Lectures on Geometric Group Theory

Cornelia Druţu and Michael Kapovich

Contents

CHAPTER 1

General preliminaries

1.1. Notation and terminology

1.1.1. General notation. Given a set X we denote by $\mathcal{P}(X)$ 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*. 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 the denote the identity map $X \to X$. For a map $f : X \to Y$ and a subset $A \subset X$, we let $f|A$ or $f|_A$ denote the restriction of f to A. We will use the notation $|E|$ or card (E) to denote cardinality of a set E.

The Axiom of Choice (AC) plays an important part in many of the arguments of this book. We discuss AC in more detail in Section **??**, 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 and the Axiom of Choice.

We will use the notation $\overline{A}$ and $cl(A)$ for the closure of a subset A in a topological space X. 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.

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

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

Given a non-empty set X, we denote by 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 $x \in A$.

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 dist(x, A) for the *minimal distance* from x to A, i.e.,

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

If $A, B \subset X$ are two subsets A, B, we let

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

denote the *Hausdorff distance* between A and B in X. See Section 1.4 for further details on this distance and its generalizations.

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. $\mathcal{N}_R(A) = \{x \in X : \mathrm{dist}(x, A) < R\}$. 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

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

We will use the notation $[A, B]$ to denote a geodesic segment connecting point A to 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 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 with the given vertices. Precise definitions of geodesic segments and triangles will be given in Section 1.3.3.

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.

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 $+$.

If two groups G and G' are isomorphic we write $G \simeq G'$.

A surjective homomorphism is called an *epimorphism*, while an injective homomorphism is called a *monomorphism*. 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}.

1.1.2. Direct and inverse limits of spaces and groups. 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 $f_{ij} \circ f_{jk}$.

The *direct limit* of the direct system 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 1.3. Every group G is the direct limit of the directed family $G_i, i \in I$, 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 1.4. 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 $\{H_i, f_{ij} : i, j \in I\}$ 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 $i \in I$,

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

EXERCISE 1.5. 1. Let $H \leqslant G$ be a subgroup. Then $|G : H| \leq n$ if and only if the following holds: For every subset $\{g_0, \ldots, g_n\} \subset G$, there exist g_i, g_j so that $g_i g_j^{-1} \in H$.

2. Suppose that G is the direct limit of a family of groups $G_i, i \in I$. 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 $\leq n$. Let the group H be the direct limit of the family $\{H_i : i \in I\}$ and $\phi : H \to G$ be the monomorphism as in Exercise 1.4. Show that

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

1.1.3. 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 1.6. 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 $a, b, c, d, e > 0$ such that

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

for every $x \in X$, $x \geqslant x_0$, for some fixed x_0.

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

The equivalence class of a numerical 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 1.7. 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 $a, b > 0$ such that $f(x) \leq ag(bx)$ for every $x \in X$, $x \geq x_0$ for some fixed x_0.

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 relation, in which case the Dehn function is zero, or another presentation that yields a linear Dehn function.

EXERCISE 1.8. 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 $f \asymp g$.

1.2. 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, 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 $\iota(e)$. As with general cell complexes and simplicial complexes, we will frequently conflate a graph with its *geometric realization*, i.e., the underlying topological space.

CONVENTION 1.9. In this book, unless we state otherwise, all graphs are assumed to be unoriented.

Note that in the definition of a graph we allow for *monogons*[1] (i.e. edges connecting a vertex to itself) and *bigons*[2] (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.

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 cycle homeomorphic to the circle, i.e., a simple loop in Γ.

A *simplicial tree* is a simply-connected simplicial graph.

An *isomorphism* of graphs is an isomorphism of the corresponding cell complexes, i.e., it is a homeomorphism $f : \Gamma \to \Gamma'$ so that the images of the edges of Γ are edges of Γ' and images of vertices are vertices. We use the notation $\mathrm{Aut}(\Gamma)$ for the group of automorphisms of a graph Γ.

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.

A *directed (or oriented) graph* Γ consists of the following data:

- a set V called *set of vertices* of the graph;
- a set $\overline{E}$ called the *set of edges* of the graph;
- two maps $o : \overline{E} \to V$ and $t : \overline{E} \to V$, called respectively the *head (or origin) map* and the *tail map*.

Then, for every $x, y \in V$ we define the set of oriented edges connecting x to y:

$$E_{(x,y)} = \{\bar{e} : (o(\bar{e}), t(\bar{e})) = (x, y)\}.$$

A directed graph is called *symmetric* if for every subset $\{u, v\}$ of V the sets $E_{(x,y)}$ and $E_{(y,x)}$ have the same cardinality. For such graphs, interchanging the maps t and o induces an automorphism of the directed graph, which fixes V.

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

[2]Also known as *digons*.

A symmetric directed graph $\overline{\Gamma}$ is equivalent to a unoriented graph Γ with the same vertex set, *via* the following replacement procedure: Pick an involutive bijection $\beta : \overline{E} \to \overline{E}$, which induces bijections $\beta : E_{(x,y)} \to E_{(y,x)}$ for all $x, y \in V$. We then get the equivalence relation $e \sim \beta(e)$. The quotient $E = \overline{E}/\sim$ is the edge-set of the graph Γ, where the incidence map ι is defined by $\iota([e]) = \{o(e), t(e)\}$. The unoriented graph Γ thus obtained, is called the *underlying unoriented graph* of the given directed graph.

EXERCISE 1.10. Describe the converse to this procedure: Given a graph Γ, construct a symmetric directed graph $\overline{\Gamma}$, so that Γ is the underlying graph of $\overline{\Gamma}$.

DEFINITION 1.11. Let $F \subset V = V(\Gamma)$ be a set of vertices in a (unoriented) 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 $V \setminus F$.

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. For graphs without bigons, the edge-boundary can be identified with the set of vertices $v \in V \setminus F$ adjacent to a vertex in F, in other words, with $\partial_V(V \setminus F)$.

For graphs having a uniform upper bound C on the valency of vertices, cardinalities of the two types of boundaries are *comparable*

$$|\partial_V F| \leqslant |E(F, F^c)| \leqslant C|\partial_V F|\,. \tag{1.1}$$

DEFINITION 1.12. 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 1.13. 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?

1.3. Topological and metric spaces

1.3.1. Topological spaces. Lebesgue covering dimension. Given two topological spaces, we let $C(X; Y)$ denote the space of all continuous maps $X \to Y$; set $C(X) := C(X; \mathbb{R})$. We always endow the space $C(X; Y)$ with the compact-open topology.

DEFINITION 1.14. Two subsets A, V of a topological space X are said to be *separated by a function* if there exists a continuous function $\rho = \rho_{A,V} : X \to [0, 1]$ so that

1. $\rho|A \equiv 0$
2. $\rho|V \equiv 1$.

A topological space X is called *perfectly normal* if every two disjoint closed subsets of X can be separated by a function.

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 covering $\mathcal{U} = \{U_i : i \in I\}$ 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 covering $\mathcal{V}$ is called a *refinement* of a covering $\mathcal{U}$ if every $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$.

DEFINITION 1.15. 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 1.16. If M is a n-dimensional topological manifold, then n equals the covering dimension of M. See e.g. [**Nag83**].

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

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

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

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

The function dist is called *metric* or *distance function*. Occasionally, it will be convenient to 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.).

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

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

We will see some examples of ultrametric spaces in Section 1.8.

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

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

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

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

EXERCISE 1.17. Show that the Euclidean plane E^2 satisfies the *parallelogram identity*: If A, B, C, D are vertices of a parallelogram P in E^2 with the diagonals $[AC]$ and $[BD]$, then

$$d^2(A,B) + d^2(B,C) + d^2(C,D) + d^2(D,A) = d^2(A,C) + d^2(B,D), \tag{1.2}$$

i.e., sum of squares of the sides of P equals the sum of squares 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

$$d((x_1, y_1), (x_2, y_2))^2 = d(x_1, x_2)^2 + d(y_1, y_2)^2. \tag{1.3}$$

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 1.18. *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

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

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. $\square$

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.

A topological space is called *locally compact* if for every $x \in X$ there exists a basis of neighborhoods of x consisting of *relatively compact* subsets of X, i.e., subsets with compact closure. A metric space is locally compact if and only if for every $x \in X$ there exists $\varepsilon = \varepsilon(x) > 0$ such that the closed ball $\overline{B}(x, \varepsilon)$ is compact.

DEFINITION 1.19. 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 $x, x' \in X$

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

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 $x, x' \in X$,

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

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 $f(x) = y$.

1.3.3. Length metric spaces. Throughout these notes 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

$$\text{length}(\mathfrak{p}) = \sup_{a=t_0<t_1<\cdots<t_n=b} \sum_{i=0}^{n-1} \mathrm{dist}(\mathfrak{p}(t_i), \mathfrak{p}(t_{i+1})) \tag{1.4}$$

where the supremum is taken over all possible partitions of $[a, b]$ and all integers n. By the definition and triangle inequalities in X, $\text{length}(\mathfrak{p}) \geqslant \mathrm{dist}(\mathfrak{p}(a), \mathfrak{p}(b))$.

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

EXERCISE 1.20. 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

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

Similarly, if (M, g) is a connected Riemannian manifold and dist is the Riemannian distance function, then the two notions of length, given by equations (2.1) and (1.4), coincide for smooth paths.

EXERCISE 1.21. 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 1.22. Show that dist_ℓ is a metric on X with values in $[0, \infty]$.

Suppose that $\mathfrak{p}$ is a path realizing the infimum in the definition of distance $\mathrm{dist}_\ell(x, y)$. We will (re)parameterize such $\mathfrak{p}$ by its arc-length; the resulting path $\mathfrak{p} : [0, D] \to (X, \mathrm{dist}_\ell)$ is called a *geodesic segment* in (X, dist_ℓ).

EXERCISE 1.23. $\mathrm{dist} \leqslant \mathrm{dist}_\ell$.

DEFINITION 1.24. A metric space (X, dist) such that $\mathrm{dist} = \mathrm{dist}_\ell$ is called a *length (or path) metric space*.

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 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 $\mathbb{R}$.

DEFINITION 1.25. 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 1.26. Prove that for (X, dist_ℓ) the two notions of geodesics agree.

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 $[A, B], [B, C], [C, A]$ in X. These segments are called *edges* of T. Later on, in Chapters 7 and 8 we will use *generalized* triangles, where some edges are geodesic rays or, even, complete geodesics. The corresponding vertices generalized triangles will be *points of the ideal boundary* of X.

EXAMPLES 1.27.
(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 2.1.3) 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 1.3.4) is a geodesic metric space.

EXERCISE 1.28. 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 (1.3).

THEOREM 1.29 (Hopf–Rinow Theorem [**Gro07**]). *If a length metric space is complete and locally compact, then it is geodesic and proper.*

EXERCISE 1.30. 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.

1.3.4. Graphs as length spaces. Let Γ be a connected graph. Recall that we are conflating Γ and its geometric realization, so 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 $x, y \in \Gamma$, $\mathrm{dist}(x, y)$ is

$$\inf_{\mathfrak{p}} (\mathrm{length}(\mathfrak{p})),$$

where the infimum is taken over all fractional edge-paths $\mathfrak{p}$ in Γ connecting x to y.

EXERCISE 1.31. 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 metric dist is called the *standard* metric on Γ.

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 1.32. 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 $(\Gamma, \mathrm{dist}_\ell)$.

1.4. Hausdorff and Gromov-Hausdorff distances. Nets

Given subsets A_1, A_2 in a metric space (X, d), define the *minimal distance* between these sets as

$$\mathrm{dist}(A_1, A_2) = \inf\{d(a_1, a_2) : a_i \in A_i, i = 1, 2\}.$$

The *Hausdorff (pseudo)distance* between subsets $A_1, A_2 \subset X$ is defined as

$$\mathrm{dist}_{Haus}(A_1, A_2) := \inf\{R : A_1 \subset \mathcal{N}_R(A_2), A_2 \subset \mathcal{N}_R(A_1)\}.$$

Two subsets of X are called *Hausdorff-close* if they are within finite Hausdorff distance from each other.

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 $K(X)$ of compact subsets of a metric space X. This topology extends to the set $C(X)$ 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 $C \in C(X)$ to be

$$\{Z \in C(X) : \mathrm{dist}_{Haus}(Z \cap K, C \cap K) < \epsilon\}.$$

This system of neighborhoods generates a topology on $C(X)$, called *Chabauty topology*. Thus, a sequence $C_i \in C(X)$ converges to a closed subset $C \in C(X)$ 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 topology of Hausdorff–convergence.

M. Gromov defined in [**Gro81**, section 6] the *modified Hausdorff pseudo-distance* (also called the *Gromov–Hausdorff pseudo-distance*) on the class of proper metric spaces:

$$\begin{aligned}(1.5)\qquad \mathrm{dist}_{GHaus}((X,d_X),(Y,d_Y)) = \inf_{(x,y)\in X\times Y} \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric} \\ \text{dist on } M = X \sqcup Y, \text{ such that } \mathrm{dist}(x,y) < \varepsilon, \mathrm{dist}|_X = d_X, \mathrm{dist}|_Y = d_Y \text{ and} \\ B(x,1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y,1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}\,.\end{aligned}$$

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

$$\begin{aligned}(1.6)\qquad \mathrm{dist}_{\tilde{H}}((X,d_X,x_0),(Y,d_Y,y_0)) = \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric} \\ \text{dist on } M = X \sqcup Y \text{ such that } \mathrm{dist}(x_0,y_0) < \varepsilon,\ \mathrm{dist}|_X = d_X, \mathrm{dist}|_Y = d_Y, \\ B(x_0,1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y_0,1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}\,.\end{aligned}$$

This pseudo-distance becomes a metric when restricted to the class of proper pointed metric spaces.

Still, as before, to simplify the terminology we shall refer to all three pseudo-distances as 'distances' or 'metrics.'

EXAMPLE 1.33. The real line $\mathbb{R}$ with the standard metric and the planar circle of radius r, $\mathcal{C}(O,r)$, with the length metric, are at modified Hausdorff distance

$$\varepsilon_0 := \frac{4}{\sqrt{\pi^2 r^2 + 16} + \pi r}.$$

Since both are homogeneous spaces, it suffices to prove that the pointed metric spaces $(\mathbb{R}, 0)$ and $(\mathcal{C}(O,r), N)$, where N is the North pole, are at the distance ε_0 with respect to the modified Hausdorff distance with respect to these base-points.

To prove the upper bound we glue $\mathbb{R}$ and $\mathcal{C}(O,r)$ by identifying isometrically the interval $\left[-\frac{\pi}{2}r\,,\ \frac{\pi}{2}r\right]$ in $\mathbb{R}$ to the upper semi-circle (see Figure 1.1), and we endow the graph M thus obtained with its length metric dist. Note that the use of pseudo-metrics on M in the definition of the modified Hausdorff pseudo-distance allows for points $x \in X$ and $y \in Y$ to be identified. The minimal $\varepsilon > 0$ such that in (M, dist)

$$\left[-\frac{1}{\varepsilon}\,,\ \frac{1}{\varepsilon}\right] \subset \mathcal{N}_\varepsilon(\mathcal{C}(O,r)) \text{ and } B(N, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(\mathbb{R})$$

is ε_0 defined above. This value is the positive solution of the equation

$$\frac{\pi}{2} r + \varepsilon = \frac{1}{\varepsilon}. \tag{1.7}$$

For the lower bound consider another metric dist' on $\mathbb{R} \vee \mathcal{C}(O,r)$ which coincides with the length metrics on both $\mathbb{R}$ and $\mathcal{C}(O,r)$. Let ε' be the smallest $\varepsilon > 0$ such that $\mathrm{dist}'(0,N) < \varepsilon$ and $\left[-\frac{1}{\varepsilon}\,,\ \frac{1}{\varepsilon}\right] \subset \mathcal{N}_\varepsilon(\mathcal{C}(O,r)),\ B(N, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(\mathbb{R})$ in the metric dist$'$. Let x', y' be the nearest points in $\mathcal{C}(O,r)$ to $-\frac{1}{\varepsilon'}$ and $\frac{1}{\varepsilon'}$, respectively. Since $\mathrm{dist}'(x', y') \leqslant \pi r$, it follows that $\frac{2}{\varepsilon'} \leqslant \pi r + 2\varepsilon'$. The previous inequality implies that $\varepsilon' \geqslant \varepsilon_0$.

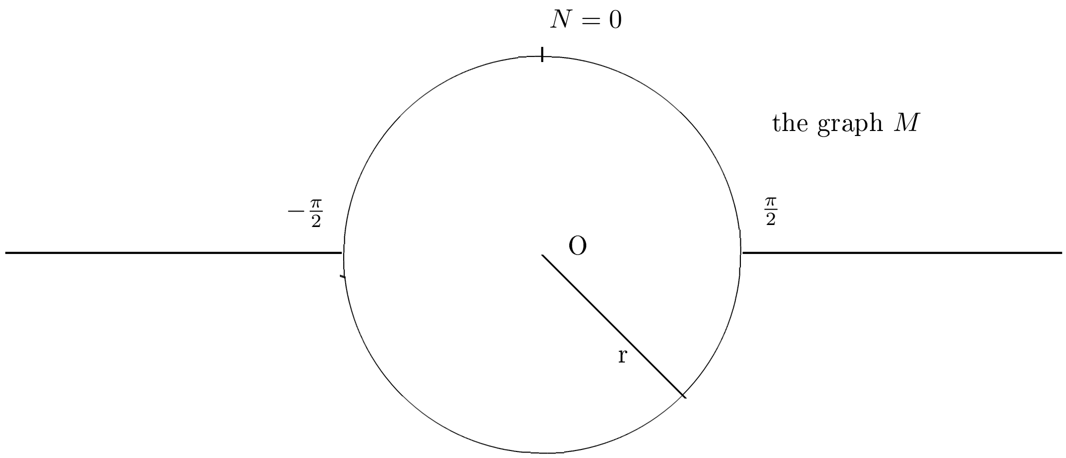


FIGURE 1.1. Circle and real line glued along an arc of length πr.

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

DEFINITION 1.34. 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.

DEFINITION 1.35. 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 1.36. *A maximal δ–separated set in X is a δ–separated net in X.*

PROOF. Let N be a maximal δ–separated set 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 $\mathrm{dist}(x, y) < \delta$. $\square$

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

EXERCISE 1.37. Prove that if (X, dist) is compact then every separated net in X is finite; hence, every separated set in X is finite.

Definition 1.38 (Rips complex). Let (X, d) be a metric space. For $R \geqslant 0$ we define a simplicial complex $\text{Rips}_R(X)$; its vertices are points of X; vertices $x_0, x_1, ..., x_n$ span a simplex if and only if for all i, j,

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

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

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

1.5. Lipschitz maps and Banach-Mazur distance

1.5.1. 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 $x, x' \in X$

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

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

Exercise 1.39. Show that every L-Lipschitz path $\mathfrak{p} : [0, 1] \to X$ is rectifiable and $\text{length}(\mathfrak{p}) \leqslant L$.

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

Theorem 1.40 (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 $f|B(x, \epsilon)$ is Lipschitz. We let $\text{Lip}_{\text{loc}}(X; Y)$ denote the space of locally Lipschitz maps $X \to Y$. We set $\text{Lip}_{\text{loc}}(X) := \text{Lip}_{\text{loc}}(X; \mathbb{R})$.

Exercise 1.41. Fix a point p in a metric space (X, dist) and define the function dist_p by $\text{dist}_p(x) := \text{dist}(x, p)$. Show that this function is 1-Lipschitz.

Lemma 1.42 (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. *φ is positive on $B(p, R)$ and zero on $X \setminus B(p, R)$.*
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 $\varphi := \zeta \circ \text{dist}_p$. □

Lemma 1.43 (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. *$\text{Supp}(\eta_i) \subset \overline{B(x_i, R_i)}$, $\quad \forall i \in I$.*

PROOF. For each i define the bump-function using Lemma 1.42:

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

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. $\square$

REMARK 1.44. Since the collection of balls $\{B_i\}$ is locally finite, it is clear that the function

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

is bounded on compact sets in X, however, in general, it is unbounded on X. We refer the reader to the equation (1.8) for the definition of $\mathrm{Lip}(\eta_i)$.

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

PROPOSITION 1.45. $\mathrm{Lip}_{\mathrm{loc}}(X)$ *is a dense subset in* $C(X)$*, the space of continuous functions* $X \to \mathbb{R}$*, equipped with the compact-open topology (topology of uniform convergence on compacts).*

PROOF. Fix a base-point $o \in X$ and let A_n denote the annulus

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

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

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

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 $x \in X$. $\square$

A relative version of Proposition 1.45 also holds:

PROPOSITION 1.46. *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. *$g|A = f|A$.*

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 1.18. According to Proposition 1.45, 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. $\square$

1.5.2. Bi–Lipschitz maps. The Banach-Mazur distance. A map $f : X \to Y$ 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 1.47. Suppose that X, Y are connected Riemannian manifolds (M, g), (N, h) (see Section 2.1.3). 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 1.48.
(1) If d_1, d_2 are metrics on $\mathbb{R}^n$ defined by two norms on $\mathbb{R}^n$, then d_1, d_2 are bi-Lipschitz equivalent.
(2) Two left-invariant Riemannian metrics on a connected real Lie group define bi-Lipschitz equivalent distance functions.

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

$$\mathrm{Lip}(f) := \inf\{L : f \text{ is } L\text{–Lipschitz}\} \tag{1.8}$$

EXAMPLE 1.49. 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 invertible linear maps $T : V \to W$.

THEOREM 1.50 (John's Theorem, see e.g. [**Ver11**], Theorem 2.1). *For every pair of n-dimensional normed vector spaces V, W, $\mathrm{dist}_{BM}(V, W) \leqslant \log(n)$.*

EXERCISE 1.51. 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.$\mathrm{Lip}(f + g) \leqslant \mathrm{Lip}(f) + \mathrm{Lip}(g)$.

2. $\mathrm{Lip}(fg) \leqslant \mathrm{Lip}(f)\|g\| + \mathrm{Lip}(g)\|f\|$.

3.

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

Note that in case when f is a smooth function on a Riemannian manifold, these formulae follow from the formulae for the derivatives of the sum, product and ratio of two functions.

1.6. Hausdorff dimension

We recall the concept of *Hausdorff dimension* for metric spaces. Let K be a metric space and $\alpha > 0$. The *α–Hausdorff measure* $\mu_\alpha(K)$ is defined as

$$\lim_{r\to 0} \inf \sum_{i=1}^{N} r_i^\alpha, \tag{1.9}$$

where the infimum is taken over all countable coverings of K by balls $B(x_i, r_i)$, $r_i \leqslant r$ $(i = 1, \ldots, N)$. The motivation for this definition is that the volume of the Euclidean r-ball of dimension $a \in \mathbb{N}$ is r^a (up to a uniform constant); hence, Lebesgue measure of a subset of $\mathbb{R}^a$ is (up to a uniform constant) estimated from above by the a-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.

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

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

EXERCISE 1.52. Verify that the Hausdorff dimension of the Euclidean space $\mathbb{R}^n$ is n.

We will need the following theorem:

THEOREM 1.53 (L. Sznirelman; see also [**HW41**]). *Suppose that X is a proper metric space; then the covering dimension* $\dim(X)$ *is at most the Hausdorff dimension* $\dim_H(X)$.

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

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

An immediate consequence of Proposition 1.46 is the following.

Corollary 1.54. *For every function $f \in C(X, A; B^n)$ and an open set $U \subset X$ containing A, there exists a sequence of functions $g_i \in C(X, A; B^n)$ so that for all $i \in \mathbb{N}$:*

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

For a continuous map $f : X \to B^n$ define $A = A_f$ as

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

Definition 1.55. The map f is *essential* if it is homotopic rel. A to a map $f' : X \to S^{n-1}$. An *inessential map* is the one which is not essential.

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

Theorem 1.56 (P. S. Alexandrov, see Theorem III.5 in [**Nag83**]). $\dim(X) < n$ *if and only if every continuous map $f : X \to B^n$ is inessential.*

We are now ready to prove Theorem 1.53. Suppose that $\dim_H(X) < n$. We will prove that $\dim(X) < n$ as well. We need to show that every continuous map $f : X \to B^n$ is inessential. Let D denote the annulus $\{x \in \mathbb{R}^n : 1/2 \leqslant |x| < 1\}$. Set $A := f^{-1}(S^{n-1})$ and $U := f^{-1}(D)$.

Take the sequence g_i given by Corollary 1.54. 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. It follows that $g_i(X)$ misses a point y in $B\left(0, \frac{1}{3}\right)$. Composing g_i with the retraction $B^n \setminus \{y\} \to S^{n-1}$ we get a map $f' : X \to S^{n-1}$ which is homotopic to f rel. A. Thus f is inessential and, therefore, $\dim(X) < n$. $\square$

1.7. 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.

A *norm* 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*.

We will say that a norm $|\cdot|$ is *nonarchimedean* if it satisfies the *ultrametric* inequality

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

We say that $|\cdot|$ is *archimedean* if there exists an isometric monomorphism $R \hookrightarrow \mathbb{C}$. 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. (By a theorem of Gelfand–Tornheim, if a normed field F contains $\mathbb{R}$ as subfield then F is isomorphic, as a field, either to $\mathbb{R}$ or to $\mathbb{C}$.)

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

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

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

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

REMARK 1.57. 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 $|\cdot| : F \setminus \{0\} \to (0, \infty)$ 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 ν, then the subset

$$O_\nu = \{x \in F : \nu(x) \geqslant 0\}$$

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

EXERCISE 1.58. 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 $m, n \in Z$ and m is not divisible by p. Note that p is a uniformizer of $\mathbb{Q}_p$.

REMARK 1.59. We will not use the common notation $\mathbb{Z}_p$ for O_p, in order to avoid the confusion with finite cyclic groups.

EXERCISE 1.60. 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 $a_k \in \{0, \dots, p-1\}$. 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. In other words, $\nu(x) = -n$ is the valuation. In particular, the ring O_p is identified with the set of series

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

REMARK 1.61. In other words, one can describe p-adic numbers as left-infinite sequences of (base p) digits

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

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

EXERCISE 1.62. 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 $\nu(0) = \infty$ and for nonzero f let $\nu(f)$ be the least n so that $a_n \neq 0$. In other words, $\nu(f)$ is the order of vanishing of f at $0 \in R$.

EXERCISE 1.63. 1. Verify that ν is a valuation on R. Define $|f| := e^{-\nu(f)}$.

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 $\nu(f)$ 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 1.64. 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?

LEMMA 1.65. *$\mathbb{Q}_p$ 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, it suffices 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

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

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

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

EXERCISE 1.66. Show that O_p is homeomorphic to the Cantor set. Hint: Verify that O_p is totally disconnected and perfect.

1.8. 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 $V = F^n$ over a normed field F, with the standard basis $e_1, \ldots, e_n$. We equip V with the usual Euclidean/hermitian norm in the case F is archimedean and with the max-norm

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

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

Exercise 1.67. Suppose that F is nonarchimedean. Show that the metric $|v - w|$ on V satisfies the ultrametric triangle inequality.

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

Exercise 1.68. If F is a nonarchimedean local field, show that the group K is compact with respect to the subset topology induced from $Mat_n(F) = F^{n^2}$.

Lemma 1.69. *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 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:

Theorem 1.70 (Singular Value Decomposition Theorem). *If F is archimedean, then every matrix $M \in End(V)$ admits a* singular valued 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 now prove an analogue of the singular value decomposition in the case of nonarchimedean normed fields:

Theorem 1.71 (Smith Normal Form Theorem). *Let F be a field with discrete norm and uniformizer π and ring of integers O. Then every matrix $M \in Mat_n(F)$ 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 $d_i \in F$ 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 F. We now 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 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 $A \in End(F^n)$. 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 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. $\square$

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 $Mat_n(F)$ 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 1.72. Deduce the Cartan decomposition in $F = \mathbb{R}$ or $F = \mathbb{C}$, from the statement that given any Euclidean/hermitian bilinear form q on $V = F^n$, there exists a basis orthogonal with respect to q and orthonormal with respect to the standard inner product

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

We now turn our discussion to projective spaces. The F-projective space $P = F\mathbb{P}^n$ is the quotient of $F^{n+1} \setminus \{0\}$ by the action of $F^\times$ *via* scalar multiplication. We let $[v]$ denote the projection of a nonzero vector $v \in V = F^{n+1}$ to $F\mathbb{P}^n$. The j-th affine coordinate patch on P is the affine subspace $A_j \subset V$,

$$A_j = (x_1, \ldots, 1, \ldots, x_{n+1}),$$

where 1 appears in the j-th coordinate.

NOTATION 1.73. Given a nonzero vector $v \in V$ let $[v]$ denote the projection of v to the projective space $\mathbb{P}(V)$; similarly, for a subset $W \subset V$ we let $[W]$ denote the image of $W \setminus \{0\}$ under the canonical projection $V \to \mathbb{P}(V)$. Given an invertible linear map $g : V \to V$, we will retain the notation g for the induced projective map $\mathbb{P}(V) \to \mathbb{P}(V)$.

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

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

to be orthonormal. Then

$$|v \wedge w|^2 = |v|^2 |w|^2 - \langle v, w\rangle \langle w, v\rangle.$$

Note that if u, v are unit vectors with $\angle(v, w) = \varphi$, then $|v \wedge w| = |\sin(\varphi)|$.

In the case when 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 $v = (x_1, \ldots, x_{n+1})$, $w = (y_1, \ldots, y_{n+1})$.

LEMMA 1.74. *Suppose that u is a unit vector and $v \in V$ is such that $|u_i - v_i| \leqslant \epsilon$ for all i. Then*

$$|v \wedge w| \leqslant 2(n+1)\epsilon.$$

PROOF. We will consider the archimedean case since the nonarchimedean case is similar. For every i let $\delta_i = v_i - u_i$. Then

$$|u_i v_j - u_j v_i|^2 \leqslant |u_i \delta_j - u_j \delta_i|^2 \leqslant 4\epsilon^2$$

Thus,

$$|u \wedge v|^2 \leqslant 4(n+1)^2 \epsilon^2. \quad \square$$

DEFINITION 1.75. The *chordal metric* on $P = F\mathbb{P}^n$ is defined by

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

In the nonarchimedean case this definition is due to A. Néron [**N$\acute{6}$4**].

EXERCISE 1.76. 1. If F is nonarchimedean, show that the group $GL(n+1, O)$ preserves the chordal metric.

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

3. If $F = \mathbb{C}$, show that the unitary group preserves 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 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 1.77. *If 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 F^n, 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. We, then, 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 1.78. *$d([x], [y]) = \mathrm{dist}([x], [y])$ 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,

$$|x_i - y_i| = |x_i y_{n+1} - x_j y_{n+1}| \leqslant \max_j |x_i y_j - x_j y_i| \leqslant d([x], [y]),$$

which implies that

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

We will now prove the opposite inequality:

$$\forall i, j \quad |x_i y_j - x_j y_i| \leqslant a := |x - y|.$$

There exist $z_i, z_j \in F$ 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 $x_i x_j \neq 0$, leaving the exceptional cases to the reader. Then,

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

Computing $x_i y_j - x_j y_i$ using the new variables z_i, z_j, we obtain:

$$|x_i y_j - x_j y_i| = |x_i x_j (1 + z_j) - x_i x_j (1 + z_i)| = |x_j x_j (z_j - z_i)| \leqslant$$

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

since $x_i, x_j \in O$.

2. Suppose that (1) does not happen. Since $d([x],[y]) \leqslant 1$ and $\mathrm{dist}([x],[y]) = 1$ (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

$$|x_i y_j - x_j y_i| = |y_j|$$

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,

$$d([x],[y]) \geqslant \frac{|x_i y_j - x_j y_i|}{|x_i| \cdot |y_j|} = \frac{|y_j - x_j|}{|y_j|}.$$

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. □

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 1.79. *Chordal metric is a metric on P in both archimedean and nonarchimedean cases.*

EXERCISE 1.80. Suppose that 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 F is local, verify that P is compact.
3. If the norm on F is complete, show that the metric space (P, d) is complete.
4. If H is a hyperplane in $V = F^{n+1}$, given as $\operatorname{Ker} f$, where $f : V \to F$ is a linear function, show that

$$\operatorname{dist}([v], [H]) = \frac{|f(v)|}{\|v\| \, \|f\|}.$$

1.9. Kernels and distance functions

A *kernel* on a set X is a symmetric map $\psi : X \times X \to \mathbb{R}_+$ such that $\psi(x, x) = 0$. Fix $p \in X$ and define the associated *Gromov kernel*

$$k(x, y) := \frac{1}{2} \left(\psi(x, p) + \psi(p, y) - \psi(x, y) \right).$$

If X were a metric space and $\psi(x, y) = \operatorname{dist}^2(x, y)$, then this quantity is just the Gromov product in X where distances are replaced by their squares (see Section 8.3 for the definition of Gromov product in metric spaces). Clearly,

$$\forall x \in X, \quad k(x, x) = \psi(x, p).$$

DEFINITION 1.81. 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 $\lambda \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \geqslant 0. \tag{1.10}$$

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:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \leqslant 0. \tag{1.11}$$

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 (1.10) becomes simply $\Psi(f, f)$, where

$$\lambda_i := f(x_i), \quad \operatorname{Supp}(f) \subset \{x_1, \ldots, x_n\} \subset X.$$

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 1.82. 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 (1.11) equals

$$q(\lambda) = \lambda^T M \lambda,$$

a symmetric bilinear form on $\mathbb{R}^n$. Then, the condition (1.11) 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 1.83. *A kernel ψ on X is conditionally negative definite if and only if there exists a map $F : X \to \mathcal{H}$ to a Hilbert space so that*

$$\psi(x, y) = \|F(x) - F(y)\|^2.$$

PROOF. 1. Suppose that the map F exists. Then, for every $p = x_0 \in X$, the associated Gromov kernel $k(x, y)$ equals

$$k(x, y) = \langle F(x), F(y) \rangle,$$

and, hence, for every finite subset $\{x_0, x_1, \ldots, x_n\} \subset X$, the corresponding matrix with the entries $k(x_i, x_j)$ 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

$$\sum_{i=0}^{n} f(x_i) = 0. \tag{1.12}$$

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

$$\sum_{i,j=1}^{n} \left(|y_0 - y_i|^2 + |y_0 - y_j|^2 - |y_i - y_j|^2\right) f(x_i) f(x_j) = \tag{1.13}$$

$$2\sum_{i,j=1}^{n} k(x_i,x_j)f(x_i)f(x_j) \geqslant 0.$$

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 (1.13), we conclude that

$$\sum_{i,j=0}^{n} \psi(x_i,x_j)f(x_i)f(x_j) \leqslant 0. \tag{1.14}$$

In other words, the kernel ψ on X is conditionally negative semidefinite.

2. Suppose that ψ is conditionally negative definite. Fix $p \in X$ and define the Gromov kernel

$$k(x,y) := (x,y)_p := \frac{1}{2}\left(\psi(x,p)+\psi(p,y)-\psi(x,y)\right).$$

The key to the proof is:

LEMMA 1.84. *k is a positive semidefinite kernel on X.*

PROOF. Consider a subset $X_0 = \{x_1,\dots,x_n\} \subset X$ 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 $f : \{x_0,\dots,x_n\} \to \mathbb{R}$ satisfies (1.12) 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 1.83 (run in the reverse) then shows that

$$\sum_{i,j=1}^{n} k(x_i,x_j)f(x_i)f(x_j) \geqslant 0.$$

Thus, k 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),

$$K(g, g) \geqslant 0.$$

On the other hand,

$$K(f, f) = F(g, g) + 2cK(g, \delta_p) + c^2 K(\delta_p, \delta_p) = F(g, g),$$

since the other two terms vanish (as $k(x, p) = 0$ for every $x \in X$). Thus, K is positive semidefinite. □

Now, consider the vector space $V = V(X)$ equipped with the positive semi-definite 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; we obtain:

$$\langle F(x), F(y)\rangle = k(x, y).$$

Let us verify now that

$$\langle F(x) - F(y), F(x) - F(y)\rangle = \psi(x, y). \tag{1.15}$$

The left hand side of this expression equals

$$\langle F(x), F(x)\rangle + \langle F(y), F(y)\rangle - 2k(x, y) = \psi(x, p) + \psi(y, p) - 2k(x, y).$$

Then, the equality (1.15) follows from the definition of the Gromov kernel k. □

According to [**Sch38**], for every conditionally negative definite kernel $\psi : X \times X \to \mathbb{R}_+$ and every $0 < \alpha \leqslant 1$, the power ψ^α is also a conditionally negative definite kernel.

CHAPTER 2

Geometric preliminaries

2.1. Differential and Riemannian geometry

In this book we will use some elementary Differential and Riemannian geometry, basics of which are reviewed in this section. All the manifolds that we consider are second countable.

2.1.1. 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.

Recall that, given a smooth n–dimensional manifold M, a *k–dimensional submanifold* is a closed 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 $\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ 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 $\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is of the form*

$$\overline{f}(x_1, \dots, x_n, \dots, x_m) = (x_1, \dots, x_n).$$

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 dimension $m - n$.*

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.*

Sard's theorem has an important quantitative improvement due to Y. Yomdin which we will describe below. Let B be the closed unit ball in $\mathbb{R}^{n-1}$. Consider a C^n-smooth function $f : B \to \mathbb{R}$. For every multi-index $i = (i_1, \dots, i_k)|$ set $|i| := k$, and for $k \leqslant n$ let

$$\partial^i f := \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}$$

be the *i-th mixed partial derivative of f*. Let

$$\|\partial^i f\| := \max_x |\partial^i f(x)|.$$

Define the C^n-*norm* of f as

$$\|f\|_{C^n} := \max_{i, 0 \leqslant |i| \leqslant n+1} \|\partial^i f\|.$$

Given $\epsilon > 0$ let $E_\epsilon \subset \mathbb{R}$ denote the set

$$\{y \in \mathbb{R} : \exists x \in f^{-1}(y), |\nabla f(x)| < \epsilon\}.$$

Thus, the set E_ϵ consists of "almost" critical values of f. Yomdin's theorem informally says that for small ϵ the set E_ϵ is small. Below is the precise statement.

THEOREM 2.7 (Y. Yomdin, [**Yom83**]). *There exists a constant $c = c(n, \|f\|_{C^n})$ so that for every C^n-smooth function $f : B \to \mathbb{R}$, and every $\epsilon \in (0, 1)$ the set E_ϵ can be covered by at most c/ϵ intervals of length $\epsilon^{n/(n-1)}$. In particular:*

1. Lebesgue measure of E_ϵ is at most

$$c\epsilon^{\frac{1}{n-1}}.$$

2. Whenever an interval $J \subset \mathbb{R}$ has length $\ell > c\epsilon^{1/(n-1)}$, there exists a subinterval $J' \subset J \setminus E_\epsilon$, so that J' has length at least

$$\frac{c}{\epsilon}\left(\ell - c\epsilon^{1/(n-1)}\right).$$

2.1.2. Smooth partition of unity.

DEFINITION 2.8. 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 2.9. *Every open cover $\mathcal{U}$ as above admits a smooth partition of unity.*

2.1.3. Riemannian metrics. A *Riemannian metric* 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 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 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 *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 U 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 $v \in T_x U$ 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 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 computation of the pull-back metric $f^*(g)$), 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 particular 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 $T_p N$ of $T_p M$, 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 quasi-conformal groups, Chapter **??**, section **??**.

