comment. Math. Helv. **56** (1981), no. 4, 581–598.

[Bro81b] ———, *Some remarks on bounded cohomology*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, Princeton, N.J., 1981, pp. 53–63.

[Bro82a] ———, *Amenability and the spectrum of the Laplacian*, Bull. Amer. Math. Soc. (N.S.) **6** (1982), no. 1, 87–89.

[Bro82b] K. Brown, *Cohomology of groups*, Graduate Texts in Math., vol. 87, Springer, 1982.

[Bro87] K.S. Brown, *Finiteness properties of groups*, J. Pure Appl. Algebra **44** (1987), 45–75.

[BS69] J. Bell and A. Slomson, *Models and Ultraproducts*, North-Holland, Amsterdam, 1969.

[BS78] R. Bieri and R. Strebel, *Almost finitely presented soluble groups*, Comm. Math. Helv. **53** (1978), 258–278.

[BS80] ———, *Valuations and finitely presented metabelian groups*, Proc. London Math. Soc. **41** (1980), no. 3, 439–464.

[BS97a] W. Ballmann and J. Światkowski, *On l^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes*, Geom. Funct. Anal. **7** (1997), no. 4, 615–645.

[BS97b] I. Benjamini and O. Schramm, *Every graph with a positive Cheeger constant contains a tree with a positive Cheeger constant*, Geom. Funct. Anal. **7** (1997), 403–419.

[BS99] W. Ballman and J. Swiatkowski, *On groups acting on nonpositively curved cubical complexes*, LEnseignement Math. **45** (1999), 51–81.

[BS00] M. Bonk and O. Schramm, *Embeddings of Gromov hyperbolic spaces*, Geom. Funct. Anal. **10** (2000), no. 2, 266–306.

[BT24] S. Banach and A. Tarski, *Sur la décomposition des ensembles de points en parties respectivement congruentes*, Fundamenta Mathematicae **6** (1924), 244–277.

[BT02] J. Burillo and J. Taback, *Equivalence of geometric and combinatorial Dehn functions*, New York J. Math. **8** (2002), 169–179 (electronic).

[Bur99] J. Burillo, *Dimension and fundamental groups of asymptotic cones*, J. London Math. Soc. (2) **59** (1999), 557–572.

[Bus65] H. Busemann, *Extremals on closed hyperbolic space forms*, Tensor (N.S.) **16** (1965), 313–318.

[Bus82] P. Buser, *A note on the isoperimetric constant*, Ann. Sci. École Norm. Sup. (4) **15** (1982), no. 2, 213–230.

[Bus10] ———, *Geometry and spectra of compact Riemann surfaces*, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2010, Reprint of the 1992 edition.

[BV98] M. Burger and A. Valette, *Idempotents in complex group rings: theorems of Zalesskii and Bass revisited*, Journal of Lie Theory (1998), 219–228.

[BW11] M. Bridson and H. Wilton, *On the difficulty of presenting finitely presentable groups*, Groups Geom. Dyn. **5** (2011), no. 2, 301–325.

[BZ88] Yu. Burago and V. Zalgaller, *Geometric inequalities*, Grundlehren der mathematischen Wissenschaften, vol. 285, Springer-Verlag, 1988.

[Cai61] S. S. Cairns, *A simple triangulation method for smooth manifolds*, Bull. Amer. Math. Soc. **67** (1961), 389–390.

[Cal08] D. Calegari, *Length and stable length*, Geom. Funct. Anal. **18** (2008), no. 1, 50–76.

[Cal09] ———, *scl*, MSJ Memoirs, vol. 20, Mathematical Society of Japan, Tokyo, 2009.

[CAPV14] M. Carette, S. Arnt, T. Pillon, and A. Valette, *The Haagerup property is not invariant under quasi-isometry*, Preprint, arXiv:1403.5446, 2014.

[Cas86] J.W. Cassels, *Local fields*, Cambridge University Press, 1986.

[Cas10] C. Cashen, *Quasi-isometries between tubular groups*, Groups, Geometry and Dynamics **4** (2010), 473–516.

[CC92] J. Cannon and D. Cooper, *A characterization of cocompact hyperbolic and finite-volume hyperbolic groups in dimension three*, Trans. Amer. Math. Soc. **330** (1992), 419–431.

[CCJ+01] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, and A. Valette, *Groups with the Haagerup property. Gromov's a-T-menability*, Progress in Mathematics, vol. 197, Birkhäuser, 2001.

[CDH10] I. Chatterji, C. Druţu, and F. Haglund, *Kazhdan and Haagerup properties from the median viewpoint*, Adv. Math. **225** (2010), no. 2, 882–921.

[CDP90] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et théorie des groupes. Les groupes hyperboliques de Gromov*, Lec. Notes Math., vol. 1441, Springer-Verlag, 1990.

[CGT82] J. Cheeger, M. Gromov, and M. Taylor, *Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds*, J. Differential Geom. **17** (1982), no. 1, 15–53.

[Cha01] I. Chavel, *Isoperimetric inequalities*, Cambridge Tracts in Mathematics, vol. 145, Cambridge University Press, Cambridge, 2001, Differential geometric and analytic perspectives.

[Cha06] ———, *Riemannian geometry*, second ed., Cambridge Studies in Advanced Mathematics, vol. 108, Cambridge University Press, Cambridge, 2006, A modern introduction.

[Che70] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in analysis (Papers dedicated to Salomon Bochner, 1969), Princeton Univ. Press, Princeton, N. J., 1970, pp. 195–199.

[Che95] S. Chern, *Complex manifolds without potential theory (with an appendix on the geometry of characteristic classes)*, second ed., Universitext, Springer-Verlag, New York, 1995.

[Cho80] C. Chou, *Elementary amenable groups*, Illinois J. Math. **24** (1980), no. 3, 396–407.

[Cho96] R. Chow, *Groups quasi-isometric to complex hyperbolic space*, Trans. Amer. Math. Soc. **348** (1996), no. 5, 1757–1769.

[CJ94] A. Casson and D. Jungreis, *Convergence groups and Seifert-fibered* 3*-manifolds*, Inventiones Mathematicae **118** (1994), 441–456.

[CL83] D. J. Collins and F. Levin, *Automorphisms and Hopficity of certain Baumslag-Solitar groups*, Arch. Math. (Basel) **40** (1983), no. 5, 385–400.

[Com84] W. W. Comfort, *Topological groups*, Handbook of set-theoretic topology, North-Holland, Amsterdam, 1984, pp. 1143–1263.

[Cor92] K. Corlette, *Archimedean superrigidity and hyperbolic geometry*, Ann. of Math. (2) **135** (1992), no. 1, 165–182.

[CR13] L. Carbone and E. Rips, *Reconstructing group actions*, Internat. J. Algebra Comput. **23** (2013), no. 2, 255–323.

[CSC93] Th. Coulhon and L. Saloff-Coste, *Isopérimétrie pour les groupes et variétés*, Rev. Math. Iberoamericana **9** (1993), 293–314.

[CSGdlH98] T. Ceccherini-Silberstein, R. Grigorchuk, and P. de la Harpe, *Décompositions paradoxales des groupes de Burnside*, C. R. Acad. Sci. Paris Sér. I Math. **327** (1998), no. 2, 127–132.

[Cut01] N. J. Cutland, *Loeb measures in practice: Recent advances*, Lecture Notes in Mathematics, vol. 1751, Springer-Verlag, Berlin, 2001.

[CY75] S. Y. Cheng and S. T. Yau, *Differential equations on Riemannian manifolds and their geometric applications*, Comm. Pure Appl. Math. **28** (1975), no. 3, 333–354.

[Dah03a] F. Dahmani, *Combination of convergence groups*, Geometry and Topology **7** (2003), 933–963.

[Dah03b] ———, *Les groupes relativement hyperboliques et leurs bords*, Ph.D. thesis, University of Strasbourg, 2003.

[Dav08] M. Davis, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series, vol. 32, Princeton University Press, Princeton, NJ, 2008.

[Day50] M.M. Day, *Means for the bounded functions and ergodicity of the bounded representations of semi-groups*, Trans. Amer. Math. Soc. **69** (1950), 276–291.

[Day57] ———, *Amenable semigroups*, Illinois J. Math. **1** (1957), 509–544.

[dC92] M. P. do Carmo, *Riemannian geometry*, Birkhäuser Boston Inc., Boston, MA, 1992.

[dC06] Y. de Cornulier, *Strongly bounded groups and infinite powers of finite groups*, Comm. Algebra **34** (2006), 2337–2345.

[dC08] ———, *Dimension of asymptotic cones of lie groups*, J. Topology **1** (2008), no. 2, 343–361.

[dC09] ———, *Asymptotic cones of Lie groups and cone equivalences*, preprint, arXiv:0907.2546, 2009.

[dCTV08] Y. de Cornulier, R. Tessera, and A. Valette, *Isometric group actions on Banach spaces and representations vanishing at infinity*, Transform. Groups **13** (2008), no. 1, 125–147.

[DD89] W. Dicks and M. J. Dunwoody, *Groups acting on graphs*, Cambridge University Press, Cambridge-New York, 1989.

[DD14] K. Dekimpe and J. Dere, *Expanding maps and non-trivial self-covers of infra-nilmanifolds*, Preprint, arXiv:1407.810, 2014.

[dDW84] L. Van den Dries and A. Wilkie, *Gromov's theorem on groups of polynomial growth and elementary logic*, J. Algebra **89** (1984), 349–374.

[Del77] P. Delorme, 1*-cohomologie des représentations unitaires des groupes de Lie semi-simples et résolubles. Produits tensoriels et représentations*, Bull. Soc. Math. France **105** (1977), 281–336.