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 length($\mathfrak{p}$) by length($\mathfrak{p}([a, b])$).

Then, provided that M is connected, one defines the Riemannian *distance function*

$$\text{dist}(p, q) = \inf_{\mathfrak{p}} \text{length}(\mathfrak{p}),$$

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 diffeomorphism, then it preserves the Riemannian distance function.

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,

$$\text{dist}(\mathfrak{p}(t_1), \mathfrak{p}(t_2)) = \text{length}(\mathfrak{p}([t_1, t_2])) = c|t_1 - t_2|.$$

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 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 $\gamma'(0) = v$. The *injectivity radius* $InjRad(p)$ 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**].

2.1.4. Riemannian volume. For every n-dimensional Riemannian manifold (M, g) one defines the *volume element* (or *volume density*) denoted dV (or dA if M is 2-dimensional). Given n vectors $v_1, \ldots, v_n \in T_pM$, $dV(v_1 \wedge \ldots \wedge v_n)$ 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, then its volume density is given by

$$\rho^n(x)dx_1 \dots dx_n.$$

Thus, every Riemannian manifold has a canonical measure, given by the integral of its volume form

$$mes(E) = \int_A dV.$$

THEOREM 2.10 (Generalized Rademacher's theorem). *Let $f : M \to N$ be a Lipschitz map of Riemannian manifolds. Then f is differentiable almost everywhere.*

EXERCISE 2.11. Deduce Theorem 2.10 from Theorem 1.40 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 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, so 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). Definition 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 a domain 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 $\zeta_\alpha = \eta_\alpha \circ \varphi_\alpha$, $f_\alpha = f \circ \varphi_\alpha$,

$$Vol(f) = \sum_{\alpha\in J} \int_{U_\alpha} \zeta_\alpha \sqrt{|\det(G_{f_\alpha})|} dx_1 \ldots dx_n$$

In particular, if f is 1-1 and $\Sigma = f(N)$, then

$$Vol(f) = Vol(\Sigma).$$

REMARK 2.12. The formula for $Vol(f)$ makes sense when $f : N \to M$ is merely Lipschitz, in view of Theorem 2.10.

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 2.13. (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,

$$Vol(f(\Omega)) = |\int_\Omega J_f dx_1 \ldots dx_n| \leqslant Vol(f) \leqslant Vol(\Omega).$$

Hint: Use that under the linear map $A = d_x f$, the image of every r-ball is contained in 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 2.14. 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 2.15. *Consider a compact Riemannian manifold M^m, a submersion $f : M^m \to N^n$, and a point $p \in N$. 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, $\varphi_x : U_x \to \overline{U}_x$, with U_x containing x, and a chart of N, $\psi_x : V_x \to \overline{V}_x$ 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 $\overline{U}_x$ is an open cube in $\mathbb{R}^m$. Therefore, $\overline{V}_x$ is also a cube in $\mathbb{R}^n$, and $\overline{U}_x = \overline{V}_x \times \overline{Z}_x$, where $\overline{Z}_x$ 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 $l \in \{1, \dots, k\}$. Let $(g_{ij}(y))_{1 \leqslant i,j \leqslant n}$ be the matrix-valued function on $\overline{U}_l$, defining the pull-back by φ_l of the Riemannian metric on M.

Since g_{ij} is continuous, there exists a neighborhood $\overline{W}_l$ of $\bar{p} = \psi_l(p)$ such that for every $\bar{x} \in W_l$ and for every $\bar{t} \in \overline{Z}_l$ we have,

$$(1-\epsilon)^2 \leqslant \frac{\det\left[g_{ij}(\bar{x},\bar{t})\right]_{n+1\leqslant i,j\leqslant m}}{\det\left[g_{ij}(\bar{p},\bar{t})\right]_{n+1\leqslant i,j\leqslant m}} \leqslant (1+\epsilon)^2 .$$

Recall that the volumes of $M_x \cap U_i$ and of $M_p \cap U_l$ are obtained by integrating respectively $(\det\left[g_{ij}(\bar{x},\bar{t})\right]_{n+1\leqslant i,j\leqslant k})^{1/2}$ and $(\det\left[g_{ij}(\bar{p},\bar{t})\right]_{n+1\leqslant i,j\leqslant k})^{1/2}$ 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 $x \in \bigcap_{i=1}^k \psi_i^{-1}(\overline{W}_l)$,

$$1-\epsilon \leqslant \frac{Vol(M_x)}{Vol(M_p)} \leqslant 1+\epsilon .$$

$\square$

Finally, we recall an important formula for volume computations:

THEOREM 2.16 (Coarea formula, see e.g. Theorem 6.3 **[Cha06]**). *Let U be an open connected subset with compact closure $\overline{U}$ in a Riemannian manifold M and let $f : U \to (0, \infty)$ be a smooth submersion with a continuous extension to $\overline{U}$ such that f restricted to $\overline{U} \setminus U$ is constant. For every $t \in (0,\infty)$ let $\mathcal{H}_t$ denote the level set $f^{-1}(t)$, and let $\mathrm{d}A_t$ be the Riemannian area density induced on $\mathcal{H}_t$.*

Then, for every function $g \in L^1(U)$,

$$\int_U g\,|\mathrm{grad} f|\mathrm{d}V = \int_0^\infty \mathrm{d}t \int_{\mathcal{H}_t} g\,\mathrm{d}A_t$$

where $\mathrm{d}V$ is the Riemannian volume density of M

2.1.5. Growth function and Cheeger constant. In this section we present two basic notions initially introduced in Riemannian geometry and later adapted and used in group theory and in combinatorics.

Given a Riemannian manifold (M, g) and a point $x_0 \in M$, we define the *growth function*

$$\mathfrak{G}_{M,x_0}(r) := Vol\, B(x_0, r),$$

the volume of the metric ball of radius r and center at x in (M, g)

REMARKS 2.17. (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 group of isometries of M is cobounded, i.e., there exists κ such that the $\mathrm{Isom}(M)$-orbit of $B(x_0, \kappa)$ equals M. In this case, 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 **??** for the detailed discussion of volume growth and its relation to group growth.

EXERCISE 2.18. Assume again that the action $\mathrm{Isom}(M) \curvearrowright M$ is cobounded and that (M, g) is complete.

(1) Prove that the growth function is *almost sub-multiplicative*, that is:

$$\mathfrak{G}_{M,x_0}((r+t)\kappa) \leqslant \mathfrak{G}_{M,x_0}(r\kappa)\mathfrak{G}_{M,x_0}((t+1)\kappa).$$

(2) Prove that the growth function of M is at most exponential, that is there exists $a > 1$ such that

$$\mathfrak{G}_{M,x_0}(x) \leqslant a^x, \text{ for every } x \geqslant 0.$$

DEFINITION 2.19. An *isoperimetric inequality* in a manifold M is an inequality satisfied by all open submanifolds Ω with compact closure and smooth boundary, of the form

$$Vol(\Omega) \leq f(\Omega) g\left(Area \partial\Omega\right),$$

where f and g are real-valued functions, g defined on $\mathbb{R}_+$.

DEFINITION 2.20. 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 submanifolds with compact closure and smooth boundary.

If in particular $h(M) \geq \kappa > 0$ then the following isoperimetric inequality holds in M:

$$Vol(\Omega) \leqslant \frac{1}{\kappa} Area(\partial\Omega) \text{ for every } \Omega.$$

This notion 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 increase at the same rate, and, moreover, that their behavior is essentially determined by the fundamental group only. See Proposition **??** 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 11.12. This was proved much later by R. Brooks [**Bro81, 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 the quoted 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.

2.1.6. 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 *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.

THEOREM 2.21 (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.

Manifolds of bounded geometry. A (complete) Riemannian manifold M is said to have *bounded geometry* if there are constants a, b and $\epsilon > 0$ so that:
1. Sectional curvature of M varies in the interval $[a, b]$.
2. Injectivity radius of M is $\geqslant \epsilon$.

The numbers a, b, ϵ are called *geometric bounds* on M. For instance, every compact Riemannian manifold M has bounded geometry, every covering space of M (with pull-back Riemannian metric) also has bounded geometry.

THEOREM 2.22 (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 2.23. *Let M be a Riemannian manifold of bounded geometry and let $\mathcal{U} = \{B_i = B(x_i, r_i) : i \in I\}$ 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.

Curvature and volume.

Below we describe without proof certain consequences of uniform lower and upper bounds on the sectional curvature on the growth of volumes of balls, that will be used in the sequel. The references for the result below are [**BC01**, Section 11.10], [**CGT82**], [**Gro86**], [**G$\ddot{\text{6}}$0**]. See also [**GHL04**], Theorem 3.101, p. 140.

Below we will use the following notation: For $\kappa \in \mathbb{R}$, $A_\kappa(r)$ and $V_\kappa(r)$ will 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 Riemannian manifold M. Likewise, $V(x, r)$ will denote the volume of the geodesic ball centered at x and of radius r in M.

THEOREM 2.24 (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. The, 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 $(0, \rho_x)$.*

The results (1) in the theorem above are also true if the Ricci curvature of M is at least $(n-1)a$.

Theorem 2.24 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 2.24, (1), is the following theorem which will be useful in the proof of quasi-isometric invariance of positivity of the Cheeger constant:

THEOREM 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 $\mathcal{O}_1 \mathcal{O}_2$ separated by $\mathcal{H}$, we have:*

$$\min \left[Vol(\mathcal{O}_1) \, , \, Vol(\mathcal{O}_2) \right] \leqslant \lambda Area \left[\mathcal{H} \cap B(x,r) \right].$$

2.1.7. Harmonic functions. For the detailed discussion of the material in this section we refer the reader to [**Li04**] 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

$$E(f) = \int_M |df|^2 dV = \int_M |\nabla f|^2 dV.$$

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.

THEOREM 2.26 (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 2.27. 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 $B = B(p,r) \subset M$,

$$E(h|B) \leqslant E(u), \quad \forall u : \bar{B} \to \mathbb{R}, u|_{\partial B} = h|_{\partial B}.$$

Equivalently, for every relatively compact open subset $\Omega \subset M$ with smooth boundary

$$E(h|B) \leqslant E(u), \quad \forall u : \bar{\Omega} \to \mathbb{R}, u|_{\partial \Omega} = h|_{\partial \Omega}.$$

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:

$$\Delta u = \operatorname{div} \nabla u$$

Here for a vector field X on M, the *divergence* $\operatorname{div} X$ is a function on M satisfying

$$\operatorname{div} X dV = L_X dV,$$

where L_X is the *Lie derivative* along the vector field X:

$$L_X : \Omega^k(M) \to \Omega^k(M),$$

$$L_X(\omega) = i_X d\omega + d(i_X \omega),$$

$$i_X : \Omega^{\ell+1}(M) \to \Omega^\ell(M), \quad i_X(\omega)(X_1, \ldots, X_\ell) = \omega(X, X_1, \ldots, X_\ell).$$

In local coordinates (assuming that M is n-dimensional):

$$\operatorname{div} X = \sum_{i=1}^{n} \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left(\sqrt{|g|} X_i \right)$$

where

$$|g| = \det((g_{ij})),$$

and

$$(\nabla u)^i = \sum_{j=1}^{n} g^{ij} \frac{\partial u}{\partial x_j}$$

and $(g^{ij}) = (g_{ij})^{-1}$, the inverse matrix of the metric tensor. Thus,

$$\Delta u = \sum_{i,j=1}^{n} \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left(g^{ij} \sqrt{|g|} \frac{\partial u}{\partial x_j} \right).$$

In terms of the Levi–Civita connection ∇ on M,

$$\Delta(u) = Trace(H(u)), \quad H(u)(X_i, X_j) = \nabla_{X_i} \nabla_{X_j}(u) - \nabla_{\nabla_{X_i} X_j}(u),$$

$$Trace(H) = \sum_{i,j=1}^{n} g^{ij} H_{ij},$$

where X_i, X_j are vector fields on M.

If $M = \mathbb{R}^n$ with the flat metric, then Δ is the usual Laplace operator:

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} u.$$

THEOREM 2.28 (Yau's gradient estimate). *Suppose that M^n is a complete n-dimensional Riemannian manifold with Ricci curvature $\geqslant a$. Then for every harmonic function h on M, every $x \in M$ with $InjRad(x) \geqslant \epsilon$,*

$$|\nabla h(x)| \leqslant h(x) C(\epsilon, n).$$

THEOREM 2.29 (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.*

THEOREM 2.30 (Maximum Principle). *Let $\Omega \subset M$ be a relatively compact domain with smooth boundary and $h : M \to \mathbb{R}$ be a harmonic function. Then $h|\overline{\Omega}$ attains maximum on the boundary of Ω and, moreover, if $h|\Omega$ attains its maximum at a point of Ω, then h is constant.*

2.1.8. Alexandrov curvature and $CAT(\kappa)$ spaces. In the more general setting of 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 less used.

For a given $\kappa \in \mathbb{R}$, 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 7, it is the *upper half-plane with the rescaled hyperbolic metric:*

$$X_\kappa = \left(U^2, |\kappa|^{-1} \frac{dx^2 + dy^2}{y^2} \right).$$

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 §7.10) the following:

Lemma 2.31. *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 lengths of edges a, b and c, and it 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* $\tilde{P} \in [\tilde{A}, \tilde{B}]$, so that

$$d(A, P) = d(\tilde{A}, \tilde{P}).$$

Thus, for $P \in [A, B], Q \in [B, C]$ we define κ–comparison points $\tilde{P}, \tilde{Q} \in \tilde{\Delta}$.

Definition 2.32. 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

$$\mathrm{dist}_{X_\kappa}\left(\tilde{P}, \tilde{Q}\right) \geqslant \mathrm{dist}_X (P, Q) .$$

Definition 2.33.

(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) We say that 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 2.33 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 2.34. *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 2.35. 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). On the other hand, Hilbert spaces provide natural examples of complete CAT(0) metric spaces.

EXERCISE 2.36. Let X be a simplicial tree with a path-metric d. Show that (X, d) is $CAT(-\infty)$.

In the case of non-positive curvature there exists a local-to-global result.

THEOREM 2.37 (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 2.38. 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. We say that a function $F : X \times X \to \mathbb{R}$ is *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 $[a_0, a_1]$ and $[b_0, b_1]$ in X, the points $a_s \in [a_0, a_1]$ and $b_s \in [b_0, b_1]$ such that $\mathrm{dist}(a_0, a_s) = s\mathrm{dist}(a_0, a_1)$ and $\mathrm{dist}(b_0, b_s) = s\mathrm{dist}(b_0, b_1)$ satisfy

$$\mathrm{dist}(a_s, b_s) \leq (1 - s)\mathrm{dist}(a_0, b_0) + s\mathrm{dist}(a_1, b_1)\,. \tag{2.3}$$

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 sup-graph

$$\{(x, y, t) \in X^2 \times \mathbb{R} : f(x, y) \geqslant t\}$$

is convex.

PROPOSITION 2.39. *A geodesic metric space X is $CAT(0)$ if and only if the distance on X is convex.*

PROOF. Consider two geodesics $[a_0, b_0]$ and $[a_1, b_1]$ in X. On the geodesic $[a_0, b_1]$ consider the point c_s such that $\mathrm{dist}(a_0, c_s) = s\mathrm{dist}(a_0, b_1)$. The fact that the triangle with edges $[a_0, a_1]$, $[a_0, b_1]$ and $[a_1, b_1]$ is $CAT(0)$ and the Thales theorem in $\mathbb{R}^2$, imply that $\mathrm{dist}(a_s, c_s) \leq s\mathrm{dist}(a_1, b_1)$. The same argument applied to the triangle with edges $[a_0, b_1], [a_0, b_0], [b_0, b_1]$, implies that $\mathrm{dist}(c_s, b_s) \leq (1 - s)\mathrm{dist}(a_0, b_0)$. The inequality (2.3) follows from

$$\mathrm{dist}(a_s, b_s) \leqslant \mathrm{dist}(a_s, c_s) + \mathrm{dist}(c_s, b_s)\,.$$

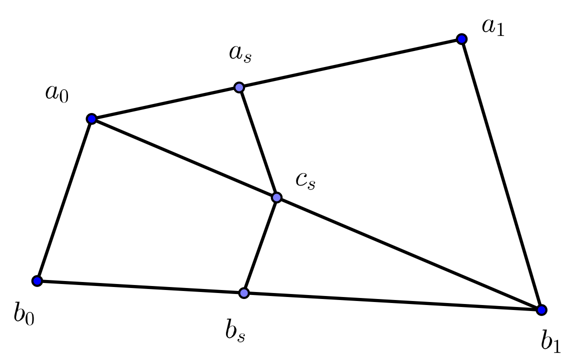


FIGURE 2.1. Argument for convexity of the distance.

Conversely, assume that (2.3) is satisfied.

In the special case when $a_0 = a_1$, this implies the comparison property in Definition 2.32 when one of the two points P, Q is a vertex of the triangle. When $a_0 = b_0$, (2.3) again implies the comparison property when $\frac{\mathrm{dist}(A,P)}{\mathrm{dist}(A,B)} = \frac{\mathrm{dist}(B,Q)}{\mathrm{dist}(B,C)}$.

We now consider the general case of two points $P \in [A, B]$ and $Q \in [B, C]$ such that $\frac{\mathrm{dist}(A,P)}{\mathrm{dist}(A,B)} = s$ and $\frac{\mathrm{dist}(B,Q)}{\mathrm{dist}(B,C)} = t$ with $s < t$. Consider $B' \in [A, B]$ such that $\mathrm{dist}[A, B'] = \frac{s}{t}\mathrm{dist}[A, B]$. Then, according to the above, $\mathrm{dist}(B', C) \leqslant \mathrm{dist}(\widetilde{B'}, \widetilde{C})$, and $\mathrm{dist}(P, Q) \leqslant t\mathrm{dist}(B', C) \leqslant t\mathrm{dist}(\widetilde{B'}, \widetilde{C}) = \mathrm{dist}(\widetilde{P}, \widetilde{Q})$. $\square$

COROLLARY 2.40. *Every $CAT(0)$-space X is uniquely geodesic.*

PROOF. It suffices to apply the inequality (2.3) to any geodesic bigon, that is, in the special case when $a_0 = b_0$ and $a_1 = b_1$. □

2.1.9. 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 2.41. *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 $[x, x'] \subset X$; 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 (1.2)):

$$d^2(\tilde{x}, \tilde{x}') + 4\, d^2(\tilde{a}, \tilde{p}) = 2\left(d^2(\tilde{a}, \tilde{x}) + d^2(\tilde{a}, \tilde{x}')\right).$$

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 Cartan in the context of Riemannian manifolds of nonpositive curvature:

THEOREM 2.42. *Let X be a complete $CAT(0)$ metric space and $G \curvearrowright X$ be a group acting isometrically with bounded orbits. 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 2.43. *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 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 $\mathcal{H}$.

2.1.10. Ideal boundary, horoballs and horospheres. In this section we define the ideal boundary of a metric space. This is a particularly significant object 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 2.44. 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 $\alpha \in \partial_\infty X$ 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 $\rho(\infty) = \alpha$.

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 2.45. 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 2.46. *If X is locally compact then for every point $x \in X$ and every point $\alpha \in \partial_\infty X$ there exists a unique geodesic ray ρ with $\rho(0) = x$ and $\rho(\infty) = \alpha$. We will also use the notation $[x, \alpha)$ for the ray ρ.*

PROOF. Let $r : [0, \infty) \to X$ be a geodesic ray with $r(\infty) = \alpha$. For every $n \in \mathbb{N}$, according to Corollary 2.40, 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 $\mathrm{dist}(x, r(0))$ 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 $\mathrm{dist}(x, r(0))$ 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 $\mathrm{dist}(\rho_1(t), \rho_2(t)) \leqslant M$, for every $t \geqslant 0$. Consider $t \in [0, \infty)$, and $\varepsilon > 0$ arbitrarily small. Convexity of the distance function implies that

$$\mathrm{dist}(\rho_1(t), \rho_2(t)) \leqslant \varepsilon \mathrm{dist}(\rho_1(t/\varepsilon), \rho_2(t/\varepsilon)) \leqslant \varepsilon M .$$

It follows that $\operatorname{dist}(\rho_1(t), \rho_2(t)) = 0$ and, hence, $\rho_1 = \rho_2$. $\square$

In particular, for a fixed point $p \in X$ one can identify the set $\bar{X} := X \sqcup \partial_\infty X$ with the set of geodesic segments and rays with initial point p. In what follows, we will equip $\bar{X}$ with the topology induced from the compact-open topology on the space of these segments and rays.

EXERCISE 2.47.
(1) Prove that the embedding $X \hookrightarrow \bar{X}$ is a homeomorphism to its image.
(2) Prove that the topology on $\bar{X}$ 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 $x \in \bar{X}$) 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 $\bar{X}$ is compact.

Consider a geodesic ray $r : [0, \infty) \to X$, and an arbitrary point $x \in X$. The function $t \mapsto \operatorname{dist}(x, r(t)) - t$ is decreasing (due to the triangle inequality) and bounded from below by $-\operatorname{dist}(x, r(0))$. Therefore, there exists a limit

$$f_r(x) := \lim_{t \to \infty} [\operatorname{dist}(x, r(t)) - t] \,. \tag{2.4}$$

DEFINITION 2.48. The function $f_r : X \to \mathbb{R}$ thus defined, is called the *Busemann function for the ray r.*

For the proof of the next lemma see e.g. [**Bal95**], Chapter 2, Proposition 2.5.

LEMMA 2.49. *If r_1 and r_2 are two asymptotic rays then $f_{r_1} - f_{r_2}$ 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 2.50. Show that f_r is linear with slope -1 along the ray r. In particular,

$$\lim_{t \to \infty} f_r(t) = -\infty.$$

DEFINITION 2.51. A sublevel set of a Busemann function, $f_r^{-1}(-\infty, a]$ is called a *(closed) horoball with center (or footpoint)* $\alpha = r(\infty)$; we sometime denote such a set $\overline{B}(\alpha)$. A level set $f_r^{-1}(a)$ of a Busemann function is called a *horosphere with footpoint* α, it is denoted $H(\alpha)$. Lastly, an open sublevel set $f_r^{-1}(-\infty, a)$ is called an open horoball with footpoint $\alpha = r(\infty)$, and denoted $B(\alpha)$.

LEMMA 2.52. *Let r be a geodesic ray and let B be the open horoball $f_r^{-1}(-\infty, 0)$. Then $B = \bigcup_{t \geqslant 0} B(r(t), t)$.*

PROOF. Indeed, if for a point x, $f_r(x) = \lim_{t\to\infty}[\text{dist}(x, r(t)) - t] < 0$, 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 $t \mapsto \text{dist}(x, r(t)) - t$ is decreasing, it follows that for every $t \geqslant s$,

$$\text{dist}(x, r(t)) - t \leqslant \delta_s \, .$$

Whence, $f_r(x) \leqslant \delta_s < 0$. $\square$

LEMMA 2.53. *Let X be a $CAT(0)$ space. Then every Busemann function on X is convex and 1-Lipschitz.*

PROOF. Note that distance function on any metric space is 1-Lipschitz (by the triangle inequality). Since Busemann functions are limits of normalized distance functions, it follows that Busemann functions are 1-Lipschitz as well. (This part does not require $CAT(0)$ assumption.) Similarly, since distance function is convex, Busemann functions are also convex as limits of normalized distance functions. $\square$

Furthermore, if X is a Hadamard manifold, then every Busemann function f_r is smooth, with gradient of constant norm 1, see [**BGS85**].

LEMMA 2.54. *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 $f = f_r$ be a Busemann function. Consider the closed horoball

$$\bar{B} = \{x : f(x) \leqslant t\}.$$

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

$$f(y) < f(x) = t, \quad \forall y \in [x, r(s)] \setminus \{x\}.$$

Therefore, x belongs to the closure of the open horoball B, which implies that $\bar{B}$ is the closure of B. $\square$

EXERCISE 2.55. 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

$$f_r(x) = -x \cdot u.$$

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 $\bar{B}$. Can this happen in the case of Hadamard manifolds?

2.2. Bounded geometry

In this section we review several notions of bounded geometry for metric spaces.

2.2.1. Riemannian manifolds of bounded geometry.

DEFINITION 2.56. We say a Riemannian manifold M has *bounded geometry* if it is connected, it has uniform upper and lower bounds for the sectional curvature, and a uniform lower bound for the injectivity radius $InjRad(x)$ (see Section 2.1.3).

Probably the correct terminology should be "uniformly locally bounded geometry", but we prefer shortness to an accurate description.

A connected Riemannian manifold without boundary, so that the isometry group of M acts cocompactly on M (see section 3.1.1), has bounded geometry.

REMARK 2.57. In the literature the condition of bounded geometry on a Riemannian manifold is usually weaker, e.g. that 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 $\mathbb{R}^n$ (**[Gro93]**, §0.5.A_3) or that the Ricci curvature has a uniform lower bound (**[Cha06]**, **[Cha01]**).

For the purposes of this book, the restricted condition in Definition 2.56 suffices.

In what follows we keep the notation $V_\kappa(r)$ from Theorem 2.24 to designate the volume of a ball of radius r in the n–dimensional space of constant sectional curvature κ.

LEMMA 2.58. *Let M be complete n–dimensional Riemannian manifold with bounded geometry, let $a \leqslant b$ and $\rho > 0$ be such that the sectional curvature is at least a and at most b, and that at every point the injectivity radius is larger than ρ.*

(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 2.24, 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 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 $\bigcap_{x\in F} B(x, \delta)$ 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

$$V_a\left(\frac{3\delta}{2}\right) \geqslant Vol\left[B_M\left(y, \frac{3\delta}{2}\right)\right] \geqslant \text{card}\, F\, V_b\left(\frac{\delta}{2}\right)\,.$$

□

2.2.2. Metric simplicial complexes of bounded geometry. 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:

$$\Delta^k = (\mathbb{R}_+)^{k+1} \cap \{x_0 + \ldots + x_n = 1\}.$$

Thus, for each m-simplex σ^m and its face σ^k, the inclusion $\sigma^k \to \sigma^m$ 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. 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 $[a_i, a_{i+1}]$ so that $\mathfrak{p}|[a_i, a_{i+1}]$ is a piecewise-linear path whose image is contained in a single simplex of X. Lengths of such paths are defined using metric on simplices of X. Then

$$d(x, y) = \inf_{\mathfrak{p}} \operatorname{length}(\mathfrak{p})$$

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 2.59. Verify that the standard metric is complete and that X is a geodesic metric space.

DEFINITION 2.60. A metric simplicial complex X has *bounded geometry* if it is connected and if there exist $L \geqslant 1$ and $N < \infty$ so 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.

EXAMPLES 2.61.

- 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 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 naturally in the context of Riemannian manifolds with bounded geometry. Given a simplicial complex X, we will equip it with the *standard metric*, where each simplex is isometric to a Euclidean simplex with unit edges.

THEOREM 2.62 (See Theorem 1.14, [**Att94**]). *Let M be an n-dimensional Riemannian manifold of bounded geometry with geometric bounds a, b, ϵ. Then M admits a triangulation X of bounded geometry (whose geometric bounds depend only on n, a, b, ϵ) and an L-bi-Lipschitz homeomorphism $f : X \to M$, where $L = L(n, a, b, \epsilon)$.*

Another procedure of approximation of Riemannian manifolds by simplicial complexes will be described in Section 5.3.

CHAPTER 3

Algebraic preliminaries

3.1. Geometry of group actions

3.1.1. Group actions. Let G be a group or a semigroup and E be a set. An *action of G on E on the left* is a map

$$\mu : G \times E \to E, \quad \mu(g,a) = g(a),$$

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 $x \in E$.

REMARK 3.1. If, in addition, G is a group, then the two properties above imply that

$$\mu(g, \mu(g^{-1}, x)) = x$$

for all $g \in G$ and $x \in E$.

An *action of G on E on the right* is a map

$$\mu : E \times G \to E, \quad \mu(a,g) = (a)g,$$

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 $x \in E$.

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 E is always to the left, and it is often denoted $G \curvearrowright E$.

If E is a metric space, an *isometric action* is an action so that $\mu(g,\cdot)$ is an isometry of E for each $g \in G$.

A group action $G \curvearrowright X$ 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,

$$\begin{array}{ccc} X & \xrightarrow{g} & X \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ Y & \xrightarrow{g} & Y \end{array}$$

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.

If G is a topological group and E is a topological space, a *continuous action* of G on E is a continuous map μ satisfying the above *action* axioms.

A topological group 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 3.2. Suppose that X is locally compact and $G \curvearrowright X$ is proper. Show that the quotient X/G is Hausdorff.

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 3.3. If $G \curvearrowright X$ is cocompact then X/G (equipped with the quotient topology) is compact.

The following is a converse to the above exercise:

LEMMA 3.4. *Suppose that X is locally compact and $G \curvearrowright X$ is such that 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 covering of Y. Since Y is compact, this open covering has a finite subcovering

$$\{p(U_{x_i} : i = 1, \ldots, n\}$$

The union

$$C := \bigcup_{i=1}^{n} cl(U_{x_i})$$

is compact in X and projects onto Y. Hence, $G \cdot C = X$. □

In the context of non-proper metric space the concept of cocompact group action is replaced with the one of *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 *quasi-homogeneous* 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.

3.1.2. Lie groups. References for this section are [**Hel01, OV90, War83**].

A *Lie group* is a group G which has 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. Actually, every Lie group G can be made into a real analytic manifold with real analytic group operations. We will assume that G is a real n-dimensional manifold, although one can also consider *complex Lie groups*.

EXAMPLE 3.5. Groups $GL(n, \mathbb{R}), SL(n, \mathbb{R})$, $O(n)$, $O(p, q)$ are (real) Lie groups. Every countable discrete group (a group with discrete topology) is 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 signature (p, q). The most important, for us, case is $O(n, 1) \cong O(1, n)$. The group $PO(n, 1) = O(n, 1)/\pm I$ is the group of isometries of the hyperbolic n-space.

EXERCISE 3.6. 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 structure of a Lie algebra, called the *Lie algebra* $\mathfrak{g}$ of the group G.

EXAMPLE 3.7. 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 3.8. $\mathfrak{u}(n) \oplus i\mathfrak{u}(n) = Mat_n(\mathbb{C})$, is the Lie algebra of the group $GL(n, \mathbb{C})$.

THEOREM 3.9. *For every finite-dimensional real Lie algebra $\mathfrak{g}$ there exists unique, up to isomorphism, simply-connected Lie group G whose Lie algebra is isomorphic to $\mathfrak{g}$.*

Every Lie group G has a left-invariant Riemannian metric. Indeed, pick a positive-definite inner product $\langle \cdot, \cdot \rangle_e$ on $V = T_e G$. For every $g \in G$ we consider the left multiplication $L_g : G \to G, L_g(x) = gx$. Then $L_g : G \to G$ is a smooth map and the action of G on itself *via* left multiplication is simply-transitive. We define the inner product $\langle \cdot, \cdot \rangle_g$ on $T_g G$ as the image of $\langle \cdot, \cdot \rangle_e$ under the derivative $Dg : T_e G \to T_g G$.

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 $V = T_e G$ 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. Therefore we have the homomorphism

$$Ad : G \to GL(V).$$

LEMMA 3.10. *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 kernel of Ad is contained in the center of G. □

THEOREM 3.11 (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.*

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, i.e., $\mathfrak{g}$ is nonabelian and contains no ideals.

EXAMPLE 3.12. 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 splits as a direct sum

$$\mathfrak{g} = \oplus_{i=1}^{k}\mathfrak{g}_i,$$

where each $\mathfrak{g}_i$ is a simple Lie algebra.

3.1.3. Haar measure and lattices.

DEFINITION 3.13. 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 3.14. 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 3.15. 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 3.16. *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 3.17. *Every discrete subgroup $\Gamma < G$ admits a* measurable fundamental set*, i.e., a measurable subset of $D \subset G$ such that*

$$\bigcup_{\gamma \in \Gamma} \gamma D = G, \quad \mu(\gamma D \cap D) = 0, \quad \forall \gamma \in \Gamma \setminus 1.$$

PROOF. Since $\Gamma < G$ is discrete, there exists an open neighborhood V of $1 \in G$ such that $\Gamma \cap V = \{1\}$. Let $U \subset V$ be another open neighborhood of $1 \in G$ such that $UU^{-1} \subset V$. Then for $\gamma \in \Gamma$ we have

$$\gamma u = u', u \in U, u' \in U \Rightarrow \gamma = u'u^{-1} \in U \Rightarrow \gamma = 1.$$

In other words, Γ-images of U are pairwise disjoint. Since G is a second countable, there exists a countable subset

$$E = \{g_i \in G : i \in \mathbb{N}\}$$

so that

$$G = \bigcup_i U g_i.$$

Clearly, each set

$$W_n := U g_n \setminus \bigcup_{i<n} \Gamma U g_i$$

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 $x \in U g_n$. Therefore,

$$G = \bigcup_{n=1}^{\infty} \left(U g_n \setminus \bigcup_{i<n} U g_i \right).$$

Next,

$$\Gamma \cdot D = \bigcup_{n=1}^{\infty} \left(\Gamma U g_n \setminus \bigcup_{i<n} \Gamma U g_i \right) =$$

$$\Gamma \cdot \bigcup_{n=1}^{\infty} \left(U g_n \setminus \bigcup_{i<n} U g_i \right) \supset \bigcup_{n=1}^{\infty} \left(U g_n \setminus \bigcup_{i<n} U g_i \right) = G.$$

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

$$U g_n \cap \gamma U g_n \neq \emptyset \Rightarrow U \cap \gamma U \neq \emptyset \Rightarrow \gamma = 1.$$

Thus, for all $\gamma \in \Gamma \setminus \{1\}$, $\gamma D \cap D = \emptyset$. $\square$

Let $D \subset G$ be a measurable fundamental set for a lattice $\Gamma < G$. Then

$$0 < \mu(D) = \mu(\Gamma \backslash G) < \infty$$

since Γ is a lattice. For every $g \in G$, Dg is again a fundamental set for Γ and, thus, $\mu(D) = \mu(Dg)$. Hence, $\mu(D) = \mu(Dg) = c\nu(D)$. It follows that $c = 1$. Thus, μ is also a right Haar measure. $\square$

3.1.4. 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 $G \subset \mathrm{Isom}(X)$ is called *discrete* if it is discrete with respect to the subset topology.

Exercise 3.18. Suppose that X is proper. Show that the following are equivalent for a subgroup $G \subset \mathrm{Isom}(X)$:
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 3.19. 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 3.20. *Suppose that a group G acts geometrically on a proper metric space X. Then $G \backslash X$ has a metric defined by*

$$\mathrm{dist}(\bar{a}, \bar{b}) = \inf\{\mathrm{dist}(p,q)\ ;\ p \in Ga\,,\ q \in Gb\} = \inf\{\mathrm{dist}(a,q)\ ;\ q \in Gb\}\,, \tag{3.1}$$

where $\bar{a} = Ga$ and $\bar{b} = Gb$.

Moreover, this metric induces the quotient topology of $G \backslash X$.

PROOF. The infimum in (3.1) is attained, i.e. there exists $g \in G$ such that

$$\operatorname{dist}(\bar{a}, \bar{b}) = \operatorname{dist}(a, gb).$$

Indeed, take $g_0 \in G$ arbitrary, and let $R = \operatorname{dist}(a, gb)$. Then

$$\operatorname{dist}(\bar{a}, \bar{b}) = \inf\{\operatorname{dist}(a, q) \; ; \; q \in Gb \cap \overline{B}(a, R)\}.$$

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. □

3.2. Complexes and group actions

3.2.1. Simplicial complexes. As we expect the reader to be familiar with basics of algebraic topology, we will discuss simplicial complexes and (in the next section) cell complexes only very briefly.

We will use the notation $X^{(i)}$ to denote the i-th skeleton of the simplicial complex 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 $\mathfrak{p}(t) = F(x, t), x \in X$.

Let X be a simplicial complex. Recall that besides 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.

3.2.2. Cell complexes. A cell complex (or 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

$$U_n := \sqcup_{j\in J} D_j^n,$$

a (possibly empty) disjoint union of closed n-balls D_j^n. Suppose that for each D_j^n we have a continuous *attaching map* $e_j : \partial D_j^n \to X^{(n-1)}$. This defines a map $e = e^n : \partial U_n \to X^{(n-1)}$ and an equivalence relation $x \equiv y = e(x)$, $x \in U, y \in X^{(n-1)}$. We then declare $X^{(n)}$ to be the quotient space of $X^{(n-1)} \sqcup U_n$ with the quotient topology with respect to the above equivalence relation. We will use the notation D_j^n/e_j the image of D^n in X^n, i.e., the quotient $D_j^n/\equiv$. We then equip

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

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 the image of each D^n in X is closed). By abuse of terminology, both the balls D_j^n and their projections to X are called *n-cells* in X. Similarly, we will conflate X and its underlying topological space.

EXERCISE 3.21. A subset $K \subset X$ is compact if and only if is closed and contained in a finite union of cells.

Regular and almost regular cell complexes. A cell complex X is said to be *regular* if every attaching map e_j is 1-1. For instance, every simplicial complex is a regular cell complex. A regular cell complex is called *triangular* if every cell is naturally isomorphic to a simplex. (Note that X itself need not be simplicial since intersections of cells could be unions of simplices.) A cell complex X is *almost regular* if the boundary S^{n-1} of every cell D_j^n is given structure of a regular cell complex K_j so that the attaching map e_j is 1-1 on every cell in S^{n-1}. Almost regular 2-dimensional cell complexes (with a single vertex) appear naturally in the context of group presentations, see Definition 4.79.

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 $X' = Y$ where every cell is a simplex. 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 $e_j : \partial D_j^n \to X^{(n-1)}$. We take the preimage of the regular cell complex structure of Y_{n-1} under e_j to be a *refinement* L_j of the regular cell complex structure K_j on S^{n-1}. We then define a regular cell complex

M_j on D_j^n by conning off every cell in L_j from the origin $o \in D_j^n$. Then cells in M_j are the cones $Cone_{o_j}(s)$, where s's are cells in L_j.

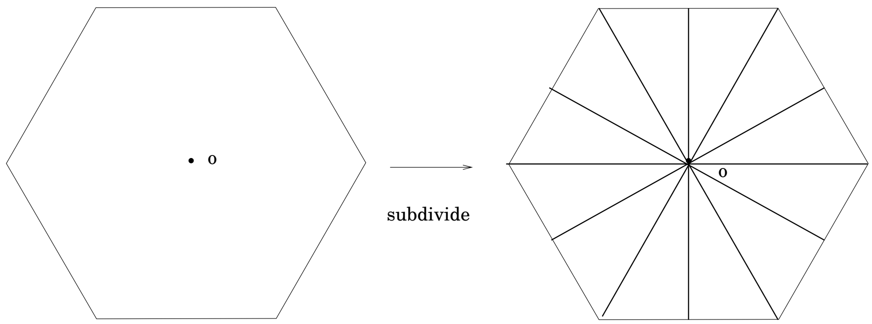


FIGURE 3.1. Barycentric subdivision of a 2-cell.

Since, by the induction assumption, every cell in Y_{n-1} is a simplex, its preimage s in S^{n-1} is also a simplex, this $Cone_o(s)$ is a simplex as well. We then attach each cell D_j^n to Y_n by the original attaching map e_j. 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_j 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.

G-cell complexes. Let X be a cell complex and G be a group. We say that X is a *G-cell complex* (or that we have a *cellular* action $G \curvearrowright X$) if G acts on X by homeomorphisms and for every n we have a G-action $G \curvearrowright U_n$ so that the attaching map e^n is G-equivariant.

DEFINITION 3.22. 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.

An action $G \curvearrowright X$ on a simplicial complex is called *simplicial* if it sends simplices to simplices and is linear on each simplex.

Assuming that X is naturally isomorphic to a simplicial complex and $G \curvearrowright X$ is without inversions, without loss of generality we may assume that $G \curvearrowright X$ is linear on every simplex in X.

The following is immediate from the definition of X'', since barycentric subdivisions are canonical:

LEMMA 3.23. *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 3.24. *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. $\square$

3.2.3. Borel construction. Recall that 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

$$g \times [g_0, \ldots, g_n] \to [gg_0, \ldots, gg_n].$$

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 G is finite then each skeleton $E(G)^i$ is compact.
2. If $G_1 < G_2$ 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 construction of $G \curvearrowright \widehat{X}$ is called the *Borel Construction*. We first explain the construction in the case 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 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 embeddings

$$\widehat{X}_{\sigma_1} = \sigma_1 \times E(G_{\sigma_1}) \supset \sigma_1 \times E(G_{\sigma_2}) \subset \mathrm{Supp}(\sigma_2) \times E(G_{\sigma_2}).$$

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

$$\partial D^n \times E(G_\sigma) \to \hat{X}_{n-1}.$$

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 3.25. *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 $\widehat{X}$.*

3.2.4. 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 3.26. 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.

Note that we allow the complex Y to be infinite-dimensional.

EXERCISE 3.27. A group G is finitely-presented if and only if it has type $\mathbf{F}_2$.

In view of Lemma 3.25, we obtain:

COROLLARY 3.28. *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. $\square$

COROLLARY 3.29. *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. $\square$

3.3. Subgroups

Given two subgroups H, K in a group G we denote by HK the subset

$$\{hk\,;\ h \in H,\, k \in K\} \subset G.$$

Recall that 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 either H or K is a normal subgroup, the set HK defined above 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 3.30. 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 3.31. A *subnormal descending series*indexsubnormal 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 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 3.32. *If G is a group, $N \lhd G$, and $A \lhd B \leqslant G$, then BN/AN is isomorphic to $B/A(B \cap N)$.*

DEFINITION 3.33. 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 3.34. *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 3.32

$$H_{ij}/H_{ij+1} = (K_j \cap H_i)H_{i+1}/(K_{j+1} \cap H_i)H_{i+1} \simeq K_j \cap H_i/(K_{j+1} \cap H_i)(K_j \cap H_{i+1})\,.$$

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 3.35. 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 3.36. 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 3.37. A group G is said to have property * *virtually* if a finite index subgroup H of G has the property *.

The following properties of finite index subgroups will be useful.

LEMMA 3.38. *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. $\square$

Proposition 3.39. *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. $\square$

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.

3.4. Equivalence relations between groups

DEFINITION 3.40. (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 $F_i \lhd H_i$, $i = 1, 2$, 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 3.41. 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 3.42. *All the relations in Definition 3.40 are equivalence relation between groups.*

PROOF. The fact that weak commensurability is an equivalence relation is immediate. It suffices to prove that virtual isomorphism is am equivalence relation. The only non-obvious property is transitivity. We need

LEMMA 3.43. *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

$$F_i \lhd H_i < G_i, |G_i : H_i| < \infty, |F_i| < \infty, \quad i = 1, 2, 3,$$

and

$$F_2' \lhd H_2' < G_2, |G_2 : H_2'| < \infty, |F_2'| < \infty,$$

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 3.43 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 abstract commensurators of G with respect to an 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 3.44. Show that $Comm_G(\Gamma)$ is a subgroup of G.

EXERCISE 3.45. Show that for $G = SL(n, \mathbb{R})$ and $\Gamma = SL(n, \mathbb{Z})$, $Comm_G(\Gamma)$ contains $SL(n, \mathbb{Q})$.

3.5. Commutators, commutator subgroup

DEFINITION 3.46. The *commutator of two elements* h, k in a group G is

$$[h, k] = hkh^{-1}k^{-1}\,.$$

Note that:

- two elements h, k commute (i.e., $hk = kh$) if and only if $[h, k] = 1$.
- $hk = [h, k]kh\,$;

Thus, the commutator $[h, k]$ 'measures de degree of non-commutativity' of the elements h and k. In Lemma **??** 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 3.47. 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 3.48. 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 3.49.

(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 $\psi : G \to G$, $\psi(S) = S$.

(2) follows from the equivalence $xy = yx \Leftrightarrow [x,y] = 1$, and (3) is an immediate consequence of (2).

(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 $S^1 \subset \mathbb{C}$ 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 3.50. Find the commutator subgroup and the abelianization for the finite dihedral group D_{2n} and for the infinite dihedral group D_∞.

Exercise 3.51. 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 4.7.

Note that it is not necessarily true that the commutator subgroup G' of G consists entirely of commutators $\{[x, y] : x, y \in G\}$ (see [**Vav**] for some finite group examples). 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**] for further details. For instance, if G is the free group on two generators (see Definition 4.16), then every nontrivial element of G' has stable commutator norm greater than 1.

3.6. 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 *spit* as a direct product of its normal subgroups $N_i \lhd G, i = 1, \ldots, k$, 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 $(n_1, \ldots, n_k), n_i \in N_i, i = 1, \ldots, k$ 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 3.52. (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 3.53. (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

$$\dots G_{n-1} \stackrel{\varphi_{n-1}}{\longrightarrow} G_n \stackrel{\varphi_n}{\longrightarrow} G_{n+1} \dots$$

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:

$$\{1\} \longrightarrow N \stackrel{\varphi}{\longrightarrow} G \stackrel{\psi}{\longrightarrow} H \longrightarrow \{1\}\,. \tag{3.2}$$

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 3.54. 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 3.55. (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 $\mathbb{Z}_2 = \{id, (12)\}$.

(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 3.56. (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 a linear map in $O(n)$. 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 3.57. Prove statement (2) of Proposition 3.56. Note that $\mathbb{R}^n$ is identified to the group of translations of the n-dimensional affine space *via* the map $b \mapsto T_b$.

In sections 3.10 and 3.11 we discuss semidirect products and short exact sequences in more detail.

3.7. 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 3.58. 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 semi-direct product

$$\left(\bigoplus_{h\in H} G\right) \rtimes_\varphi H\,.$$

DEFINITION 3.59. The semidirect product $(\bigoplus_{h\in H} G) \rtimes_\varphi H$ 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*.

3.8. Group 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.

Let G be a group and let M, N be left G-modules; then $Hom_G(M, N)$ denotes the subspace of G-invariants in the G-module $Hom(M, N)$, where G acts on homomorphisms $u : M \to N$ by the formula:

$$(gu)(m) = g \cdot u(g^{-1}m).$$

If C_* is a chain complex and A is a G-module, then $Hom_G(C_*, A)$ is the chain complex formed by subspaces $Hom_G(C_k, A)$ 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 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 tuple. Then $C_* = C_*(G)$ is the simplicial chain complex of the simplicial complex defining the Milnor's classifying space EG of the group G (see Section 3.2.3). The dual cochain complex C^* is defined by:

$$C^k = Hom(C_k, A), \quad \delta_k(f)((g_0, \ldots, g_{k+1})) = f(\partial_{k+1}(g_0, \ldots, g_{k+1})), f \in C^k.$$

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 $(C^*(G, A))^G = Hom_G(C_*, A)$. If A is a nontrivial G-module then the $Hom_G(C_*, A)$ is still isomorphic to a certain natural cochain complex based on the simplicial complex $C_*(BG)$ (cochain complex with twisted coefficients, or coefficients in a certain sheaf), but the definition is more involved and we will omit it.

DEFINITION 3.60. 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,

$$H^*(G, A) = \mathrm{Ker}(\delta_k)/\operatorname{Im}(\delta_{k-1}), \quad H^i(G, A) = Z^i(G, A)/B^i(G, A).$$

In particular, if A is a trivial G-module, then $H^*(G, A) = H^*(BG, A)$.

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

$$(1, g_1, g_1 g_2, \ldots, g_1 \cdots g_k)$$

(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,

$$\delta_{k-1} f(1, g_1, g_1 g_2, \ldots, g_1 \cdots g_k) = \delta_{k-1} f([g_1|g_2|\ldots|g_k])$$

we get

$$\delta_{k-1} f(1, g_1, g_1 g_2, \ldots, g_1 \cdots g_k) =$$

$$f(g_1, \ldots, g_1 \cdots g_k) - f(1, g_1 g_2, \ldots, g_1 \cdots g_k) + f(1, g_1, g_1 g_2 g_3, \ldots, g_1 \cdots g_k) - \ldots =$$

$$g_1 \cdot f(1, g_2, \ldots, g_2 \cdots g_k) - f([g_1 g_2|g_3|\ldots|g_k]) + f([g_1|g_2 g_3|g_4|\ldots|g_k]) - \ldots =$$

$$g_1 \cdot f([g_2|\ldots|g_k]) - f([g_1 g_2|g_3|\ldots|g_k]) + f([g_1|g_2 g_3|g_4|\ldots|g_k]) - \ldots$$

Thus,

$$\delta_{k-1} f([g_1|g_2|\ldots|g_k]) = g_1 \cdot f([g_2|\ldots|g_k]) - f([g_1 g_2|g_3|\ldots|g_k]) +$$

$$f([g_1|g_2 g_3|g_4|\ldots|g_k]) - \ldots$$

Then, we let $\bar{C}^k$ $(k \geq 1)$ denote the abelian group of functions f sending k-tuples $[g_1|\ldots|g_k]$ 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) $\delta_1(f)([g_1, g_2]) = g_1 \cdot f([g_2]) - f([g_1 g_2]) + f([g_1]).$
(3) $\delta_2(f)([g_1|g_2|g_3]) = g_1 \cdot f([g_2|g_3]) - f([g_1 g_2|g_3]) + f([g_1|g_2 g_3]) - f([g_1|g_2]).$

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) $Z^1(G, A) = \{f : G \to A | f(g_1 g_2) = f(g_1) + g_1 \cdot f(g_2)\}.$
(3) $B^2(G, A) = \{h : G \times G \to A | \exists f : G \to A, h(g_1, g_2) = f(g_1) - f(g_1 g_2) + g_1 \cdot f(g_2)\}.$
(4) $Z^2(G, A) = \{f : G \times G \to A | g_1 \cdot f(g_2, g_3) - f(g_1, g_2) = f(g_1 g_2, g_3) - f(g_1, g_2 g_3)\}.$

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

$$f(g_1 g_2) = f(g_1) \cdot g_2 + g_1 \cdot f(g_2).$$

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*. 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 3.61. A function $d : G \to N$ is called a *derivation* if

$$d(g_1 g_2) = d(g_1) \star g_1 \cdot d(g_2), \quad \forall g_1, g_2 \in G.$$

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 3.62. Verify that every principal derivation is indeed a derivation.

EXERCISE 3.63. 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 **??**. In section (§3.10) we will discuss derivations in the context of semidirect products, while in §3.11 we explain how 2nd cohomology group $H^2(G,A)$ can be used to describe central co-extensions.

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, 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 3.64. 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. 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.*

Group rings. The *(integer) group ring* $\mathbb{Z}G$ of a group G is the set of formal sums $\sum_{g \in G} m_g\, g$, where m_g are integers which are equal to zero for all but finitely many values of g. Then $\mathbb{Z}G$ is a ring when 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 group ring is an integral domain. Both $\mathbb{Z}$ and G embed as subsets of $\mathbb{Z}G$ by identifying every $m \in \mathbb{Z}$ with $m1_G$ and every $g \in G$ with $1g$. Every homomorphism between groups $\varphi : G \to H$ 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 $o : \mathbb{Z}G \to \mathbb{Z}$, 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
$$(\ell_1 \star \ell_2)\cdot m = \ell_1\cdot(\ell_2\cdot m), \tag{3.3}$$
where $\star$ denotes the multiplication operation in L.

Similarly, the ring M is the *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
$$m\cdot(\ell_1 \star \ell_2) = (m\cdot \ell_1)\cdot \ell_2. \tag{3.4}$$

Lastly, M is an *L-bimodule* if M has structure of both left and right L-module.

DEFINITION 3.65. Let M be an L-bimodule. A *derivation* (with respect to this bimodule structure) is a map $d : L \to M$ so 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 $L := \mathbb{Z}G, M := \mathbb{Z}H$, 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:
$$(\sum_i a_i g_i)\cdot(\sum_j b_i h_i) := \sum_{i,j} a_i b_i g_i\cdot h_i, \quad a_i \in \mathbb{Z}, b_j \in \mathbb{Z},$$
where $g\cdot h = \varphi(g)(h)$ for $g \in G, h \in H$. We define the structure of right L-module on M by:
$$(m, \ell) \mapsto m o(\ell) = o(\ell) m, \quad o(\ell) \in \mathbb{Z}$$
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 3.66. 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 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, theory of derivations was developed in [**Fox53**].

LEMMA 3.67. *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 $X \cup X^{-1}$. We then 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. □

NOTATION 3.68. To each generator $x_i \in X$ we associate a derivation ∂_i, called *Fox derivative*, defined by $\partial_i x_j = \delta_{ij} \in \mathbb{Z} \subset \mathbb{Z}G$. In particular,
$$\partial_i(x_i^{-1}) = -x_i^{-1}.$$

PROPOSITION 3.69. *Suppose that $G = F_r$ is free group of rank $r < \infty$. Then every derivation $d \in Der(\mathbb{Z}G, \mathbb{Z}G)$ can be written as a sum*
$$d = \sum_{i=1}^{r} k_i \partial_i\,, \quad \textit{where } k_i = d(x_i) \in \mathbb{Z}\,.$$
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 3.66 (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 $\partial_i x_j = \delta_{ij}$. □

3.10. Derivations and split extensions

Components of homomorphisms to semidirect products.

DEFINITION 3.70. 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 $F : L \to G$ is 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 3.71. Show that d is indeed a derivation.

EXERCISE 3.72. 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 co-extensions.

DEFINITION 3.73. 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 *co-extension of* H *by* N.

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:

$$\begin{tikzcd}
1 \arrow[r] & N_1 \arrow[r] \arrow[d, "f_1"'] & G_1 \arrow[r] \arrow[d, "f_2"'] & H_1 \arrow[r] \arrow[d, "f_3"'] & 1 \\
1 \arrow[r] & N_2 \arrow[r] & G_2 \arrow[r] & H_2 \arrow[r] & 1
\end{tikzcd}$$

We now use the notion of isomorphism of exact sequences to reinterpret the notion of split extension.

PROPOSITION 3.74. *Consider a short exact sequence*

$$1 \to N \xrightarrow{\iota} G \xrightarrow{\pi} Q \to 1\,. \tag{3.5}$$

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 $N \rtimes Q$;*

(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 $gh^{-1} \in \operatorname{Ker} \pi = \operatorname{Im} \iota$.

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 $H \cap \operatorname{Im} \iota = \{1\}$, whence π restricted to H is injective. □

Remark 3.75. Every sequence with free nonabelian group Q splits: Construct a section $\sigma : Q \to G$ by sending each free generator x_i of Q to an element $\tilde{x}_i \in G$ so that $\pi(\tilde{x}_i) = x_i$. In particular, every group which admits an epimorphism to a free nonabelian group F, also contains a subgroup isomorphic to F.

Examples 3.76.

(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 (see Definition 4.16) and let F_n' be its commutator subgroup (see Definition 3.47). Note that the abelianization of F_n as defined in Proposition 3.49, (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

$$1 \to A \xrightarrow{\iota} G \xrightarrow{\pi} Q \to 1, \tag{3.6}$$

where A is an abelian group. Recall that the set of derivations $Der(Q, A)$ has natural structure of an abelian group.

Remarks 3.77.

(1) The short exact sequence (3.6) uniquely defines an action of Q in 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 (3.6) 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 (3.6) 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 (3.6), 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 3.70). 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 (3.6) 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 (3.6).

Exercise 3.78. 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. We say that two splittings σ_1, σ_2 are *A-conjugate* if they differ by a principal shearing automorphism:

$$\sigma_2(q) = a\sigma_1(q)a^{-1}, \forall q \in Q,$$

where $a \in A$. 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 3.79. *A-conjugacy classes of splittings of the short exact sequence* (3.6) *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 $F : L \to G = A \rtimes Q$ 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 3.80. *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. □

3.11. Central co-extensions and 2-nd cohomology

We restrict ourselves to the case of central co-extensions (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 $k+1$-connected cell complex.

Let G be a group and A an abelian group. A *central co-extension* 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: A function f gives rise to a central co-extension 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 §3.8. We will refer to f simply as a *cocycle*.

Two central co-extensions are said to be equivalent if there exist an isomorphism τ making the following diagram commutative:

$$\begin{tikzcd}
1 \arrow[r] & A \arrow[r] \arrow[d, "id"'] & \tilde{G}_1 \arrow[r] \arrow[d, "\tau"'] & G \arrow[r] \arrow[d, "id"'] & 1 \\
1 \arrow[r] & A \arrow[r] & \tilde{G}_2 \arrow[r] & G \arrow[r] & 1
\end{tikzcd}$$

For instance, a co-extension is trivial, meaning equivalent to the product $A \times G$, if and only if the central co-extension splits. We will use the notation $\mathbb{E}(G, A)$ to denote the set of equivalence classes of co-extensions. 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, $c \in B^2(G,A)$. 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 co-extensions

$$A \to G_i \xrightarrow{r_i} G$$

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 co-extension 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 co-extensions corresponds to the sum of cocycles and the trivial element is represented by the cocycle $f = 0$.

To summarize:

THEOREM 3.81 (See Chapter IV in [**Bro82b**].). *There exists an isomorphism of abelian groups*

$$H^2(K(G,1),A) \cong H^2(G,A) \to \mathbb{E}(G,A).$$

Co-extensions and group presentations. Below we describe the isomorphism in Theorem 3.81 in terms of generators and relators, which will require familiarity with some of the material in Chapter 4.

Start with a presentation $\langle \mathcal{X} | \mathcal{R} \rangle$ of the group G and let Y^2 denote the corresponding presentation complex (see Definition 4.80). Embed Y^2 in a 3-connected cell complex Y by attaching appropriate 3-cells to Y^2. Then $H^2(Y,A) \cong H^2(G,A)$. Each cohomology class $[\zeta] \in H^2(G,A)$ is realized by a cocycle $\zeta \in Z^2(Y,A)$, which will assigns elements of A to each 2-cell in Y. The 2-cells c_i of Y are indexed by the defining relators $R_i, i \in I$, of G. By abusing the notation, we set $\zeta(R_i) := \zeta(c_i)$, so that $\zeta(R_i^{-1}) = -\zeta(c_i)$. Given such ζ, define the group $\tilde{G} = \tilde{G}_\zeta$ by the presentation

$$\tilde{G} = \left\langle \tilde{\mathcal{X}} = \mathcal{X} \cup A | [a,x] = 1, \forall a \in A, \forall x \in \tilde{\mathcal{X}}; R_i(\zeta(R_i))^{-1} = 1, i \in I \right\rangle.$$

In particular, if w is a word in the alphabet $\mathcal{X}$, which is a product of conjugates of the relators $R_{i_j}^{t_j}, t_j = \pm 1$, then

$$w \cdot \left(\sum_j t_j \zeta(c_{i_j}) \right) = 1 \tag{3.7}$$

in $\tilde{G}$.

Clearly, we have the epimorphism $r : \tilde{G} \to G$ which sends every $a \in A \subset \tilde{\mathcal{X}}$ to $1 \in G$. We need to identify the kernel r. We have a homomorphism $\iota : A \to \tilde{G}$, defined by $a \to a \in A \subset \tilde{\mathcal{X}}, a \in A$. Furthermore, $\iota(A)$ is a central subgroup of $\tilde{G}$, hence, $Ker(r) = \iota(A)$, since the homomorphism r amounts to dividing $\tilde{G}$ by $\tilde{A}$.

We next show that ι is injective. Let $\tilde{Y}$ denote the presentation complex $\tilde{Y}$ for $\tilde{G}$; the homomorphism $r : \tilde{G} \to G$ is induced by the map $F : \tilde{Y} \to Y$ which collapses each loop corresponding to $a \in A$ to the vertex of Y and sends 2-cells corresponding to the relators $[x, a], x \in X$, to the base-point in Y. So far we did not use the assumption that ζ is a cocycle, i.e., that $\zeta(\sigma) = 0$ whenever σ is the boundary o a 3-cycle in Y. Suppose that $\iota(a) = 1 \in \tilde{G}, a \in A$. Then the loop α in $\tilde{Y}$ corresponding to a bounds a 2-disk $\tilde{\sigma}$ in $\tilde{Y}$. The image of this disk under f is a spherical 2-cycle σ in Y since F is constant on α. The spherical cycle σ is null-homologous since Y is 2-connected, $\sigma = \partial\xi$, $\xi \in C^3(Y, A)$. Since ζ is a cocycle, $0 = \zeta(\partial\xi) = \zeta(\sigma)$. Thus, equation (3.7), implies that $a = \zeta(\sigma) = 0$ in A. This means that ι is injective.

Suppose the cocycle ζ is a coboundary, $\zeta = \delta\eta$, where $\eta \in C^1(Y^1, A)$, i.e., η yields a homomorphism $\eta' : G \to A, \eta'(x_k) = a_k$. We then define a map $s : G \to \tilde{G}$ by $s(x_k) = x_k a_k$. Then relations $R_i = \zeta(R_i)$ imply that $s(R_i) = 1$ in $\tilde{G}$, so the co-extension defined by ζ splits and, hence, is trivial.

We, thus, have a map from $H^2(Y, A)$ to the set $\mathbb{E}(G, A)$.

If, $\zeta \in Z^2(Y, A)$ maps to a trivial co-extension $\tilde{G} \to G$ of G by A, this means that we have a section $s : G \to \tilde{G}$. Then, for every generator $x_k \in X$ of the group G, we have $s(x_k) = x_k a_k$, for some $a_k \in A$. Thus, we define a 1-cochain $\eta \in C^1(Y^1, A)$ by $\eta(x_k) = a_k$, where we identify x_k with a 1-cell in Y^1. Then the same arguments as above, run in the reverse, imply that $\zeta = \delta\eta$ and, hence $[\zeta] = 0 \in H^2(Y, A)$.

Example 3.82. Let G be the fundamental group of a genus $p \geqslant 1$ closed oriented surface S. Take the standard presentation of G, so that S is the (aspherical) presentation complex. Let $A = \mathbb{Z}$ and take $[\zeta] \in H^2(G, \mathbb{Z}) \cong H^2(S, \mathbb{Z})$ be the class Poincaré dual to the fundamental class of S. Then for the unique 2-cell c in S corresponding to the relator

$$R = [a_1, b_1] \cdots [a_p, b_p],$$

we have $\zeta(c) = -1 \in \mathbb{Z}$. The corresponding group $\tilde{G}$ has the presentation

$$\langle a_1, b_1, \ldots, a_p, b_p, t | \quad [a_1, b_1] \cdots [a_p, b_p]t, [a_i, t], [b_i, t], i = 1, \ldots, p\rangle .$$

The conclusion, thus, is that a group G with nontrivial 2-nd cohomology group $H^2(G, A)$ admits nontrivial central co-extensions with the kernel A. How does one construct groups with nontrivial $H^2(G, A)$? Suppose that G admits an aspherical presentation complex Y so 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 then shows that if A is an abelian group which admits an epimorphism to $\mathbb{Z}$, then $H^2(G, A) \neq 0$ provided that $\chi(Y) \geqslant 2$ as before.

3.12. 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 3.83. 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 3.84. *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 $\varphi : G \to \Phi$, so 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 $g \notin G_i$. Since G is finitely generated, it contains a normal subgroup of finite index $N_i \lhd G$, so that $N_i \leqslant G_i$. Indeed, we can take

$$N_i := \bigcap_{x \in S} G_i^x$$

where S is a finite generating set of G and G_i^x denotes the subgroup xG_ix^{-1}. Then N_i is invariant under all inner automorphisms of G and, hence, is normal in G. 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. $\square$

EXAMPLE 3.85. The group $G = GL(n, \mathbb{Z})$ is residually finite. Indeed, we take subgroups $G_p \leqslant G$, $G_p = \mathrm{Ker}(\varphi_p)$, $\varphi_p : G \to GL(n, \mathbb{Z}_p))$. If $g \in G$ is a nontrivial element, we consider its nonzero off-diagonal entry $g_{ij} \neq 0$. Then $g_{ij} \neq 0 \bmod p$, whenever $p > |g_{ij}|$. Thus, $\varphi_p(g) \neq 1$ and G is residually finite.

COROLLARY 3.86. *Free group of rank 2 F_2 is residually finite. Every free group of (at most) countable rank is residually finite.*

PROOF. We will see in Example 4.38 that F_2 embeds in $SL(2, \mathbb{Z})$. Furthermore, every free group of (at most) countable rank embeds in F_2. Now, the assertion follows from the above example. $\square$

The simple argument for $GL(n, \mathbb{Z})$ is a model for a proof of a harder theorem:

THEOREM 3.87 (A. I. Mal'cev [**Mal40**]). *Let G be a finitely generated subgroup of $GL(n, R)$, where R is a commutative ring with unity. Then G is residually finite.*

Mal'cev's theorem is complemented by the following result, known as *Selberg Lemma* [**Sel60**]:

THEOREM 3.88 (Selberg Lemma). *Let G be a finitely generated subgroup of $GL(n, F)$, where F is a field of characteristic zero. Then G contains a torsion-free subgroup of finite index.*

We refer the reader to [**Rat94**, §7.5] and [**Nic**] for the proofs. Note that Selberg Lemma fails for fields of positive characteristic, see e.g. [**Nic**].

3.13. Appendix by B. Nica: Proofs of Malcev's Theorem and Selberg Lemma

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*.

REMARK 4.1. In French, the terminology for finitely generated groups is *groupe de type fini*. On the other hand, in English, *group of finite type* is a much stronger requirement than finite generation (typically, this means that the group has type $\mathbf{F}_\infty$).

EXERCISE 4.2. Show that every finitely generated group is countable.

EXAMPLES 4.3.
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 4.4. Prove that the transposition (12) and the cycle $(12\dots n)$ generate the permutation group S_n.

REMARKS 4.5.
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 $\{n_1, .., n_k\}$ for N, and a finite generating set $\{g_1 N, ..g_m N\}$ 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.

REMARK 4.6. 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 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 $g_1, \ldots, g_k$ 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 4.8. 1. Let F be a non-abelian free group (see Definition 4.16). 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 $x_1, \ldots, x_n$; 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 finitely generated (Lemma 4.75 or Theorem 5.29).

Below we describe a finite generating set for the group $GL(n, \mathbb{Z})$. In the proof we use *elementary matrices* $N_{i,j} = I_n + E_{i,j}$ $(i \ne j)$; 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 4.9. *The group $GL(n, \mathbb{Z})$ is generated by*

$$s_1 = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \quad s_2 = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

$$s_3 = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \quad s_4 = \begin{pmatrix} -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

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 $\varphi(12 \ldots n) = s_1$, $\varphi(12) = s_2$. 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 4.4, the permutation $(12 \ldots n)$ 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 $\langle s_1 , \dots , s_4\rangle \subset GL(n, \mathbb{Z})$, 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, $C_1 (N_{1,2}d_2g) < C_1(g)$. By the induction assumption, there exists an element p of $\langle s_1 , \dots , s_4\rangle$ 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 $\langle s_1, \dots, s_4\rangle$ 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 $\langle s_1, \dots s_4\rangle$. This can be done for instance by repeating the argument in Step 2 with multiplications on the right. □

The wreath product (see Definition 3.59) is a useful construction of a finitely generated group from two finitely generated groups:

EXERCISE 4.10. 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 4.11. 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 4.12. 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 $(k_1, \ldots, k_n)$ in $\mathbb{Z}^n$ such that $\gcd(k_1, \ldots, k_n) = 1$.)

DEFINITION 4.13. *A group G is said to have* bounded generation property *(or is* boundedly generated*) if there exists a finite subset $\{t_1, \ldots, t_m\} \subset G$ 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 $k_1, k_2, \ldots, k_m$ are integers.*

Clearly, all finitely generated abelian groups have the bounded generation property, and so are all the finite groups. On the other hand, the nonabelian free f 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 **??**.

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).

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,

$$1, a_2 a_1, a_1 a_2 a_1^{-1}$$

are reduced, while

$$a_2 a_1 a_1^{-1} a_3$$

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.

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

$$u a_i a_i^{-1} v \sim uv, \quad u a_i^{-1} a_i v \sim uv$$

where $u, v \in X^*$.

PROPOSITION 4.14. *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 $w \sim a_1 u$. If $a_1 \neq b_1^{-1}$ then $a_1 u$ is reduced. If $a_1 = b_1^{-1}$ then $a_1 u \sim b_2 \cdots b_k$ 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 $v = w$. $\square$

EXERCISE 4.15. 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 e. The reduced path $\mathfrak{p}^*$ in T corresponding to the reduction w^* of w is the unique geodesic in T connecting e 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 4.14 implies that $X^*/\sim$ can be identified with $F(X)$.

DEFINITION 4.16. 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 $F(X)$.

The set $F(X)$ with the product defined in Definition 4.16 is indeed a group. The inverse of a reduced word

$$w = a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_k}^{\epsilon_k}$$

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 ww^{-1} project to the empty word 1 in F.

REMARK 4.17. A free group of rank at least two is not abelian. Thus *free non-abelian* means free of rank at least two.

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 4.18 (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 $F(X)$ define

$$\Phi(a_1 \cdots a_n) = \varphi(a_1) \cdots \varphi(a_n).$$

Set $\Phi(e) := 1$, 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 $F(X)$, $\Psi(w) = \Psi(a_1) \cdots \Psi(a_n) = \varphi(a_1) \cdots \varphi(a_n) = \Phi(w)$. $\square$

Corollary 4.19. *Every group is the quotient of a free group.*

Proof. Apply Proposition 4.18 to the group G and the set $X = G$. $\square$

Lemma 4.20. *A short exact sequence $1 \to N \to G \xrightarrow{r} F(X) \to 1$ always 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 a monomorphism. $\square$

Corollary 4.21. *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 4.18, the inclusion map $i : S \to G$ extends uniquely to an epimorphism $\pi_S : F(S) \to G$. The elements of $\operatorname{Ker} \pi_S$ 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 $\operatorname{Ker} \pi_S$ is normally generated by R (i.e. $\langle\langle R \rangle\rangle = \operatorname{Ker} \pi_S$) 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 $\langle S | R \rangle$.

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 4.22. (1) $\langle a_1, \ldots, a_n \mid [a_i, a_j], 1 \leqslant i, j \leqslant n\rangle$ is a finite presentation of $\mathbb{Z}^n$;

(2) $\big\langle x, y \mid x^n, y^2, yxyx\big\rangle$ is a presentation of the finite dihedral group D_{2n};

(3) $\big\langle x, y \mid x^2, y^3, [x, y]\big\rangle$ is a presentation of the cyclic group $\mathbb{Z}_6$.

Let $\langle X|R\rangle$ 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 4.23. *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 4.24. The group $\bigoplus_{x\in X} \mathbb{Z}_2$ has presentation

$$\big\langle x \in X | x^2, [x, y], \forall x, y \in X\big\rangle\,.$$

PROPOSITION 4.25 (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, so 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 to $F(X)$, 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 $\operatorname{Ker} \pi_X$ 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

$$\operatorname{proj} : F(X)/N \to F(X)/\operatorname{Ker} \pi_X\,.$$

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 $\varphi : F(S)/\operatorname{Ker} \pi_S \to F(X)/N$.

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 $\operatorname{proj} \circ \varphi : F(S)/\operatorname{Ker} \pi_S \to F(X)/\operatorname{Ker} \pi_X$. 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

$$\operatorname{proj} \circ \varphi\,(q(x)\operatorname{Ker} \pi_S) = \operatorname{proj}(xN) = x\operatorname{Ker} \pi_X.$$

This means that modulo the two isomorphisms mentioned above, the map $\operatorname{proj} \circ \varphi$ 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. $\square$

Proposition 4.25 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 4.26. *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 4.25 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.

Consider $n_j = r_j(S)$, 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 $n_1, \ldots, n_k$. $\square$

PROPOSITION 4.27. *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 $\bar{Y} = \{yN \mid y \in Y\}$ 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 $\left\langle \bar{Y} \mid \rho_1, \ldots, \rho_m \right\rangle$ 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 $\operatorname{Ker}\varphi$, 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 $\operatorname{Ker}\varphi$ is trivial, hence φ is an isomorphism, equivalently that $K = \operatorname{Ker}\pi_S$. This implies that $\operatorname{Ker}\pi_S$ is normally generated by the finite set of relators listed in (4.2) and (4.3). $\square$

We continue with a list of finite presentations of some important groups:

EXAMPLES 4.28. (1) *Surface groups*:

$$G = \langle a_1, b_1, \ldots, a_n, b_n | [a_1, b_1] \cdots [a_n, b_n] \rangle ,$$

is the fundamental group of the closed connected oriented surface of genus n, see e.g. [**Mas91**].

(2) *Right–angled Artin groups (RAAGs).* Let $\mathcal{G}$ be a finite graph with the vertex set $V = \{x_1, \ldots, x_n\}$ 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 4.29. 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

$$A_\Gamma := \left\langle x_i \in V | \underbrace{x_i x_j \cdots}_{m(e) \text{ terms}} = \underbrace{x_j x_i \cdots}_{m(e) \text{ terms}} , \text{ whenever } e = [x_i, x_j] \in E \right\rangle .$$

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 $x \in V(\Gamma)$, subject to Artin relators and, in addition, relators

$$x^{n_x}, \quad x \in V(\Gamma).$$

Note that, in the case $n_x = 2$ for all $x \in V(\Gamma)$, 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 $x \in V(\Gamma)$, subject to the relators

$$x^{n_x}, \quad x \in V(\Gamma);$$

$$(xy)^{m(e)}, e = [x,y] \in E(\Gamma).$$

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:*

$$H_{2n+1}(\mathbb{Z}) := \langle x_1, \ldots, x_n, y_1, \ldots, y_n, z \mid$$

$$[x_i, z] = 1, [y_j, z] = 1, [x_i, x_j] = 1, [y_i, y_j] = 1, [x_i, y_j] = z^{\delta_{ij}}, 1 \leqslant i, j \leqslant n \big\rangle .$$

(8) *Baumslag–Solitar groups:*

$$BS(p,q) = \big\langle a, b | ab^p a^{-1} = b^q \big\rangle .$$

EXERCISE 4.30. Show that $H_{2n+1}(\mathbb{Z})$ is isomorphic to the group appearing in Example **??**, (**??**).

OPEN PROBLEM 4.31. 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?

An important feature of finitely presented groups is provided by the following theorem, see e.g. [**Hat02**]:

THEOREM 4.32. *Every finitely generated group is the fundamental group of a smooth compact manifold of dimension* 4.

Presentations $G = \langle X|R \rangle$ 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.

Algorithmic problems in the combinatorial group theory.

Word Problem. Let $G = \langle X|R\rangle$ 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 $G = \langle X|R\rangle$ 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 $G = \langle X|R\rangle$, 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 $G = \langle X|R\rangle$ 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 Novikov, Boone and 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 $G = \langle X|R\rangle$, 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. 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 8.143). Furthermore, there are examples [**BH05**] of finitely-presented groups with solvable conjugacy problem but unsolvable simultaneous conjugacy problem for every $n \geq 2$.

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 4.33. *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. $\square$

PROPOSITION 4.34. *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 $\langle X|R\rangle$ 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 $g = 1$ in G.

The point of residual finiteness is that, eventually, one of the machines stops and we conclude that g is trivial or not. $\square$

Laws in groups.

DEFINITION 4.35. An *identity* (or *law*) is a non-trivial reduced word $w = w(x_1, \ldots, x_n)$ in n letters $x_1, \ldots, x_n$ and their inverses. A group G is said to *satisfy the identity (law)* $w(x_1, \ldots, x_n) = 1$ if the equality is satisfied in G whenever $x_1, \ldots, x_n$ are replaced by arbitrary elements in G.

EXAMPLES 4.36 (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 (**??**).

(3) Free Burnside groups. The *free Burnside group*
$$B(n, m) = \langle x_1, \ldots, x_n \mid w^n \text{ for every word } w \text{ in } x_1^{\pm 1}, \ldots, x_n^{\pm 1} \rangle.$$
It is known that these groups are infinite for sufficiently large m (see [**Ady79**], [**Ol'91**], [**Iva94**], [**Lys96**], [**DG**] and references therein).

Note that free nonabelian groups (and, hence, groups containing them) do not satisfy any law.

4.4. Ping-pong lemma. Examples of free groups

LEMMA 4.37 (Ping–pong, or Table–tennis, lemma). *Let X be a set, and let $g : X \to X$ and $h : X \to X$ be two bijections. If A, B are two non-empty subsets of X, such that $A \not\subset B$ and*

$$g^n(A) \subset B \text{ for every } n \in \mathbb{Z} \setminus \{0\}\,,$$

$$h^m(B) \subset A \text{ for every } m \in \mathbb{Z} \setminus \{0\}\,,$$

then g, h generate a free subgroup of rank 2 in the group $\mathrm{Bij}(X)$ *with the binary operation given by composition $\circ$.*

PROOF. *Step 1.* Let w be a non-empty reduced word in $\{g, g^{-1}, h, h^{-1}\}$. We want to prove that w is not equal to the identity in $\mathrm{Bij}(X)$. We begin by noting that it is enough to prove this when

$$w = g^{n_1}h^{n_2}g^{n_3}h^{n_4}\dots g^{n_k}, \text{ with } n_j \in \mathbb{Z} \setminus \{0\}\, \forall j \in \{1, \dots, k\}\,. \tag{4.4}$$

Indeed:

- If $w = h^{n_1}g^{n_2}h^{n_3}\dots g^{n_k}h^{n_{k+1}}$, then gwg^{-1} is as in (4.4), and $gwg^{-1} \neq \mathrm{id} \Rightarrow w \neq \mathrm{id}$.
- If $w = g^{n_1}h^{n_2}g^{n_3}h^{n_4}\dots g^{n_k}h^{n_{k+1}}$, then for any $m \neq -n_1$, $g^m w g^{-m}$ is as in (4.4).
- If $w = h^{n_1}g^{n_2}h^{n_3}\dots g^{n_k}$, then for any $m \neq n_k$, $g^m w g^{-m} \neq \mathrm{id}$ is as in (4.4).