[Del96] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), no. 3, 661–682.

[DF66] D. S. Dummit and R. M. Foote, *Abstract Algebra, 3rd Edition*, Allyn and Bacon, Inc., Boston, 1966.

[DG] T. Delzant and M. Gromov, *Groupes de Burnside et géométrie hyperbolique*, preprint.

[DG08] F. Dahmani and D. Groves, *The isomorphism problem for toral relatively hyperbolic groups*, Publ. Math. of IHES **107** (2008), 211–290.

[DG11] F. Dahmani and V. Guirardel, *The isomorphism problem for all hyperbolic groups*, Geom. Funct. Anal. **21** (2011), 223–300.

[DGP11] F. Dahmani, V. Guirardel, and P. Przytycki, *Random groups do not split*, Mathematische Annalen **349** (2011), 657–673.

[dlH00] P. de la Harpe, *Topics in geometric group theory*, Chicago Lectures in Mathematics, University of Chicago Press, 2000.

[dlHGCS99] P. de la Harpe, R. I. Grigorchuk, and T. Ceccherini-Silʹberstaĭn, *Amenability and paradoxical decompositions for pseudogroups and discrete metric spaces*, Proc. Steklov Inst. Math. **224** (1999), 57–97.

[dlHV89] P. de la Harpe and A. Valette, *La propriété (T) de Kazhdan pour les groupes localemant compacts*, Astérisque, vol. 175, Société Mathématique de France, 1989.

[Dod84] J. Dodziuk, *Difference equations, isoperimetric inequality and transience of certain random walks*, Trans. Amer. Math. Soc. **284** (1984), no. 2, 787–794.

[DP98] A. Dyubina and I. Polterovich, *Structures at infinity of hyperbolic spaces*, Russian Math. Surveys **53** (1998), no. 5, 1093–1094.

[DP01] ———, *Explicit constructions of universal $\mathbb{R}$–trees and asymptotic geometry of hyperbolic spaces*, Bull. London Math. Soc. **33** (2001), 727–734.

[DPT15] T. Dymarz, I. Peng, and J. Taback, *Bilipschitz versus quasi-isometric equivalence for higher rank lamplighter groups*, New York J. Math. **21** (2015), 129–150.

[DR09] M. Duchin and K. Rafi, *Divergence of geodesics in Teichmüller space and the Mapping Class group*, GAFA **19** (2009), 722–742.

[DR13] W. Dison and T. Riley, *Hydra groups*, Comment. Math. Helv. **88** (2013), no. 3, 507–540.

[Dre84] A. W. M. Dress, *Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces*, Adv. in Math. **53** (1984), no. 3, 321–402.

[Dru01] C. Druţu, *Cônes asymptotiques et invariants de quasi-isométrie pour des espaces métriques hyperboliques*, Ann. Inst. Fourier (Grenoble) **51** (2001), no. 1, 81–97.

[Dru09] ———, *Relatively hyperbolic groups: geometry and quasi-isometric invariance*, Comment. Math. Helv. **84** (2009), no. 3, 503–546.

[DS84] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, vol. 22, Mathematical Association of America, Washington, DC, 1984.

[DS88] N. Dunford and J. T. Schwartz, *Linear operators. Part I. General theory.*, John Wiley & Sons, Inc., New York, 1988.

[DS99] M. J. Dunwoody and M. Sageev, *JSJ splittings for finitely presented groups over slender groups*, Invent. Math. **135** (1999), no. 1, 25–44.

[DS05a] C. Druţu and M. Sapir, *Relatively Hyperbolic Groups with Rapid Decay Property*, Int. Math. Res. Notices **19** (2005), 1181–1194.

[DS05b] ———, *Tree-graded spaces and asymptotic cones of groups*, Topology **44** (2005), 959–1058, with an appendix by D. Osin and M. Sapir.

[DS07] ———, *Groups acting on tree-graded spaces and splittings of relatively hyperbolic groups*, Adv. Math. **217** (2007), 1313–1367.

[DSS95] W. A. Deuber, M. Simonovits, and V. T. Sós, *A note on paradoxical metric spaces*, Studia Sci. Math. Hungar. **30** (1995), no. 1-2, 17–23.

[Dun85] M. J. Dunwoody, *The accessibility of finitely presented groups*, Inventiones Mathematicae **81** (1985), 449–457.

[Dun93] ———, *An inaccessible group*, Geometric group theory, Vol. 1 (Sussex, 1991) (G. A. Niblo and M. A. Roller, eds.), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 75–78.

[Dym10] T. Dymarz, *Bilipschitz equivalence is not equivalent to quasi-isometric equivalence for finitely generated groups*, Duke Math. J. **154** (2010), no. 3, 509–526.

[Dyu00] A. Dyubina, *Instability of the virtual solvability and the property of being virtually torsion-free for quasi-isometric groups*, Internat. Math. Res. Notices **21** (2000), 1097–1101.

[Ebe72] P. Eberlein, *Geodesic flow on certain manifolds without conjugate points*, Transaction of AMS **167** (1972), 151– 170.

[ECH+92] D.B.A. Epstein, J. Cannon, D.F. Holt, S. Levy, M.S. Paterson, and W.P. Thurston, *Word Processing and Group Theory*, Jones and Bartlett, 1992.

[EF97a] D. B. A. Epstein and K. Fujiwara, *The second bounded cohomology of word-hyperbolic groups*, Topology **36** (1997), no. 6, 1275–1289.

[EF97b] A. Eskin and B. Farb, *Quasi-flats and rigidity in higher rank symmetric spaces*, J. Amer. Math. Soc. **10** (1997), no. 3, 653–692.

[Efr53] V. A. Efremovič, *The proximity geometry of riemannian manifolds*, Uspehi Matem. Nauk (N.S.) **8** (1953), 189.

[EFW12] A. Eskin, D. Fisher, and K. Whyte, *Coarse differentiation of quasi-isometries I: Spaces not quasi-isometric to Cayley graphs*, Ann. of Math. (2) **176** (2012), no. 1, 221–260.

[EFW13] ———, *Coarse differentiation of quasi-isometries II: Rigidity for Sol and lamplighter groups*, Ann. of Math. (2) **177** (2013), no. 3, 869–910.

[Eng95] R. Engelking, *General Topology*, Sigma Series in Pure Mathematics, vol. 6, Heldermann Verlag, 1995.

[EO73] P. Eberlein and B. O'Neill, *Visibility manifolds*, Pacific J. Math. **46** (1973), 45–109.

[Ers03] A. Erschler, *On Isoperimetric Profiles of Finitely Generated Groups*, Geometriae Dedicata **100** (2003), 157–171.

[Ers04] ———, *Not residually finite groups of intermediate growth, commensurability and non-geometricity*, J. Algebra **272** (2004), no. 1, 154–172.

[Ers06] ———, *Piecewise automatic groups*, Duke Math. J. **134** (2006), no. 3, 591–613.

[Esk98] A. Eskin, *Quasi-isometric rigidity of nonuniform lattices in higher rank symmetric spaces*, J. Amer. Math. Soc. **11** (1998), no. 2, 321–361.

[ET64] V. Efremovich and E. Tihomirova, *Equimorphisms of hyperbolic spaces*, Izv. Akad. Nauk SSSR **28** (1964), 1139– 1144.

[Far97] B. Farb, *The quasi-isometry classification of lattices in semisimple Lie groups*, Math. Res. Letters **4** (1997), no. 5, 705–717.

[Far98] ———, *Relatively hyperbolic groups*, Geom. Funct. Anal. **8** (1998), no. 5, 810–840.

[Fed69] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.

[FF60] H. Federer and W. H. Fleming, *Normal and integral currents*, Ann. of Math. (2) **72** (1960), 458–520.

[FH74] J. Faraut and K. Harzallah, *Distances hilbertiennes invariantes sur un espace homogène*, Ann. Institut Fourier **3** (1974), no. 24, 171–217.

[FH94] W. Fulton and J. Harris, *Representation theory: A first course*, Springer, 1994.

[FJ03] R. Fleming and J. Jamison, *Isometries on Banach spaces: Function spaces*, Monographs and Surveys in Pure and Applied Mathematics, CRC Press, 2003.

[FK13] K. Fujiwara and M. Kapovich, *On quasihomomorphisms with noncommutative targets*, Preprint, 2013.

[FM98] B. Farb and L. Mosher, *A rigidity theorem for the solvable Baumslag–Solitar groups*, Invent. Math. **131** (1998), no. 2, 419–451, With an appendix by Daryl Cooper.

[FM99] ———, *Quasi-isometric rigidity for the solvable Baumslag–Solitar groups, II*, Invent. Math. **137** (1999), no. 3, 613–649.

[FM00] ———, *On the asymptotic geometry of abelian-by-cyclic groups*, Acta Math. **184** (2000), no. 2, 145–202.

[FM11] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton University Press, 2011.

[Fol99] G. Folland, *Real analysis: Modern techniques and their applications*, Pure and Applied Mathematics, Wiley–Interscience, 1999.

[For73] E. Formanek, *Idempotents in noetherian group rings*, Canad. J. Math. **25** (1973), 366–369.

[Fox53] R. H. Fox, *Free differential calculus. I. Derivation in the free group ring*, Ann. of Math. (2) **57** (1953), 547–560.

[FP06] K. Fujiwara and P. Papasoglu, *JSJ-decompositions of finitely presented groups and complexes of groups*, Geom. Funct. Anal. **16** (2006), no. 1, 70–125.

[Fri60] A. A. Fridman, *On the relation between the word problem and the conjugacy problem in finitely defined groups*, Trudy Moskov. Mat. Obšč. **9** (1960), 329–356.