Step 2. If w is as in (4.4) then

$$w(A) \subset g^{n_1}h^{n_2}g^{n_3}h^{n_4}\dots g^{n_{k-2}}h^{n_{k-1}}(B) \subset g^{n_1}h^{n_2}g^{n_3}h^{n_4}\dots g^{n_{k-1}}(A) \subset \dots \subset$$

$$g^{n_1}(A) \subset B\,.$$

If $w = \mathrm{id}$, then it would follow that $A \subset B$, a contradiction. $\square$

EXAMPLE 4.38. For any integer $k \geqslant 2$ the matrices

$$g = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \text{ and } h = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$$

generate a free subgroup of $SL(2, \mathbb{Z})$.

1st proof. The group $SL(2, \mathbb{Z})$ acts on the upper half plane $\mathbb{H}^2 = \{z \in \mathbb{C} \mid \Im(z) > 0\}$ by linear fractional transformations $z \mapsto \frac{az+b}{cz+d}$. The matrix g acts as a horizontal translation $z \mapsto z + k$, while

$$h = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Therefore h acts as represented in Figure 4.1, where h sends the interior of the disk bounded by C to the exterior of the disk bounded by C'. We apply Lemma 4.37 to g, h and the subsets A and B represented below, i.e. A is the strip

$$\{z \in \mathbb{H}^2 \ :\ -\frac{k}{2} < \mathrm{Re}\, z < \frac{k}{2}\}$$

and B is the complement of its closure, that is

$$B = \{z \in \mathbb{H}^2 \ :\ \mathrm{Re}\, z < -\frac{k}{2} \text{ or } \mathrm{Re}\, z > \frac{k}{2}\}.$$

Hence $g^n(A) \subset B$ and $h^n(B) \subset A$ for all $n \neq 0$. Therefore, the claim follows from lemma 4.37.

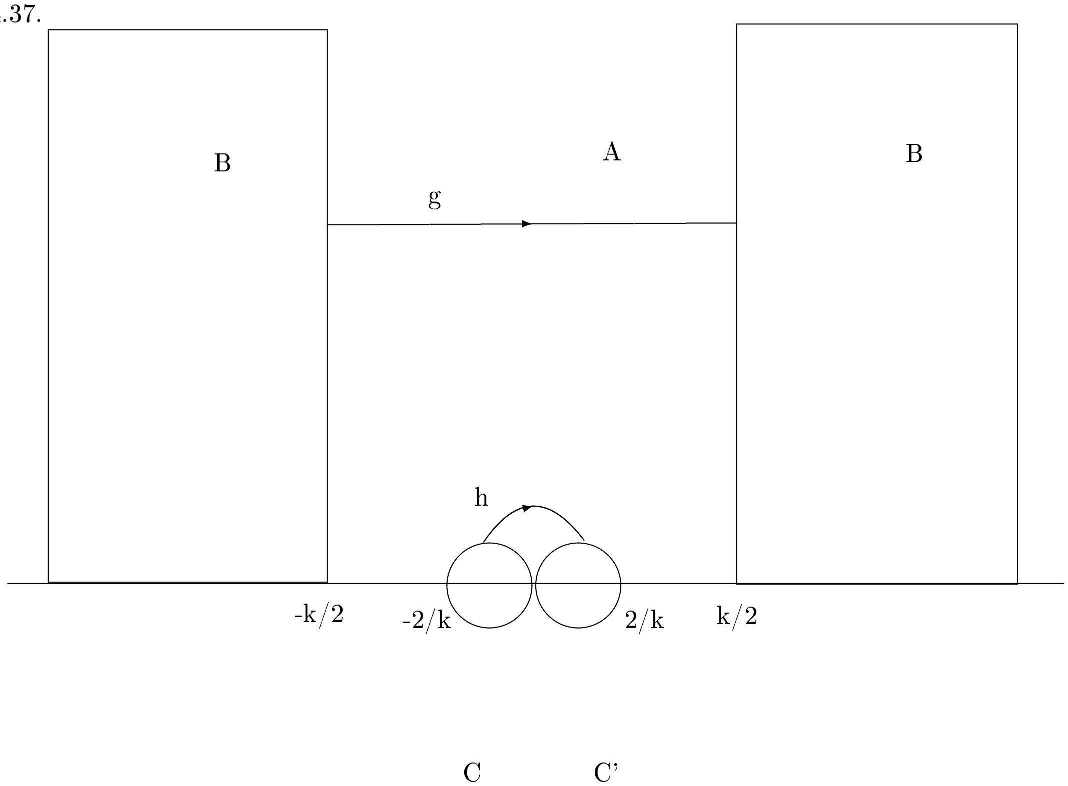


FIGURE 4.1. Example of ping-pong.

2nd proof. The group $SL(2,\mathbb{Z})$ also acts linearly on $\mathbb{R}^2$, and we can apply Lemma 4.37 to g, h and the following subsets of $\mathbb{R}^2$

$$A = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \ : \ |x| < |y| \right\} \text{ and } B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \ : \ |x| > |y| \right\} .$$

□

REMARK 4.39. The statement in the Example above no longer holds for $k = 1$. Indeed, in this case we have

$$g^{-1}hg^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} .$$

Thus, $(g^{-1}hg^{-1})^2 = I_2$, and, hence, the group generated by g, h is not free.

Lemma 4.37 extends to the case of several bijections as follows.

LEMMA 4.40 (The generalized Ping-pong lemma). *Let X be a set, and let $g_i : X \to X$, $i \in \{1, 2, \ldots, k\}$, be bijections. Suppose that $A_1, \ldots, A_k$ are non-empty subsets of X, such that $\bigcup_{i=2}^{k} A_i \not\subset A_1$ and that for every $i \in \{1, 2, \ldots, k\}$*

$$g_i^n \left(\bigcup_{j \neq i} A_j \right) \subset A_i \text{ for every } n \in \mathbb{Z} \setminus \{0\}.$$

Then $g_1, \ldots, g_k$ generate a free subgroup of rank k in the group of bijections $\mathrm{Bij}(X)$.

PROOF. Consider a non-trivial reduced word w in $\{g_1^{\pm 1}, \ldots, g_k^{\pm 1}\}$. As in the proof of Lemma 4.37, without loss of generality we may assume that the word w begins with g_1^a and ends with g_1^b, where $a, b \in \mathbb{Z} \setminus \{0\}$. We apply w to $\bigcup_{i=2}^{k} A_i$, and obtain that the image is contained in A_1. If $w = \mathrm{id}$ in $Bij(X)$, it would that $\bigcup_{i=2}^{k} A_i \subset A_1$, a contradiction. $\square$

4.5. Ping-pong on a projective space

We will frequently use Ping-Pong lemma in the case when X is a projective space. Since this application of the ping-pong argument is the key for the proof of the Tits' Alternative, we explain it here in detail.

Let V be a finite dimensional space over a *normed field* $\mathbb{K}$, which is either $\mathbb{R}, \mathbb{C}$ or has discrete norm and uniformizer π, as in §1.7. We endow the projective space $\mathbb{P}(V)$ with the metric d as in §1.8.

LEMMA 4.41. *Every $g \in GL(n, \mathbb{K})$ induces a bi-Lipschitz transformation of $P(\mathbb{K}^n)$ with Lipschitz constant $\leqslant \frac{|a_1|^2}{|a_n|^2}$, where $a_1, \ldots, a_n$ are the singular values of g and*

$$|a_1| \geqslant \ldots \geqslant |a_n|.$$

PROOF. According to the Cartan decomposition $g = kdk'$ and since all elements in the subgroup K act by isometries on the projective space, it suffices to prove the statement when g is a diagonal matrix A with diagonal entries $a_1, \ldots, a_n$ which are arranged in the order as above. We will do the computation in the case $\mathbb{K} = \mathbb{R}$ and leave the other cases to the reader. Given nonzero vectors $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$, we obtain:

$$|gx \wedge gy| = |\sum_{i<j} a_i a_j x_i x_j e_i \wedge e_j| \leqslant |a_1|^2 |\sum_{i<j} x_i x_j| = |a_1|^2 |x \wedge y|,$$

$$|gx| = |\sum_i a_i^2 x_i^2|^{1/2} \geqslant |a_n||x|, \quad |gy| \geqslant |a_n||y|$$

and, hence,

$$d(g[x], g[y]) \leqslant \frac{|a_1|^2}{|a_n|^2} \frac{|x \wedge y|}{|x| \cdot |y|} = \frac{|a_1|^2}{|a_n|^2} d([x], [y]).$$

$\square$

Let g be an element in $GL(n, \mathbb{K})$ such that with respect to some ordered basis $\{u_1, \ldots, u_n\}$, the matrix of g is diagonal with diagonal entries $\lambda_1, \ldots, \lambda_n$ satisfying

$$|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \geqslant |\lambda_{n-1}| > |\lambda_n| > 0.$$

Let us denote by $A(g)$ and by $H(g)$ the projection to $P(\mathbb{K}^n)$ of the span of $\{u_1\}$, respectively of the span of $\{u_2, \ldots, u_n\}$. Note that then $A(g^{-1})$ and $H(g^{-1})$

are the respective projections to $P(\mathbb{K}^n)$ of the span of $\{u_n\}$, respectively, of the span of $\{u_1, \ldots, u_{n-1}\}$. Obviously, $A(g) \in H(g^{-1})$ and $A(g^{-1}) \in H(g)$.

LEMMA 4.42. *Assume that g and h are two elements in $GL(n, \mathbb{K})$ as above, which are diagonal with respect to bases $\{u_1, \ldots, u_n\}$, $\{v_1, \ldots, v_n\}$ respectively. Assume also that the points $A(g^{\pm 1})$ are not in $H(h) \cup H(h^{-1})$, and $A(h^{\pm 1})$ are not in $H(g) \cup H(g^{-1})$. Then there exists a positive integer N such that g^N and h^N generate a free non-abelian subgroup of $GL(n, \mathbb{K})$.*

PROOF. We first claim that for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that for every $m \geqslant N$, $g^{\pm m}$ maps the complement of the ε-neighborhood of $H(g^{\pm 1})$ inside the ball of radius ε and center $A(g^{\pm 1})$.

According to Lemma 4.41, it suffices to prove the statement when $\{u_1, \ldots, u_n\}$ is the standard basis $\{e_1, \ldots, e_n\}$ of V (since we can conjugate g to a matrix diagonal with respect to the standard basis). In particular, $A(g^{\pm 1})$ is either $[e_1]$ or $[e_n]$. In the former case we take $f(x) = x \cdot e_1$, in the latter case, take $f(x) = x \cdot e_n$, so that $Ker(f) = H = H(g^{\pm 1})$. Then, for a unit vector $v = (x_1, \ldots, x_n) \in V$, according to Exercise 1.80, $\mathrm{dist}([v], [H]) = |f(v)|$. To simplify the notation, we will assume that $f(x) = x \cdot e_1$, since the other case is obtained by relabeling. Then,

$$[v] \notin \mathcal{N}_\varepsilon(H(g^{\pm 1})) \iff |x_1| \geqslant \varepsilon.$$

We have

$$|g^m v \wedge e_1| = |\sum_{i>1} \lambda_i^m x_i e_i \wedge e_1| \leqslant \sqrt{n} |\lambda_2|^m |v| = \sqrt{n} |\lambda_2|^m$$

while

$$|g^m v| \geqslant |\lambda_1|^m |x_1|,$$

which implies that

$$d(g^m[v], [e_1]) = \frac{|g^m v \wedge e_1|}{|g^m v|} \leqslant \frac{\sqrt{n}}{|x_1|} \frac{|\lambda_2|^m}{|\lambda_1|^m} \leqslant \frac{\sqrt{n}}{\varepsilon} \left(\frac{|\lambda_2|}{|\lambda_1|} \right)^m$$

The latter quantity converges to zero as $m \to \infty$, since $|\lambda_1| > |\lambda_2|$. Thus, for all large m, $d(g^m[v], [e_1]) < \varepsilon$. The same claim holds for $h^{\pm 1}$.

Now consider $\varepsilon > 0$ such that for every $\alpha \in \{g, g^{-1}\}$ and $be \in \{h, h^{-1}\}$ the points $A(\alpha)$ and $A(\beta^{\pm 1})$ are at distance at least 2ε from $H(\alpha)$. Let N be large enough so that $g^{\pm N}$ maps the complement of the ε-neighborhood of $H(g^{\pm 1})$ inside the ball of radius ε and center $A(g^{\pm 1})$, and $h^{\pm N}$ maps the complement of the ε-neighborhood of $H(h^{\pm 1})$ inside the ball of radius ε and center $A(h^{\pm 1})$.

Let $A := B(A(g), \varepsilon) \sqcup B(A(g^{-1}), \varepsilon)$ and $B := B(A(h), \varepsilon) \sqcup B(A(h^{-1}), \varepsilon)$. Clearly, $g^{kN}(A) \subseteq B$ and $h^{kN}(B) \subseteq A$ for every $k \in \mathbb{Z}$. Hence by Lemma 4.37, g^N and h^N generate a free group. $\square$

4.6. The rank of a free group determines the group. Subgroups

PROPOSITION 4.43. *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 4.18. 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 4.44. *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 4.24. 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 4.23): 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. $\square$

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. $\square$

REMARK 4.45. Proposition 4.43 implies that for every cardinal number n there exists, up to isomorphism, exactly one free group of rank n. We denote it by F_n.

THEOREM 4.46 (Nielsen–Schreier). *Any subgroup of a free group is a free group.*

This theorem will be proven in Corollary 4.70 using topological methods; see also [**LS77**, Proposition 2.11].

PROPOSITION 4.47. *The free group of rank two contains an isomorphic copy of F_k for every finite k and $k = \aleph_0$.*

PROOF. 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 Section 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 4.40 imply that $\{L_{x_k}\, ;\, k \in \mathbb{N}\}$ generate a free subgroup in $\mathrm{Bij}(F_2)$, hence so do $\{x_k\, ;\, k \in \mathbb{N}\}$ in F_2. $\square$

4.7. Free constructions: Amalgams of groups and graphs of groups

4.7.1. 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 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:

$$G_1 * G_2 = \langle G_1, G_2 | \quad \rangle$$

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 $\mathbb{Z} * \mathbb{Z}$.

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*

$$G_1 *_{H_1 \cong H_2} G_2 = \left\langle G_1, G_2 | \phi(h) h^{-1}, h \in H_1 \right\rangle .$$

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

$$G_1 *_H G_2$$

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 subgroups H_1, H_2 and an isomorphism $\phi : H_1 \to H_2$. Then the HNN extension of G *via* ϕ is defined as

$$G \star_{H_1 \cong H_2} = \left\langle G, t | t h t^{-1} = \phi(h), \forall h \in H_1 \right\rangle .$$

A common shorthand for the HNN extension is

$$G \star_H$$

where $H \cong H_1 \cong H_2$ (the two embeddings of H into G are suppressed in this notation).

EXERCISE 4.48. Suppose that H_1 and H_2 are both trivial subgroups. Then

$$G \star_{H_1 \cong H_2} \cong G * \mathbb{Z}.$$

4.7.2. 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 it has the *initial and the terminal* (possibly equal) vertices e_- and e_+. Suppose that for each edge e we are given *monomorphisms*

$$\phi_{e_+} : G_e \to G_{e_+}, \phi_{e_-} : G_e \to G_{e_-} .$$

REMARK 4.49. More generally, one can allow non-injective homomorphisms

$$G_e \to G_{e_+}, G_e \to G_{e_-},$$

but we will not consider them here.

The graph Γ together with the collection of vertex and edge groups and the monomorphisms $\phi_{e_\pm}$ is called a *graph of groups* $\mathcal{G}$.

Definition 4.50. The *fundamental group* $\pi(\mathcal{G}) = \pi_1(\mathcal{G})$ of the above graph of groups is a group G satisfying the following:

1. There is a collection of *compatible homomorphisms* $G_v \to G, G_e \to G, v \in V(\Gamma), e \in E(\Gamma)$, so that whenever $v = e_\pm$, we have the commutative diagram

$$\begin{tikzcd} & G_v \arrow[dr] & \\ G_e \arrow[ur] \arrow[rr] & & G \end{tikzcd}$$

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 $G \to H$ so that we have commutative diagrams

$$\begin{tikzcd} & G \arrow[dr] & \\ G_v \arrow[ur] \arrow[rr] & & H \end{tikzcd}$$

for all $v \in V(\Gamma)$.

Note that the above definition easily implies that $\pi(\mathcal{G})$ is unique (up to an isomorphism). For the existence of $\pi(\mathcal{G})$ see **[Ser80]** and discussion below. Whenever $G \cong \pi(\mathcal{G})$, we will say that $\mathcal{G}$ determines a *graph of groups decomposition* of G. The decomposition of G is called *trivial* if there is a vertex v so that the natural homomorphism $G_v \to G$ is onto.

Example 4.51. 1. Suppose that the graph Γ is a single edge $e = [1, 2]$, $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_2$. Then

$$\pi(\mathcal{G}) \cong G_1 \star_{H_1 \cong H_2} G_2.$$

2. Suppose that the graph Γ is a single loop $e = [1, 1]$, $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_1$. 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 $\mathcal{G}$, $\pi_1(\mathcal{G})$ exists by describing this group in terms of generators and relators in the manner similar to the definition of the amalgamated free product and HNN extension. In the next section we will see how to construct $\pi_1(\mathcal{G})$ using topology.

4.7.3. Converting graphs of groups to amalgams. Suppose that $\mathcal{G}$ is a graph of groups and $G = \pi_1(\mathcal{G})$. Our goal is to convert $\mathcal{G}$ in 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

$$G_i := \pi_1(\mathcal{G}_i), i = 1, 2, \quad G_3 := G_e.$$

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 $e = [v_1, v_2]$ so e does not separate Γ. Let $\Gamma_1 := \Gamma'$, where Γ' is obtained from Γ by removing the edge e as in Case 1. Set $G_1 := \pi_1(\mathcal{G}_1)$ as before. Then 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.

4.7.4. 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 e.g. [**Hat02**]. 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 *initial* or *final vertex*. Then assume that the maps $f_{e_\pm}$ are homeomorphisms onto the respective boundary components.

For each edge e 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 4.52. *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*. Given the above interpretation, one can easily see that for each vertex $v \in V(\Gamma)$ the canonical homomorphism $G_v \to \pi(\mathcal{G})$ is injective.

EXAMPLE 4.53. The group $F(X)$ is isomorphic to $\pi_1(\vee_{x\in X}\mathbb{S}^1)$.

4.7.5. Graphs of groups and group actions on trees. An *action* of a group G on a tree T is an action $G \curvearrowright T$ so 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 *trivial* if there is a vertex $v \in T$ fixed by the entire group G.

REMARK 4.54. Later on, we will encounter more complicated (non-simplicial) trees and actions.

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 above. 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 4.55. Suppose that Γ is a single segment $[1,2]$, 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 4.2.

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. Therefore we get the induced action $G \curvearrowright T$. If $g \in G$ preserves some $\tilde{M}_e \times (0,1)$, then it 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 4.56. Let $G = BS(p,q)$ be the Baumslag-Solitar group described in Example 4.28, (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 p and q respectively. In order to construct the cell complex $K(G,1)$ take the circle $S^1 = M_v$, the cylinder $S^1 \times [0,1]$ and attach the ends to this cylinder to M_v by the maps of the degree p and q respectively. Now, consider the associated G–tree T. Its vertices have valence

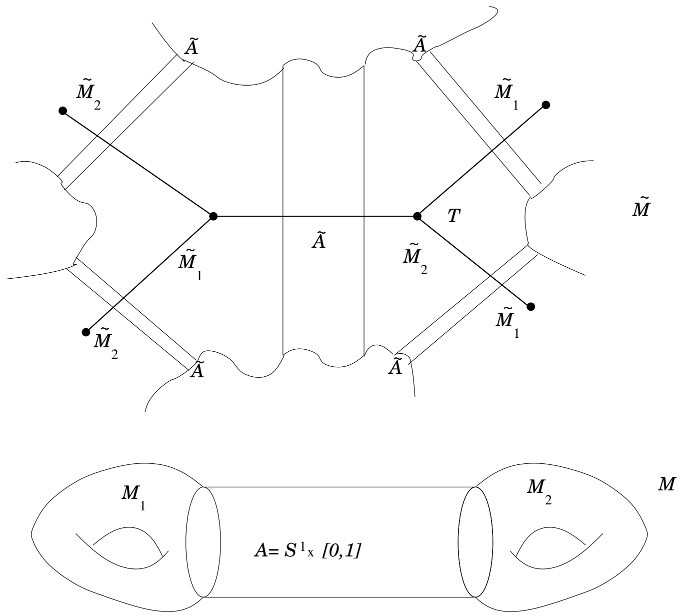


FIGURE 4.2. Universal cover of the genus 2 surface.

$p+q$: Each vertex v has q incoming and p 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 incoming vertices is *via* the group $\mathbb{Z}/q$ and on the outgoing vertices *via* the group $\mathbb{Z}/p$.

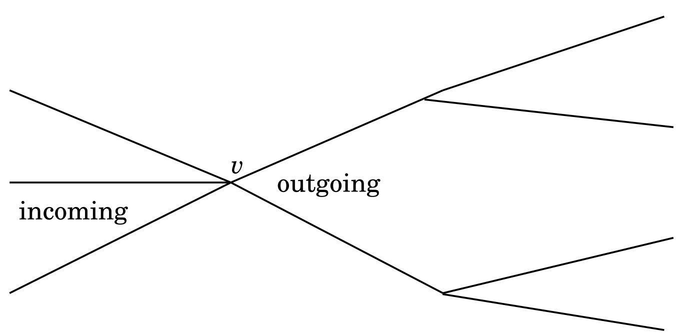


FIGURE 4.3. Tree for the group $BS(2,3)$.

LEMMA 4.57. *$G \curvearrowright T$ is trivial 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 $\tilde{v}$. $\square$

Conversely, each action of G on a simplicial tree T yields a realization of G as the fundamental group of a graph of groups $\mathcal{G}$, so that $T = T(\mathcal{G})$. Here is the construction of $\mathcal{G}$. Furthermore, a *nontrivial* action leads to a *nontrivial* graph of groups.

If the action $G \curvearrowright T$ has inversion, we replace T with its barycentric subdivision T'. Then the action $G \curvearrowright T'$ is without inversions. If $G \curvearrowright T$ were nontrivial, 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 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 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:

$$\begin{tikzcd} G_{\tilde{e}} \arrow[r] \arrow[d] & G_{\tilde{v}} \arrow[d] \\ G_{\tilde{e}'} \arrow[r] & G_{\tilde{v}'} \end{tikzcd}$$

We then 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 $\tilde{e} = [\tilde{v}, \tilde{w}]$. Then We define the embedding $G_e \to G_v$ to make the diagram

$$\begin{tikzcd} G_{\tilde{e}} \arrow[r] \arrow[d] & G_{\tilde{v}} \arrow[d] \\ G_e \arrow[r] & G_v \end{tikzcd}$$

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 just the reverse 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 trivial.

DEFINITION 4.58. $\mathcal{G} \to (G \curvearrowright T) \to \mathcal{G}$ 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.

4.8. Cayley graphs

Finitely generated groups may be turned into geometric object 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 symmetric directed graph $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$ 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 *Cayley graph of G with respect to S*.

By abusing notation, we will also use the notation $[g, h] = \overline{gh}$ for the edge $\{g, h\}$.

Since S is a generating set of G, it follows 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 $\mathrm{Cayley}(G, S)$ 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 4.59. 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 4.60. In this book, unless stated otherwise, all Cayley graphs are for finite generating sets S.

Much of the discussion in this section though remains valid for arbitrary generating sets, including infinite ones.

REMARK 4.61. 1. Every group acts on itself by 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 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 4.62. 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 4.63. Consider $\mathbb{Z}^2$ with set of generators

$$S=\{a=(1,0),b=(0,1),a^{-1}=(-1,0),b^{-1}=(0,-1)\}.$$

The Cayley graph $\mathrm{Cayley}(G,S)$ 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

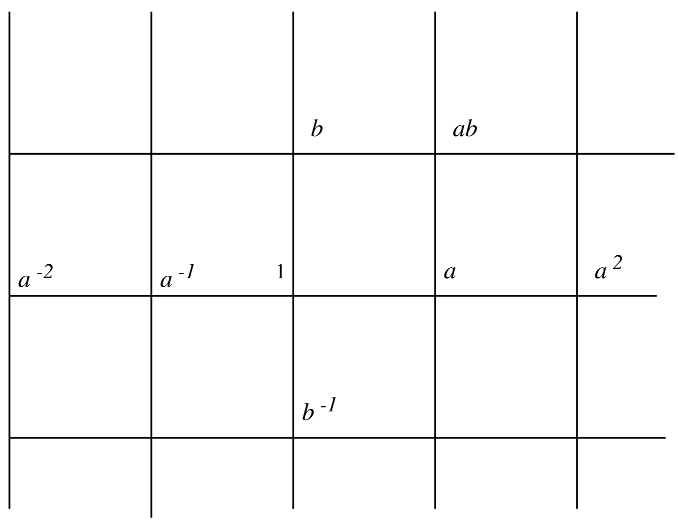


FIGURE 4.4. Cayley graph of $\mathbb{Z}^2$.

The Cayley graph of $\mathbb{Z}^2$ with respect to the set of generators $\{\pm(1,0),\pm(1,1)\}$ has the same set of vertices as the above, but the vertical lines must be replaced by diagonal lines.

EXAMPLE 4.64. Let G be the free group on two generators a,b. Take $S=\{a,b,a^{-1},b^{-1}\}$. The Cayley graph $\mathrm{Cayley}(G,S)$ is the 4-valent tree (there are four edges incident to each vertex).

See Figure 4.5.

THEOREM 4.65. *Fundamental group of every connected graph Γ is free.*

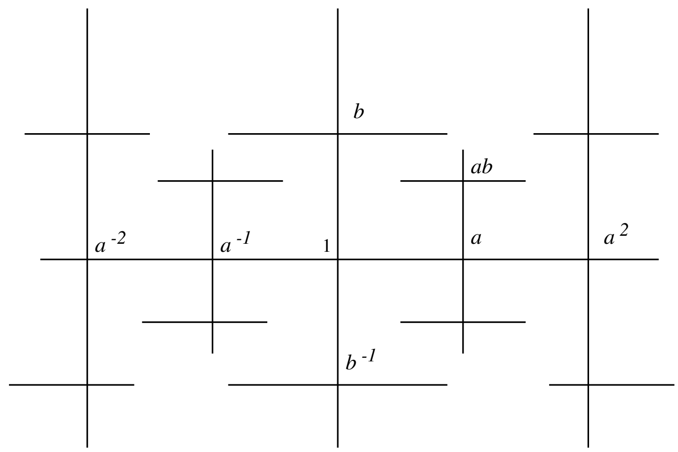


FIGURE 4.5. Free group.

PROOF. By axiom of choice, 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 that $G = \pi_1(\Gamma)$ is free, with the free generators indexed by the set $\mathcal{E}$. $\square$

COROLLARY 4.66. *A connected graph is simply connected if and only if the graph is a tree.*

COROLLARY 4.67. *1. Every free group $F(X)$ is the fundamental group of the bouquet 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 4.65, $G = \pi_1(B)$ 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 $F(X)$. $\square$

COROLLARY 4.68. *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, Theorem 4.65, $G = \pi_1(\Gamma)$ is free. See [**Ser80**] for another proof and more general discussion of group actions on trees. $\square$

Remark 4.69. The concept of a simplicial tree generalizes to the one of a real tree. 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. [**Kap01**] for a proof.

Corollary 4.70 (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. $\square$

Exercise 4.71. Let G and H be finitely generated groups, with S and X respective finite generating sets.

Consider the wreath product $G \wr H$ as defined in Definition 3.59, endowed with the finite generating set canonically associated to S and X described in Exercise 4.10. For every function $f : H \to G$ denote by $\operatorname{supp} f$ 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

$$\operatorname{dist}((f,h),(g,k)) = \sum_{x \in H} \operatorname{dist}_S(f(x), g(x)) + \operatorname{Length}(\operatorname{supp} g^{-1}f\,; h, k)\,, \tag{4.5}$$

where $\operatorname{Length}(\operatorname{supp} g^{-1}f\,; h, k)$ is the length of the shortest path in $\operatorname{Cayley}(H, X)$ starting in h, ending in k and whose image contains the set $\operatorname{supp} g^{-1}f$.

Thus we succeeded in assigning to every finitely generated group G a metric space $\operatorname{Cayley}(G, S)$. The problem, however, is that this assignment $G \to \operatorname{Cayley}(G, S)$ is far from canonical: different generating sets could yield completely different Cayley graphs. For instance, the trivial group has the presentations:

$$\langle \ \mid \ \rangle\,, \quad \langle a | a \rangle\,, \quad \left\langle a, b | ab, ab^2 \right\rangle, \ldots,$$

which give rise to the non-isometric Cayley graphs:

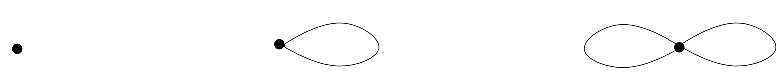


Figure 4.6. *C*ayley graphs of the trivial group.

The same applies to the infinite cyclic group:

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

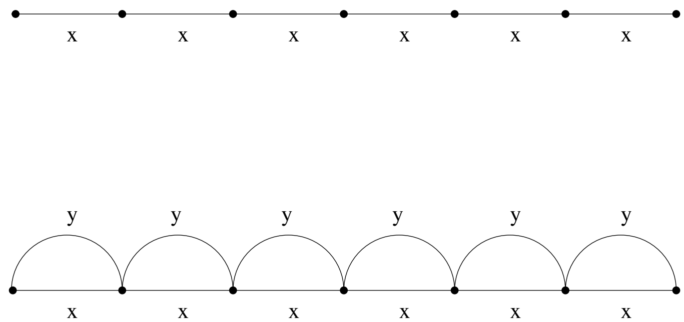


FIGURE 4.7. Cayley graphs of $\mathbb{Z} = \langle x | \rangle$ and $\mathbb{Z} = \left\langle x, y | xy^{-1} \right\rangle$.

finite group looks like a "fuzzy dot"; every Cayley graph of $\mathbb{Z}$ looks like a "fuzzy line," etc. Therefore, although non-isometric, they "look alike".

EXERCISE 4.72. (1) Prove that if S and $\bar{S}$ are two finite generating sets of G then the word metrics dist_S and $\mathrm{dist}_{\bar{S}}$ on G are bi-Lipschitz equivalent, i.e. there exists $L > 0$ such that

$$\frac{1}{L}\mathrm{dist}_S(g, g') \leqslant \mathrm{dist}_{\bar{S}}(g, g') \leqslant L\mathrm{dist}_S(g, g')\,, \forall g, g' \in G\,. \tag{4.6}$$

(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 4.73. 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 4.74. 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 your 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, the latter grows exponentially fast as R goes to infinity.

The goal of the rest of the book is to make sense of this "fuzzy math".

In Section 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 4.75. *A* finite index *subgroup of a finitely generated group is finitely generated.*

PROOF. It follows from Theorem 5.29. 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 $G = H \sqcup \bigsqcup_{i=1}^{k} Hg_i$ 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 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 $|h|_X \leqslant |h|_S = n$. □

4.9. 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 section 2.1.4 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 Euclidean simplicial complexes, i.e., simplicial complexes where each k-simplex is equipped with the metric of the Euclidean simplex where every edge has unit length. In order to compute n-volume of a map f, first compute volumes of restrictions $f|\Delta_i$, for all n-dimensional simplices and then add up the results.

4.9.1. Simplicial and combinatorial volumes of maps. Suppose that X, Y are simplicial complexes equipped with *standard metrics* and $f : X \to Y$ is a simplicial map, i.e., a map which sends every simplex to simplex so that the restriction is linear. Then the n-dimensional simplicial volume $sVol_n(f)$ of f is just the number of n-dimensional simplices in the domain X. Note that this, somewhat strange, concept, is independent of the map f but is, nevertheless, useful. The more natural concept is the one of the *combinatorial volume* of the map f, namely,

$$cVol_n(f) = \sum_{\Delta} \frac{1}{c_n} Vol(f(\Delta))$$

where the sum is taken over all n-simplices in X and c_n is the volume of the Euclidean simplex with unit edges. In other words, $cVol_n$ counts the number of n-simplices in X which are not mapped by f to simplices of lower dimension.

Both definitions extend in the context of cellular maps of cell-complexes.

DEFINITION 4.76. Let X, Y be n-dimensional almost regular cell complexes. A cellular map $f : X \to Y$ is said to be *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.

For instance, simplicial map of simplicial complexes is regular.

We define the *combinatorial* n-volume $cVol_n(f)$ of f to be the total number of n-cells in X which are not collapsed by f. The combinatorial 2-volume is called *area*. Thus, this definition agrees with the notion of combinatorial volume for simplicial maps.

Geometric volumes of maps. Similarly, suppose that X, Y are regular n-dimensional cell complexes. We define smooth structure on each open n-cell in X and Y by using the identification of these cells with the open n-dimensional Euclidean balls of unit volume, coming from the regular cell complex structure on X and Y.

We say that a cellular map $f : X \to Y$ is *smooth* if for every $y \in Y$ which belongs to an open n-cell, f is smooth at every $x \in f^{-1}(y)$. At points $x \in f^{-1}(y)$ for such y we have a continuous function $|J_f(x)|$. We declare $|J_f(x)|$ to be zero at all points $x \in X$ which map to $Y^{(n-1)}$. Then we again define the *geometric volume* $Vol(f)$ by the formula (2.2) where the integral is taken over all open n-cells in X. We extend this definition to the case where f is not smooth over some open m-cells by setting $Vol(f) = \infty$ in this case. In the case when $n = 2$, $Vol(f)$ is called the *area* of f and denoted $Area(f)$.

We now assume that X is an n-dimensional finite regular cell complex and $Z \subset X$ is a subcomplex of dimension $n-1$. The example we will be primarily interested in is when X is the 2-disk and Z is its boundary circle.

LEMMA 4.77 (Regular cellular approximation). *After replacing X with its subdivision if necessary, every cellular map $f : X \to Y$ is homotopic, rel. Z, to a smooth regular map $h : X \to Y$ so that*

$$Vol(h) = cVol_n(h) \leqslant cVol_n(f)$$

i.e., the geometric volume equals the combinatorial volume for the map h.

PROOF. First, without loss of generality, we may assume that f is smooth. For each open n-cell σ° in Y we consider components U of $f^{-1}(\sigma^\circ)$. If for some U and $p \in \sigma^\circ$, $f(U) \subset \sigma^\circ \setminus p$, then we compose $f|cl(U)$ with the retraction of σ to its boundary from the point p. The resulting map f_1 is clearly cellular, homotopic to f rel. Z and its n-volume is at most the n-volume of f (for both geometric and combinatorial volumes). Moreover, for every component U of $f_1^{-1}(\sigma^\circ)$, $f_1(U) = \sigma^\circ$. We let $m(f_1, \sigma)$ denote the number of components of $f^{-1}(\sigma^\circ)$.

Our next goal is to replace f_1 with a new (cellular) map f_2 so that f_2 is 1-1 on each U as above. By Sard's theorem, for every n-cell σ in Y there exists a point $p = p_\sigma \in \sigma^\circ$ which is a regular value of f_1. Let $V = V_\sigma \subset \sigma^\circ$ be a small closed ball whose interior contains p and so that f_1 is a covering map over V. Let $\rho_\sigma : \sigma \to \sigma$ denote the retraction of σ to its boundary which sends V diffeomorphically to σ° and which maps $\sigma \setminus V$ to the boundary of σ. Let $\rho : Y \to Y$ be the map whose restriction to each closed n-cell σ is ρ_σ and whose restriction to $Y^{(n-1)}$ is the identity map. Then we replace f_1 with the composition $f_2 := \rho \circ f_1$. It is clear that the new map f_2 is cellular and is homotopic to f_1 rel. Z. Moreover, f_2 is a trivial covering over each open n-cell in Y. By construction, we have:

$$Vol_n(f_2) = \sum_\sigma m(f_1, \sigma) Vol_n(\sigma) = \sum_\sigma m(f_2, \sigma) Vol_n(\sigma) \leqslant Vol_n(f), \tag{4.7}$$

where the sum is taken over all n-cells σ in Y. Furthermore, for each n-cell σ, $f_2^{-1}(\sigma^\circ)$ is a disjoint union of open n-balls, each of which is contained in an open n-cell in X. Moreover, the restriction of f_2 to the boundary of each of these balls factors as the composition

$$e_\sigma \circ g$$

where g is a homeomorphism to the Euclidean ball B^n and $e_\sigma : \partial B^n \to Y^{(n-1)}$ is the attaching map of the cell σ. We then subdivide the cell complex X so that the closure of each $f_2^{-1}(\sigma^\circ)$ is a cell. Then $h := f_2$ is the required regular map. The required equality (and inequality) of volumes is an immediate corollary of the equation (4.7). $\square$

4.9.2. Topological interpretation of finite-presentability.

LEMMA 4.78. *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

$$\langle X|R\rangle = \langle x_1, \ldots, x_n | r_1, \ldots, r_m\rangle .$$

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_X 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 $\alpha_i : S^1 \to Y^1$ as the attaching maps. Let Y be the resulting cell complex. It is clear from the construction that Y is *almost regular*.

We obtain a homomorphism $\phi : F_X \to \pi_1(Y^1) \to \pi_1(Y)$. 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 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 is now again follows from Seifert-Van Kampen theorem that X is a presentation complex for a finite presentation of G: The generators x_i are the loops γ_i and the relators are the 2-cells (or, rather, their attaching maps $S^1 \to X^1$). $\square$

DEFINITION 4.79. The complex Y constructed in this proof is called the *presentation complex* of G associated with the presentation $\langle X|R\rangle$.

DEFINITION 4.80. The 2-dimensional complex Y constructed in the first part of the above proof is called the *presentation complex* of the presentation

$$\langle x_1, \ldots, x_n | r_1, \ldots, r_m\rangle .$$

4.9.3. Van Kampen diagrams and Dehn function. Van Kampen diagrams of relators. Suppose that $\langle X|R\rangle$ is a (finite) presentation of a group G and Y be the corresponding presentation complex. Suppose that $w \in \langle\langle R\rangle\rangle < F_X$ is a relator in this presentation. Then w corresponds to a null-homotopic loop λ_w in the 1-skeleton $Y^{(1)}$ of Y. Let $f : D^2 \to Y$ be an extension of $\lambda_w : S^1 \to Y$. By the cellular approximation theorem (see e.g. [**Hat02**]), after subdivision of D^2 as a regular cell complex, we can assume that f is cellular. Note, however, that some edges in this cell complex structure on D^2 will be mapped to vertices and some 2-cells will be mapped to 1-skeleton. A *Van Kampen diagram* if an convenient (and traditional) way to keep track of these dimension reductions.

DEFINITION 4.81. We say that a contractible finite planar regular cell complex K is a *tree-graded disk* (a *tree of discs* or a *discoid*) provided that every edge of K is contained in the boundary of K. In other words, K is obtained from a finite simplicial tree by replacing some vertices with 2-cells, which is why we think of K as a "tree of discs".

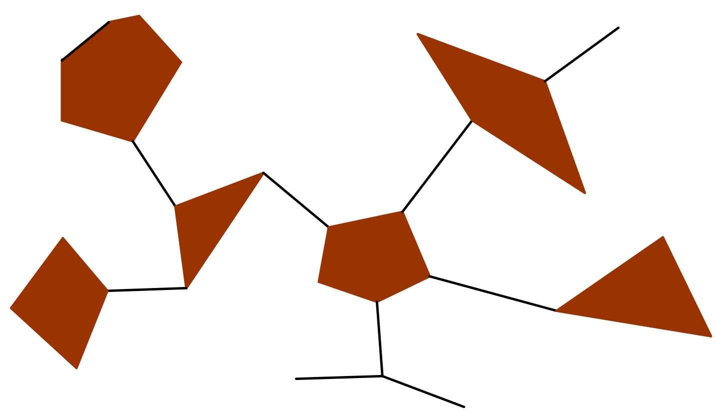


FIGURE 4.8. Example of tree-graded disk.

LEMMA 4.82. *For every w as above, there exists a tree-graded disk K, a regular cell complex structure $\tilde{K}$ on D^2, a regular cellular map $f : \tilde{K} \to Y$ extending λ_w and cellular maps $h : K \to Y, \kappa : \tilde{K} \to K$ so that: $f = h \circ \kappa$.*

PROOF. Write w as a product

$$w = v_1 \cdots v_k, \quad v_i = u_i r_i u_i^{-1}, i = 1, \ldots, k,$$

where each $r_i \in R$ is a defining relator. Then the circle S^1 admits a regular cell complex structure so that λ_w sends each vertex to the unique vertex $v \in Y$ and for every edge α_i, the based loop $f|\alpha_i$ represents the word $v_i \in F_X$. Moreover, the arcs α_i are cyclically ordered on S^1 in order of appearance of v_i in w. Furthermore, each α_i is subdivided in 3 arcs $\alpha_i^+, \beta_i, \alpha_i^-$ so that the loop $f|\alpha_i^\pm$ represents $u_i^{\pm 1}$ and $f|\beta_i$ represents r_i. We then construct a collection of pairwise disjoint arcs $\tau_i \subset D^2$ which intersect S^1 only at their end-points: For each pair α_i^+, α_i^- we connect the end-points of α_i^+ to that of α_i^- by arcs $\epsilon_i^{\pm}$. The result is a cell-complex structure $\tilde{K}$ on D^2 where every vertex is in S^1. There three types of 2-cells in $\tilde{K}$:

1 Cells A_i bounded by bigons $\gamma_i \cup \epsilon_i^-$,
2 Cells B_i bounded by rectangles $\alpha_i^+ \cup \epsilon^+ \cup \alpha_i^- \cup \epsilon^-$,
3 The rest, not containing any edges in S^1.

We now collapse each 2-cell of type (3) to a point, collapse each 2-cell of type (2) to an edge e_i (so that $\alpha_i^{\pm}$ map homeomorphically onto this edge while $\epsilon_i^{\pm}$ map to the end-points of e_i). Note that $\alpha_i^{\pm}$ with their orientation inherited from S^1 define two opposite orientations on e_i.

The result is a tree-graded disk K and a *collapsing map* $\kappa : \tilde{K} \to K$. We define a map $h : K^1 \to Y$ so that $h \circ \kappa|\alpha_i^{\pm} = \lambda_{v_i^{\pm 1}}$ while $h \circ \kappa|\beta_i = \lambda_{r_i}$. Lastly, we extend h to the 2-cells $C_i := \kappa(A_i)$ in K: $h : C_i \to Y$ are the 2-cells corresponding to the defining relators r_i. □

Definition 4.83. A map $h : K \to Y$ constructed in the above lemma is called a *Van Kampen diagram* of w in Y.

The combinatorial area $cArea(h)$ of the Van Kampen diagram $h : K \to Y$ is the number of 2-cells in K, i.e., the number k of relators r_i used to describe w as a product of conjugates of defining relators. The (algebraic) *area* of the loop λ_w in Y, denoted $A(w)$, is

$$\min_{h:K\to Y} cArea(h)$$

where the minimum is taken over all Van Kampen diagrams of w in Y. Algebraically, the area $A(w)$ is the least number of defining relators in the representation of w as the product of conjugates of defining relators. This explains the significance of this notion of area: It captures the complexity of the word problem for the presentation $\langle X|R\rangle$ of the group G.

We identify all open 2-cells in Y with open 2-disks of unit area. Our next goal is to convert arbitrary disks that bound L_w to Van Kampen diagrams. Let $f : D^2 \to Y$ be a cellular map extending λ_w, where D^2 is given structure of a regular cell complex W. By Lemma 4.77, we can replace f with a regular cellular map $f_1 : D^2 \to Y$, which is homotopic to f rel. $Z := \partial D^2$, so that $cArea(f_1) = Area(f_1) \leqslant Area(f)$.

We use the orientation induced from D^2 on each 2-cell in W. Pick a base-point $x \in \partial D^2$ which is a vertex of W. Let $\sigma_1, \ldots, \sigma_m$ be the 2-cells in W. For each 2-cell $\sigma = \sigma_i$ of W we let p_σ denote a path in $W^{(1)}$ connecting x to $\partial\sigma$. Then, by attaching the "tail" p_σ to each $\partial\sigma$ (whose orientation is induced from σ) we get an oriented loop τ_σ based at x. By abusing the notation we let τ_σ denote the corresponding elements of $\pi_1(W^{(1)}, x)$. We let $\lambda \in \pi_1(W^{(1)}, x)$ denote the element corresponding to the (oriented) boundary circle of D^2. We leave it to the reader to verify that the group $\pi_1(W^{(1)}, x)$ is freely generated by the elements τ_σ and that λ is the product

$$\prod_\sigma \tau_\sigma$$

(in some order) of the elements τ_σ where each τ_σ appears exactly once. (This can be shown, for instance, by induction on the number of 2-cells in W.) We renumber the 2-cells in W so that the above product has the form

$$\prod_\sigma \tau_\sigma = \tau_{\sigma_1} \dots \tau_{\sigma_m}$$

For each σ_i set $\phi_i := \pi_1(f_1)(\tau_{\sigma_i}) \in \pi_1(Y^{(1)}, y)$, $y = f_1(x)$. Then, the element $\pi_1(f_1)(\lambda) \in \pi_1(Y^{(1)}, y)$ (represented by the loop λ_w) is the product

$$\phi_1 \dots \phi_m \tag{4.8}$$

in the group $\pi_1(Y^{(1)}, y)$. For every 2-cell σ_i of W either σ_i is collapsed by f_1 or not. In the former case, ϕ_i represents a trivial element of the free group $\pi_1(Y^{(1)}, y)$. In the latter case, ϕ_i has the form

$$u_i r_{j(i)} u_i^{-1}$$

where $r_{j(i)} \in R$ is one of the defining relators of the presentation $\langle X|R\rangle$ and the word $u_i \in F_X$ corresponds to the loop $f_1(p_{\sigma_i})$. Therefore, we can eliminate the elements of the second type from the product (4.8) while preserving the identity

$$w = \phi_{i_1} \cdots \phi_{i_k} \in F_X.$$

This product decomposition, as we observed above, corresponds to a Van Kampen diagram $h : K \to Y$. The number k is nothing but the combinatorial area of the map f_1 above. We conclude

PROPOSITION 4.84 (Combinatorial area equals geometric area equals algebraic area).

$$A(w) = \min\{cArea(f) = Area(f) | f : D^2 \to Y\},$$

where the minimum is taken over all regular cellular maps f extending the map $\lambda_w : S^1 \to Y^{(1)}$.

DEFINITION 4.85 (Dehn function). Let G be a group with finite presentation $\langle X|R\rangle$ and the corresponding presentation complex Y. The *Dehn function of G* (with respect to the finite presentation $\langle X|R\rangle$) equals

$$Dehn(n) := \max\{A(w) : |w| \leqslant n\}$$

where w's are elements in X^* representing trivial words in G. Geometrically speaking,

$$Dehn(n) = \max_{\lambda, \ell(\lambda) \leqslant n} \min\{cArea(f) | f : D^2 \to Y, f|\partial D^2 = \lambda\}$$

where λ's are homotopically trivial regular cellular maps of the triangulated circle to Y and f's are regular cellular maps of the triangulated disk D^2 to Y.

CHAPTER 5

Coarse geometry

5.1. Quasi-isometry

We now define an important equivalence relation between metric spaces: the quasi-isometry. The quasi-isometry has two equivalent definitions: one which is easy to visualize and one which makes it easier to understand why it is an equivalence relation. We begin with the first definition, continue with the second and prove their equivalence.

DEFINITION 5.1. Two metric spaces (X, dist_X) and (Y, dist_Y) are *quasi-isometric* if and only if there exist $A \subset X$ and $B \subset Y$, separated nets, such that (A, dist_X) and (B, dist_Y) are bi-Lipschitz equivalent.

EXAMPLES 5.2.
(1) A metric space of finite diameter is quasi-isometric to a point.
(2) The space $\mathbb{R}^n$ endowed with a norm is quasi-isometric to $\mathbb{Z}^n$ with the metric induced by that norm.

Historically, quasi-isometry 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. A particular instance of this is the relationship between hyperbolic spaces and certain hyperbolic groups.

When trying to prove that the quasi-isometry relation is an equivalence relation, reflexivity and symmetry are straightforward, but when attempting to prove transitivity, the following question naturally arises:

QUESTION 5.3 ([**Gro93**], p. 23). Can a space contain two separated nets that are not bi-Lipschitz equivalent?

THEOREM 5.4 ([**BK98**]). *There exists a separated net N in $\mathbb{R}^2$ which is not bi-Lipschitz equivalent to $\mathbb{Z}^2$.*

OPEN QUESTION 5.5 ([**BK02**]). When placing a point in the barycenter of each tile of a Penrose tiling, is the resulting separated net bi-Lipschitz equivalent to $\mathbb{Z}^2$?

A more general version of this question: 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 all $z \in \mathbb{Z}^n$ such that $z + B$ intersects P. The projections of all such z on P compose a separated net. Is such a net bi-Lipschitz equivalent to $\mathbb{Z}^2$?

Fortunately there is a second equivalent way of defining the fact that two metric spaces are quasi-isometric, which is as follows. We begin by loosening up the Lipschitz concept.

Definition 5.6. 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)–quasi-isometric 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 quasi–isometric embedding does not have to be an embedding in the usual sense, however distant points have distinct images.

If X is a finite interval $[a, b]$ then an (L, C)–quasi–isometric embedding $\mathfrak{q} : X \to Y$ is called *a quasi-geodesic (segment)*. If $a = -\infty$ or $b = +\infty$ then $\mathfrak{q}$ is called *quasi-geodesic ray*. If both $a = -\infty$ and $b = +\infty$ then $\mathfrak{q}$ is called *quasi-geodesic line*. By abuse of terminology, the same names are used for the image of $\mathfrak{q}$.

An (L, C)–quasi-isometric embedding is called an (L, C)–**quasi–isometry** if it admits a **quasi–inverse** map $\bar{f} : Y \to X$ which is also an (L, C)–quasi–isometric embedding so that:

$$\mathrm{dist}_X(\bar{f}f(x), x) \leqslant C, \quad \mathrm{dist}_Y(f\bar{f}(y), y) \leqslant C \tag{5.3}$$

for all $x \in X, y \in Y$.

Two metric spaces X, Y are *quasi-isometric* if there exists a quasi-isometry $X \to Y$.

We will abbreviate *quasi-isometry*, *quasi–isometric* and *quasi-isometrically* to QI.

Exercise 5.7. Let $f_i : X \to X$ be maps so that f_3 is (L_3, A_3) coarse Lipschitz and $\mathrm{dist}(f_2, id_X) \leqslant A_2$. Then

$$\mathrm{dist}(f_3 \circ f_1, f_3 \circ f_2, \circ f_1) \leqslant L_3A_2 + A_3.$$

Definition 5.8. A metric space X is called *quasi-geodesic* if there exist constants (L, A) so that every pair of points in X can be connected by an (L, A)–quasi-geodesic.

In most cases the *quasi–isometry constants* L, C do not matter, so we shall use the words *quasi–isometries* and *quasi-isometric embeddings* without specifying constants.

Exercise 5.9.
(1) Prove that the composition of two quasi-isometric embeddings is a quasi-isometric embedding, and that the composition of two quasi-isometries is a quasi-isometry.
(2) Prove that quasi-isometry of metric spaces is an equivalence relation.

Some quasi-isometries $X \to X$ are more interesting than others. The *boring* quasi-isometries are the ones which are within finite distance from the identity:

Definition 5.10. 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

$$\mathrm{dist}(f, id_X) = \sup_{x\in X} \mathrm{dist}(f(x), x) \text{ is finite.}$$

In order to mod out the semigroup of quasi-isometries $X \to X$ 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 of quasi-isometries $X \to X$, where two quasi-isometries f, g are equivalent if and only if $\mathrm{dist}(f, g)$ is finite. In particular, the set of quasi-isometries equivalent to id_X is $\mathcal{B}(X)$. It is easy to see that the composition defines a binary operation on $QI(X)$, that the quasi-inverse defines an inverse in this group, and that $QI(X)$ is a group when endowed with these operations.

DEFINITION 5.11. The group $(QI(X)\,, \circ)$ is called the *group of quasi-isometries of the metric space* X.

There is a natural homomorphism $\mathrm{Isom}(X) \to QI(X)$. In general, this homomorphism is not injective. For instance if $X = \mathbb{R}^n$ then the kernel 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 $QI(G)$. In general, kernel K of $G \to QI(G)$ is a subgroup such that for every $k \in K$ the G-centralizer of k has finite index in G, see Lemma **??**. Thus, every finitely generated subgroup in K is *virtually central*. In particular, if $G = K$ then G is virtually abelian.

QUESTION 5.12. Is the subgroup $K \leqslant G$ always virtually central? Is it at least true that K is always virtually abelian?

The group $VI(G)$ of virtual automorphisms of G defined in Section 3.4 maps naturally to $QI(G)$ since every virtual isomorphism ϕ of G ($\phi : G_1 \overset{\cong}{\to} G_2$, where G_1, G_2 are finite-index subgroups of G) induces a quasi-isometry $f_\phi : G \to G$. Indeed, $\phi : G_1 \to G_2$ is a quasi-isometry. Since both $G_i \subset G$ are nets, ϕ extends to a quasi-isometry $f_\phi : G \to G$.

EXERCISE 5.13. Show that the map $\phi \to f_\phi$ projects to a homomorphism $VI(G) \to QI(G)$.

When G is a finitely generated group, $QI(G)$ is independent of the choice of word metric. More importantly, we will see (Corollary 5.62) that every group quasi-isometric to G admits a natural homomorphism to $QI(G)$.

EXERCISE 5.14. Show that if $f : X \to Y$ is a quasi-isometric embedding such that $f(X)$ is r-dense in Y for some $r < \infty$ then f is a quasi-isometry.

Hint: Construct a quasi-inverse $\bar{f}$ to the map f by mapping a point $y \in Y$ to $x \in X$ such that

$$\mathrm{dist}_Y(f(x), y) \leqslant r.$$

EXAMPLE 5.15. The cylinder $X = \mathbb{S}^n \times \mathbb{R}$ with a product metric is quasi-isometric to $Y = \mathbb{R}$; the quasi-isometry is the projection to the second factor.

EXAMPLE 5.16. 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 5.17. Let $\varphi : [1, \infty) \to \mathbb{R}_+$ be a differentiable function so 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 quasi-isometry $\mathbb{R}^2 \to \mathbb{R}^2$.

EXERCISE 5.18. If $f, g : X \to Y$ are within finite distance from each other, i.e.

$$\sup \operatorname{dist}(f(x), g(x)) < \infty$$

and f is a quasi-isometry, then g is also a quasi-isometry.

PROPOSITION 5.19. *Two metric spaces $(X, \operatorname{dist}_X)$ and $(Y, \operatorname{dist}_Y)$ are quasi-isometric in the sense of Definition 5.1 if and only if there exists a quasi-isometry $f : X \to Y$.*

PROOF. Assume there exists an (L, C)–quasi-isometry $f : X \to Y$. 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 $a, a' \in A$,

$$\operatorname{dist}_Y(f(a), f(a')) \leqslant L\operatorname{dist}_X(a, a') + C \leqslant \left(L + \frac{C}{\delta}\right) \operatorname{dist}_X(a, a')$$

and

$$\operatorname{dist}_Y(f(a), f(a')) \geqslant \frac{1}{L}\operatorname{dist}_X(a, a') - C \geqslant \left(\frac{1}{L} - \frac{C}{\delta}\right) \operatorname{dist}_X(a, a') =$$

$$\frac{1}{L(C+1)}\operatorname{dist}_X(a, a').$$

It follows that f restricted to A and with target B is bi-Lipschitz.

Conversely, assume that $A \subset X$ and $B \subset Y$ are two ε–separated δ–nets, and that there exists a bi-Lipschitz map $g : A \to B$ which is onto. We define a map $f : X \to Y$ as follows: for every $x \in X$ we choose one $a_x \in A$ at distance at most δ from x and define $f(x) = g(a_x)$.

N.B. The axiom of choice makes here yet another important appearance, if we do not count the episodic appearance of Zorn's Lemma, which is equivalent to the axiom of choice. Details on this axiom will be provided later on. Nevertheless, when X is proper (for instance X is a finitely generated group with a word metric) there are finitely many possibilities for a_x, so the axiom of choice need not be assumed, 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 $x, y \in X$,

$$\operatorname{dist}_Y(f(x), f(y)) = \operatorname{dist}_Y(g(a_x), g(a_y)) \leqslant L\operatorname{dist}_X(a_x, a_y) \leqslant L(\operatorname{dist}_X(x, y) + 2\varepsilon).$$

Also

$$\operatorname{dist}_Y(f(x), f(y)) = \operatorname{dist}_Y(g(a_x), g(a_y)) \geqslant \frac{1}{L}\operatorname{dist}_X(a_x, a_y) \geqslant \frac{1}{L}(\operatorname{dist}_X(x, y) - 2\varepsilon).$$

Now the proposition follows from Exercise 5.14. □

Below is yet another variation on the definition of quasi-isometry, 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 5.20. Let X and Y be metric spaces. A subset $R \subset X \times Y$ is called an (L, A)–*quasi-isometric 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 each $x, x' \in \pi_X(R)$
$$\operatorname{dist}_{Haus}(\pi_Y(R(x)), \pi_Y(R(x'))) \leqslant L\operatorname{dist}(x, x') + A.$$
3. Similarly, for each $y, y' \in \pi_Y(R)$
$$\operatorname{dist}_{Haus}(\pi_X(R(y)), \pi_X(R(y'))) \leqslant L\operatorname{dist}(y, y') + A.$$

Observe that for any (L, A)–quasi-isometric relation R, for all pair of points $x, x' \in X$, and $y \in R(x), y' \in R(x')$ we have

$$\frac{1}{L}\operatorname{dist}(x, x') - \frac{A}{L} \leqslant \operatorname{dist}(y, y') \leqslant L\operatorname{dist}(x, x') + A.$$

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 5.19, if R is an (L, A)–quasi-isometric relation between nonempty metric spaces, then it induces an (L_1, A_1)–quasi-isometry $X \to Y$. Conversely, every (L, A)–quasi-isometry is an (L_2, A_2)–quasi-isometric relation.

In some cases, in order to show that a map $f : X \to Y$ is a quasi-isometry, it suffices to check a weaker version of (5.3). 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.

DEFINITION 5.21. 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 $\operatorname{diam}(f^{-1}(B(y, R))) \leqslant \psi(R)$ 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 ψ and η are called *upper* and *lower distortion function*, respectively.

For instance, the following function is L-Lipschitz, proper, but not uniformly proper:

$$f(x) = (|x|, \arctan(x)).$$

EXERCISE 5.22. 1. Composition of uniformly proper maps is again uniformly proper.

2. If $f_1, f_2 : X \to Y$ are such that $\mathrm{dist}(f_1, f_2) < \infty$ and f_1 is uniformly proper, then so is f_2.

LEMMA 5.23. *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 quasi-isometry.*

PROOF. Construct a quasi-inverse to the map f. Given a point $y \in Y$ pick a point $\bar{f}(y) := x \in X$ such that $\mathrm{dist}(f(x), y) \leqslant r$. 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 $\mathrm{dist}(y, y') \leqslant 1$, one has:

$$\mathrm{dist}(\bar{f}(y), \bar{f}(y')) \leqslant A.$$

Pick $t > 2r + 1$ which is in the image of the lower distortion function η. Then take $A \in \eta^{-1}(t)$.

It is also clear that $f, \bar{f}$ are quasi-inverse to each other. □

LEMMA 5.24. *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 and set

$$L = \max_{s \in S}(d(s(o), o).$$

Then for every $g \in G$, $d_S(gs, g) = 1$, while

$$d(gs(o), g(o)) = d(s(o), o) \leqslant L.$$

Therefore, f is L-Lipschitz.

2. Define the function

$$\eta(n) = \min\{d(go, o) : |g| = n\}.$$

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. Thus, $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and proper. By definition of the function η, for every $g \in G$,

$$d(f(g), f(1)) = d(go, o) \geqslant \eta(d(g, 1)).$$

Since G acts on itself and on X isometrically, it follows that

$$d(f(g), f(h)) \geqslant \eta(d(g, h)), \quad \forall g, h \in G.$$

Thus, the map f is uniformly proper. □

Coarse convergence.

DEFINITION 5.25. Suppose that X is a proper metric space. A sequence (f_i) of maps $X \to Y$ is said to *coarsely uniformly converge to a map $f : X \to Y$ on compacts*, if:

There exists a number $R < \infty$ so that for every compact $K \subset X$, there exits i_K so that for all $i > i_K$,

$$\forall x \in K, \quad d(f_i(x), f(x)) \leqslant R.$$

PROPOSITION 5.26 (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, so 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 a (L_2, A_2)–Lipschitz map $f : X \to Y$, so that*

$$\overset{c}{\lim_{k\to\infty}} f_i = f.$$

Furthermore, if the maps f_i are (L_1, A_1) quasi-isometries, then f is also an (L_3, A_3) quasi-isometry.

PROOF. Let $N \subset X$ be a separated net. We can assume that $x_0 \in N$. Then the restrictions $f_i|N$ are L'-Lipschitz maps and, by the usual Arzela-Ascoli theorem, the sequence $(f_i|N)$ subconverges (uniformly on compacts) 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 $f : X \to Y$ is an (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 $x' \in N \cap B(x_0, r + R)$ so 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 argument for quasi-isometries is similar. □

5.2. Group-theoretic examples of quasi-isometries

We begin by noting that given a finitely generated group G endowed with a word metric the space $\mathcal{B}(G)$ 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 4.61).

LEMMA 5.27. *In a finitely generated group (G, dist_S) endowed with a word metric, the set of maps $\mathcal{B}(G)$ is consisting 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 quasi-isometry, which partly justifies the interest in such maps, is given by the following result, proved in the context of Riemannian manifolds first by A. Schwarz [**Šva55**] and, 13 years later, by J. Milnor [**Mil68**]. 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 in the literature it is at times 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 5.29 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 5.28 (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 quasi-isometry and its quasi-inverse.

Theorem 5.29 (Milnor–Schwarz). *Let (X, dist) be a proper geodesic metric space (which is equivalent, by Theorem 1.29, 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 map $G \to X$ given by $g \mapsto gx$ is a quasi-isometry.*

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

$$S = \{s \in G\, ;\, s \neq 1\,,\, s\overline{B} \cap \overline{B} \neq \emptyset\}\,.$$

Note that S is finite because the action of G is proper, and that $S^{-1} = S$ by the definition of S.

Step 2: Outside of the generating set.

Now consider $\inf\{\mathrm{dist}(\overline{B}, g\overline{B})\, ;\, g \in G \setminus (S \cup \{1\})\}$. For some $g \in G \setminus (S \cup \{1\})$ the distance $\mathrm{dist}(\overline{B}, g\overline{B})$ is a positive constant R, by the definition of S. The set H of elements $h \in G$ such that $\mathrm{dist}(\overline{B}, h\overline{B}) \leqslant R$ is contained in the set $\{g \in G\, ;\, g\overline{B}(x, D+R) \cap \overline{B}(x, D+R) \neq \emptyset\}$, hence it is finite. Now $\inf\{\mathrm{dist}(\overline{B}, g\overline{B})\, ;\, g \in G \setminus (S \cup \{1\})\} = \inf\{\mathrm{dist}(\overline{B}, g\overline{B})\, ;\, g \in H \setminus (S \cup \{1\})\}$ and the latter infimum is over finitely many positive numbers, therefore there exists $h_0 \in H \setminus (S \cup \{1\})$ such that $\mathrm{dist}(\overline{B}, h_0\overline{B})$ realizes that infimum, which is therefore positive. Let then $2d$ be this infimum. By definition $\mathrm{dist}(\overline{B}, g\overline{B}) < 2d$ implies that $g \in S \cup \{1\}$.

Step 3: G is finitely generated.

Consider a geodesic $[x, gx]$ and $k = \left\lfloor \frac{\mathrm{dist}(x,gx)}{d} \right\rfloor$. 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 $g = h_{k+1} = s_0 s_1 \cdots s_k$. We have thus proved that G is generated by S, consequently G is finitely generated.

Step 4: The quasi-isometry.

Since all word metrics on G are bi-Lipschitz equivalent it suffices to prove (2) for the word metric dist_S, where S is the finite generating set found as above for the chosen arbitrary point x. The space X is contained in the $2D$–tubular neighborhood of the image Gx of the map defined in (2). It therefore remains to prove that the map is a quasi-isometric embedding. The previous argument proved that $|g|_S \leqslant k + 1 \leqslant \frac{1}{d}\text{dist}(x, gx) + 1$. Now let $|g|_S = m$ and let $w = s_1' \cdots s_m'$ be a word in S such that $w = g$ in G. Then, by the triangle inequality,

$$\text{dist}(x, gx) = \text{dist}(x, s_1' \cdots s_m' x) \leqslant \text{dist}(x, s_1' x) + \text{dist}(s_1' x, s_1' s_2' x) + \ldots +$$

$$+\text{dist}(s_1' \cdots s_{m-1}' x, s_1' \cdots s_m' x) = \sum_{i=1}^{m} \text{dist}(x, s_i' x) \leqslant 2Dm = 2D|g|_S\,.$$

We have, thus, obtained that for any $g \in G$,

$$d\text{dist}_S(1, g) - d \leqslant \text{dist}(x, gx) \leqslant 2\text{dist}_S(1, g)\,.$$

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 argument, $1 \in G$ can be replaced by any element $h \in G$. □

COROLLARY 5.30. *Given M a compact connected Riemannian manifold, let $\widetilde{M}$ be its universal covering 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.

A natural question to ask is whether two infinite finitely generated groups G and H that are quasi-isometric are also bi-Lipschitz equivalent. In fact, this question was asked in [**Gro93**], p. 23. We discuss this question in Chapter **??**.

COROLLARY 5.31. *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 quasi-isometric.*
(2) *Given a finite normal subgroup N in G, the groups G and G/N are quasi-isometric.*

PROOF. (1) is a particular case of Theorem 5.29, with $G_2 = G$ and X a Cayley graph of G.

(2) follows from Theorem 5.29 applied to the action of the group G on a Cayley graph of the group G/N. □

LEMMA 5.32. *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*

$$id : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

is a quasi-isometry.

PROOF. The group G is finitely generated by Theorem 5.29, choose a word metric dist_G on G corresponding to any finite generating set. Pick a point $x_0 \in X$; then the maps

$$f_i : (G, \mathrm{dist}_G) \to (X, \mathrm{dist}_i), \quad f_i(g) = g(x_0)$$

are quasi-isometries, let $\bar{f}_i$ denote their quasi-inverses. Then the map

$$id : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

is within finite distance from the quasi-isometry $f_2 \circ \bar{f}_1$. $\square$

COROLLARY 5.33. *Let* $\mathrm{dist}_1, \mathrm{dist}_2$ *be as in Lemma 5.32. Then any geodesic γ with respect to the metric* dist_1 *is a quasi-geodesic with respect to the metric* dist_2.

LEMMA 5.34. *Let X be a proper geodesic metric space, $G \curvearrowright X$ is a geometric action. 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 quasi-isometry 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 5.24. 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 5.22). Taking $\bar{\phi} : X \to G$, a quasi-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. $\square$

Let $G \curvearrowright X, G \curvearrowright X'$ be isometric actions and let $f : X \to X'$ be a quasi-isometric embedding. We say that f is (quasi) equivariant if for every $g \in G$

$$\mathrm{dist}(g \circ f, f \circ g) \leqslant C,$$

where $C < \infty$ is independent of G.

LEMMA 5.35. *Suppose that X, X' are proper geodesic metric spaces, G, G' are groups acting geometrically on X and X' respectively and $\rho : G \to G'$ is an isomorphism. Then there exists a ρ–equivariant quasi-isometry $f : X \to X'$.*

PROOF. Pick points $x \in X$, $x' \in X'$. According to Theorem 5.29 the maps

$$G \to G \cdot x \hookrightarrow X, \quad G' \to G' \cdot x' \hookrightarrow X'$$

are quasi-isometries; therefore the map

$$f : G \cdot x \to G' \cdot x, \quad f(gx) := \rho(g)x$$

is also a quasi-isometry.

We now define a G–equivariant projection $\pi : X \to X$ such that $\pi(X) = G \cdot x$, and π is at bounded distance from the identity map on X. We start with a closed ball $\overline{B}$ in X such that $G\overline{B} = X$. Using the axiom of choice, pick a subset Δ of $\overline{B}$ intersecting each orbit of G in exactly one point. For every $y \in X$, there exists a unique $g \in G$ such that $gy \in \Delta$. Define $\pi(y) = g^{-1}x$. Clearly $\mathrm{dist}_X(y, \pi(y)) = \mathrm{dist}(gy, x) \leqslant \mathrm{diam}(\overline{B})$.

Then the map $\tilde{f}$ below is a ρ–equivariant quasi-isometry:

$$\tilde{f} : X \to X', \tilde{f} = f \circ \pi,$$

since $\tilde{f}$ is a composition of two equivariant quasi-isometries. □

COROLLARY 5.36. *Two virtually isomorphic (VI) finitely generated groups are quasi-isometric (QI).*

PROOF. Let G be a finitely generated group, $H < G$ a finite index subgroup and $F \lhd H$ a finite normal subgroup. According to Corollary 5.31, G is QI to H/F.

Recall now that two groups G_1, G_2 are virtually isomorphic if there exist finite index subgroups $H_i < G_i$ and finite normal subgroups $F_i \lhd H_i$, $i = 1, 2$, so that $H_1/F_1 \cong H_2/F_2$. Since G_i is QI to H_i/F_i, we conclude that G_1 is QI to G_2. □

The next example shows that VI is not equivalent to QI.

EXAMPLE 5.37. 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 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, so G_A acts isometrically on Sol_3 regarded as a metric space. Hence, 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 $A, B \in Hyp(2, \mathbb{Z})$ so 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 abelian subgroup of G_A along a rank 2 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, $H_B \cong \mathbb{Z}^2 \rtimes_{B^m} \mathbb{Z}$. It is easy to see that an isomorphism $H_A \to H_B$ would have to carry L_A isomorphically to L_B. However, this would imply that A^n is conjugate to B^m. Contradiction.

EXAMPLE 5.38. 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 $\pi_1(S)$ is hyperbolic, the groups G_1 and G_2 are quasi-isometric, see section 8.14.

Another example of quasi-isometry is the following.

EXAMPLE 5.39. All non-abelian free groups of finite rank are quasi-isometric 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 $F_n, 2 \leqslant n < \infty$ groups are commensurable. Indeed, let a, b denote the generators of F_2. Define the epimorphism $\rho_m : F_2 \to \mathbb{Z}_m$ by sending a to 1 and b 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 $\chi(R) = 1 - b_1(R) = 1 - \operatorname{rank}(\pi_1(R))$. 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 $F_n \hookrightarrow F_2$. Then $\chi(R) = m\chi(R_2) = -m$. Hence, $\operatorname{rank}(F) = 1 - \chi(R) = 1 + m$. Thus, for every $n = 1 + m \geqslant 2$, we have a finite-index inclusion $F_n \hookrightarrow F_2$. Since commensurability is a transitive relation which implies quasi-isometry, 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 are quasi-isometric. We denote by $\mathcal{T}_k$ the regular simplicial tree of valence k and we show that $\mathcal{T}_3$ is quasi-isometric to $\mathcal{T}_k$ for every $k \geqslant 4$.

We define a piecewise-linear map $\mathfrak{q} : \mathcal{T}_3 \to \mathcal{T}_k$ as in Figure 5.1: Sending all edges drawn in thin lines isometrically onto edges and collapsing each edge-path of length $k - 3$ (drawn in thick lines) to a single vertex. The map $\mathfrak{q}$ thus defined is surjective and it satisfies the inequality

$$\frac{1}{k-2} \operatorname{dist}(x, y) - 1 \leqslant \operatorname{dist}(\mathfrak{q}(x), \mathfrak{q}(y)) \leqslant \operatorname{dist}(x, y).$$

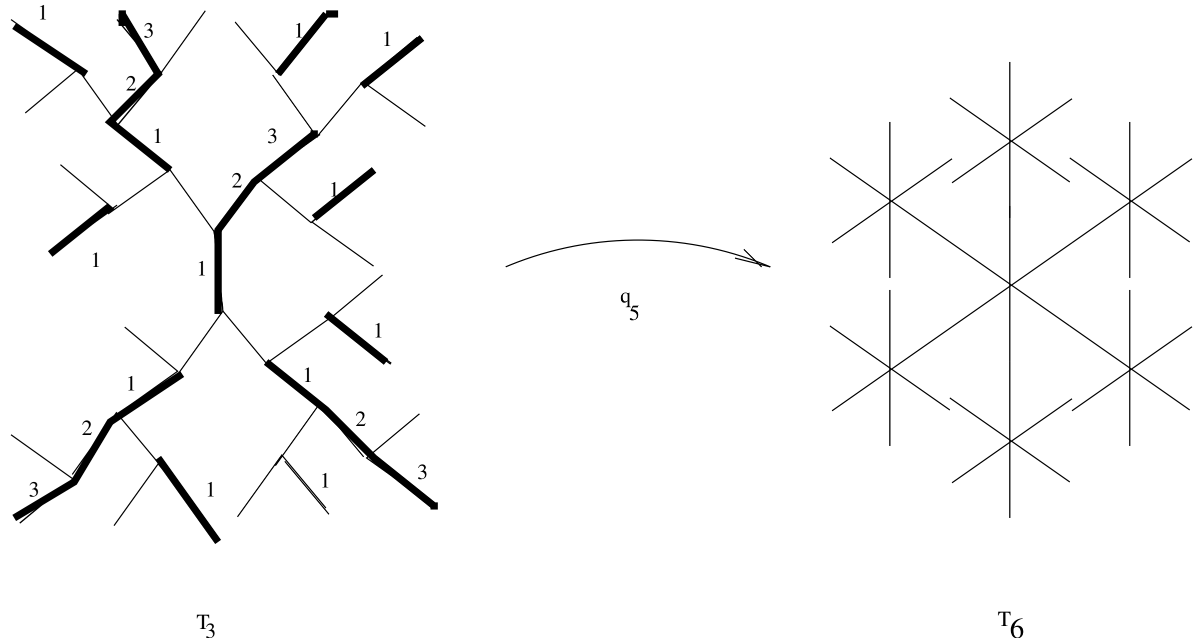


FIGURE 5.1. All regular simplicial trees are quasi-isometric.

□

5.3. 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 manifold/space, that is a simplicial graph quasi-isometric to the manifold. We begin with a few simple observations.

LEMMA 5.40. *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 $\mathcal{G}_\kappa(A)$ with the standard metric.

Theorem 5.41. (1) *Let* (X, dist) *be a proper geodesic metric space (equivalently a complete, locally compact length metric space, see Theorem 1.29). Let* N *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 metric space* (X, dist) *and the graph* $\mathcal{G}$ *are quasi-isometric. More precisely, for all* $x, y \in N$ *we have that*

$$\frac{1}{8\delta}\mathrm{dist}_X(x,y) \leqslant \mathrm{dist}_{\mathcal{G}}(x,y) \leqslant \frac{3}{\varepsilon}\mathrm{dist}_X(x,y)\,. \tag{5.4}$$

(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.*

Proof. (1) Let x, y be two fixed 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 (5.4) 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 $e_1e_2\dots e_s$, where x is the initial vertex of e_1 and y is the terminal 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

$$t_0 = 0, t_1, t_2, \dots, t_m = \mathrm{dist}_X(x,y)$$

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 $i \in \{0, 1, \dots, m-1\}$.

Let $x_i = \mathfrak{c}(t_i)$, $i \in \{0, 1, 2, \dots, m\}$. For every $i \in \{0, 1, 2, \dots, m\}$ there exists $w_i \in N$ such that $\mathrm{dist}_X(x_i, w_i) \leqslant \delta$. We note that $w_0 = x, w_m = y$. The choice of t_i implies that

$$3\delta \leqslant \mathrm{dist}_X(w_i, w_{i+1}) \leqslant 8\delta\,, \quad \text{for every } i \in \{0, \dots, m-1\}$$

In particular:

- w_i and w_{i+1} are the endpoints of an edge in $\mathcal{G}$, for every $i \in \{0, \dots, m-1\}$;
- $\mathrm{dist}_X(x_i, x_{i+1}) \geqslant \mathrm{dist}(w_i, w_{i+1}) - 2\delta \geqslant \mathrm{dist}(w_i, w_{i+1}) - \frac{2}{3}\mathrm{dist}(w_i, w_{i+1}) = \frac{1}{3}\mathrm{dist}(w_i, w_{i+1})$.

We can then write

$$\mathrm{dist}_X(x,y) = \sum_{i=0}^{m-1}\mathrm{dist}_X(x_i, x_{i+1}) \geqslant \frac{1}{3}\sum_{i=0}^{m-1}\mathrm{dist}(w_i, w_{i+1}) \geqslant \frac{\varepsilon}{3}m \geqslant \frac{\varepsilon}{3}\mathrm{dist}_{\mathcal{G}}(x,y)\,. \tag{5.5}$$

(2) According to the discussion following Definition 2.60, the graph $\mathcal{G}$ has bounded geometry if and only if its set of vertices with the induced simplicial distance is uniformly discrete. Lemma 5.40 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 2.58. 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 quasi-isometric simplicial complex) using Theorem 2.62, which implies:

THEOREM 5.42. *Every Riemannian manifold M of bounded geometry is quasi-isometric to a simplicial complex homeomorphic to M.*

5.4. Metric filling functions

In this section we define notions of loops, filling disks and minimal filling area in the setting of geodesic metric spaces, following [**Gro93**]. Let X be a geodesic metric space and $\delta > 0$ be a fixed constant. In this present setting of isoperimetric inequalities, by *loops* we always mean Lipschitz maps $\mathfrak{c}$ from the unit circle $\mathbb{S}^1$ to X. We will use the notation ℓ_X for the length of an arc in X.