[FS88] M. Freedman and R. Skora, *Strange actions of groups on spheres. II*, "Holomorphic functions and moduli", Vol. II (Berkeley, CA, 1986), Springer, New York, 1988, pp. 41–57.

[FS96] B. Farb and R. Schwartz, *The large-scale geometry of Hilbert modular groups*, J. Differential Geom. **44** (1996), no. 3, 435–478.

[FW91] M. Foreman and F. Wehrung, *The Hahn-Banach theorem implies the existence of a non-Lebesgue measurable set*, Fundam. Math. **138** (1991), 13–19.

[G6̈0] P. Günther, *Einige Sätze über das Volumenelement eines Riemannschen Raumes*, Publ. Math. Debrecen **7** (1960), 78–93.

[Gab92] D. Gabai, *Convergence groups are Fuchsian groups*, Annals of Math. **136** (1992), 447–510.

[Gab05] D. Gaboriau, *Examples of groups that are measure equivalent to the free group*, Ergodic Theory Dynam. Systems **25** (2005), no. 6, 1809–1827.

[Gab10] ———, *Orbit equivalence and measured group theory*, Proceedings of the International Congress of Mathematicians. Volume III (New Delhi), Hindustan Book Agency, 2010, pp. 1501–1527.

[Gar73] H. Garland, *p-adic curvature and cohomology of discrete subgroups of p-adic groups*, Ann. of Math. **97** (1973), 375–423.

[Gau73] C. F. Gauß, *Werke. Band VIII*, Georg Olms Verlag, Hildesheim, 1973, Reprint of the 1900 original.

[GdlH90] E. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d'apres Mikhael Gromov*, Progress in Mathematics, vol. 83, Birkhäuser, 1990.

[Gel11] T. Gelander, *Volume versus rank of lattices*, J. Reine Angew. Math. **661** (2011), 237–248.

[Gel14] ———, *Lectures on lattices and locally symmetric spaces*, Preprint, http://arxiv.org/pdf/1402.0962.pdf, 2014.

[Geo08] R. Geoghegan, *Topological methods in group theory*, Graduate Texts in Mathematics, vol. 243, Springer, New York, 2008.

[Ger87] S. Gersten, *Reducible diagrams and equations over groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 15–73.

[Ger92] ———, *Bounded cocycles and combings of groups*, Internat. J. Algebra Comput. **2** (1992), no. 3, 307–326.

[Ger93a] ———, *Isoperimetric and isodiametric functions of finite presentations*, Geometric group theory, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 79–96.

[Ger93b] ———, *Quasi-isometry invariance of cohomological dimension*, C. R. Acad. Sci. Paris Sér. I Math. **316** (1993), no. 5, 411–416.

[Ger94] ———, *Quadratic divergence of geodesics in CAT(0)-spaces*, Geom. Funct. Anal. **4** (1994), no. 1, 37–51.

[Ger09] V. Gerasimov, *Expansive convergence groups are relatively hyperbolic*, Geom. Funct. Anal. **19** (2009), no. 1, 137–169.

[Ger12] ———, *Floyd maps for relatively hyperbolic groups*, Geom. Funct. Anal. **22** (2012), no. 5, 1361–1399.

[GHL04] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, third ed., Universitext, Springer-Verlag, Berlin, 2004.

[Ghy04] E. Ghys, *Groupes aléatoires (d'après Misha Gromov,. . .)*, Astérisque (2004), no. 294, viii, 173–204.

[Gol98] R. Goldblatt, *Lectures on the hyperreals*, Graduate Texts in Mathematics, vol. 188, Springer-Verlag, New York, 1998, An introduction to nonstandard analysis.

[GP88] K. Grove and P. Petersen, *Bounding homotopy types by geometry*, Ann. of Math. **128** (1988), 195–206.

[GP10] V. Guillemin and A. Pollack, *Differential topology*, AMS Chelsea Publishing, Providence, RI, 2010, Reprint of the 1974 original.

[GP13] V. Gerasimov and L. Potyagailo, *Quasi-isometric maps and Floyd boundaries of relatively hyperbolic groups*, J. Eur. Math. Soc. (JEMS) **15** (2013), no. 6, 2115–2137.

[GPS88] M. Gromov and I. Piatetski-Shapiro, *Non-arithmetic groups in Lobacevski spaces*, Publ. Math. IHES **66** (1988), 93–103.

[Gre62] L. Greenberg, *Discrete subgroups of the Lorentz group*, Math. Scand. **10** (1962), 85–107.

[Gri83] R. I. Grigorchuk, *On the Milnor problem of group growth*, Dokl. Akad. Nauk SSSR **271** (1983), no. 1, 30–33.

[Gri84a] ———, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk SSSR Ser. Mat. **48** (1984), no. 5, 939–985.

[Gri84b] ———, *The growth rate of finitely generated groups and the theory of invariant means*, Inv. Akad. Nauk. **45** (1984), 939–986.

[Gri87] ———, *Superamenability and the problem of occurence of free semigroups*, Functional Analysis and its Applications **21** (1987), 64–66.

[Gri98] ———, *An example of a finitely presented amenable group that does not belong to the class EG*, Mat. Sb. **189** (1998), no. 1-2, 75–95.

[Gro81a] M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. IHES **53** (1981), 53–73.

[Gro81b] ———, *Hyperbolic manifolds, groups and actions*, Annals of Mathematics Studies, vol. 97, Princeton University Press, Princeton, 1981.

[Gro83] ———, *Infinite groups as geometric objects*, Proceedings of the International Congress of Mathematicians, Warsaw, Amer. Math. Soc., 1983, pp. 385–392.

[Gro86] ———, *Isoperimetric inequalities in Riemannian manifolds*, "Asymptotic Theory of Finite Dimensional Normed Spaces", Lecture Notes Math., vol. 1200, Springer-Verlag, Berlin, 1986, pp. 114–129.

[Gro87] ———, *Hyperbolic groups*, "Essays in group theory", Math. Sci. Res. Ins. Publ., vol. 8, Springer, 1987, pp. 75–263.

[Gro93] ———, *Asymptotic invariants of infinite groups*, "Geometric groups theory", volume 2, Proc. of the Symp. in Sussex 1991 (G.A. Niblo and M.A. Roller, eds.), Lecture Notes series, vol. 182, Cambridge University Press, 1993.

[Gro03] ———, *Random walk in random groups*, Geom. Funct. Anal. **13** (2003), no. 1, 73–146.

[Gro07] ———, *Metric structures for Riemannian and non-Riemannian spaces*, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2007, based on the 1981 French original, with appendices by M. Katz, P. Pansu and S. Semmes.

[Gru57] K. W. Gruenberg, *Residual properties of infinite soluble groups*, Proc. London Math. Soc. **7** (1957), 29–62.

[GS90] S. Gersten and H. Short, *Small cancellation theory and automatic groups*, Invent. Math. **102** (1990), 305–334.

[GS92] M. Gromov and R. Schoen, *Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one*, Publ. Math. of IHES **76** (1992), 165–246.

[GT83] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, vol. 224, Springer Verlag, Grundlehren der Mathematischen Wissenschaften, 1983.

[GT87] M. Gromov and W. Thurston, *Pinching constants for hyperbolic manifolds*, Inventiones Mathematicae **89** (1987), 1–12.

[Gui70] Y. Guivarc'h, *Groupes de Lie à croissance polynomiale*, C. R. Acad. Sci. Paris, Ser. A—B **271** (1970), A237–A239.

[Gui73] ———, *Croissance polynomiale et périodes des fonctions harmonique*, Bull. Soc. Math. France **101** (1973), 333–379.

[Gui77] A. Guichardet, *Étude de la* 1*-cohomologie et de la topologie du dual pour les groupes de Lie à radical abélien*, Math. Ann. **228** (1977), no. 1, 215–232.

[Haï09] P. Haïssinsky, *Géométrie quasiconforme, analyse au bord des espaces métriques hyperboliques et rigidités [d'après Mostow, Pansu, Bourdon, Pajot, Bonk, Kleiner...]*, Astérisque (2009), no. 326, Exp. No. 993, ix, 321–362 (2010), Séminaire Bourbaki. Vol. 2007/2008.

[Haï15] ———, *Hyperbolic groups with planar boundaries*, Invent. Math. **201** (2015), no. 1, 239–307.

[Hal49] M. Hall, *Coset representations in free groups*, Trans. Amer. Math. Soc. **67** (1949), 421–432. MR MR0032642 (11,322e)

[Hal64] J. Halpern, *The independence of the axiom of choice from the boolean prime ideal theorem*, Fund. Math. **55** (1964), 57–66.

[Hal76] M. Hall, *The theory of groups*, Chelsea Publishing Co., New York, 1976, Reprinting of the 1968 edition.

[Ham07] U. Hamenstädt, *Geometry of complex of curves and Teichmüller spaces*, Handbook of Teichmüller Theory, vol. 1, EMS, 2007, pp. 447–467.

[Hat02] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[Hau14] F. Hausdorff, *Bemerkung ber den inhalt von Punktmengen*, Mathematische Annalen **75** (1914), 428–434.

[Hei01] J. Heinonen, *Lectures on analysis on metric spaces*, Universitext, Springer-Verlag, New York, 2001.

[Hei05] ———, *Lectures on Lipschitz analysis*, Report. University of Jyväskylä Department of Mathematics and Statistics, vol. 100, University of Jyväskylä, Jyväskylä, 2005.

[Hel01] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Graduate Studies in Mathematics, vol. 34, Amer. Math. Soc., 2001.