A *δ-loop* in X is a triangulated circle S^1 together with a (Lipschitz) map $\mathfrak{c} : S^1 \to X$, so that for $\ell_X(c(e)) \leqslant \delta$ for every edge e of the triangulation.

A *filling disk of* $\mathfrak{c}$ is a pair consisting of a triangulation $\mathcal{D}$ of the 2-dimensional unit disk $\mathbb{D}^2$ extending the triangulation of its boundary circle S^1 and a map

$$\mathfrak{d} : \mathcal{D}^{(0)} \to X$$

extending the map $\mathfrak{c}$ restricted to the set of boundary vertices. Here $\mathcal{D}^{(0)}$ is the set of vertices in $\mathcal{D}$. Sometimes by abuse of language we call the image of the map $\mathfrak{d}$ also *filling disk of* $\mathfrak{c}$.

We next extend the map $\mathfrak{d}$ to the 1-skeleton of $\mathcal{D}$. For every edge e of $\mathcal{D}$ (not contained in the boundary circle) we pick a geodesic connecting the images of the end-points of e under $\mathfrak{d}$. For every boundary edge e of the 2-disk we use the restriction of the map $\mathfrak{d}$ to e in order to connect the images of the vertices. The triangles in X thus obtained are called *bricks*. The *length of a brick* is the sum of the lengths of its edges. The *mesh of a filling disk* is the maximum of the lengths of its bricks. By abusing the notation, we will refer to this extension of $\mathfrak{d}$ to $\mathcal{D}^{(1)}$ as a δ-filling disk as well.

A *δ-filling disk of* $\mathfrak{c}$ is a filling disk with mesh at most δ. The combinatorial area of such a disk is just the number of 2-simplices in the triangulation of D^2.

DEFINITION 5.43. The *δ-filling area of* $\mathfrak{c}$ is the minimal combinatorial area of a δ-filling disk of $\mathfrak{c}$. We will use the double notation $\mathrm{Ar}_\delta(\mathfrak{c}) = P(\mathfrak{c}, \delta)$ for the δ-filling area.

Note that Ar_δ is a function defined on the set Ω of loops and taking values in $\mathbb{Z}_+$.

We, likewise, define the *δ-filling radius function* as

$$r_\delta : \Omega \to \mathbb{R}_+,$$

$$r_\delta(\mathfrak{c}) = \inf \left\{ \max_{x \in \mathcal{D}^{(0)}} \mathrm{dist}_X(\mathfrak{d}(x), \mathfrak{c}(\mathbb{S}^1)) \; ; \; \mathfrak{d} \text{ is a } \delta - \text{filling disk of the loop } \mathfrak{c} \right\}.$$

Both functions depend on the parameter δ, and may take infinite values. In order to obtain finite valued functions, we add the hypothesis that there exists a sufficiently large μ so that for all $\delta \geqslant \mu$, every loop has a δ-filling disk. Such spaces will be called *μ-simply connected*.

EXERCISE 5.44. Show that a geodesic metric space is coarsely simply-connected in the sense of Definition 6.13 if and only if X is μ-simply connected for some μ.

In the sequel we only deal with μ-simply connected metric spaces. We occasionally omit to recall this hypothesis.

We can now define the *δ-filling function* $Ar_\delta : \mathbb{R}_+ \to \mathbb{Z}_+$, $Ar_\delta(\ell) :=$ the maximal area needed to fill a loop of length at most ℓ. For our convenience, we use in parallel the notation $P(\ell, \delta)$ for this function. We will also use the name *δ-isoperimetric function* for $Ar_\delta(\ell)$.

To get a better feel for the δ-filling function, let us relate Ar_δ with the usual area function in the case $X = \mathbb{R}^2$. Recall (see [**Fed69**]) that every loop c in $\mathbb{R}^2$ satisfies the *Euclidean isoperimetric inequality*

$$4\pi A(c) \leqslant \ell^2(c), \tag{5.6}$$

where the equality is realized in the case when c is a round circle. Suppose that $\mathfrak{c}$ is a loop in $\mathbb{R}^2$ and $\mathfrak{d} : \mathcal{D}^{(1)} \to X$ is a δ-filling disk for $\mathfrak{c}$. Then $\mathfrak{d}$ extends to a map $\mathfrak{d} : D^2 \to \mathbb{R}^2$, where we extend the restriction of $\mathfrak{d}$ to each 2-simplex σ by the least area disk bounded by the loop $\mathfrak{d}|\partial\sigma$. In view of the isoperimetric inequality (5.6) he resulting map $\mathfrak{d}$ will have area

$$Area(\mathfrak{d}) \leqslant \sum_\sigma \ell(\mathfrak{d}_{\partial\sigma}) \leqslant Ar_\delta(\mathfrak{d})\frac{\delta^2}{4\pi}, \tag{5.7}$$

where the sum is taken over all 2-simplices in $\mathcal{D}$. In general, it is impossible to estimate Ar_δ from above, however, one can do so for carefully chosen maps $\mathfrak{d}$. Namely, we will think of the map $\mathfrak{c}$ as a function f of the angular coordinate $\theta \in [0, 2\pi]$. Suppose that f is L-Lipschitz. Choose coordinates in $\mathbb{R}^2$ so that the origin is $\mathfrak{c}(0)$ and define a function

$$F(r, \theta) = r\mathfrak{c}(\theta).$$

Then F is $L' = \sqrt{1 + 4\pi^2}L$-Lipschitz. Subdivide the rectangle $[0, 1] \times [0, 2\pi]$ (the domain of F) in subrectangles of width ϵ_1 and height ϵ_2 and draw the diagonal in each rectangle. Then the restriction of F to the boundary of each 2-simplex of the resulting triangulation is a $2L'(\epsilon_1 + \epsilon_2)$-brick. Therefore, in order to ensure that F is a δ-filling of the map f, we take:

$$n = \lceil \frac{4L'}{\delta} \rceil, m = \lceil \frac{8\pi L'}{\delta} \rceil.$$

Hence, $Ar_\delta(\mathfrak{c})$ is at most

$$2nm \leqslant \frac{1}{\delta^2} 32(L')^2 = \frac{1 + 4\pi^2}{\delta^2} L^2.$$

In terms of the length ℓ of $\mathfrak{c}$,

$$Ar_\delta(\mathfrak{c}) \leqslant \frac{1 + 4\pi^2}{\delta^2 4\pi^2} \ell^2 \leqslant \frac{2}{\delta^2} \ell^2.$$

Likewise, using the radius function we define the *filling radius function* as

$$r : \mathbb{R}_+ \to \mathbb{R}_+,\ r(\ell) = \sup\{r(\mathfrak{c})\ ;\ \mathfrak{c} \text{ loop of length } \leqslant \ell\}\ .$$

Two filling functions corresponding to different δ's for a metric space, or, more generally, for two quasi-isometric metric spaces, satisfy a certain equivalence relation.

In a geodesic metric space X that is μ-simply connected, if $\mu \leqslant \delta_1 \leqslant \delta_2$ then one can easily see, by considering partitions of bricks of length at most δ_2 into bricks of length at most δ_1 that

$$A_{\delta_1}(\ell) \leqslant A_{\delta_2}(\ell) \leqslant A_{\delta_2}(\delta_1)\, A_{\delta_1}(\ell)$$

and that

$$r_{\delta_1}(\ell) \leqslant r_{\delta_2}(\ell) \leqslant r_{\delta_2}(\delta_1)\, r_{\delta_1}(\ell)\,.$$

EXERCISE 5.45. (1) Prove that if two geodesic metric spaces X_i, $i = 1, 2$, are coarsely simply connected and quasi-isometric, then their filling functions, respectively their filling radii, are asymptotically equal. Hint: Suppose that $f : X_1 \to X_2$ is an (L, A)-quasi-isometry. Start with a 1-loop $\mathfrak{c}_1 : S^1 \to X_1$, then fill-in $\mathfrak{c}_2 = f \circ \mathfrak{c}$ in X_2 using a δ_2-disk $\mathcal{D}_2$, where $\delta_2 = L + A$; then compose $\mathcal{D}_2$ with quasi-inverse to f in order to fill-in the original loop $\mathfrak{c}_1$ using a δ_1-disk $\mathcal{D}_1$, where $\delta_1 = L\delta_2 + A$. Now, argue that $Ar_{\delta_1}(\mathfrak{c}) \leqslant Ar_{\delta_2}(\mathfrak{c}_2)$).

(2) Prove that for a finitely presented group G the metric filling function for an arbitrary Cayley graph Γ_G and the Dehn function have the same order. Hint: It is clear that $Dehn(\ell) \leqslant Ar_\mu(\ell)$, where μ is the length of the longest relators of G. Use optimal Van Kampen diagrams for a loop $\mathfrak{c}$ of length ℓ, to construct μ-filling disks in Γ_G whose area is $\leqslant Dehn(\ell) + 4(\ell + 1)$.

Note that one can also define Riemannian filling functions in the context of simply-connected Riemannian manifolds M: Given a Lipschitz loop c in M one defines $Area(c)$ to be the least area of a disk in M bounding c. Then the *isoperimetric function* $IP_M(\ell)$ of the manifold M is

$$IP_M(\ell) = \sup\{A(c) : length(c) \leqslant \ell\}$$

where $\ell(c)$ is the length of c. Then, assuming that M admits a geometric action of a group G, we have

$$Ar_\delta(\ell) \approx Dehn(\ell) \approx IP_M(\ell),$$

see [**BT02**].

The order of the filling function of a metric space X is also called *the filling order of* X. Besides the fact that it is a quasi-isometry invariant, the interest of the filling order comes from the following result, a proof of which can be found for instance in [**Ger93**].

PROPOSITION 5.46. *In a finitely presented group G the following statements are equivalent.*

(S_1) *G has solvable word problem.*
(S_2) *the Dehn function of G is recursive.*
(S_3) *the filling radius function of G is recursive.*

If in a metric space X the filling function $Ar(\ell)$ 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 in Rips complex. Suppose that X is μ-connected. Instead of filling closed curves in X by δ-disks, one can fill in polygonal loops in $P = Rips_\delta(X)$ with simplicial disks. Let $\mathfrak{c}$ be a δ-loop in X. Then we have a triangulation of the circle S^1 so that $\operatorname{diam}(\mathfrak{c}(\partial e)) \leqslant \delta$ for every edge e of the triangulation. Thus, we define a loop $\mathfrak{c}_\delta$ in P by replacing arcs $\mathfrak{c}(\partial e)$ with edges of P connecting the end-points of these arcs. Then

$$\delta c-length(\mathfrak{c}_\delta)=\delta length(\mathfrak{c}_\delta)\geqslant \text{length}(\mathfrak{c})$$

since every edge of P has unit length. It is clear that for $\delta>0$ the map

$$\{\text{loops in } X \text{ of length} \leqslant \ell\}\to\{\text{loops in } P \text{ of length} \leqslant \frac{\ell}{\delta}\}$$

$$c\mapsto c_\delta$$

is surjective. Furthermore, every δ-disk $\mathcal{D}$ which fills in $\mathfrak{c}$ yields a simplicial map $\mathcal{D}_\delta: D^2\to P$ which is an extension of $\mathfrak{c}_\delta$: The maps $\mathcal{D}$ and $\mathcal{D}_\delta$ agree on the vertices of the triangulation of D^2, and for every 2-simplex σ in D^2, the map $\mathcal{D}_\delta|\sigma$ is the canonical linear extension of $\mathcal{D}|\sigma^{(0)}$ to the simplex (of dimension $\leqslant 2$) in P spanned by the vertices $\mathcal{D}(\sigma^{(0)})$. Furthermore, area is preserved by this construction:

$$cArea(\mathcal{D}_\delta)=Ar_\delta(\mathcal{D}).$$

This construction produces all simplicial disks in P bounding $\mathfrak{c}_\delta$ and we obtain

$$cArea(\mathfrak{c}_\delta)=Ar_\delta(\mathfrak{c}).$$

Summarizing all this, we obtain

$$A_{Rips_\delta(X)}(\ell)=Ar_\delta(\frac{\ell}{\delta}).$$

The same argument applies to the filling radius and we obtain:

Observation 5.47. Studying filling area and filling radius functions in X (up to the equivalence relation $\approx$) is equivalent to studying combinatorial filling area and filling radius functions in $Rips_\delta(X)$.

Besikovitch inequality. The following proposition relates filling areas of curvilinear quadrilaterals in X to the product among of separation of their sides.

Proposition 5.48 (The quadrangle or Besikovitch inequality). *Let X be a μ–simply connected geodesic metric space and let $\delta\geqslant\mu$.*

Consider a loop $\mathfrak{c}\in\Omega_X$ and its decomposition $\mathfrak{c}(\mathbb{S}^1)=\alpha_1\cup\alpha_2\cup\alpha_3\cup\alpha_4$ into four consecutive paths. Then, with the notation $d_1=\text{dist}(\alpha_1,\alpha_3)$ and $d_2=\text{dist}(\alpha_2,\alpha_4)$ we have that

$$\text{Ar}_\delta(\mathfrak{c})\geqslant\frac{2\pi}{\delta^2}d_1d_2\,.$$

Proof. Let $\mathfrak{d}:\mathcal{D}^{(1)}\to X$ be a filling disk of $\mathfrak{c}$ realizing the filling area. Consider a map $\beta: X\to\mathbb{R}^2$ defined by

$$\beta(x)=(\text{dist}(x,\alpha_1)\,,\,\text{dist}(x,\alpha_2))\,.$$

Since each of its components is a 1–Lipschitz map, the map β is $\sqrt{2}$–Lipschitz. The image $\beta(\alpha_1)$ is a vertical segment connecting the origin to a point $(0,y_1)$, with $y_1\geqslant d_2$, while $\beta(\alpha_2)$ is a horizontal segment connecting the origin to a point $(x_2,0)$, with $x_2\geqslant d_1$. Similarly, the image $\beta(\alpha_3)$ is a path to the right of the vertical line $x=d_1$ and $\beta(\alpha_4)$ another path above the horizontal line $y=d_2$. Thus, the rectangle R with the vertices $(0,0),(d_1,0),(d_1,d_2),(0,d_2)$ is separated from infinity by the curve $\beta\mathfrak{c}(S^1)$ (see Figure 5.2). In particular, the image of any extension F of $\beta\circ\mathfrak{d}$

to D^2 contains the rectangle R. Thus, $A(F) \geqslant A(R) = d_1 d_2$, hence, by inequality (5.7),

$$d_1 d_2 \leqslant \frac{2\delta^2}{4\pi} Ar_{\sqrt{2}\delta}(\beta \circ c).$$

Furthermore, since β is $\sqrt{2}$-Lipschitz,

$$Ar_{\sqrt{2}\delta}(\beta \circ \mathfrak{c}) \leqslant 2Ar_\delta(\mathfrak{c}).$$

Putting this all together, we get

$$Ar_\delta(\mathfrak{c}) \geqslant \frac{\pi}{\delta^2} d_1 d_2$$

as required. $\square$

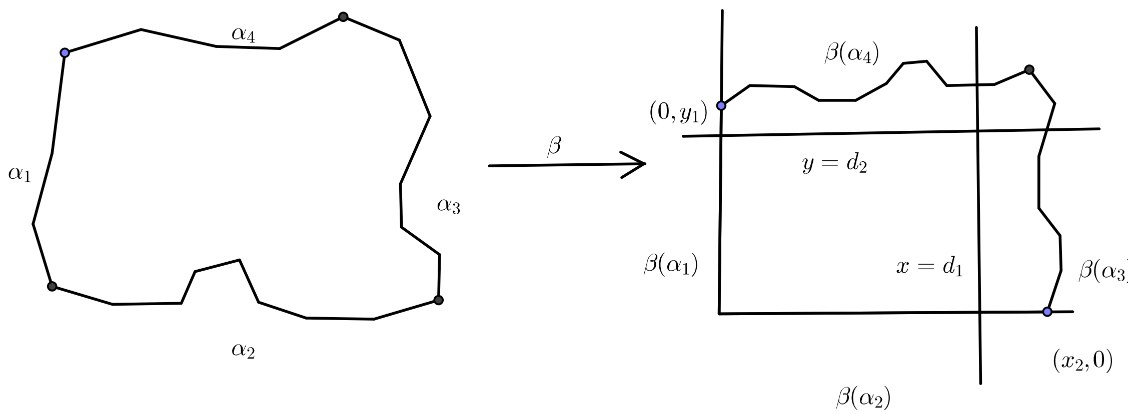


FIGURE 5.2. The map β.

Besikovitch's inequality generalizes from curvilinear quadrilaterals to curvilinear triangles: This generalization below is has interesting applications to δ-hyperbolic spaces. We first need a definition which would generalize the condition of separation of the opposite edges of a curvilinear quadrilateral.

DEFINITION 5.49. Given a topological triangle T, i.e. a loop $\mathfrak{c}$ composed of a concatenation of three paths τ_1, τ_2, τ_3, *the minimal size* (minsize) of T is defined as

$$\text{minsize}(T) = \inf\{\text{diam}\{y_1,\, y_2,\, y_3\}\ ;\ y_i \in \tau_i,\, i = 1, 2, 3\}\ .$$

PROPOSITION 5.50 (Minsize inequality). *Let X be a μ–simply connected geodesic metric space and let $\delta \geqslant \mu$.*

Given a topological triangle $T \in \Omega$, we have that

$$\text{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2\pi}{\delta^2} [\text{minsize}(T)]^2\ .$$

PROOF. As before, define a $\sqrt{2}$–Lipschitz map $\beta : X \to \mathbb{R}^2$,

$$\beta(x) = (\beta_1(x), \beta_2(x)) = (\mathrm{dist}(x, \tau_1), \quad \mathrm{dist}(x, \tau_2))$$

and note that, as in the proof of Besikovitch's inequality, β maps τ_1, τ_2 to coordinate segments, while the restriction of β to τ_3 satisfies:

$$\min(\beta_1(x), \beta_2(x)) \geqslant m,$$

where $m = \mathrm{minsize}(T)$. Therefore, the loop $\beta \circ \mathfrak{c}$ separates from infinity the square Q with the vertices $(0,0), (m,0), (m,m), (0,m)$. Then, as before,

$$m^2 \leqslant \frac{\delta^2}{2\pi} \mathrm{Ar}(\mathfrak{c})$$

and claim follows. □

The Dehn function/area filling function can be generalized to higher dimensions and n-Dehn functions, which give information about the way to fill topological spheres $\mathbb{S}^n$ with topological balls $\mathbf{B}^{n+1}$ ([**Gro93**, Chapter 5], [**ECH$^+$92**, Chapter 10], [**Pap00**]). The following result was proven by P. Papasoglou:

THEOREM 5.51 (P. Papasoglou, [**Pap00**]). *The second Dehn function of a group of type $\mathbf{F}_3$ is bounded by a recursive function.*

The condition $\mathbf{FP}_3$ is a 3-dimensional version of the condition of finite presentability of a group: A group G is of type $\mathbf{F}_3$ if there exists a finite simplicial complex K with $G = \pi_1(K)$ and $\pi_2(K) = 0$. A *basic sphere* in the 2-dimensional skeleton of K is the boundary of an oriented 3-simplex together with a path connecting its vertex to a base-point v in K.

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.

The idea of the proof of Theorem 5.51 is to produce an algorithm which, given $n \in \mathbb{N}$, finds in finite time an upper bound on the number N of basic spheres σ_j, so that (in $\pi_2(K, v)$)

$$\sum_{i=1}^{N} \sigma_j = \sigma,$$

where σ is a spherical 2-cycle in K which consists of at most n 2-dimensional simplices. The algorithm only gives a recursive bound of the second Dehn function, because the filling found by it might be not the smallest possible.

The above algorithm does not work for the ordinary Dehn function since it would require one to recognize which loops in K are homotopically trivial.

5.5. Summary of various notions of volume and area

(1) $Vol(f)$ is the Riemannian volume of a map; geometric volume of a smooth map of regular cell-complexes. For $n = 2$, $Vol(f) = Area(f)$.
(2) Combinatorial volume: $cVol_n(f)$, the number of n-simplices in the domain not collapsed by f. For $n = 2$, $cVol_2(f) = cArea(f)$.
(3) Simplicial volume: $sVol_n(f)$ is the number of n-simplices in the domain of f.

(4) Combinatorial area: $A(w)$, minimal filling combinatorial area for a trivial word w (algebraic area); algebraically speaking, it equals area of the minimal van Kampen diagram with the given boundary loop w.
(5) Coarse area: $\mathrm{A}r_\delta(\mathfrak{c})$, the δ-filling area of a δ-loop $\mathfrak{c}$ in a coarsely simply-connected metric space X.
(6) Dehn function: $Dehn_G(n)$, the Dehn function of a presentation complex Y of a group G.
(7) Isoperimetric function $IP_M(\ell)$ of a simply-connected Riemannian manifold M.

Summary of relationships between the volume/area concepts:

(1) Functions $Dehn_G(n)$ and $IP_M(\ell)$ are *approximately equivalent* to each other, provided that G acts geometrically on M; both functions are QI invariant, provided that one considers them up to approximate equivalence.
(2) $\mathrm{A}r_\delta(\mathfrak{c}) \asymp Area_P(c)$, where $P = \mathrm{Rips}_\delta(X)$ and c is the loop in P obtained from $\mathfrak{c}$ by connecting "consecutive points" by the edges in P.

5.6. Topological coupling

We first introduce Gromov's interpretation of quasi-isometry between groups using the language of topological actions.

Given groups G_1, G_2, a *topological coupling* of these groups is a metrizable locally compact topological space X together with two commuting cocompact properly discontinuous topological actions $\rho_i : G_i \curvearrowright X, i = 1, 2$. (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. 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 groups.*

PROOF. 1. Suppose that G_1 is QI to G_2. Then there exists an (L, A) quasi-isometry $q : G_1 \to G_2$. Without loss of generality, we may assume that q is L–Lipschitz. Consider the space X of such maps $G_1 \to G_2$. We will give X the topology of pointwise convergence. By 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 $f \in X$ there exist $g_1 \in G_1, g_2 \in G_2$ so that

$$g_2 \circ f(1) = 1, f \circ g_1^{-1}(1) \in B(1, A).$$

Therefore, by Arzela–Ascoli theorem, both actions 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 $f \in K$, $f(1) \in B(1, R)$. If $g_2 \in G_2$ is such that $g_2 \circ f \in K$ for some $f \in K$, then

$$g_2(B(1, R)) \cap B(1, R) \neq \emptyset. \tag{5.8}$$

Since the action of G_2 on itself is free, it follows that the collection of $g_2 \in G_2$ satisfying (5.8) 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. However, there are examples of QI groups which do not act geometrically on the same geodesic metric space, see Theorem 5.29. Nevertheless, the construction of a quasi-isometry below is pretty much the same as in the proof of Milnor-Schwarz theorem.

Since $G_i \curvearrowright X$ is cocompact, there exists a compact $K \subset X$ so that $G_i \cdot K = X$; pick a point $p \in K$. Then for each $g_i \in G_i$ there exists $\phi_i(g_i) \in G_{i+1}$ so that $g_i(p) \in \phi_i(g_i)(K)$, here and below i is taken mod 2. We have maps $\phi_i : G_i \to G_{i+1}$.

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

$$\cup_{i=1,2} \bigcup_{s \in S_i} s(K).$$

Since ρ_i are properly discontinuous actions, the sets $G_i^C := \{h \in G_i : h(C) \cap C \neq \emptyset\}$ are finite for $i = 1, 2$. 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 $g_i(p) \in g_{i+1}(K)$, $sg_i(p) \in g'_{i+1}(K)$ for some $g'_{i+1} \in G_{i+1}$. Therefore, $sg_{i+1}(K) \cap g'_{i+1}(K) \neq \emptyset$ 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, $g_{i+1}^{-1} g'_{i+1} \in G_{i+1}^C$, hence $d(g_{i+1}, g'_{i+1}) \leqslant L$. Consequently, ϕ_i is L–Lipschitz.

b. Let $\phi_i(g_i) = g_{i+1}, \phi_{i+1}(g_{i+1}) = g'_i$. Then $g_i(K) \cap g'_i(K) \neq \emptyset$ hence $g_i^{-1} g'_i \in G_i^C$. Therefore, $\mathrm{dist}(\phi_{i+1} \circ \phi_i, Id_{G_i}) \leqslant L$ and $\phi_i : G_i \to G_{i+1}$ is a quasi-isometry. $\square$

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 quasi-geodesic metric space X such that G_1, G_2 both act geometrically on X.

In view of Theorem 5.29, if two groups have a common geometric model then they are quasi-isometric. The following theorem shows that the converse is false:

Theorem 5.54 (L. Mosher, M. Sageev, K. Whyte, [**MSW03**]). *Let $G_1 := \mathbb{Z}_p * \mathbb{Z}_p, G_2 := \mathbb{Z}_q * \mathbb{Z}_q$, where p, q are distinct odd primes. Then the groups G_1, G_2 are quasi-isometric (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).

5.7. Quasi-actions

The notion of an *action* of a group on a space is replaced, in the context of quasi-isometries, by the one of *quasi-action*. 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 quasi-isometries are defined only up to "bounded error", the concept of a homomorphism has to be modified when we use quasi-isometries.

Definition 5.55. Let G be a group and X be a metric space. An (L, A)-quasi-action of G on X is a map $\phi : G \to Map(X, X)$, so that:

- $\phi(g)$ is an (L, A)-quasi-isometry of X for all $g \in G$.
- $d(\phi(1), id_X) \le A$.
- $d(\phi(g_1 g_2), \phi(g_1)\phi(g_2)) \le A$ for all $g_1, g_2 \in G$.

Thus, ϕ is "almost" a homomorphism with the error A.

By abusing notation, we will denote quasi-actions by $\phi : G \curvearrowright X$, even though, what we have is not an action.

Example 5.56. 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:

$$|\phi(g_1 g_2) - \phi(g_1) - \phi(g_2)| \le A.$$

Such maps ϕ are called *quasi-morphisms.* and they appear frequently in geometric group theory, in the context of *2nd bounded cohomology*, see e.g. [**EF97a**]. Many interesting groups do not admit nontrivial homomorphisms of $\mathbb{R}$ but admit unbounded quasi-morphisms. For instance, a hyperbolic Coxeter group G does not admit nontrivial homomorphisms to $\mathbb{R}$. However, unless G is virtually abelian, it has infinite-dimensional space of equivalence classes quasi-morphisms, where

$$\phi_1 \sim \phi_2 \iff \|\phi_1 - \phi_2\| < \infty.$$

See [**EF97a**].

Exercise 5.57. Let $QI(X)$ denote the group of (equivalence classes of) quasi-isometries $X \to X$. Show that every quasi-action determines a homomorphism $\widehat{\phi} : G \to QI(X)$ given by composing ϕ with the projection to $QI(X)$.

The *kernel* of the quasi-action $\phi : G \curvearrowright X$ is the kernel of the homomorphism $\widehat{\phi}$.

Exercise 5.58. Construct an example of a geometric quasi-action $G \curvearrowright \mathbb{R}$ whose kernel is the entire group G.

We can also define proper discontinuity and cocompactness for quasi-actions by analogy with isometric actions:

Definition 5.59. Let $\phi : G \curvearrowright X$ be a quasi-action.

1. We say that ϕ is *properly discontinuous* if for every $x \in X, R \in \mathbb{R}_+$, the set

$$\{g \in G | d(x, \phi(g)(x)) \le R\}$$

is finite. Note that if X proper and ϕ is an isometric action, this definition is equivalent to proper discontinuity of $G \curvearrowright X$.

2. We say that ϕ is *cobounded* if there exists $x \in X, R \in \mathbb{R}_+$ so that for every $x' \in X$ there exists $g \in G$ so that $d(x', \phi(g)(x)) \le R$. Equivalently, there exists R' so that $d(x, \phi(g)(x')) \le R$.

3. Lastly, we say that quasi-action ϕ is *geometric* if it is both properly discontinuous and cobounded.

Below we explain how quasi-actions 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 quasi-isometry with quasi-inverse $\bar{f}$. We then define a *conjugate* quasi-action $\phi = f^*(\psi_2)$ of G_2 on X_1 by

$$\phi(g) = \bar{f} \circ g \circ f. \tag{5.9}$$

More generally, we say that two quasi-actions $\psi_i : G \curvearrowright X_i$ are *quasi-conjugate* if there exists a quasi-isometry $f : X_1 \to X_2$, so that ψ_1 and $f^*(\psi_2)$ project to the same homomorphism

$$G \to QI(X_1).$$

LEMMA 5.60. *1. Under the above assumptions, $\phi = f^*(\psi_2)$ is a quasi-action. 2. If ψ_2 is geometric, so is ϕ.*

PROOF. 1. Suppose that f is an (L, A)-quasi-isometry. It is clear that ϕ satisfies Parts 1 and 2 of the definition, we only have to verify (3):

$$\mathrm{dist}(\phi(g_1 g_2), \phi(g_1)\phi(g_2)) = \mathrm{dist}(\bar{f} g_1 g_2 f, \bar{f} g_1 f \bar{f} g_2 f) \le LA + A$$

in view of Exercise 5.7.

2. In order to verify that ϕ is geometric, one needs to show proper discontinuity and coboundedness. We will verify the former since the proof of the latter is similar. Pick $x \in X, R \in \mathbb{R}_+$, and consider the set the set

$$G_{x,R} = \{g \in G = G_2 | d(x, \phi(g)(x)) \le R\} \subset G.$$

By definition, $\phi(g)(x) = \bar{f} g f(x)$. Thus, $d(x, g(x)) \le LR + 2A$. Hence, by proper discontinuity of the action $G \curvearrowright X_2$, the set $G_{x,R}$ is finite. □

The same construction of a conjugate quasi-action applies if $G_2 \curvearrowright X_2$ is not an action, but merely a quasi-action.

EXERCISE 5.61. Suppose that $\phi_2 : G \curvearrowright X_2$ is a quasi-action, $f : X_1 \to X_2$ is a quasi-isometry and $\phi_1 : G \curvearrowright X_1$ is the conjugate quasi-action. Then ϕ_2 is properly discontinuous (respectively, cobounded, or geometric) if and only if ϕ_1 is properly discontinuous (respectively, cobounded, or geometric).

COROLLARY 5.62. *Let G_1 and G_2 be finitely generated quasi-isometric groups and let $f : G_1 \to G_2$ be a quasi-isometry. Then:*

1. The quasi-isometry f induces (by conjugating actions and quasi-actions on G_2) an isomorphism $QI(G_2) \to QI(G_1)$ and a homomorphism $f_* : G_2 \to QI(G_1)$

2. The kernel of f_* is quasi-finite: For every $K \ge 0$, the set of $g \in G_2$ such that $\mathrm{dist}(f_*(g), \mathrm{id}_{G_1}) \leqslant K$, is finite.

PROOF. To construct f_* apply Lemma 5.60 to the isometric action $\psi_2 : G_2 \curvearrowright G_2$. Quasifiniteness of the kernel of f_* follows from proper discontinuity of the quasi-action $G_2 \curvearrowright G_1$. The isomorphism $QI(G_2) \to QI(G_1)$ is defined *via* the formula (5.9). The inverse to this homomorphism is defined by switching the roles of f and $\bar{f}$. □

REMARK 5.63. For many groups $G = G_1$, if $h : G \to G$ is an (L, A)-quasi-isometry, so that $\mathrm{dist}(f, Id_G) < \infty$, then $\mathrm{dist}(f, Id_G) \le D(L, A)$. For instance, this holds when G is a non-elementary hyperbolic group, see Lemma 8.86. (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, quasi-finite kernel of f_* above is actually finite.

The following theorem is a weak converse to the construction of a conjugate quasi-action:

THEOREM 5.64 (B. Kleiner, B. Leeb, [**KL09**]). *Suppose that $\phi : G \curvearrowright X_1$ is a quasi-action. Then there exists a metric space X_2, a quasi-isometry $f : X_1 \to X_2$ and an isometric action $\psi : G \curvearrowright X_2$, so that f quasi-conjugates ψ to ϕ.*

Thus, every quasi-action is conjugate to an isometric action, but, *a priori*, on a different metric space. The key issue of the QI rigidity is:

Can one, under some conditions, take $X_2 = X_1$?

Most proofs of QI rigidity theorems follow this route:

1. Suppose that groups G_1, G_2 are quasi-isometric. Find a "nice space" X_1 on which G_1 acts geometrically. Take a quasi-isometry $f : X_1 \to X_2 = G_2$, where $\psi : G_2 \curvearrowright G_2$ is the action by left multiplication.

2. Define the conjugate quasi-action $\phi = f^*(\psi)$ of G_2 on X_1.

3. Show that the quasi-action ϕ has finite kernel (or, at least, identify the kernel, prove that it is, say, abelian).

4. Extend, if necessary, the quasi-action $G_2 \curvearrowright X_1$ to a quasi-action $\hat{\phi}$ on a larger space $\hat{X}_1$.

5. Show that $\hat{\phi}$ has the same projection to $QI(\hat{X}_1)$ as a isometric action $\phi' : G_2 \curvearrowright \hat{X}_1$ by verifying, for instance, that $\hat{X}_1$ has very few quasi-isometries, namely, every quasi-isometry of X is within finite distance from an isometry. (Well, maybe no all quasi-isometries of $\hat{X}_1$, but the ones which extend from X_1.) Then conclude either that $G_2 \curvearrowright \hat{X}_1$ is geometric, or, that the isometric actions of G_1, G_2 are commensurable, i.e., the images of G_1, G_2 in $\mathrm{Isom}(\hat{X}_2)$ have a common finite-index subgroup.

We will see how R. Schwarz's proof of QI rigidity for nonuniform lattices follows this line of arguments: X_1 will be a truncated hyperbolic space and $\hat{X}_1$ is 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 [**KL98**]; 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 quasi-isometries, e.g. if $X_1 = \mathbb{H}^n$ or $X_1 = \mathbb{CH}^n$. Then, instead, one shows that every geometric quasi-action $G_2 \curvearrowright X_1$ is quasi-conjugate to a geometric (isometric!) action. We will see such a proof in the case of Sullivan–Tukia rigidity theorem for uniform lattices in $\mathrm{Isom}(\mathbb{H}^n), n \geq 3$. Similar arguments apply in the case of groups quasi-isometric to the hyperbolic plane.

Not all quasi-isometric 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 polynomial growth; the argument in the second case is less direct (see Theorem **??**), but the key fact is that geometric condition of having infinitely many ends is equivalent to the algebraic condition that a group splits over a finite subgroup.

CHAPTER 6

Coarse topology

The goal of this section is to provide tools of algebraic topology for studying quasi-isometries and other concepts of the geometric group theory. The class of *metric cell complexes with bounded geometry* provides a class of spaces for which application of algebraic topology is possible.

6.1. Ends of spaces

In this section we review the oldest coarse topological notion, the one of *ends* of a topological space. Let X be a connected, locally path-connected topological space which admits an exhaustion by compact subsets, i.e., an increasing family of compact subsets $\{K_i\}_{i\in I}$, where I is an ordered set,

$$K_i \subset K_j, \quad i \leqslant j,$$

so that

$$\bigcup_{i\in I} K_i = X.$$

The key example to consider is when X is a proper metric space, $K_i = \overline{B}(o,i)$, $i \in \mathbb{N}$ and $o \in X$ is a fixed point. We will refer to this as the *standard example*. (An important special case to keep in mind is the Cayley graph of a finitely-generated group, where o is a vertex.) For each $K = K_i$ we let $K^c = X \setminus K$.

We then let J denote the set whose elements are connected components of various K_i^c. The set J has the partial order: $C \leqslant C'$ iff $C' \subset C$. Thus, the "larger" C's are the ones which correspond to bigger K's.

Definition 6.1. The set $Ends(X) = \epsilon(X)$ of *ends* of X, is the set of unbounded (from above) increasing chains in the poset J. Every such chain is called an *end* of X.

In the standard example, each end is a sequence of connected nonempty sets

$$C_1 \supset C_2 \supset C_3 \supset \ldots$$

where each C_i is a component of K_i^c.

Equivalently, since we assumed that X is locally path-connected, each element of J is an element of the set $\pi_0(K_i^c)$ for some i. Thus, we have the inverse system of sets $\{\pi_0(K_i^c)\}$ indexed by I, where

$$f_{i,j} : \pi_0(K_j^c) \to \pi_0(K_i^c), i \leqslant j,$$

is the map induced by the inclusion $K_j^c \subset K_i^c$. Then there is a natural bijection between the inverse limit

$$\pi_0^\infty(X) = \varprojlim \pi_0(K_i^c)$$

of this system and the set of ends $\epsilon(X)$: Choosing an element σ of $\pi_0(K_j^c)$ is equivalent to choosing the connected component of K_i^c which gives rise to σ. Note that if X is a Cayley graph, then each $\pi_0(K_i^c)$ is a finite set.

We say that a family of points $(x_i)_{i\in I}$, $x_i \in C_i$, $C_i \subset K_i^c$, *represents* the corresponding end of X, since each x_i represents an element of $\pi_0(K_i^c)$. We will use the notation $x_\bullet$ for this end.

We next topologize $\epsilon(X)$. We equip each $\pi_0(K_i^c)$ with the discrete topology (which makes sense in view of the Cayley graph example) and then put the initial topology on the inverse limit as explained in Section 1.1.

Concretely, one describes this topology as follows. Pick some $C \in J$, which is a component of K_i^c. Then C defines a subset $\epsilon_C \subset X$, which consists of ends which are represented by those families (x_j) so that, $x_j \in C$ for all $j \geqslant i$. These sets form a basis of the inverse limit topology on $\epsilon(X)$ described above. Since $\epsilon(X)$ is the inverse limit of sets with discrete topology, the space $\epsilon(X)$ is totally disconnected. Furthermore, clearly, $\epsilon(X)$ is Hausdorff.

EXERCISE 6.2. 1. The above topology on $\epsilon(X)$ defines a compactification $\bar{X} = X \cup \epsilon(X)$ of the topological space X.

2. Let G be a group of homeomorphisms of X. Then the action of G on X extends to a topological action of G on $\bar{X}$.

REMARK 6.3. 1. Some of the sets ϵ_C could be empty: They correspond to the sets C which are relatively compact. This, of course, means that one should discard such sets C when thinking about the ends of X.

2. There is a terminological confusion here coming from the literature in differential geometry and geometric analysis, where X is a smooth manifold: An analyst would call each set C an end of X.

EXAMPLE 6.4. 1. Every compact topological space X has empty set of ends. Conversely, if $\epsilon(X) = \emptyset$, then X is compact.

2. If $X = \mathbb{R}$, then $\epsilon(X)$ is a 2-point set. If $X = \mathbb{R}^n, n \geqslant 2$, then $\epsilon(X)$ is a single point.