[Hem78] J. Hempel, *3-manifolds*, Annals of Mathematics Studies, Princeton University Press, 1978.

[Hem87] ———, *Residual finiteness for* 3*-manifolds*, "Combinatoral group theory and topology" (S.M. Gersten and J.R. Stallings, eds.), Annals of Mathematics Studies, vol. 111, Princeton University Press, Princeton, 1987, pp. 379–396.

[Hig40] G. Higman, *The units of group-rings*, Proc. London Math. Soc. (2) **46** (1940), 231–248.

[Hir38] K. A. Hirsch, *On infinite soluble groups, II*, Proc. Lond. Math. Soc. **44** (1938), 336–344.

[Hir76] M. Hirsch, *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer, 1976.

[HJ99] K. Hrbacek and T. Jech, *Introduction to set theory*, third ed., Monographs and Textbooks in Pure and Applied Mathematics, vol. 220, Marcel Dekker Inc., New York, 1999.

[HK95] J. Heinonen and P. Koskela, *Definitions of quasiconformality*, Inventiones Math. **120** (1995), 61–79.

[HK98] ———, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math. **181** (1998), 1–61.

[HNN49] G. Higman, B. H. Neumann, and H. Neumann, *Embedding theorems for groups*, J. London Math. Soc. **24** (1949), 247–254. MR MR0032641 (11,322d)

[HP74] E. Hille and R. Phillips, *Functional analysis and semi-groups*, American Mathematical Society, Providence, R. I., 1974, Third printing of the revised edition of 1957, American Mathematical Society Colloquium Publications, Vol. XXXI.

[Hru12] E. Hrushovski, *Stable group theory and approximate subgroups*, Jour. Amer. Math. Soc. (2012), 189–243.

[HS53] G. Hochschild and J.-P. Serre, *Cohomology of group extensions*, Trans. Amer. Math. Soc. **74** (1953), 110–134.

[HŚ08] F. Haglund and J. Świątkowski, *Separating quasi-convex subgroups in 7-systolic groups*, Groups Geom. Dyn. **2** (2008), no. 2, 223–244.

[Hum75] J. E. Humphreys, *Linear algebraic groups*, Springer-Verlag, New York, 1975, Graduate Texts in Mathematics, No. 21.

[Hum97] ———, *Reflection groups and Coxeter groups*, Cambridge University Press, 1997, Cambridge Studies in Advanced Mathematics, Vol. 29.

[Hun80] T. W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York, 1980.

[HW41] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1941.

[HW12] F. Haglund and D. Wise, *A combination theorem for special cube complexes*, Ann. of Math. (2) **176** (2012), no. 3, 1427–1482.

[IM01] T. Iwaniec and G. Martin, *Geometric function theory and non-linear analysis*, Oxford Mathematical Monographs, Oxford University Press, New York, 2001.

[IS98] S. V. Ivanov and P. E. Schupp, *On the hyperbolicity of small cancellation groups and one-relator groups*, Trans. Amer. Math. Soc. **350** (1998), no. 5, 1851–1894.

[Isa90] N. A. Isachenko, *Uniformly quasiconformal discontinuous groups that are not isomorphic to Möbius groups*, Dokl. Akad. Nauk SSSR **313** (1990), no. 5, 1040–1043.

[Iva88] N. Ivanov, *Automorphisms of Teichmüller modular groups*, Topology and geometry—Rohlin Seminar, Lecture Notes in Math., vol. 1346, Springer, Berlin, 1988, pp. 199–270.

[Iva92] ———, *Subgroups of Teichmüller Modular Groups*, Translations of Math. Monographs, vol. 115, AMS, 1992.

[Iva94] S. V. Ivanov, *The free Burnside groups of sufficiently large exponents*, Internat. J. Algebra Comput. **4** (1994), no. 1-2, ii+308 pp.

[Iva96] N. Ivanov, *Actions of Moebius transformations on homeomorphisms: stability and rigidity*, Geom. Funct. Anal. **6** (1996), 79–119.

[JA95] M. Bridson J. Alonso, *Semihyperbolic groups*, Proc. London Math. Soc. (3) **70** (1995), no. 1, 56–114.

[Jec03] Th. Jech, *Set Theory: The Third Millennium Edition, Revised and Expanded*, Springer, 2003.

[Joh48] F. John, *Extremum problems with inequalities as subsidiary conditions*, Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, Interscience Publishers, Inc., New York, N. Y., 1948, pp. 187–204.

[JR88] W. Jaco and J. H. Rubinstein, *PL minimal surfaces in 3-manifolds*, J. Diff. Geom. **27** (1988), no. 3, 493–524.

[JR89] ———, *PL equivariant surgery and invariant decomposition of 3-manifolds*, Advances in Math. **73** (1989), 149–191.

[JŚ03] T. Januszkiewicz and J. Świątkowski, *Hyperbolic Coxeter groups of large dimension*, Comment. Math. Helv. **78** (2003), no. 3, 555–583.

[JŚ06] ———, *Simplicial nonpositive curvature*, Publ. Math. Inst. Hautes Études Sci. (2006), no. 104, 1–85.

[JvN35] P. Jordan and J. von Neumann, *On inner products in linear metric spaces*, Ann. of Math. **36** (1935), 719–723.

[Kak41] S. Kakutani, *Concrete representation of abstract (L)-spaces and the mean ergodic theorem*, Ann. of Math. (2) **42** (1941), 523–537.

[Kap69] I. Kaplansky, *Fields and rings*, The University of Chicago Press, 1969.

[Kap70] ———, *"Problems in the theory of rings" revisited*, Amer. Math. Monthly **77** (1970), 445–454.

[Kap92] M. Kapovich, *Intersection pairing on hyperbolic 4-manifolds*, preprint, 1992.

[Kap96] I. Kapovich, *Detecting quasiconvexity: algorithmic aspects*, Geometric and computational perspectives on infinite groups (Minneapolis, MN and New Brunswick, NJ, 1994), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 25, Amer. Math. Soc., Providence, RI, 1996, pp. 91–99.

[Kap01] M. Kapovich, *Hyperbolic manifolds and discrete groups*, Birkhäuser Boston Inc., Boston, MA, 2001.

[Kap05] ———, *Representations of polygons of finite groups*, Geom. Topol. **9** (2005), 1915–1951.

[Kap07] ———, *Kleinian groups in higher dimensions*, "Geometry and Dynamics of Groups and Spaces. In memory of Alexander Reznikov", vol. 265, Birkhauser, Progress in Mathematics, 2007, pp. 485–562.

[Kap14] ———, *Energy of harmonic functions and Gromov's proof of the Stallings' theorem*, Georgian Math. Journal **21** (2014), 281–296.

[Kaz75] D. Kazhdan, *On arithmetic varieties*, Lie groups and their representations, Halsted, 1975, pp. 151–216.

[KB02] I. Kapovich and N. Benakli, *Boundaries of hyperbolic groups*, Combinatorial and geometric group theory (New York, 2000/Hoboken, NJ, 2001), Contemp. Math., vol. 296, Amer. Math. Soc., Providence, RI, 2002, pp. 39–93.

[Kei76] J. Keisler, *Foundations of Infinitesimal Calculus*, Prindel-Weber-Schmitt, Boston, 1976.

[Kei10] ———, *The ultraproduct construction*, Ultrafilters across mathematics, Contemp. Math., vol. 530, Amer. Math. Soc., Providence, RI, 2010, pp. 163–179.

[Kel72] G. Keller, *Amenable groups and varieties of groups*, Illinois J. Math. **16** (1972), 257–268.

[Kes59] H. Kesten, *Full Banach mean values on countable groups*, Math. Scand. **7** (1959), 146–156.

[KK00] M. Kapovich and B. Kleiner, *Hyperbolic groups with low-dimensional boundary*, Ann. Sci. École Norm. Sup. (4) **33** (2000), no. 5, 647–669.

[KK05] ———, *Coarse Alexander duality and duality groups*, Journal of Diff. Geometry **69** (2005), 279–352.

[KK09] ———, *Appendix to "Lacunary hyperbolic groups", by A. Olshanskii, D. Osin, M. Sapir*, Geometry and Topology **13** (2009), 2132–2137.

[KKL98] M. Kapovich, B. Kleiner, and B. Leeb, *Quasi-isometries and the de Rham decomposition*, Topology **37** (1998), 1193–1212.

[KL95] M. Kapovich and B. Leeb, *On asymptotic cones and quasi-isometry classes of fundamental groups of nonpositively curved manifolds*, Geom. Funct. Anal. **5** (1995), no. 3, 582–603.

[KL97] ———, *Quasi-isometries preserve the geometric decomposition of haken manifolds*, Invent. Math. **128** (1997), no. 2, 393–416.

[KL98a] ———, 3*-manifold groups and nonpositive curvature*, Geom. Funct. Anal. **8** (1998), no. 5, 841–852.

[KL98b] B. Kleiner and B. Leeb, *Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings*, Math. Publ. of IHES **86** (1998), 115–197.

[KL01] ———, *Groups quasi-isometric to symmetric spaces*, Comm. Anal. Geom. **9** (2001), no. 2, 239–260.

[KL08] B. Kleiner and J. Lott, *Notes on Perelman's papers*, Geometry & Topology **12** (2008), no. 5, 2587–2855.

[KL09] B. Kleiner and B. Leeb, *Induced quasi-actions: A remark*, Proc. Amer. Math. Soc. **137** (2009), no. 5, 1561–1567.

[KL12] I. Kapovich and A. Lukyanenko, *Quasi-isometric co-Hopficity of non-uniform lattices in rank-one semi-simple Lie groups*, Conform. Geom. Dyn. **16** (2012), 269–282.