3. If X is a binary (i.e., tri-valent) tree then $\epsilon(X)$ is homeomorphic to the Cantor set.

See Figure 6.1 for an example. The space X in this picture has 5 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.

Analogously, one defines *higher homotopy groups* $\pi_k^\infty(X, x^\bullet)$ at infinity of X, $k \geqslant 1$. We now assume that the set I is the set of natural numbers with the usual order. For each end $x_\bullet \in \epsilon(X)$ pick a representing sequence $(x_i)_{i\in I}$. For each $i \leqslant j$, pick a path p_{ij} in K_i^c connecting x_i to x_j. The concatenation of such paths is a proper map $p : \mathbb{R}_+ \to X$. The proper homotopy class of p is denoted $x^\bullet$. Given p, we then have the inverse system of group homomorphisms

$$\pi_k(K_j^c, x_j) \to \pi_k(K_i^c, x_i), i \leqslant j,$$

induced by inclusion maps of the components $C_j \hookrightarrow C_i$, where $x_i \in C_i, x_j \in X_j$. Note that the paths p_{ij} are needed here since we are using different base-points for the homotopy groups.

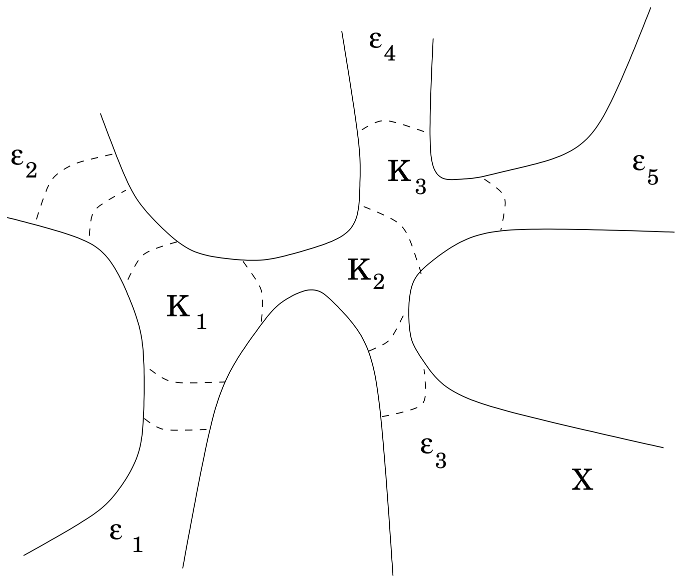


FIGURE 6.1. *E*nds of X.

The group $\pi_k^\infty(X, x^\bullet)$ then is the inverse limit

$$\varprojlim \pi_k(K_i^c, x_i).$$

EXERCISE 6.5. Verify that this construction depends only on $x^\bullet$ and not on the paths p_{ij}.

For the rest of the book, we will not need π_k^∞ for $k > 0$.

PROPOSITION 6.6. *If $f : X \to Y$ is an (L, A)–quasi-isometry of proper geodesic metric spaces then f induces a homeomorphism $\epsilon(X) \to \epsilon(Y)$.*

PROOF. For geodesic metric spaces, path-connectedness is equivalent to connectedness. Since f is a quasi-isometry, for each bounded subset $K \subset X$, the image $f(K)$ is again bounded. Note that f need not map connected sets to connected sets since f is not required to be continuous. nevertheless, we have

LEMMA 6.7. *The open $A' = A + 1$-neighborhood $\mathcal{N}_{A'}(f(C))$ is connected for every connected subset $C \subset X$.*

PROOF. For points $x, x' \in C$, and every $\delta > 0$ there exists a *chain* $x_0 = x, x_1, ..., x_n = x'$, so that $x_i \in C$ and $\mathrm{dist}(x_i, x_{i+1}) \leqslant \delta$, $i = 0, ..., n-1$. Then we obtain a chain $y_i = f(x_i)$, $i = 0, ..., n$, so that

$$\mathrm{dist}(y_i, y_{i+1}) \leqslant \delta' = L\delta + A$$

It follows that a geodesic segment $[y_i y_{i+1}]$ is contained in $\mathcal{N}_{\delta'}(f(C))$. Hence, the δ'–neighborhood of $f(C)$ is path-connected for every $\delta > 0$. We conclude that $\mathcal{N}_{A'}(f(C))$ is connected by taking $\delta = 1$. □

Without loss of generality, we may assume that $K_i = \overline{B}(o, i)$ is a closed metric ball in X and $i \in \mathbb{N}$. We define a map $\epsilon(f) : \epsilon(X) \to \epsilon(Y)$ as follows. Set $R := A + 1$. Suppose that $\eta \in \epsilon(X)$ is represented by a nested sequence (C_i), where C_i is a connected component of $X \setminus K_i$, $K_i \subset X$ is compact. By reindexing our system of compacts K_i, without loss of generality we may assume that for each i, $\mathcal{N}_R(C_i) \subset C_{i-1}$. Thus we get a nested sequence of connected subsets $\mathcal{N}_R(f(C_i)) \subset Y$ each of which is contained in a connected component V_i of the complement to the bounded subset $f(K_{i-1}) \subset Y$. Thus we send η to the end $\epsilon(f)(\eta)$ represented by (V_i). By considering the quasi-inverse $\bar{f}$ to f, we see that $\epsilon(f)$ has the inverse map $\epsilon(\bar{f})$. It is also clear from the construction that both $\epsilon(f)$ and $\epsilon(\bar{f})$ are continuous. □

If G is a finitely generated group then the space of ends $\epsilon(G)$ is defined to be the set of ends of its Cayley graph. The previous lemma implies that $\epsilon(G)$ does not depend on the choice of a finite generating set and that quasi-isometric groups have homeomorphic sets of ends.

THEOREM 6.8 (Properties of $\epsilon(X)$). *1. The topological space $\epsilon(X)$ is compact, Hausdorff and totally disconnected; $\epsilon(X)$ is empty if and only if X is compact.*

2. Suppose that G is a finitely-generated group. Then $\epsilon(G)$ consists of 0, 1, 2 points or has cardinality of continuum. In the latter case the set $\epsilon(G)$ is perfect: Each point is a limit point.

3. $\epsilon(G)$ is empty iff G is finite. $\epsilon(G)$ consists of 2-points iff G is virtually (infinite) cyclic.

4. $|\epsilon(G)| > 1$ iff G splits nontrivially over a finite subgroup.

COROLLARY 6.9. *1. If G is quasi-isometric to $\mathbb{Z}$ then G contains $\mathbb{Z}$ as a finite index subgroup.*

2. Suppose that G splits nontrivially as $G_1 \star G_2$ and G' is quasi-isometric 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.

Note that we already know that $\epsilon(X)$ is Hausdorff and totally-disconnected. Compactness of $\epsilon(X)$ follows from the fact that each K^c has only finitely many components which are not relatively compact. Properties 2 and 3 in Theorem 6.8 are also relatively easy, see for instance [**BH99**, Theorem 8.32] for the detailed proofs. The hard part of this theorem is

THEOREM 6.10. *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 **??** and **??** we prove Theorem 6.10 first for finitely-presented and then for all finitely-generated groups. We will also prove QI rigidity of the class of virtually free groups.

6.2. Rips complexes and coarse homotopy theory

6.2.1. Rips complexes. Let X be a uniformly discrete metric space (see Definition 1.19). Recall that the R-Rips complex of X is the simplicial complex whose vertices are 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

$$\iota_{R_1,R_2} : \mathrm{Rips}_{R_1}(X) \to \mathrm{Rips}_{R_2}(X)$$

and

$$\iota_{R_1,R_2} = \iota_{R_2,R_3} \circ \iota_{R_1,R_2}$$

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 Section 2.2.2, we metrize (connected) Rips complexes $\mathrm{Rips}_R(X)$ using the *standard length metric* on simplicial complexes. Then, each embedding ι_{R_1,R_2} is isometric on every simplex and 1-Lipschitz overall. Note that the assumption that X is uniformly discrete implies that $\mathrm{Rips}_R(X)$ is a simplicial complex of bounded geometry (Definition 2.60) for every R.

EXERCISE 6.11. 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.

The following simple observation explains why Rips complexes are useful for analyzing quasi-isometries:

LEMMA 6.12. *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 $R \geqslant 0$.*

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'; we extend this map linearly to a map $\sigma \to \sigma'$. It is clear that this extension defines 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 6.13. 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 for each r there exists $R \geqslant r$ so that the mapping $\mathrm{Rips}_r(X) \to \mathrm{Rips}_R(X)$ induces trivial maps of the i-th homotopy groups

$$\pi_i(\mathrm{Rips}_r(X), x) \to \pi_i(\mathrm{Rips}_R(X), x)$$

for all $0 \leqslant i \leqslant k$ and $x \in X$.

In particular, X is *coarsely simply-connected* if it is coarsely 1-connected.

In other words, 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 general 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 6.14. *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 disk of diameter $\leqslant r + c$ within $\mathrm{Rips}_{r+c}(X)$.*

PROOF. Pick a vertex $x \in \sigma$. Then $\mathrm{Rips}_{r+c}(X)$ contains a simplicial cone $\tau(\sigma)$ over σ with vertex at x. Clearly, $\mathrm{diam}(\tau) \leqslant r + c$. □

PROPOSITION 6.15. *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. We give the product $\mathrm{Rips}_{r_1}(X) \times I$ the *standard* structure of a simplicial complex with the vertex set $X \times \{0, 1\}$ (by triangulating the each $k+1$-dimensional prisms $\sigma \times I$, where σ are simplices in X, this triangulation has in $\leqslant (k+1)$ top-dimensional simplices); we equip this complex with the standard metric.

The map F of the zero-skeleton of $\mathrm{Rips}_{r_1}(X) \times I$ is, of course, $F(x, 0) = f(x), F(x, 1) = g(x)$. Let $\sigma \subset \mathrm{Rips}_{r_1}(X) \times I$ be an i–simplex. Then $\mathrm{diam}(F(\sigma^0)) \leqslant r_3 = r_2 + c$, where σ^0 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 6.16. Maps $f, g : X \to Y$ are *coarsely homotopic* if for all r_1, r_2 so 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

$$f, g : \mathrm{Rips}_{r_1} \to \mathrm{Rips}_{r_3}(Y)$$

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 :

$Y \to X$ so 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 6.15.

COROLLARY 6.17. *Let $f, g : X \to Y$ be L–Lipschitz maps within finite distance from each other. Then they are coarsely homotopic.*

COROLLARY 6.18. *If $f : X \to Y$ is a quasi-isometry, 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 6.19. *Coarse k-connectedness is a QI invariant.*

PROOF. Suppose that X' is coarsely k-connected and $f : X \to X'$ is an L–Lipschitz quasi-isometry with L–Lipschitz quasi-inverse $\bar{f} : X' \to X$. 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. $\square$

6.3. Metric cell complexes

We now introduce a concept which generalizes simplicial complexes, where the notion of bounded geometry does not imply finite-dimensionality.

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 $X^{(1)}$ us a graph, and, hence, can be equipped with the standard metric dist. Then the map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$ in general need not be a quasi-isometry. 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 quasi-isometry.

Even though this assumption could be avoided in what follows, restricting to complexes satisfying this axiom will make our discussion more intuitive.

Our first goal 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 $v \in X^{(0)}$ we set $c(v) = \{v\}$. Suppose that c is defined on $X^{(i)}$. For each $i+1$-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 6.20. 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*. In particular, a metric cell complex is a special case of a *metric chain complex* defined in [**KK05**].

We now say that the *diameter* $\mathrm{diam}(\sigma)$ of a cell σ in X is the diameter of $c(\sigma)$.

EXAMPLE 6.21. Take a 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 6.22. A metric cell complex X is said to have *bounded geometry* if there exists a collections of increasing functions $\phi_k(r)$ and numbers $D_k < \infty$ so 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 $D_k = D_{k,X}$, $k = 1, 2, 3, \ldots..$

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

$$D_X = \sup_{k>0} D_{k,X}. \tag{6.1}$$

EXERCISE 6.23. 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 Section 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 $x_0 = u, x_1, \ldots, x_m = v$, so that $d(x_i, x_{i+1}) \leqslant D_1$ for every i. In particular, if X is connected, the identity map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$ is D_1-Lipschitz.

3. Let $X^{(0)} := G$ be a finitely-generated group with its word metric, X be the Cayley graph of G. Then X is a metric cell complex of bounded geometry.

As a trivial example, consider spheres S^n with the usual cell structure (single 0-cell and single n-cell). Thus, the cellular embeddings $S^n \hookrightarrow S^{n+1}$ give rise to an infinite-dimensional cell complex S^∞. This complex has bounded geometry (since it has only one cell in every dimension). Therefore, the concept of metric cell complexes is more flexible than the one of simplicial complexes.

EXERCISE 6.24. 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 the union of the cells σ in X so 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 $S \subset X$ so that for every $\sigma \in S$, there exists a vertex $\tau \in W$ so 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, $\mathrm{diam}(f(\sigma)) \leqslant L$. In particular, $f : (X^{(0)}, d) \to (Y^{(0)}, d)$ is L/D_0-Lipschitz as a map of metric spaces.

EXERCISE 6.25. 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 6.26. Construct examples of a cellular map $f : X \to Y$ between metric graphs of bounded geometry, so that the restriction $f|X^{(0)}$ is L–Lipschitz but f is not L'–Lipschitz, for any $L' < \infty$.

A map $f : X \to Y$ of metric cell complexes is called *uniformly proper* if f is cellular, L-Lipschitz for some $L < \infty$ and $f|X^{(0)}$ is uniformly proper: There exists a proper monotonically increasing function $\eta(R)$ so that

$$\eta(d(x, x')) \leqslant d(f(x), f(x'))$$

for all $x, x' \in X$. The function $\eta(R)$ is called the *distortion function* of f.

We will now relate metric cell complexes of bounded geometry to simplicial complexes of bounded geometry:

EXERCISE 6.27. 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 quasi-isometry 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. $f : X^{(0)} \to Y^{(0)}, \bar{f} : Y^{(0)} \to X^{(0)}$ are quasi-inverse to each other:

$$d(f \circ \bar{f}, Id) \leqslant A, \quad d(\bar{f} \circ f, Id) \leqslant A.$$

Hint: Apply the usual construction which converts a finite-dimensional CW-complex to a simplicial complex.

Recall that quasi-isometries are not necessarily continuous. In order to use algebraic topology, we, thus, have to approximate quasi-isometries by cellular maps in the context of metric cell complexes. In general, this 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.

DEFINITION 6.28. A metric cell complex X is said to be *uniformly contractible* if there exists a continuous function $\psi(R)$ so 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 $\psi_k(R)$ so that for every $x \in X^{(0)}$ the map

$$\mathbf{B}(x, R) \to \mathbf{B}(x, \psi_k(R))$$

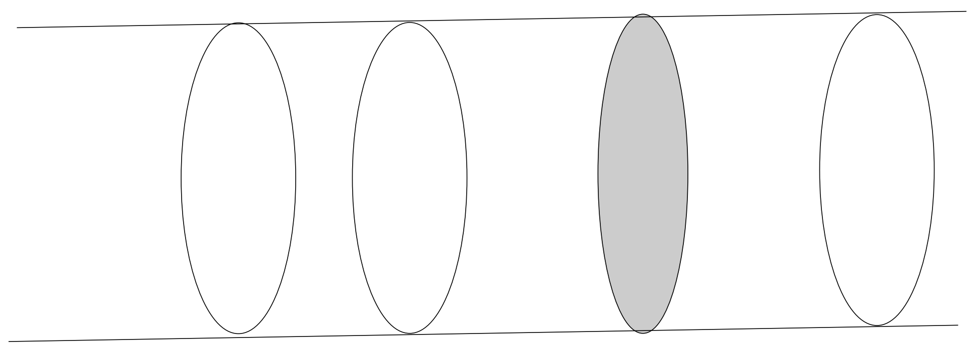
induces trivial map on π_i, $0 \leqslant i \leqslant k$.

We will refer to ψ, ψ_k as the *contractibility functions* of X.

EXAMPLE 6.29. 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 D 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 6.30. Show that Y is not uniformly contractible.



Y

FIGURE 6.2. Contructible but not uniformly constructible space.

We will see, nevertheless, in Lemma 6.34, that under certain assumptions (presence of a cocompact group action) contractibility implies uniform contractibility.

The following proposition is a metric analogue of the cellular approximation theorem:

PROPOSITION 6.31. *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 $g : X \to Y$, 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, $g(X) \subset \bar{\mathcal{N}}_{L'}(f(X^{(0)}))$.*

PROOF. The proof of this proposition is a prototype of most of the proofs which appear in this chapter. 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 $f_k : X^{(k)} \to Y^{(k)}$ so that

1. $\operatorname{diam}(f(\sigma)) \leqslant C_k$ for every k-cell σ.
2. $\operatorname{diam}(f(\partial\tau)) \leqslant C'_{k+1}$ for every $(k+1)$–cell τ in X.

Base of 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 $\operatorname{dist}(f(x), f(x')) \leqslant LD_1$, where $D_1 = D_{1,X}$ is the upper bound on the diameter 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 σ be a $(k+1)$–cell in X. Note that

$$\operatorname{diam}(f(\partial\sigma)) \leqslant C'_{k+1}$$

by the induction hypothesis. Then, using uniform contractibility of Y, we extend f to σ so that $\operatorname{diam}(f(\sigma)) \leqslant C_{k+1}$ where $C_{k+1} = \psi(C'_k)$. Let us verify that the extension $f : X^{k+1} \to Y^{k+1}$ 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

$$L' := \operatorname{diam}(f(\sigma)) \leqslant \max_{i=1,\dots,n} C_i.$$

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 the above proposition can be *relativized*:

LEMMA 6.32. *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 and geometric bounds on X and contractibility function of Y.*

PROOF. The proof is the same induction on skeleta argument as in Proposition 6.31. □

COROLLARY 6.33. *Suppose that X, Y are as above and $f_0, f_1 : X \to Y$ are L–Lipschitz cellular maps so that $\operatorname{dist}(f_0, f_1) \leqslant C$ in the sense that $d(f_0(x), f_1(x)) \leqslant C$ for all $x \in X^{(0)}$. 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 6.24). Then the required extension $f : W \to Y$ of this map exists by Lemma 6.32. □

6.4. Connectivity and coarse connectivity

Our next goal is to find a large supply of examples of metric spaces which are coarsely k-connected.

LEMMA 6.34. *If X is a finite-dimensional m-connected complex which admits a geometric (properly discontinuous cocompact) cellular group action $G \curvearrowright X$, then X is uniformly m-connected.*

PROOF. Existence of geometric action $G \curvearrowright X$ implies that X is locally finite. Pick a base-vertex $x \in X$ and let $r < \infty$ be such that G-orbit of $B(x,r) \cap X^{(0)}$ is the entire $X^{(0)}$. Therefore, if $C \subset X$ has diameter $\leqslant R/2$, there exists $g \in G$ so that $C' = g(C) \subset \mathbf{B}(x, r+R)$.

Since C is finite, $\pi_1(C')$ is finitely-generated. Thus, simple connectivity of X implies that there exists a finite subcomplex $C'' \subset X$ so 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 so that the map $H_i(C') \to H_i(C'')$ is zero. To summarize, there exists a finite subcomplex C'' in X containing C', so 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 $R' < \infty$ be such

$$C'' \subset \mathbf{B}(x, r + R + R').$$

Hence, the map

$$\pi_k(\mathbf{B}(x, r+R)) \to \pi_k(\mathbf{B}(x, r + R + R'))$$

is trivial for all $k \leqslant m$. Set $\psi(k, r) = \rho = r + R'$. Therefore, if $C \subset X$ is a subcomplex of diameter $\leqslant R/2$, then maps

$$\pi_k(C) \to \pi_k(\mathcal{N}_\rho(C))$$

are trivial for all $k \leqslant m$. □

THEOREM 6.35. *Suppose that X is a uniformly n-connected metric cell complex of bounded geometry. Then $Z := X^{(0)}$ is coarsely n-connected.*

PROOF. Let $\gamma : S^k \to \mathrm{Rips}_R(Z)$ 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 S^k is given a triangulation τ so that γ sends simplices of S^k to simplices in $\mathrm{Rips}_R(Z)$ and the restriction of γ to each simplex is a linear map.

LEMMA 6.36. *There exists a cellular map $\gamma' : (S^k, \tau) \to X$ which agrees with γ on the vertex set of τ and so that $\mathrm{diam}(\gamma'(S^k)) \leqslant R'$, where $R' \geqslant R$ depends only on R and contractibility functions $\psi_i(k, \cdot)$ of X, $i = 0, ..., k$.*

PROOF. We construct γ' by induction on skeleta of (S^k, τ). 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. Suppose we constructed γ' on i-skeleton τ^i of τ so that $\mathrm{diam}(\gamma'(\tau^i)) \leqslant R_i = R_i(R, \psi(k, \cdot))$. 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, using uniform contractibility of X we extend $\geq'$ to σ, so that the resulting map satisfies

$$\mathrm{diam}(\gamma'(\sigma)) \leqslant \psi(i+1, R_i),$$

which implies that the image is contained in $\mathbf{B}(\gamma(v), 2\psi(i+1, R_i))$, where v is a vertex of σ. Thus,

$$\mathrm{diam}(\gamma'(\tau^{i+1}) \leqslant R_{i+1} := R + \psi(i+1, R_i).$$

Now, lemma follows by induction. Figure 6.3 illustrates the proof in the case $k = 1$. □

Since X is k-connected, the map γ' extends to a cellular map $\gamma' : D^{k+1} \to X^{(k+1)}$, where D^{k+1} is a triangulated disk whose triangulation τ extends the triangulation τ of S^k. Our next goal is to "push" γ' to a map $\gamma'' : D^{k+1} \to \mathrm{Rips}_{R'}(Z)$ relative to the boundary, where we want $\gamma''|S^k$. Let σ be a simplex D^{k+1}. A simplicial map is determined by images of vertices. By definition of the number R',

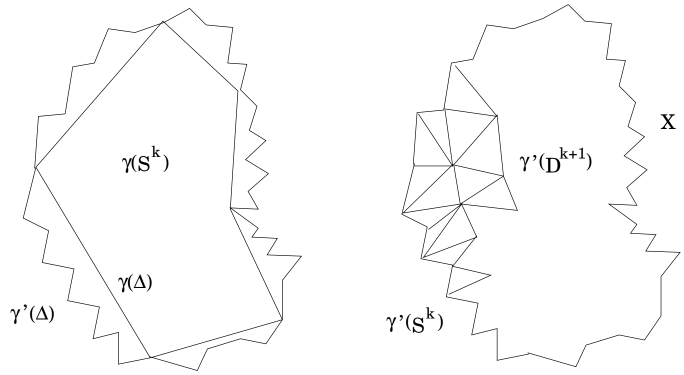


FIGURE 6.3.

images of vertices of σ under γ' are within distance $\leqslant R'$ from each other. Therefore, we have a canonical extension of $\gamma'|\sigma^{(0)}$ to a map $\sigma \to \mathrm{Rips}_{R'}(Z)$. If $\sigma_1 \subset \sigma_2$, then $\gamma'' : \sigma_1 \to \mathrm{Rips}_{R'}(Z)$ agrees with the restriction of $\gamma'' : \sigma_2 \to \mathrm{Rips}_{R'}(Z)$, since maps are determined by their vertex values. We thus obtain a simplicial map $D^{k+1} \to \mathrm{Rips}_{R'}(Z)$ which, by construction of γ' and γ'', agrees with γ on the boundary sphere.

Thus, the inclusion map $\mathrm{Rips}_R(Z) \to \mathrm{Rips}_{R'}(Z)$ induces trivial maps on k-th homotopy groups, $0 \leqslant k \leqslant n$. □

As a simple illustration of this theorem, consider the case $n = 0$.

COROLLARY 6.37. *If a bounded geometry metric cell complex X is connected, then X is quasi-isometric to a connected metric graph (with the standard metric).*

PROOF. By connectivity of X, for every pair of vertices $x, y \in Z$, there exists a path $\mathfrak{p}$ in X connecting x to y, so that $\mathfrak{p}$ is a concatenation of 1-cells in X. Since X has bounded geometry diameter of each 1-cell is $\leqslant R = D_1$, where D_1 is a geometric bound on X as in Definition 6.22. Therefore, consecutive vertices of X which appear in $\mathfrak{p}$ are within distance $\leqslant R$ from each other. It follows that $\Gamma = \mathrm{Rips}_R(Z)$ is connected. Without loss of generality, we may assume that $R \geqslant 1$. Then the map $\iota : Z \to \mathrm{Rips}_R(Z)$ (sending Z to the vertex set of the Rips complex) is 1-Lipschitz. It is also clear that this map is a R^{-1}-quasi-isometric embedding. Thus, ι is an $(R, 1)$-quasi-isometry. □

We saw, so far, how to go from uniform k-connectivity of a metric cell complex X to coarse k-connectivity of its 0-skeleton. Our main goal now is to go in the opposite direction. This, of course, may require modifying the complex X. The simplest instance of the "inverse" relation is

EXERCISE 6.38. Suppose that Z is a coarsely connected uniformly discrete metric space. Then Z is the 0-skeleton of a connected metric graph Γ of bounded geometry so that the inclusion map is a quasi-isometry. Hint: Γ is the 1-skeleton of a connected Rips complex $\mathrm{Rips}_R(Z)$. Bounded valence property comes from the uniform discreteness assumption on Z.

Below we consider the situation $k \geqslant 1$ in the group-theoretic context, starting with $k = 1$.

Lemma 6.39. *Let G be a finitely-generated group with word metric. Then G is coarsely simply-connected if and only if* $\mathrm{Rips}_R(G)$ *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 $\mathrm{Rips}_R(G)$ is simply-connected for all sufficiently large R. Our argument is similar to the proof of Theorem 6.35. Note that 1-skeleton of $\mathrm{Rips}_1(G)$ is just the Cayley graph of G. Using coarse simple connectivity of G, we find $D \geqslant 1$ such that the map

$$\pi_1(\mathrm{Rips}_1(G)) \to \pi_1(\mathrm{Rips}_D(G))$$

is trivial. We claim that for all $R \geqslant D$ the Rips complex $\mathrm{Rips}_R(G)$ is simply-connected. Let $\gamma \subset \mathrm{Rips}_R(G)$ be a polygonal loop. For every edge $\gamma_i := [x_i, x_{i+1}]$ of γ we let $\gamma_i' \subset \mathrm{Rips}_1(X)$ denote a geodesic path from x_i to x_{i+1}. Then, by the triangle inequality, γ_i' has length $\leqslant R$. Therefore, all the vertices of γ_i' are contained in the ball $\mathbf{B}(x_i, R) \subset G$ and, hence, they span a simplex in $\mathrm{Rips}_R(G)$. Thus, the paths γ_i, γ_i' are homotopic in $\mathrm{Rips}_R(G)$ rel. their end-points. Let γ' denote the loop in $\mathrm{Rips}_1(G)$ which is the concatenation of the paths γ_i'. Then, by the above observation, γ' is freely homotopic to γ in $\mathrm{Rips}_R(G)$. On the other hand, γ' is null-homotopic in $\mathrm{Rips}_R(G)$ since the map

$$\pi_1(\mathrm{Rips}_1(G)) \to \pi_1(\mathrm{Rips}_R(G))$$

is trivial. We conclude that γ is null-homotopic in $\mathrm{Rips}_R(G)$ as well. □

Corollary 6.40. *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 4.80). Then the universal cover X of Y is simply-connected. Hence, by Lemma 6.34, X is uniformly simply-connected and hence by Theorem 6.35, the group G is coarsely simply-connected.

Conversely, suppose that G is coarsely simply-connected. Then, by Lemma 6.39, 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 3.28, G admits a properly discontinuous, free cocompact action on another simply-connected cell complex Z. Therefore, G is finitely-presented. □

We now proceed to $k \geqslant 2$. Recall (see Definition 3.26) 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.

Example 6.41 (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 direct product of $\mathbb{F}_2$ with itself n times. Define a homomorphism $\phi : G \to \mathbb{Z}$ which sends each generator a_i, b_i of G to the same generator of $\mathbb{Z}$. Let $K := Ker(\phi)$. Then K is of type $\mathbf{F}_{n-1}$ but not of type $\mathbf{F}_n$.

Analogously to Corollary 6.40 we obtain:

Theorem 6.42 (See 1.C2 in [**Gro93**]). *Type $\mathbf{F}_n$ is a QI invariant.*

Proof. Our argument is similar to the proof of Corollary 6.40, 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 (possibly infinite-dimensional) $n-1$-connected cell complex X so that the quotient of each skeleton is a finite complex. By combining Lemma 6.34 and Theorem 6.35, we see that the group G is coarsely $n-1$-connected. It remains, then to prove

Proposition 6.43. *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 6.40). The proof below follows [**KK05**].

Our goal is to build the complex X on which G would act 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 $f: X^{(i)} \to Y_{R_i} = \mathrm{Rips}_{R_i}(G)$ and equivariant "deformation retractions" $\rho_i : Y^{(i)}_{R_i} \to X^{(i)}, i = 0, ..., n$, which are G-equivariant cellular maps so that composition $h_i = \rho_i \circ f_i : X^{(i)} \to X^{(i)}$ is homotopic to the identity for $i = 0, ..., n-1$. We first explain the construction in the case when G is torsion-free and then show how to modify the construction for groups with torsion.

Torsion-free case. In this case G-action on every Rips complex is free and cocompact. The construction is by induction on i.

$i=0$. We let $X^{(0)} = G, R_0 = 0$ and let $f_0 = \rho_0 : G \to G$ be the identity map.

$i=1$. We let $R_1 = 1$ and let $X_1 = Y^{(1)}_{R_1}$ be the Cayley graph of G. Again $f_1 = \rho_1 = Id$.

$i=2$. According to Lemma 6.39, there exists R_2 so that Y_R is simply-connected for all $R \geqslant R_2$. We then take $X_2 := Y^{(2)}_{R_2}$. Again, we let $f_2 = \rho_2 = Id$.

$i \Rightarrow i+1$. Suppose now that $3 \leqslant i \leqslant n-1$, $X^{(i)}, f_i, \rho_i$ are constructed and R_i chosen; we will construct $X^{(i+1)}, f_{i+1}, \rho_{i+1}$.

We first construct $X^{(i+1)}$.

Lemma 6.44. *There are finitely many spherical i-cycles $\sigma_\alpha, \alpha \in A'$, in $X^{(i)}$ such that their G-orbits generate $\pi_i(X^{(i)})$.*

Proof. Let $R' > R = R_i$ be such that the map

$$Y_R = \mathrm{Rips}_R(G) \to Y_{R'} = \mathrm{Rips}_{R'}(G)$$

induces zero map on π_k, $k = 0, ..., i$. Let $\tau_\alpha : S^i \to (Y_R)^{(i)}, \alpha \in A$, denote the attaching maps of the $i+1$-cells $\hat{\tau}_\alpha$ in $Y^{(i+1)}_{R'}$, these maps are just simplicial homeomorphic embeddings from the boundary S^i of the standard $i+1$-simplex to the boundaries of the $i+1$-simplices in $Y^{(i)}_{R'}$. Since the map $H_i(Y_R) \to H_i(Y_{R+1})$ is zero, the spherical cycles $\tau_\alpha, \alpha \in A$, generate the image of the map

$$\eta_i : H_i(Y^{(i)}_R) \to H_i(Y^{(i)}_{R'}).$$

Since the action of G on Y_R is cocompact, there are finitely many of these spherical cycles $\{\tau_\alpha, \alpha \in A'\}$, whose G-images generate the entire image of η_i. We then let

$\sigma_\alpha := \rho_i(\tau_\alpha), \alpha \in A'$. We claim that this finite set of spherical cycles does the job. Note that for every $\sigma \in \pi_i(X^{(i)})$,

$$[f(\sigma)] = \sum_{\alpha \in A'} \sum_{g \in G} z_{g,\alpha} \cdot g([\tau_\alpha)]), \quad g \in G, z_{g,\alpha} \in \mathbb{Z},$$

in the group $H_i(Y_{R'})$. Applying the retraction ρ_i and using the fact that $h_i = \rho_i \circ f_i$ is homotopic to the identity, we get

$$\sigma = \sum_{\alpha \in A'} \sum_{g \in G} z_{g,\alpha} \cdot g([\sigma_\alpha)]). \quad \square$$

We now equivariantly attach $i+1$-cells $\hat{\sigma}_{g,\alpha}$ along the spherical cycles $g(\sigma_\alpha), \alpha \in A'$. We let $X^{(i+1)}$ denote the resulting complex and we extend the G-action to $X^{(i+1)}$ in obvious fashion. It is clear that $G \curvearrowright X^{(i+1)}$ is properly discontinuous, free and cocompact. By the construction $X^{(i+1)}$ is i-connected.

We next construct maps f_{i+1} and ρ_{i+1}. To construct the map $f_{i+1} : X^{(i+1)} \to Y_{R'}$ we extend $f_i|\sigma_{1,\alpha}$ to $\hat{\sigma}_{1,\alpha}$ using the fact that the map

$$\pi_i(Y_R) \to \pi_i(Y_{R'})$$

is trivial. We extend f_{i+1} to the rest of the cells $\hat{\sigma}_{g,\alpha}, \alpha \in A'$, by G-equivariance. We extend ρ_i to each $g\hat{\tau}_\alpha$ using the attaching map $g\hat{\sigma}_\alpha$. We extend the map to the rest of $Y_{R'}^{(i+1)}$ by induction on the skeleta, G-equivariance and using the fact that $X^{(i+1)}$ is i-connected. Lastly, we observe that h_{i+1} is homotopic to the identity. Indeed, for each $i+1$-cell $g(\hat{\sigma}_\alpha)$, the map $f_i(g\sigma_\alpha)$ is homotopic to $g\tau_\alpha$ in $Y_{R'}$ (as $\pi_i(Y_R) \to \pi_i(Y_{R'})$ is zero) and $f_{i+1}(g\hat{\tau}_\alpha) = g(\hat{\sigma}_\alpha)$. (Note that we do not claim that h_n is homotopic to the identity.)

If $n < \infty$, this construction terminates after finitely many steps, otherwise, it takes infinitely many steps. In either case, the result is $n-1$-connected complex X and a free action $G \curvearrowright X$ which is cocompact on each skeleton. This concludes the proof in the case of torsion-free groups G.

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 would still exist, we would be unable to construct equivariant maps $\rho_i : \mathrm{Rips}_R(G) \to X^{(i)}$. 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 were to have f_i and ρ_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 Z_R denote the barycentric subdivision of $Y_R^{(i)} = \mathrm{Rips}_R(G)^{(i)}$. Then G acts on Z_R without inversions (see Definition 3.22). Let $\widehat{Z_R}$ denote the regular cell complex obtained by applying the *Borel construction* to Z_R, see section 3.2. The complex $\widehat{Z_R}$ is infinite-dimensional if G has torsion, but this does not cause problems since at each step of induction we work only with finite skeleta. The action $G \curvearrowright Z_R$ lifts to a free (properly discontinuous) action $G \curvearrowright \widehat{Z_R}$ which is cocompact on each skeleton. We then can apply the arguments from the torsion-free case to the complexes $\widehat{Z_R}$ instead of $\mathrm{Rips}_R(G)$. The key is that, since the action of G on $\widehat{Z_R}$ is free, the construction of the equivariant retractions $\rho_i : Y_{R_i}^{(i)} \to X^{(i)}$ 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 $\widehat{Z_R} \to Z_R$ is homotopy-equivalence, 2-skeleton of $\widehat{Z_R}$ is simply-connected for the same values of R. $\square$

This finishes the proof of Theorem 6.42 as well. $\square$

There are other group-theoretic finiteness conditions, for instance, the condition $\mathbf{F}P_n$ which is a cohomological analogue of the finiteness condition $\mathbf{F}_n$. 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 R.

THEOREM 6.45. *Let R be a commutative ring with neutral element. Then the property of being $\mathbf{FP}_n$ over R is QI invariant.*

QUESTION 6.46. 1. Is the homological dimension of a group QI 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 geometric dimension QI invariant for torsion-free groups?

Note that cohomological dimension is known to equal geometric dimension, except there could be groups satisfying

$$2 = cd(G) \le gd(G) \le 3,$$

see [**Bro82b**]. On the other hand,

$$cd(G) \le hd(G) \le cd(G) + 1,$$

see [**Bie76a**]. Here cd stands for cohomological dimension, gd is the geometric dimension and hd is the homological dimension.

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 6.43 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 6.47. *Under the above assumptions, there exist numbers L, L', A, function $R = R(r)$ which depend only on the distortion function of f and on the geometry of X and Y so 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 $\operatorname{dist}(\rho \circ f, id_X) \leqslant A$. *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 $id : V \to \mathcal{N}_R(V)$.

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 $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ so that

$$\eta(d(x, x')) \leqslant d(f(x), f(x')), \forall x, x' \in X^{(0)}.$$

Let A_0, A_1 denote numbers such that

$$\eta(t) > 0, \quad \forall t > A_0,$$

$$\eta(t) > 2r + D_1, \quad \forall t > A_1,$$

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 $\operatorname{dist}(y, f(x))$ 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

$$\operatorname{dist}(\rho \circ f, id_{X^{(0)}}) \leqslant A := A_0.$$

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, $\operatorname{diam}(\rho(\partial\sigma)) \leqslant A_1$. Indeed, if $\partial\sigma = \{y_1, y_2\}$, $d(y_1, y_2) \leqslant D_1$ 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 6.31. This proves (1).

Part (2) follows from Corollary 6.33. To prove Part (3), observe that $h = f \circ \rho : \bar{\mathcal{N}}_r(V) \to V$ is L''-Lipschitz (see Exercise 6.25), $\operatorname{dist}(h, Id) \leqslant r$. Now, (3) follows from Corollary 6.33 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 6.32. □

Corollary 6.48. *There exists a function $\alpha(r) \geqslant r$ so 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 6.49. *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 6.48 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. $\square$

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 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 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 6.50. 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.

We will present in this section two applications of Poincaré duality to the coarse topology of X.

Coarse surjectivity

THEOREM 6.51. *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 6.50. Let $\rho : V \to X$ be a retraction constructed in Lemma 6.47. By Lemma 6.47, 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. $\square$

COROLLARY 6.52. *Let X, Y be as above an $f : X^{(0)} \to Y^{(0)}$ is a quasi-isometric embedding. Then f is a quasi-isometry.*

PROOF. Combine Proposition 6.31 with Theorem 6.51. $\square$

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*, $inrad(C)$, of C to be the supremum of radii of metric balls $\mathbf{B}(x, R)$ in X contained in C. A component C is called *shallow* if $inrad(C)$ is $< \infty$ and *deep* if $inrad(C) = \infty$.

EXAMPLE 6.53. Suppose that W is compact. Then deep complementary components of C_R are components of infinite diameter. These are the components which appears as neigborhoods 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 6.54. The simple properly embedded curve Γ in $\mathbb{R}^2$ need not coarsely separate $\mathbb{R}^2$ (see Figure 6.4). A straight line in $\mathbb{R}^2$ coarsely separates $\mathbb{R}^2$.