[Kle10] B. Kleiner, *A new proof of Gromov's theorem on groups of polynomial growth*, J. Amer. Math. Soc. **23** (2010), no. 3, 815–829.

[KM12] J. Kahn and V. Markovic, *Immersing almost geodesic surfaces in a closed hyperbolic three manifold*, Ann. of Math. (2) **175** (2012), no. 3, 1127–1190.

[KS97] N. Korevaar and R. Schoen, *Global existence theorems for harmonic maps to non-locally compact spaces*, Comm. Anal. Geom. **5** (1997), no. 2, 333–387.

[KS08] I. Kapovich and P. Schupp, *On group-theoretic models of randomness and genericity*, Groups Geom. Dyn. **2** (2008), no. 3, 383–404.

[KSS06] I. Kapovich, P. Schupp, and V. Shpilrain, *Generic properties of Whitehead's algorithm and isomorphism rigidity of random one-relator groups*, Pacific J. Math. **223** (2006), no. 1, 113–140.

[KSTT05] L. Kramer, S. Shelah, K. Tent, and S. Thomas, *Asymptotic cones of finitely presented groups*, Adv. Math. **193** (2005), no. 1, 142–173.

[Kun80] K. Kunen, *Set Theory: An Introduction to Independence Proofs*, Elsevier, 1980.

[KW92] V. Kaimanovich and W. Woess, *The Dirichlet problem at infinity for random walks on graphs with a strong isoperimetric inequality*, Probab. Theory Related Fields **91** (1992), 445–466.

[Laf04] J.-F. Lafont, *Rigidity results for certain 3-dimensional singular spaces and their fundamental groups*, Geom. Dedicata **109** (2004), 197–219.

[Lan64] S. Lang, *Algebraic numbers*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-Palo Alto-London, 1964.

[Lan00] Urs Lang, *Higher-dimensional linear isoperimetric inequalities in hyperbolic groups*, Internat. Math. Res. Notices (2000), no. 13, 709–717.

[Lan02] S. Lang, *Algebra*, Addison-Wesley Publishing Company, 2002.

[Lee12] S. R. Lee, *Geometry of Houghton's groups*, Preprint, arXiv:1212.0257, 2012.

[Leh87] O. Lehto, *Univalent functions and Teichmüller spaces*, Springer, 1987.

[Li12] P. Li, *Geometric analysis*, Cambridge Studies in Advanced Mathematics, vol. 134, Cambridge University Press, Cambridge, 2012.

[ŁRN51] J. Łoś and C. Ryll-Nardzewski, *On the application of Tychonoff's theorem in mathematical proofs*, Fund. Math. **38** (1951), 233–237.
[LS77] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*, Springer-Verlag, Berlin, 1977, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.
[LT79] J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces. II: Function spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], vol. 97, Springer-Verlag, Berlin, 1979.
[LT92] P. Li and L. Tam, *Harmonic functions and the structure of complete manifolds*, J. Differential Geom. **35** (1992), 359–383.
[Lux62] W. A. J. Luxemburg, *Two applications of the method of construction by ultrapowers to anaylsis*, Bull. Amer. Math. Soc. **68** (1962), 416–419.
[Lux67] ———, *Beweis des satzes von Hahn-Banach*, Arch. Math (Basel) **18** (1967), 271–272.
[Lux69] ———, *Reduced powers of the real number system and equivalents of the Hahn-Banach extension theorem*, Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967), Holt, Rinehart and Winston, New York, 1969, pp. 123–137.
[Lys96] I.G. Lysenok, *Infinite Burnside groups of even exponent*, Izvestiya Akad. Nauk. SSSR Ser., Mat **60** (1996), no. 3, 453–654.
[Mac] N. Macura, *CAT(0) spaces with polynomial divergence of geodesics*, preprint 2011.
[Mag39] W. Magnus, *On a theorem of Marshall Hall*, Ann. of Math. (2) **40** (1939), 764–768.
[Mal40] A. I. Mal'cev, *On isomorphic matrix representations of infinite groups*, Mat. Sb. **8** (1940), 405–422.
[Mal49a] ———, *Generalized nilpotent algebras and their associated groups*, Mat. Sbornik N.S. **25(67)** (1949), 347–366.
[Mal49b] ———, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk. SSSR Ser., Mat **13** (1949), 9–32, Amer. Math. Soc. Translation No. 39 (1951).
[Mar86] G. Martin, *Discrete quasiconformal groups which are not quasiconformal conjugates of Mobius groups*, Ann. Ac. Sci. Fenn. **11** (1986), 179–202.
[Mar91] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag, Berlin, 1991.
[Mar13] V. Markovic, *Criterion for Cannon's conjecture*, Geom. Funct. Anal. **23** (2013), no. 3, 1035–1061. MR 3061779
[Mas91] W. S. Massey, *A basic course in algebraic topology*, Graduate Texts in Mathematics, vol. 127, Springer, 1991.
[McM98] C. T. McMullen, *Lipschitz maps and nets in Euclidean space*, Geom. Funct. Anal. **8** (1998), no. 2, 304–314.
[ME] M. Sapir M. Ershov, G. Golan, *The Tarski numbers of groups*, Preprint, arXiv:1401.2202, 2014.
[Mer10] S. Merenkov, *A Sierpinsky carpet with the co-hopfian property*, Inventiones Math. **180** (2010), 361–388.
[Mes72] S. Meskin, *Nonresidually finite one-relator groups*, Trans. Amer. Math. Soc. **164** (1972), 105–114.
[Mil68a] J. Milnor, *Growth of finitely generated solvable groups*, J. Diff. Geom. **2** (1968), 447–449.
[Mil68b] ———, *A note on curvature and fundamental group*, J. Diff. Geom. **2** (1968), 1–7.
[Mil68c] ———, *Problem 5603*, Amer. Math. Monthly **75** (1968), 685–686.
[Mil76] J. J. Millson, *On the first betti number of a constant negatively curved manifold*, Ann. of Math. **104** (1976), no. 2, 235–247.
[Mil79] ———, *Real vector bundles with discrete structure group*, Topology **18** (1979), no. 1, 83–89.
[Mil12] J.S. Milne, *Group theory*, http://www.jmilne.org/math/CourseNotes/GT.pdf, 2012.
[Min01] I. Mineyev, *Straightening and bounded cohomology of hyperbolic groups*, Geom. Funct. Anal. **11** (2001), no. 4, 807–839.
[Min10] Yair Minsky, *The classification of Kleinian surface groups. I. Models and bounds*, Ann. of Math. (2) **171** (2010), no. 1, 1–107.
[Mj14a] M. Mj, *Cannon-Thurston maps for surface groups*, Ann. of Math. (2) **179** (2014), no. 1, 1–80.

[Mj14b] ———, *Ending laminations and Cannon-Thurston maps*, Geom. Funct. Anal. **24** (2014), no. 1, 297–321, With an appendix by Shubhabrata Das and Mj.

[MK] J. Porti M. Kapovich, B. Leeb, *A Morse Lemma for quasigeodesics in symmetric spaces and euclidean buildings*, Preprint, 2014.

[MN82] J. Meakin and K. S. S. Nambooripad, *Coextensions of regular semigroups by rectangular bands. I*, Trans. Amer. Math. Soc. **269** (1982), no. 1, 197–224.

[MNLGO92] J. C. Mayer, J Nikiel, and L. G. L. G. Oversteegen, *Universal spaces for $\mathbb{R}$-trees*, Trans. Amer. Math. Soc. **334** (1992), no. 1, 411–432.

[Mon06] N. Monod, *An invitation to bounded cohomology*, International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 1183–1211.

[Mor24] M. Morse, *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of AMS **26** (1924), 25–60.

[Mor38] C. B. Morrey, *On the solution of quasilinear elliptic partial differential equations*, Transactions of AMS **43** (1938), 126–166.

[Mos65] G. D. Mostow, *Quasiconformal mappings in n-space and the rigidity of hyperbolic space-forms*, Publ. Math. IHES **34** (1965), 53–104.

[Mos73] ———, *Strong Rigidity of Locally Symmetric Spaces*, Annals of Math. Studies, no. 78, Princeton Univ. Press, 1973.

[Mou88] G. Moussong, *Hyperbolic Coxeter groups*, Ph.D. thesis, The Ohio State University, 1988.

[MR81] J. J. Millson and M. S. Raghunathan, *Geometric construction of cohomology for arithmetic groups. I*, Proc. Indian Acad. Sci. Math. Sci. **90** (1981), no. 2, 103–123.

[MR03] C. Maclachlan and A. Reid, *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics, vol. 219, Springer-Verlag, New York, 2003.

[MR08] M. Mj and L. Reeves, *A combination theorem for strong relative hyperbolicity*, Geom. Topol. **12** (2008), no. 3, 1777–1798.

[MS86] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, Berlin, 1986, With an appendix by M. Gromov.

[MSW03] L. Mosher, M. Sageev, and K. Whyte, *Quasi-actions on trees. I. Bounded valence*, Ann. of Math. (2) **158** (2003), no. 1, 115–164.

[Mun75] J. R. Munkres, *Topology: a first course*, Prentice-Hall Inc., 1975.

[Mur05] A. Muranov, *Diagrams with selection and method for constructing boundedly generated and boundedly simple groups*, Comm. Algebra **33** (2005), no. 4, 1217–1258.

[MY81] W. Meeks and S.T. Yau, *The equivariant Dehn's lemma and equivariant loop theorem*, Comment. Math. Helv. **56** (1981), 225– 239.

[MY02] I. Mineyev and G. Yu, *The Baum-Connes conjecture for hyperbolic groups*, Invent. Math. **149** (2002), no. 1, 97–122.

[MZ74] D. Montgomery and L. Zippin, *Topological transformation groups*, Robert E. Krieger Publishing Co., Huntington, N.Y., 1974, Reprint of the 1955 original.

[Nag83] J. Nagata, *Modern dimension theory*, revised ed., Sigma Series in Pure Mathematics, vol. 2, Heldermann Verlag, Berlin, 1983.

[Nag85] ———, *Modern general topology*, second ed., North Holland Mathematical Library, vol. 33, North Holland, 1985.

[Nér64] A. Néron, *Modèles minimaux des variétes abèliennes sur les corps locaux et globaux*, Math. Publ. of IHES **21** (1964), 5–128.

[New68] B. B. Newman, *Some results on one-relator groups*, Bull. Amer. Math. Soc. **74** (1968), 568–571.

[Nib04] G. Niblo, *A geometric proof of Stallings' theorem on groups with more than one end*, Geometriae Dedicata **105** (2004), no. 1, 61–76.

[Nic12] Bogdan Nica, *The Mazur-Ulam theorem*, Expo. Math. **30** (2012), no. 4, 397–398.

[Nic13] B. Nica, *Proper isometric actions of hyperbolic groups on L^p-spaces*, Compos. Math. **149** (2013), no. 5, 773–792.

[Nos91] G. Noskov, *Bounded cohomology of discrete groups with coefficients*, Leningrad Math. Journal **2** (1991), no. 5, 1067–1084.

[Nov58] P. S. Novikov, *On the algorithmic insolvability of the word problem in group theory*, American Mathematical Society Translations, Ser 2, Vol. 9, American Mathematical Society, Providence, R. I., 1958, pp. 1–122.

[Now15] P. Nowak, *Group actions on banach spaces*, Handbook of Group Actions. vol. II, Advanced Lectures in Mathematics (ALM) 32, International Press, Somerville; Higher Education Press, Beijing, 2015, pp. 121–149.

[NR97] W. Neumann and L. Reeves, *Central extensions of word hyperbolic groups*, Ann. of Math. **145** (1997), 183–192.

[NS14] B. Nica and J. Spakula, *Superbolicity*, Preprint, arXiv:1408.0250, 2014.

[NY11] P. Nowak and G. Yu, *Large scale geometry*, European Mathematical Society, 2011.

[Ol′80] A. Yu. Ol′shanskiĭ, *On the question of the existence of an invariant mean on a group*, Uspekhi Mat. Nauk **35** (1980), no. 4, 199–200.

[Ol′91a] ———, *Geometry of defining relations in groups*, Mathematics and its Applications (Soviet Series), vol. 70, Kluwer Academic Publishers Group, Dordrecht, 1991.

[Ol′91b] ———, *Hyperbolicity of groups with subquadratic isoperimetric inequalities*, Intl. J. Alg. Comp. **1** (1991), 282–290.

[Ol′91c] ———, *Periodic quotient groups of hyperbolic groups*, Mat. Sb. **182** (1991), no. 4, 543–567.

[Ol′92] ———, *Almost every group is hyperbolic*, Internat. J. Algebra Comput. **2** (1992), no. 1, 1–17.

[Ol′95] ———, SQ-*universality of hyperbolic groups*, Mat. Sb. **186** (1995), no. 8, 119–132.

[Oll04a] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. **14** (2004), no. 3, 595–679.

[Oll04b] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. **14** (2004), no. 3, 595–679.

[Oll05] ———, *A January 2005 invitation to random groups*, Ensaios Matemáticos [Mathematical Surveys], vol. 10, Sociedade Brasileira de Matemática, Rio de Janeiro, 2005.

[Oll07] ———, *Some small cancellation properties of random groups*, Internat. J. Algebra Comput. **17** (2007), no. 1, 37–51.

[OOS09] A. Yu. Ol′shanskiĭ, D. V. Osin, and M. V. Sapir, *Lacunary hyperbolic groups*, Geom. Topol. **13** (2009), no. 4, 2051–2140, With an appendix by M. Kapovich and B. Kleiner.

[Ore51] O. Ore, *Some remarks on commutators*, Proc. Amer. Math. Soc. **2** (1951), 307–314.

[OS01] A. Yu. Ol′shanskiĭ and M. V. Sapir, *Length and area functions on groups and quasi-isometric Higman embeddings*, International Journal of Algebra and Computation **11** (2001), 137–170.

[OS02] ———, *Non-amenable finitely presented torsion-by-cyclic groups*, Publ. Math. IHES **96** (2002), 43–169.

[Osa13] D. Osajda, *A construction of hyperbolic Coxeter groups*, Comment. Math. Helv. **88** (2013), no. 2, 353–367. MR 3048190

[Osi01] D. Osin, *Subgroup distortions in nilpotent groups*, Comm. Algebra **29** (2001), no. 12, 5439–5463.

[Osi06] ———, *Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems*, Mem. Amer. Math. Soc. **179** (2006), no. 843, vi+100pp.

[Osi10] ———, *Small cancellations over relatively hyperbolic groups and embedding theorems*, Ann. of Math. **172** (2010), 1–39.

[OV90] A. Onishchik and E. Vinberg, *Lie groups and algebraic groups*, Springer, 1990.

[OW80] D. Ornstein and B. Weiss, *Ergodic theory of amenable group actions. I. The Rohlin lemma*, Bull. Amer. Math. Soc. (N.S.) **2** (1980), no. 1, 161–164.

[OW11] Y. Ollivier and D. Wise, *Cubulating random groups at density less than* $1/6$, Trans. Amer. Math. Soc. **363** (2011), no. 9, 4701–4733.

[Oza15] N. Ozawa, *A functional analysis proof of Gromov's polynomial growth theorem*, Preprint, 2015.

[Pan83] P. Pansu, *Croissance des boules et des géodésiques fermées dan les nilvariétés*, Ergodic Th. & Dyn. Sys. **3** (1983), 415–455.

[Pan89] ———, *Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un*, Ann. of Math. (2) **129** (1989), no. 1, 1–60.

[Pan95] ———, *Cohomologie* L^p*: invariance sous quasi-isométrie*, preprint, `http://www.math.u-psud.fr/~pansu/liste-prepub.html`, 1995.

[Pan07] ———, *Cohomologie* L^p *en degré 1 des espaces homogènes*, Potential Anal. **27** (2007), no. 2, 151–165.

[Pap95a] P. Papasoglu, *Homogeneous trees are bi-Lipschitz equivalent*, Geom. Dedicata **54** (1995), no. 3, 301–306.

[Pap95b] ———, *On the subquadratic isoperimetric inequality*, Geometric group theory, vol. 25, de Gruyter, Berlin-New-York, 1995, R. Charney, M. Davis, M. Shapiro (eds), pp. 193–200.

[Pap95c] ———, *Strongly geodesically automatic groups are hyperbolic*, Invent. Math. **121** (1995), no. 2, 323–334.

[Pap96] ———, *An algorithm detecting hyperbolicity*, Geometric and computational perspectives on infinite groups (Minneapolis, MN and New Brunswick, NJ, 1994), vol. 25, Amer. Math. Soc., Providence, RI, 1996, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., pp. 193–200.

[Pap98] ———, *Quasi-flats in semihyperbolic groups*, Proc. Amer. Math. Soc. **126** (1998), 1267–1273.

[Pap00] ———, *Isodiametric and isoperimetric inequalities for complexes and groups*, J. London Math. Soc. (2) **62** (2000), no. 1, 97–106.

[Pap03] ———, *Notes On Hyperbolic and Automatic Groups*, Lecture Notes, based on the notes of M. Batty, http://www.math.ucdavis.edu/∼kapovich/280-2009/hyplectures_papasoglu.pdf, 2003.

[Pap05] ———, *Quasi-isometry invariance of group splittings*, Ann. of Math. **161** (2005), no. 2, 759–830.

[Par08] J. Parker, *Hyperbolic spaces*, vol. 2, Jyväskylä Lectures in Mathematics, 2008.

[Pau91a] F. Paulin, *Outer automorphisms of hyperbolic groups and small actions on* **R**-*trees*, Arboreal group theory (Berkeley, CA, 1988), Math. Sci. Res. Inst. Publ., vol. 19, Springer, New York, 1991, pp. 331–343.

[Pau91b] ———, *Outer automorphisms of hyperbolic groups and small actions on* $\mathbb{R}$*–trees*, Arboreal group theory (R. Alperin, ed.), Publ. M.S.R.I., vol. 19, Springer, 1991, pp. 331–343.

[Pau96] ———, *Un groupe hyperbolique est déterminé par son bord*, J. London Math. Soc. (2) **54** (1996), no. 1, 50–74.

[Paw91] J. Pawlikowski, *The Hahn-Banach theorem implies the Banach-Tarski paradox*, Fundamenta Mathematicae **138** (1991), no. 1, 21–22.

[Ped95] E.K. Pedersen, *Bounded and continuous control*, Proceedings of the conference "Novikov conjectures, index theorems and rigidity" volume II, Oberwolfach 1993, LMS Lecture Notes Series, vol. 227, Cambridge University Press, 1995, pp. 277–284.

[Pen11a] I. Peng, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups I*, Geom. Topol. **15** (2011), no. 4, 1883–1925.

[Pen11b] ———, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups II*, Geom. Topol. **15** (2011), no. 4, 1927–1981.

[Pin72] D. Pincus, *Independence of the prime ideal theorem from the Hahn Banach theorem*, Bull. Amer. Math. Soc. **78** (1972), 766–770.

[Pin74] ———, *The strength of the Hahn-Banach theorem*, Victoria Symposium on Nonstandard Analysis (Univ. Victoria, Victoria, B.C., 1972), Springer, Berlin, 1974, pp. 203–248. Lecture Notes in Math., Vol. 369.

[Pit98] C. Pittet, *On the isoperimetry of graphs with many ends*, Colloq. Math. **78** (1998), 307–318.

[Pla68] V. P. Platonov, *A certain problem for finitely generated groups*, Dokl. Akad. Nauk BSSR **12** (1968), 492–494.

[Poi95] F. Point, *Groups of polynomial growth and their associated metric spaces*, J. Algebra **175** (1995), no. 1, 105–121.

[Pos37] B. Pospíšil, *Remark on bicompact spaces*, Ann. of Math. (2) **38** (1937), no. 4, 845–846.

[Pra73] G. Prasad, *Strong rigidity of Q–rank one lattices*, Invent. Math. **21** (1973), 255–286.

[PW02] P. Papasoglu and K. Whyte, *Quasi-isometries between groups with infinitely many ends*, Comment. Math. Helv. **77** (2002), no. 1, 133–144.

[Rab58] M. O. Rabin, *Recursive unsolvability of group theoretic problems*, Ann. of Math. (2) **67** (1958), 172–194.

[Rag72] M. Raghunathan, *Discrete subgroups of lie groups*, Springer, 1972.

[Rag84] ———, *Torsion in cocompact lattices in coverings of Spin*$(2,n)$, Math. Ann. **266** (1984), no. 4, 403–419.
[Rat06] J.G. Ratcliffe, *Foundations of hyperbolic manifolds*, Springer, Second Edition, 2006.
[Res89] Yu. G. Reshetnyak, *Space mappings with bounded distortion*, Translations of Mathematical Monographs, vol. 73, American Mathematical Society, Providence, RI, 1989, Translated from the Russian by H. H. McFaden.
[RF] A. Sisto R. Frigerio, J.-F. Lafont, *Rigidity of high dimensional graph manifolds*, Monograph to appear in Astérisque.
[Rin61] W. Rinow, *Die innere Geometrie der metrischen Räume*, Die Grundlehren der mathematischen Wissenschaften, Bd. 105, Springer-Verlag, Berlin, 1961.
[Rip82] E. Rips, *Subgroups of small cancellation groups*, Bull. London Math. Soc. **14** (1982), no. 1, 45–47.
[Rob47] R.M. Robinson, *On the decomposition of spheres*, Fund. Math. **34** (1947), 246–260.
[Roe03] J. Roe, *Lectures on coarse geometry*, University Lecture Series, vol. 31, American Mathematical Society, Providence, RI, 2003.
[Ros74] J. Rosenblatt, *Invariant measures and growth conditions*, Trans. Amer. Math. Soc. **193** (1974), 33–53.
[Roy68] H. Royden, *Real Analysis*, Macmillan, New York, 1968.
[RR85] H. Rubin and J. E. Rubin, *Equivalents of the axiom of choice. II*, Studies in Logic and the Foundations of Mathematics, vol. 116, North-Holland Publishing Co., Amsterdam, 1985.
[RS94] E. Rips and Z. Sela, *Structure and rigidity in hyperbolic groups. I*, Geom. Funct. Anal. **4** (1994), no. 3, 337–371.
[RS97] ———, *Cyclic splittings of finitely presented groups and the canonical JSJ decomposition*, Ann. of Math. **146** (1997), 53–109.
[Rud87] W. Rudin, *Real and complex analysis*, McGraw-Hill International editions, 1987.
[Sal12] A. Sale, *Short conjugators and compression exponents in free solvable groups*, preprint, arXiv:1202.5343v1, 2012.
[Sap14] M. Sapir, *Combinatorial algebra: Syntax and semantics*, Springer Verlag, 2014.
[Sap15] ———, *On groups with locally compact asymptotic cones*, International Journal of Algebra and Computation **25** (2015), 37–40.
[Sau06] R. Sauer, *Homological invariants and quasi-isometry*, Geom. Funct. Anal. **16** (2006), no. 2, 476–515.
[SBR02] M.. Sapir, J.-C. Birget, and E. Rips, *Isoperimetric and isodiametric functions of groups*, Ann. of Math. (2) **156** (2002), no. 2, 345–466.
[Sch38] I.J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.
[Sch96a] R. Schwartz, *Quasi-isometric rigidity and Diophantine approximation*, Acta Math. **177** (1996), no. 1, 75–112.
[Sch96b] ———, *The quasi-isometry classification of rank one lattices*, Publ. Math. IHES **82** (1996), 133–168.
[Sch97] ———, *Symmetric patterns of geodesics and automorphisms of surface groups*, Invent. Math. **128** (1997), no. 1, 177–199.
[Sch13] V. Schur, *A quantitative version of the morse lemma and quasi-isometries fixing the ideal boundary*, Journal of Functional Analysis **264** (2013), 815–836.
[Sco73] P. Scott, *Compact submanifolds of 3-manifolds*, J. London Math. Sco. **7** (1973), 246–250.
[Sco78] ———, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. (2) **17** (1978), no. 3, 555–565.
[Sco83] ———, *The geometry of* 3*-manifolds*, Bull. of the LMS **15** (1983), 401–487.
[Sel60] A. Selberg, *On discontinuous groups in higher-dimensional symmetric spaces*, Contributions to Function Theory (K. Chandrasekhadran, ed.), Tata Inst. of Fund. Research, Bombay, 1960, pp. 147–164.
[Sel92] Z. Sela, *Uniform embeddings of hyperbolic groups in Hilbert spaces*, Israel J. Math. **80** (1992), no. 1-2, 171–181.
[Sel95] ———, *The isomorphism problem for hyperbolic groups. I*, Ann. of Math. (2) **141** (1995), no. 2, 217–283.

[Sel97a] ______, *Structure and rigidity in (Gromov) hyperbolic groups and discrete groups in rank* 1 *Lie groups. II*, Geom. Funct. Anal. **7** (1997), no. 3, 561–593.

[Sel97b] ______, *Structure and rigidity in hyperbolic groups and discrete groups in rank 1 Lie groups, II*, Geom. Funct. Anal. **7** (1997), 561–593.

[Sel99] ______, *Endomorphisms of hyperbolic groups. I. The Hopf property*, Topology **38** (1999), no. 2, 301–321.

[Sel01] ______, *Diophantine geometry over groups. I. Makanin-Razborov diagrams*, Publ. Math. IHES **93** (2001), 31–105.

[Sel03] Z. Sela, *Diophantine geometry over groups. II. Completions, closures and formal solutions*, Israel J. Math. **134** (2003), 173–254.

[Sel04] ______, *Diophantine geometry over groups. IV. An iterative procedure for validation of a sentence*, Israel J. Math. **143** (2004), 1–130.

[Sel05a] ______, *Diophantine geometry over groups. III. Rigid and solid solutions*, Israel J. Math. **147** (2005), 1–73.

[Sel05b] ______, *Diophantine geometry over groups.* V_1*. Quantifier elimination. I*, Israel J. Math. **150** (2005), 1–197.

[Sel06a] ______, *Diophantine geometry over groups.* V_2*. Quantifier elimination. II*, Geom. Funct. Anal. **16** (2006), no. 3, 537–706.

[Sel06b] ______, *Diophantine geometry over groups. VI. The elementary theory of a free group*, Geom. Funct. Anal. **16** (2006), no. 3, 707–730.

[Sel09] ______, *Diophantine geometry over groups. VII. The elementary theory of a hyperbolic group*, Proc. Lond. Math. Soc. (3) **99** (2009), no. 1, 217–273.

[Sel13] ______, *Diophantine geometry over groups VIII: Stability*, Ann. of Math. (2) **177** (2013), no. 3, 787–868.

[Ser80] J. P. Serre, *Trees*, Springer, New York, 1980.

[Sha00] Y. Shalom, *Rigidity of commensurators and irreducible lattices*, Invent. Math. **141** (2000), no. 1, 1–54.

[Sha04] ______, *Harmonic analysis, cohomology, and the large-scale geometry of amenable groups*, Acta Math. **192** (2004), no. 2, 119–185.

[She78] S. Shelah, *Classification theory and the number of non-isomorphic models*, Studies in Logic and the Foundations of Mathematics, vol. 92, North-Holland Publishing Co., 1978.

[Sil] L. Silbermann, *Equivalence between properties (T) and (FH)*, unpublished notes.

[Šir76] V. L. Širvanjan, *Imbedding of the group* $B(\infty, n)$ *in the group* $B(2, n)$, Izv. Akad. Nauk SSSR Ser. Mat. **40** (1976), no. 1, 190–208, 223.

[Siu80] Y. T. Siu, *The complex-analyticity of harmonic maps and the strong rigidity of compact Kähler manifolds*, Ann. of Math. (2) **112** (1980), 73–111.

[Sol11] Y. Solomon, *Substitution tilings and separated nets with similarities to the integer lattice*, Israel J. Math. **181** (2011), 445–460.

[SS05] E. Stein and R. Shakarchi, *Real analysis*, Princeton Lectures in Analysis, III, Princeton University Press, 2005, Measure theory, integration, and Hilbert spaces.

[ST10] Y. Shalom and T. Tao, *A finitary version of Gromov's polynomial growth theorem*, Geom. Funct. Anal. **20** (2010), no. 6, 1502–1547.

[Sta68] J. Stallings, *On torsion-free groups with infinitely many ends*, Ann. of Math. **88** (1968), 312–334.

[Sta83] ______, *Topology of finite graphs*, Inv. Math. **71** (1983), 551–565.

[Sta00] ______, *Commutative rings and groups*, UC Berkeley Lecture Notes, 2000.

[Str74] R. A. Struble, *Metrics in locally compact groups*, Compositio Math. **28** (1974), 217–222.

[Str06] G. Strang, *Linear algebra and its applications*, Thomson, Brooks/Cole, 2006.

[Sul81] D. Sullivan, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, Riemann surfaces and related topics, Proceedings of the 1978 Stony Brook Conference, Ann. Math. Studies, vol. 97, Princeton University Press, 1981, pp. 465–496.

[Sul14] H. Sultan, *Hyperbolic quasi-geodesics in CAT(0) spaces*, Geom. Dedicata **169** (2014), 209–224.

[Šva55] A. S. Švarc, *A volume invariant of coverings*, Dokl. Akad. Nauk SSSR (N.S.) **105** (1955), 32–34.

[SW79] P. Scott and T. Wall, *Topological methods in group theory*, "Homological Group Theory", London Math. Soc. Lecture Notes, vol. 36, 1979, pp. 137– 204.

[SW05] M. Sageev and D. Wise, *The Tits alternative for* CAT(0) *cubical complexes*, Bull. London Math. Soc. **37** (2005), no. 5, 706–710.

[Swa93] G. A. Swarup, *A note on accessibility*, Geometric group theory, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 204–207.

[SX12] N. Shanmugalingam and X. Xie, *A rigidity property of some negatively curved solvable Lie groups*, Commentarii Mathematici Helvetici **87** (2012), 805–823.

[SY94] R. Schoen and S.-T. Yau, *Lectures on differential geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, I, International Press, Cambridge, MA, 1994, Lecture notes prepared by Wei Yue Ding, Kung Ching Chang, Jia Qing Zhong and Yi Chao Xu, Translated from the Chinese by Ding and S. Y. Cheng, Preface translated from the Chinese by Kaising Tso.

[Tao08] T. Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no. 5, 547–594.

[Tar38] A. Tarski, *Algebraische fassung des massproblems*, Fund. Math. **31** (1938), 47–66.

[Tar86] ———, *Collected papers. Vol. 1*, Contemporary Mathematicians, Birkhäuser Verlag, Basel, 1986, 1921–1934, Edited by Steven R. Givant and Ralph N. McKenzie.

[Tho02] B. Thornton, *Asymptotic cones of symmetric spaces*, Ph.D. Thesis, University of Utah, http://www.math.utah.edu/theses/2002/thornton/blake-thornton-dissertation.pdf, 2002.

[Thu97] W. Thurston, *Three-dimensional geometry and topology, I*, Princeton Mathematical Series, vol. 35, Princeton University Press, 1997.

[Tit72] J. Tits, *Free subgroups in linear groups*, Journal of Algebra **20** (1972), 250–270.

[Tuk81] P. Tukia, *A quasiconformal group not isomorphic to a Möbius group*, Ann. Acad. Sci. Fenn. Ser. A I Math. **6** (1981), 149–160.

[Tuk85] ———, *Differentiability and rigidity of Moebius groups*, Inventiones Mathematicae **82** (1985), 555–578.

[Tuk86] ———, *On quasiconformal groups*, J. Analyse Math. **46** (1986), 318–346.

[Tuk88] ———, *Homeomorphic conjugates of Fuchsian groups*, J. Reine Angew. Math. **391** (1988), 1–54.

[Tuk94] ———, *Convergence groups and Gromov's metric hyperbolic spaces*, New Zealand J. Math. **23** (1994), no. 2, 157–187.

[TW93] C. Thomassen and W. Woess, *Vertex-transitive graphs and accessibility*, J. Combin. Theory Ser. B **58** (1993), no. 2, 248–268.

[Uml91] K.A. Umlauf, *über Die Zusammensetzung Der Endlichen Continuierlichen Transformationsgruppen, Insbesondre Der Gruppen Vom Range Null*, Nabu Press, 2010, 1891, Inaugural-Dissertation, Universität Leipzig.

[Väi71] J. Väisälä, *Lectures on n-dimensional quasiconformal mappings*, Lecture Notes in Math., vol. 229, Springer, 1971.

[Väi05] ———, *Gromov hyperbolic spaces*, Expo. Math. **23** (2005), no. 3, 187–231.

[Väi85] ———, *Quasi-Möbius maps*, J. Analyse Math. **44** (1984/85), 218–234.

[Var99] N. Varopoulos, *Distance distortion on Lie groups*, Random walks and discrete potential theory (Cortona, 1997), Sympos. Math., XXXIX, Cambridge Univ. Press, Cambridge, 1999, pp. 320–357.

[Vas12] S. Vassileva, *The Magnus embedding is a quasi-isometry*, Internat. J. Algebra Comput. **22** (2012), no. 8, 1240005.

[Ver82] A. Vershik, *Amenability and approximation of infinite groups*, Selecta Math. Soviet. **2** (1982), no. 4, 311–330, Selected translations.

[Vie27] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Mathematische Annalen **97** (1927), 454–472.

[Vit05] G. Vitali, *Sul problema della misura dei gruppi di punti di una retta*, Gamberini e Parmeggiani, 1905, Bologna.

[vN28] J. von Neumann, *Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre*, Math. Ann. **99** (1928), 373–391.

[vN29] ———, *Zur allgemeinen theorie des masses*, Fund. math. **13** (1929), 73–116.

[VSCC92] N. Th. Varopoulos, L. Saloff-Coste, and T. Coulhon, *Analysis and geometry on groups*, Cambridge Tracts in Mathematics, vol. 100, Cambridge University Press, Cambridge, 1992.
[Vuo88] M. Vuorinen, *Conformal geometry and quasiregular mappings*, Lecture Notes in Math, vol. 1319, Springer, 1988.
[Wag85] S. Wagon, *The Banach-Tarski paradox*, Cambridge Univ. Press, 1985.
[War83] F. W. Warner, *Foundations of differentiable manifolds and lie groups*, Graduate Texts in Mathematics, 94, Springer-Verlag, 1983.
[War12] E. Warner, *Ultraproducts and foundations of higher order Fourier analysis*, Thesis, Princeton University, 2012.
[Wat82] Y. Watatani, *Property (T) of Kazhdan implies property (FA) of Serre*, Math. Japonica **27** (1982), 97–103.
[Weh73] B.A.F. Wehrfritz, *Infinite linear groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 76, Springer, 1973.
[Wei84] B. Weisfeiler, *On the size and structure of finite linear groups*, Preprint, math. arxiv, 1203.1960, 1984.
[Wen08] S. Wenger, *Gromov hyperbolic spaces and the sharp isoperimetric constant*, Invent. Math. **171** (2008), no. 1, 227–255.
[Wen11] S. Wenger, *Nilpotent groups without exactly polynomial Dehn function*, J. Topol. **4** (2011), no. 1, 141–160.
[Why99] K. Whyte, *Amenability, bilipschitz equivalence, and the von Neumann conjecture*, Duke Math. J. **99** (1999), 93–112.
[Why01] ———, *The large scale geometry of the higher Baumslag-Solitar groups*, Geom. Funct. Anal. **11** (2001), no. 6, 1327–1343.
[Woe00] W. Woess, *Random walks on infinite graphs and groups*, Cambridge University Press, 2000.
[Wol68] J. Wolf, *Growth of finitely generated solvable groups and curvature of riemannian manifolds*, J. Diff. Geom. **2** (1968), 421–446.
[Wor07] K. Wortman, *Quasi-isometric rigidity of higher rank S-arithmetic lattices*, Geometry & Topology **11** (2007), 995–1048.
[WW75] J.H. Wells and L.R. Williams, *Embeddings and extensions in analysis*, Springer-Verlag, New York-Heidelberg, 1975.
[Wys88] J. Wysocz´anski, *On uniformly amenable groups*, Proc. Amer. Math. Soc. **102** (1988), no. 4, 933–938.
[Xie06] X. Xie, *Quasi-isometric rigidity of Fuchsian buildings*, Topology **45** (2006), 101–169.
[Xie12] ———, *Quasisymmetric maps on the ideal boundary of a negatively curved solvable Lie group*, Mathematische Annalen **353** (2012), 727–746.
[Xie13] ———, *Quasisymmetric maps on reducible Carnot groups*, Pacific Journal of Mathematics **265** (2013), 113–122.
[Xie14a] ———, *Large scale geometry of negatively curved* $\mathbb{R}^n$, Geometry & Topology **18** (2014), 831–872.
[Xie14b] ———, *Some examples of quasiisometries of nilpotent lie groups*, Journal für die reine und angewandte Mathematik (2014).
[Yam04] A. Yaman, *A topological characterisation of relatively hyperbolic groups*, J. Reine Angew. Math. (Crelle's Journal) **566** (2004), 41–89.
[Yu00] G. Yu, *The Coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert spaces*, Invent. Math. **139** (2000), 201–240.
[Yu05] ———, *Hyperbolic groups admit proper isometric actions on* ℓ^p*-spaces*, Geom. Funct. Anal. **15** (2005), no. 5, 1144–1151.
[Zer04] E. Zermelo, *Beweis, dass jede Menge wohlgeordnet werden kann*, Math. Ann. **59** (1904), 514–516.
[Zim84] R. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Math, vol. 81, Birkhauser, 1984.
[Żuk96] A. Żuk, *La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres*, C. R. Acad. Sci. Paris Sér. I Math. **323** (1996), no. 5, 453–458.
[Żuk03] ———, *Property (T) and Kazhdan constants for discrete groups*, Geom. Funct. Analysis **13** (2003), 643–670.

Index