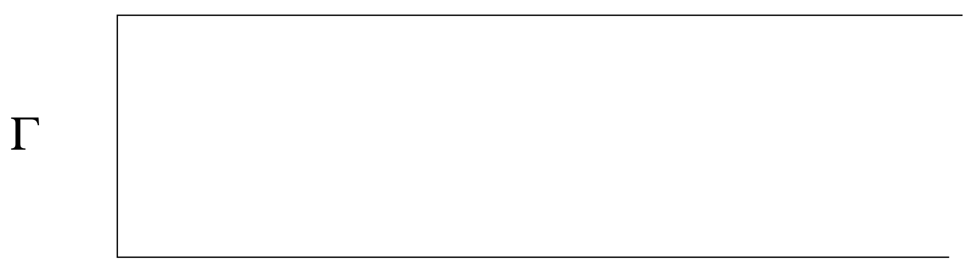


FIGURE 6.4. A separating curve which does not coarsely separate the plane.

THEOREM 6.55. *Suppose that X, Y are uniformly contractible simplicial complexes of bounded geometry which are homeomorphic to $\mathbb{R}^{n-1}$ and $\mathbb{R}^n$ respectively. Then for each uniformly proper cellular 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 $H_c^{n-1}(X) \neq 0$.

Recall that in Section 6.5 we constructed a system of retractions $\rho_R : \mathcal{N}_R(V) \to X$, $R \in \mathbb{N}$, and proper homotopy-equivalences $f \circ \rho \equiv Id$ and $\rho_R \circ f|_{\mathcal{N}_R(V)} \equiv Id : \mathcal{N}_R(V) \to \mathcal{N}_{\alpha(R)}(V)$. 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 6.47. 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:

$$\begin{tikzcd}
H_c^*(\bar{\mathcal{N}}_{R_2}(V)) \arrow[r, "P"] \arrow[d, "rest_{R_1,R_2}"'] & H_{n-*}(Y, C_{R_2}) \arrow[d, "proj_{R_1,R_2}"] \\
H_c^*(\bar{\mathcal{N}}_{R_1}(V)) \arrow[r, "P"] & H_{n-*}(Y, C_{R_1})
\end{tikzcd}$$

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 6.5. Observe that for every r, ω_r is non-zero, since $f^* \circ \rho^* = id$ 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. Thus, each 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

$$C'_R \subset C'_r, \quad C''_R \subset C''_r.$$

Since this could be done for arbitrarily large r, R, we conclude that components C'_r, C''_r are both deep.

The same argument run in the reverse implies that there are exactly two deep complementary components. □

We refer to [**FS96**], [**KK05**] for further discussion and generalization of coarse separation and coarse Poincaré/Alexander duality.

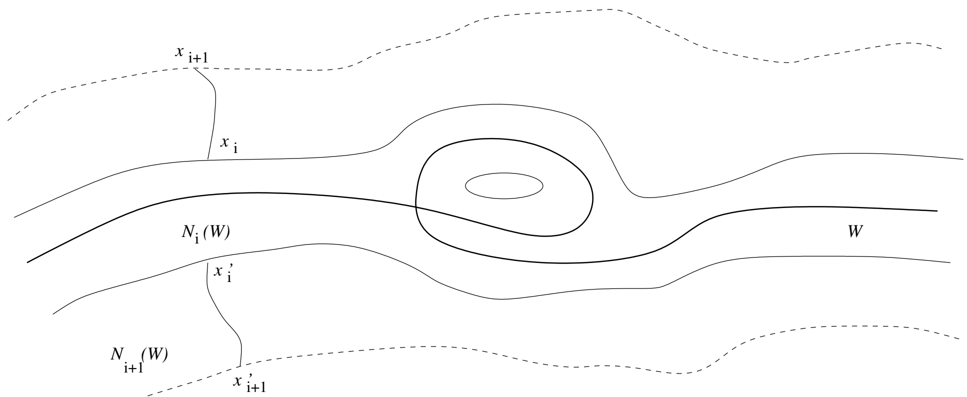


FIGURE 6.5. Coarse separation.

CHAPTER 7

Hyperbolic Space

The real hyperbolic space is the oldest and easiest example of hyperbolic space. A good reference for hyperbolic spaces in general is [**?**]. The real-hyperbolic space has its origin in the following classical question that has challenged the geometers for nearly 2000 years:

QUESTION 7.1. 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.

7.1. Moebius transformations

We will think of the sphere S^n as the 1-point compactification of $\mathbb{R}^n$. Accordingly, we will regard the 1-point compactification of a hyperplane in $\mathbb{R}^n$ as a *round sphere* (of infinite radius) and the 1-points compactification of a line in $\mathbb{R}^n$ as a *round circle* (of infinite radius). Recall that the *inversion* in the r-sphere $\Sigma_r = \{x : \|x\| = r\}$ is the map

$$J_\Sigma : x \mapsto \frac{r^2 x}{\|x\|^2}, \quad J_\Sigma(0) = \infty, \quad J_\Sigma(\infty) = 0.$$

One defines the inversion J_Σ in the sphere $\Sigma = \{x : \|x - a\| = r\}$ by the formula

$$T_a \circ J_{\Sigma_r} \circ T_{-a}$$

where T_a is the translation by the vector a. Inversions map round spheres to round spheres and round circles to circles; inversions also preserve the Euclidean angles, and the *cross-ratio*

$$[x, y, z, w] := \frac{|x - y|}{|y - z|} \cdot \frac{|z - w|}{|w - x|},$$

see e.g. [**Rat94**, Theorem 4.3.1]. We will regard the reflection in a Euclidean hyperplane as an inversion (such inversion fixes ∞).

Definition 7.2. A *Moebius transformation* of $\mathbb{R}^n$ (or, rather, S^n) is a composition of finitely many inversions in $\mathbb{R}^n$. The group of all Moebius transformations of $\mathbb{R}^n$ is denoted $Mob(\mathbb{R}^n)$ or $Mob(S^n)$.

In particular, Moebius transformations preserve angles, cross-ratios and map 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 $\mathbb{R}^n$ is the composition of at most n inversions (reflections), since every rotation in $\mathbb{R}^2$ is the composition of two reflections. Every dilation $x \mapsto \lambda x, \lambda > 0$ is the composition of two inversions in spheres centered at 0.

Lemma 7.3. *The subgroup $Mob_{\infty,0}(\mathbb{R}^n)$ of $Mob(\mathbb{R}^n)$ fixing ∞ and 0 equals the group $CO(n) = \mathbb{R}_+ \cdot O(n)$.*

Proof. We just observed that $CO(n)$ is contained in $Mob_{\infty,0}(\mathbb{R}^n)$. We, thus, need to prove the opposite inclusion. Consider the coordinate lines $L_1, ..., L_n$ in $\mathbb{R}^n$. Then every $g \in Mob_\infty(\mathbb{R}^n)$ sends these lines to pairwise orthogonal lines $L_1', ..., L_n'$ through the origin (since Moebius transformations map circles to circles and preserve angles). By postcomposing g with an element of $O(n)$, we can assume that g preserves each coordinate line L_n and, furthermore, preserves the orientation on this line. By postcomposing g with dilation we can also assume that g maps the unit vector e_1 to itself. Thus, g maps the unit sphere Σ_1 to the round sphere which is orthogonal to the coordinate lines and passes through the point e_1. Hence, $d(\Sigma_1) = \Sigma_1$. We claim that such g is the identity. Indeed, if L is a line through the origin, then the line $g(L)$ has the same angles with L_i as L for each $i = 1, ..., n$. Thus, $g(L) = L$ for every such L. By considering intersections of these lines with Σ_1, we conclude that g restricts to the identity on Σ_1. It remains to show that g is the identity on every sphere centered at the origin. Equivalently, we need to show that g is the identity on the line L_1.

Let $x \in L_1$ be outside of Σ_1 and let L be a line in the x_1x_2-plane through x and tangent to Σ_1 at a point y. Then $g(L)$ is also a line through $g(x), y$, tangent to Σ_1 at y. Since g preserves the orientation on L_1, $g(L) = L$ and, hence, $g(x) = x$. We leave the case of points $x \in L_1$ contained inside Σ_1 to the reader. $\square$

Example 7.4. Let us construct a Moebius transformation σ sending the unit ball $\mathbf{B}^n = B(0,1) \subset \mathbb{R}^n$ to the upper half-space $U^n = \mathbb{R}^n_+$,

$$\mathbb{R}^n_+ = \{(x_1, ...x_n) : x_n > 0\}.$$

We take σ to be the composition of translation $x \mapsto x + e_n$, where $e_n = (0, ..., 0, 1)$, inversion J_Σ, where $\Sigma = \partial \mathbf{B}^n$, translation $x \mapsto x - \frac{1}{2}e_n$ and, lastly, the similarity $x \to 2x$. 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 $-e_n$.

Note that the map σ sends the origin $0 \in \mathbf{B}^n$ to the point $e_n \in U^n$.

Low-dimensional Moebius transformations. Suppose now that $n = 2$. The group $SL(2, \mathbb{C})$ acts on the extended complex plane $S^2 = \mathbb{C} \cup \infty$ by *linear-fractional transformations*:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}. \tag{7.1}$$

Note that the matrix $-I$ lies in the kernel of this action, thus, the above 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 $S^2 = \mathbb{C} \cup \infty$ *via* the map $[z : w] \mapsto z/w$, the above action of $SL(2,\mathbb{C})$ 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 7.5. Show the group $PSL(2,\mathbb{C})$ acts faithfully on S^2.

EXERCISE 7.6. Prove that the subgroup $SL(2,\mathbb{R}) \subset SL(2,\mathbb{C})$ preserves the upper half-plane $U^2 = \{z : Im(z) > 0\}$. Moreover, $SL(2,\mathbb{R})$ is the stabilizer of U^2 in $SL(2,\mathbb{C})$.

EXERCISE 7.7. Prove that any 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 7.8. *$PSL(2,\mathbb{C})$ is the subgroup $Mob_+(S^2)$ of Moebius transformations of 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 7.7). Note that $j(z)$ is the composition of the complex conjugation with the inversion in the unit circle. Thus, $PSL(2,\mathbb{C}) \subset Mob_+(S^2)$. Conversely, let $g \in Mob(S^2)$ 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 Moebius transformation, it follows that g is also a Moebius transformation. $\square$

7.2. Real hyperbolic space

Upper half-space model. We equip $U^n = \mathbb{R}^n_+$ with the Riemannian metric

$$ds^2 = \frac{dx^2}{x_n^2} = \frac{dx_1^2 + ... + dx_n^2}{x_n^2} \tag{7.2}$$

The Riemannian manifold (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 $U^2 = \{z | Im(z) > 0\}$, $z = x + iy$, and

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

Note that the hyperbolic Riemannian metric ds^2 on 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 U^n given by the formula

$$\pi : (x_1, ..., x_n) \mapsto (0, ..., 0, x_n).$$

EXERCISE 7.9. 1. Verify that $d_x\pi$ does not increase the length of tangent vectors $v \in T_x\mathbb{H}^n$ for every $x \in \mathbb{H}^n$.

2. Verify that for a unit vector $v \in T_x\mathbb{H}^n$, $\|d_x\pi(v)\| = 1$ if and only if v is "vertical", i.e., it has the form $(0, ..., 0, v_n)$.

EXERCISE 7.10. Suppose that $p = ae_n, q = be_n$, where $0 < a < b$. Let α be the vertical path $\alpha(t) = (1 - t)p + tq$, $t \in [0, 1]$ connecting p to q. Show that α is the shortest path (with respect to the hyperbolic metric) connecting p to q in $\mathbb{H}^n$. In particular, α is a hyperbolic geodesic and

$$d(p, q) = \log(b/a).$$

Hint: Use the previous exercise.

We note that the metric ds^2 on $\mathbb{H}^n$ is clearly invariant under the "horizontal" Euclidean translations $x \mapsto x + v$, where $v = (v_1, ..., v_{n-1}, 0)$ (since they preserve the Euclidean metric and the x_n-coordinate). Similarly, ds^2 is invariant under the dilations

$$h : x \mapsto \lambda x, \lambda > 0$$

since h scales both numerator and denominator in (7.2) 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, compositions of such isometries of $\mathbb{H}^n$ act transitively on $\mathbb{H}^n$, which means that $\mathbb{H}^n$ is a *homogeneous Riemannian manifold*.

EXERCISE 7.11. 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 7.12. Show that the inversion $J = J_\Sigma$ in the unit sphere Σ centered at the origin, is an isometry of $\mathbb{H}^n$, i.e., $ds_B^2 = J^*(ds^2)$. The proof is easy but (somewhat) tedious calculation, which is best done using *calculus* interpretation of the pull-back Riemannian metric.

EXERCISE 7.13. 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 $\mathbf{B}^n := \{x : |x| < 1\}$ in $\mathbb{R}^n$. We equip $\mathbf{B}^n$ with the Riemannian metric

$$ds_B^2 = 4\frac{dx_1^2 + ... + dx_n^2}{(1 - |x|^2)^2}.$$

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_B^2 (since its elements preserve $|x|$ and, hence, the denominator of ds_B^2). The two models of the hyperbolic space are related by the Moebius transformation $\sigma : \mathbf{B}^n \to U^n$ defined in the previous section.

EXERCISE 7.14. Show that $ds_B^2 = \sigma^*(ds^2)$. The proof is again a straightforward calculation similar to the Exercise 7.12. Namely, first, pull-back ds^2 *via* dilatation $x \to 2x$, then apply pull-back *via* the translation $x \mapsto x - \frac{1}{2}e_n$, etc. Thus, σ is an isometry of the Riemannian manifolds $(\mathbf{B}^n, ds_B^2), (U^n, ds^2)$.

LEMMA 7.15. *The group $O(n)$ is the stabilizer of 0 in the group of isometries of $(\mathbf{B}^n, ds_B^2)$.*

PROOF. Note that if $g \in \mathrm{Isom}(\mathbf{B}^n)$ fixes 0, then its derivative at the origin dg_0 is an orthogonal transformation u. Thus, $h = u^{-1}g \in \mathrm{Isom}(\mathbf{B}^n)$ has the property $dh_0 = Id$. Therefore, for every geodesic γ in $\mathbb{H}^n$ so that $\gamma(0) = 0$, $dh(\gamma'(0)) = \gamma'(0)$. Since 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. Since $\mathbf{B}^n$ is complete, for every $q \in \mathbf{B}^n$ there exists a geodesic hyperbolic γ connecting p to q. Thus, $h(q) = q$ and, therefore, $g = u \in O(n)$. $\square$

COROLLARY 7.16. *The stabilizer of the point $p = e_n \in U^n$ in the group* $\mathrm{Isom}(\mathbb{H}^n)$ *is contained in the group of Moebius transformations.*

PROOF. Note that σ sends $0 \in B^n$ to $p = e_n \in U^n$, and σ is Moebius. Thus, $\sigma : \mathbf{B}^n \to U^n$ conjugates the stabilizer $O(n)$ of 0 in $\mathrm{Isom}(\mathbf{B}^n, ds_B^2)$ to the stabilizer $K = \sigma^{-1}O(n)\sigma$ of p in $\mathrm{Isom}(U^n, ds^2)$. Since $O(n) \subset Mob(S^n), \sigma \in Mob(S^n)$, claim follows. $\square$

COROLLARY 7.17. *a.* $\mathrm{Isom}(\mathbb{H}^n)$ *equals the group $Mob(\mathbb{H}^n)$ of Moebius transformations of 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 two models of $\mathbb{H}^n$ differ by a Moebius transformation, it suffices to work with 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 $h \in Mob(\mathbb{H}^n) \cap \mathrm{Isom}(\mathbb{H}^n)$ so 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 $Mob(\mathbb{H}^n) \to Mob(S^{n-1})$ is injective. Every $g \in Mob(S^{n-1})$ 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 $T_p\mathbb{H}^n$. $\square$

LEMMA 7.18. *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 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 $\sigma : \mathbf{B}^n \to U^n$ 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 U^n. Since $\mathrm{Isom}(\mathbb{H}^n)$ acts transitively on $U\mathbb{H}^n$, there exists a hyperbolic isometry g so that the hyperbolic geodesic $\beta = g \circ \alpha$ satisfies: $\beta(0) = p = e_n$ and the vector $\beta'(0)$ has the form $e_n = (0, ..., 0, 1)$. We already know that the curve $\gamma(t) = e^t e_n$ is a hyperbolic geodesic, see Exercise 7.10. Furthermore, $\gamma'(0) = e_n$ and $\gamma(0) = p$. Thus, $\beta = \gamma$ is a (generalized) 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. $\square$

COROLLARY 7.19. *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 7.10, the vertical segment is the unique length-minimizing path between such p and q. $\square$

COROLLARY 7.20. *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). $\square$

We will refer to the submanifolds $H \subset \mathbb{H}^n$ as *hyperbolic subspaces.*

EXERCISE 7.21. 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., disjoint from L).

EXERCISE 7.22. Prove that

- the unit sphere S^{n-1} is the ideal boundary (in the sense of Definition 2.44) of the hyperbolic space $\mathbb{H}^n$ in the unit ball model;

- the extended Euclidean space $\mathbb{R}^{n-1} \cup \{\infty\} = S^{n-1}$ is the ideal boundary of the hyperbolic space $\mathbb{H}^n$ in the upper half-space model.

Note that the Moebius transformation $\sigma : \mathbf{B}^n \to U^n$ carries the ideal boundary of $\mathbf{B}^n$ to the ideal boundary of U^n. Note 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 7.20 and 7.33 that $\mathbb{H}^n$ has sectional curvature -1, therefore all the considerations in Section 2.1.8, in particular those concerning the ideal boundary, apply to it. Later on, in Section 8.9 of Chapter 8, we will give another more intrinsic definition of ideal boundaries, for metric hyperbolic spaces in the sense of Gromov.

Lorentzian model of $\mathbb{H}^n$. We refer the reader to [**Rat94**] and [**Thu97**] for the material below.

Consider the *Lorentzian space* $\mathbb{R}^{n,1}$ which is $\mathbb{R}^{n+1}$ equipped with the quadratic form

$$q(x) = x_1^2 + \ldots + x_n^2 - x_{n+1}^2.$$

Let H denote the upper sheet of the 2-sheeted hyperboloid in $\mathbb{R}^{n,1}$:

$$x_1^2 + \ldots + x_n^2 - x_{n+1}^2 = -1, x_{n+1} > 0.$$

Restriction of q to the tangent bundle of H is positive-definite and defines a Riemannian metric ds^2 on H. We identify the unit ball $\mathbf{B}^n$ in $\mathbb{R}^n$ with the ball

$$\{(x_1, \ldots, x_n, 0) : x_1^+ \ldots + x_n^2 < 1\} \subset \mathbb{R}^{n+1}.$$

Let $\pi : H \to \mathbf{B}^n$ denote the radial projection from the point $-e_{n+1}$:

$$\pi(x) = tx - (1-t)e_{n+1}, \quad t = \frac{1}{x_{n+1}+1}.$$

One then verifies that

$$\pi : (H, ds^2) \to \mathbb{H}^n = \left(\mathbf{B}^n, \frac{4dx^2}{(1-|x|^2)^2}\right)$$

is an isometry.

The stabilizer $PO(n,1)$ of H in $O(n,1)$ acts isometrically on H. Furthermore, $PO(n,1)$ is the entire isometry group of (H, ds^2). Thus, $\mathrm{Isom}(\mathbb{H}^n) \cong PO(n,1) \subset SO(n,1)$; in particular, the Lie group $\mathrm{Isom}(\mathbb{H}^n)$ is linear.

7.3. Hyperbolic trigonometry

In this section we consider geometry of triangles in the hyperbolic plane. We refer to [**?, Rat94, 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 of $\mathbb{H}^2$. Such triangles are easiest to introduce by using 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 S^1 to be the same.*

The vertices of T which happen to be points of the boundary circle S^1 are called the *ideal vertices* of T. The *angle* of T at its ideal vertex is just the Euclidean angle. 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 $\mathbb{H}^n$.

EXERCISE 7.23. If A is an ideal vertex of a hyperbolic triangle T, then T has zero angle at A. Hint: It suffices to consider the case when $A = 0$ and the side $[A, B]$ of T is contained in the vertical line L. Show that the side $[A, C]$ of T is a circular arc tangent to L at A.

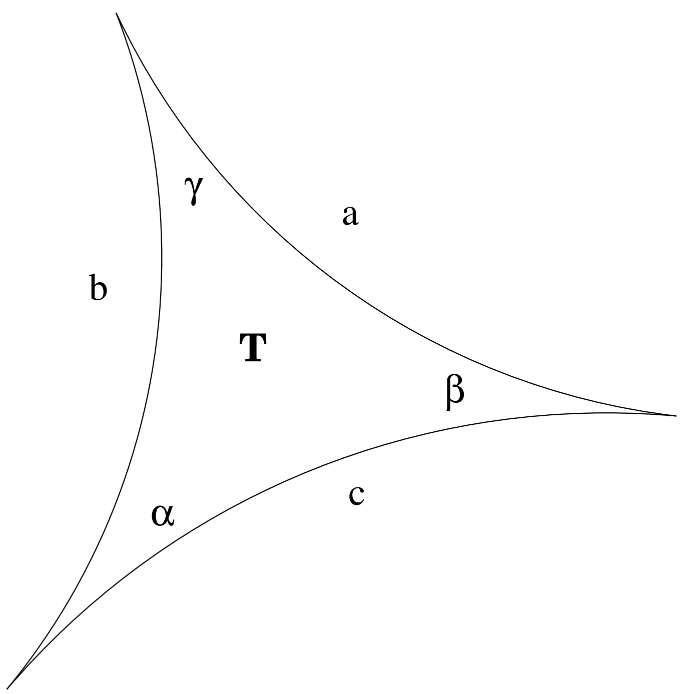


FIGURE 7.1. Geometry of a general hyperbolic triangle.

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 7.1.

a. Hyperbolic Sine Law:

$$\frac{\sinh(a)}{\sin(\alpha)} = \frac{\sinh(b)}{\sin(\beta)} = \frac{\sinh(c)}{\sin(\gamma)}. \tag{7.3}$$

b. Hyperbolic Cosine Law:

$$\cosh(c) = \cosh(a)\cosh(b) - \sinh(a)\sinh(b)\cos(\gamma) \tag{7.4}$$

c. Dual Hyperbolic Cosine Law:

$$\cos(\gamma) = -\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)\cosh(c) \tag{7.5}$$

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:

$$\cosh(c) = \cosh(a)\cosh(b), \tag{7.6}$$

$$\cos(\alpha) = \sin(\beta)\cosh(a), \tag{7.7}$$

$$\cos(\alpha) = \frac{\tanh b}{\tanh c} \tag{7.8}$$

In particular,

$$\cos(\alpha) = \frac{\cosh(a)\sinh(b)}{\sinh(c)}. \tag{7.9}$$

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 (7.6) and (7.4) we obtain the *First Variation Formula*:

$$c'(0) = \frac{\cosh(a)\sinh(b)}{\sinh(c)} b'(0) = \cos(\alpha)b'(0). \tag{7.10}$$

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 7.24. *$a + b - c \geqslant ma$, where*

$$m = \min\{|1 - \cos(\beta)|, |1 - \cos(\gamma)|\}.$$

PROOF. We let $g(t)$ denote the unit speed parameterizations of the segment $[BC]$, so that $g(0) = C, g(a) = B$. Let $c(t)$ denote the distance $\mathrm{dist}(A, g(t))$ (so 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,

$$a + b - c = f(a) \geqslant ma \quad \Box$$

EXERCISE 7.25. [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 $A = A_1 = A_2$, $[A_1, B_1] \subset [A_2, B_2]$, $\alpha_1 = \alpha_2$ and $\gamma_1 = \gamma_2 = \pi/2$. See Figure 7.2. Then $a_1 \leqslant a_2$. Hint: Use either (7.8).

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.

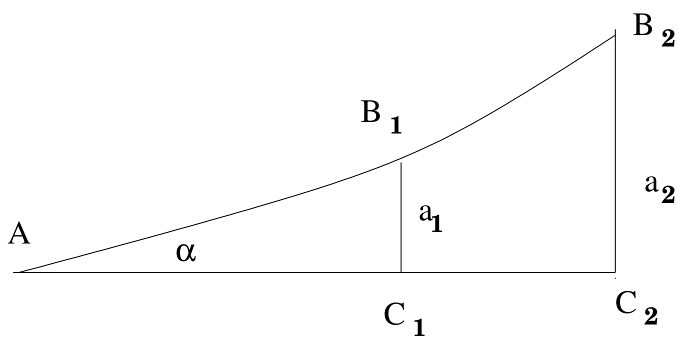


FIGURE 7.2. *Monotonicity of distance.*

7.4. 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 7.26. Angles of hyperbolic triangles are continuous functions on $Tri(\mathbb{H}^n)$.

EXERCISE 7.27. Every hyperbolic triangle T in $\mathbb{H}^n$ is contained in (the compactification of) a 2-dimensional hyperbolic subspace $H \subset \mathbb{H}^n$. 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 $\mathbb{H}^2$ in several (2, 3 or 4) convex regions, one of which has the property that its boundary is the triangle T. 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 $H \subset \mathbb{H}^n$ containing T. We will retain the notation $\blacktriangle$ for solid triangles in $\mathbb{H}^n$.

EXERCISE 7.28. 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 7.3.

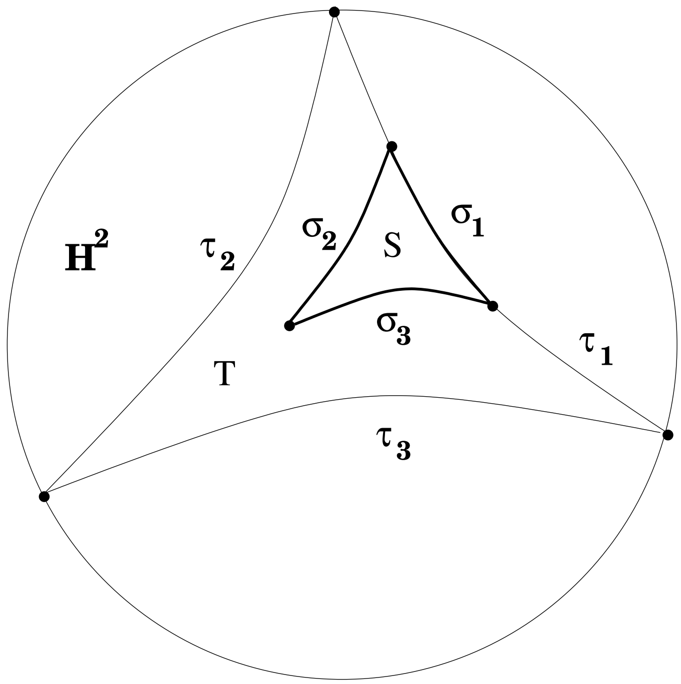


FIGURE 7.3. Triangles in the hyperbolic plane.

LEMMA 7.29. $\mathrm{Isom}(\mathbb{H}^2)$ *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 S^1 and, hence, equals to the restriction of an isometry of $\mathbb{H}^2$. $\square$

COROLLARY 7.30. *All ideal hyperbolic triangles are congruent to each other.*

EXERCISE 7.31. 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.

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

$$Area(T_{0,0,\alpha}) = \pi - \alpha. \tag{7.11}$$

In particular, area of the ideal triangle equals π.

LEMMA 7.32. $Area(T_{\alpha,\beta,\gamma}) = \pi - (\alpha + \beta + \gamma)$.

PROOF. The proof given here is due to Gauss, it appears in the letter from Gauss to Bolyai, see [?]. 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 7.4. Using additivity of area and equation (7.11), we obtain the area formula for T. $\square$

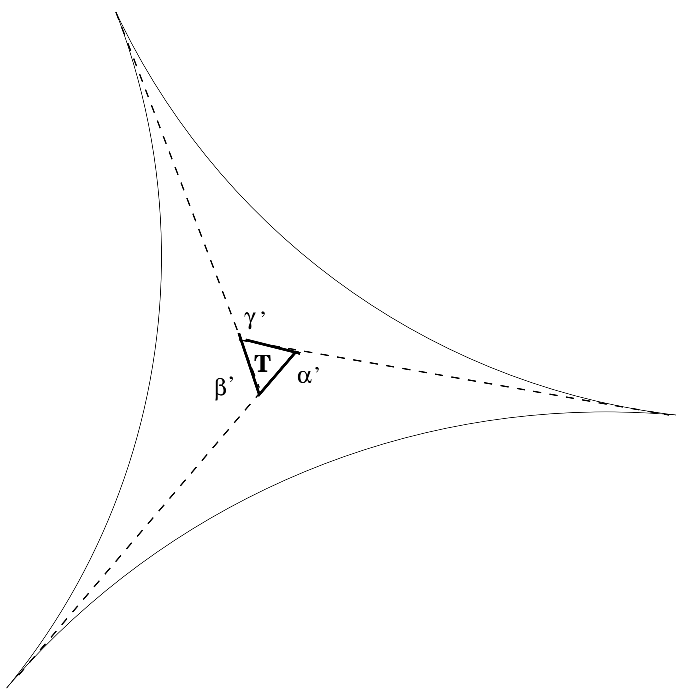


FIGURE 7.4. Computation of area of the triangle T.

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 2.1.6). Since $\mathbb{H}^2 \subset \mathbb{H}^n$ is a totally-geodesic isometric embedding (in the sense of Riemannian geometry), κ is the same for $\mathbb{H}^n$ and $\mathbb{H}^2$.

COROLLARY 7.33. *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 7.32 with Gauss-Bonnet formula (Theorem 2.21) we conclude that $\kappa = -1$. □

Note that scaling properties of the sectional curvature (see Section 2.1.6) imply that sectional curvature of

$$\left(U^n, \frac{adx^2}{x_n^2}\right)$$

equals $-a^2$ for every $a > 0$.

7.5. Distance function on $\mathbb{H}^n$

We begin by defining the following quantities:

$$\operatorname{dist}(z, w) = \operatorname{arccosh}\left(1 + \frac{|z-w|^2}{2 \operatorname{Im} z \operatorname{Im} w}\right) \; z, w \in U^2 \tag{7.12}$$

and, more generally,

$$\operatorname{dist}(p, q) = \operatorname{arccosh}\left(1 + \frac{|p-q|^2}{2 p_n q_n}\right) \; p, q \in U^n \tag{7.13}$$

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 7.34. *dist is invariant under* $\operatorname{Isom}(\mathbb{H}^n) = Mob(U^n)$.

PROOF. First, it is clear that *dist* is invariant under the group $Euc(U^n)$ of Euclidean isometries which preserve U^n. Next, any two points in U^n belong to a vertical half-plane in U^n. Applying elements of $Euc(U^n)$ to this half-plane, we can transform it to the coordinate half-plane $U^2 \subset U^n$. 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, *dist*. Now, the assertion follows from Exercise 7.7 and Lemma 7.8. □

Recall that $d(p, q)$ denotes the hyperbolic distance between points $p, q \in U^n$.

PROPOSITION 7.35. *$dist(p, q) = d(p, q)$ 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 7.10,

$$dist(p, q) = \int_1^b \frac{dt}{t} = \log(b), \quad \exp(d(p, q)) = b.$$

On the other hand, the formula (7.12) yields:

$$dist(p,q) = \operatorname{arccosh}\left(1 + \frac{(b-1)^2}{2b}\right).$$

Hence,

$$\cosh(dist(p,q)) = \frac{e^{dist(p,q)} + e^{-dist(p,q)}}{2} = 1 + \frac{(b-1)^2}{2b}.$$

Now, the equality $dist(p,q) = d(p,q)$ 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 7.18, 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 $d(p,q) = dist(p,q)$ follows from the special case of points on the y-axis. □

EXERCISE 7.36. Deduce from (7.12) that

$$\ln\left(1 + \frac{|z-w|^2}{2\operatorname{Im} z \operatorname{Im} w}\right) \leq d(z,w) \leq \ln\left(1 + \frac{|z-w|^2}{2\operatorname{Im} z \operatorname{Im} w}\right) + \ln 2$$

for all points $z, w \in U^2$.

7.6. 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

$$S_h(p,R) = \{x \in \mathbb{H}^n : d(x,p) = R\}.$$

EXERCISE 7.37. 1. Prove that $S_h(e_n, R) \subset \mathbb{H}^n = U^n$ equals the Euclidean sphere of center $\cosh(R)e_n$ and radius $\sinh(R)$. *Hint.* It follows immediately from the distance formula (7.12).

2. Suppose that $S = S(x,R) \subset U^n$ is a Euclidean sphere with Euclidean radius R and the center x so that $x_n = a$. Then $S = S_h(p,r)$, where the hyperbolic radius r equals

$$\frac{1}{2}\left(\log(a+R) - \log(a-R)\right).$$

Since group generated by dilations and horizontal translations acts transitively on 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_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 7.38. *Suppose that $B(x_1, R_1) \subset B(x_2, R_2)$ 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, length of γ is $2R_1$, therefore, $R_1 \leqslant R_2$. □

7.7. Horoballs and horospheres in $\mathbb{H}^n$

Consider the unit ball model $\mathbf{B}^n$ of $\mathbb{H}^n$, α a point in the ideal boundary (here identified with the unit sphere S^{n-1}) and r a geodesic ray with $r(\infty) = \alpha$, i.e. according to Lemma 7.18, an arc of circle orthogonal to S^{n-1} in α with the other endpoint x in the interior of $\mathbf{B}^n$. By Lemma 2.52, the open horoball $B(\alpha)$ defined by the inequality $f_r < 0$, where f_r is the Busemann function for the ray r, equals the union of open balls $\bigcup_{t\geqslant 0} B(r(t), t)$. The discussion in Section 7.6, in particular Exercise 7.37, implies that each ball $B(r(t), t)$ is a Euclidean ball with center in a point $r(T_t)$ with $T_t > t$. Therefore, the above union is the open Euclidean ball with boundary tangent to S^{n-1} at α, and containing the point x. According to Lemma 2.54, the closed horoball and the horosphere defined by $f_r \leqslant 0$ and $f_r = 0$, respectively, are the closed Euclidean ball and the boundary sphere, both with the point α removed.

We conclude that the set of horoballs (closed or open) with center α is the same as the set of Euclidean balls (closed or open) tangent to S^{n-1} at α, with the point α removed.

Applying the map $\sigma : \mathbf{B}^n \to U^n$ to horoballs and horospheres in $\mathbf{B}^n$, we obtain horoballs and horospheres in the upper-half space model 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). It is then clear that hyperbolic isometries carry horoballs/horospheres to horoballs/horospheres.

Recall that $\sigma(-e_n) = \infty$. Therefore, every horosphere in $\mathbf{B}^n$ centered at $-e_n$ is sent by σ to an $n-1$-dimensional Euclidean subspace E of U^n whose compactification contains the point ∞. Hence, E has to be a horizontal Euclidean subspace, i.e., a subspace of the form

$$\{x \in U^n : x_n = t\}$$

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 ds^2 to every horosphere is isometric to the flat metric on $\mathbb{R}^{n-1}$.

EXERCISE 7.39. 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 f_r for the ray r is given by

$$f_r(x_1, \dots, x_n) = -\log(x_n).$$

7.8. $\mathbb{H}^n$ 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 $\mathfrak{g}$ through p, $\sigma_p(\mathfrak{g}(t)) = \mathfrak{g}(-t)$. Let us verify that such 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 $\sigma_m(p) = q$. Besides being homogeneous, symmetric spaces also admit large discrete isometry groups: For every symmetric space X, there exists a subgroup $\Gamma \subset \mathrm{Isom}(X)$ which acts geometrically on X.

Details on symmetric spaces can be found for instance in [**Hel01**] and [**?**]. The *rank* of a symmetric space X is the largest number r so that X contains a totally-geodesic submanifold $F \subset X$ which is isometric to an open disk in $\mathbb{R}^r$.

We note that in the unit ball model of $\mathbb{H}^n$ we clearly have the symmetry σ_p with respect to $p = 0$, namely, $\sigma_0 : x \mapsto -x$. 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 7.40. 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 Section 7.4 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.

It turns out that besides real-hyperbolic space $\mathbb{H}^n$, there are three other families of rank 1 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). The rank 1 symmetric spaces X are also characterized among symmetric spaces by the property that any two segments of the same length are congruent in X. Below is a brief discussion of these spaces, we refer to Mostow's book [**?**] and Parker's survey [**?**] for a more detailed discussion.

In all four cases, the symmetric X will appear as a projectivization of a certain cone equipped with a hermitian form $\langle \cdot, \cdot \rangle$ and the distance function in X will be given by the formula:

$$\cosh^2(\mathrm{dist}(p, q)) = \frac{\langle p, q \rangle \langle q, p \rangle}{\langle p, p \rangle \langle q, q \rangle}, \tag{7.14}$$

where $p, q \in C$ represent points in X.

Complex-hyperbolic space. Consider $\mathbb{C}^{n+1}$ equipped with the Hermitian bilinear form

$$\langle v, w \rangle = \sum_{k=1}^{n} v_k \bar{w}_k - v_{n+1} \bar{w}_{n+1}.$$

The group $U(n, 1)$ is the group of complex-linear automorphisms of $\mathbb{C}^{n+1}$ preserving this bilinear form. Consider the negative light cone

$$C = \{v : \langle v, v \rangle < 0\} \subset \mathbb{C}^{n+1}.$$

Then the *complex-hyperbolic space* $\mathbb{CH}^n$ is the projectivization of C. 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 $p \in C$ be such that $\langle p, p \rangle = 1$; tangent space at the projection of p to X is the projection of the orthogonal complement $p^\perp$ in $\mathbb{C}^{n+1}$. Let $v, w \in \mathbb{C}^{n+1}$ be such that $\langle p, v \rangle = 0, \langle p, w \rangle$. Then set

$$(v, w)_p := -Im \langle v, w \rangle .$$

This determines a $PU(n, 1)$-invariant Riemannian metric on X. The corresponding distance function (7.14) will be G-invariant.

Quaternionic-hyperbolic space. 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 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 multiplication action) is the n-dimensional *quaternionic-projective space* PV. Analogously to the case of real and complex hyperbolic spaces, we consider the negative light cone

$$C = \{\mathbf{q} \in V : \langle q, q \rangle < 0\}.$$

The group G acts naturally on $PC \subset PV$ 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). The space PC is called the n-dimensional *quaternionic-hyperbolic space* $\mathbf{H}\mathbb{H}^n$

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, so one cannot talk about free $\mathbf{O}$-modules; we refer the reader to [**?**, **?**] for the details.

7.9. 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 7.41. $\Delta(T) = Inrad(T)$.

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 7.38. 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 $[A, B], [B, C]$ 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 7.25. □

Thus, we need to estimate inradius of hyperbolic triangles from above. Recall that by Exercise 7.28, 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 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 7.37, its hyperbolic radius equals $\log(3)/2$. By combining these observations with Exercise 7.27, we obtain

PROPOSITION 7.42. *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 7.43. *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 7.5. 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 sides $\tau_i = [A_{j-1}, A_{j+1}], j = 1, 2, 3$ modulo 3, according to their orientation. Then, by the Exercise 7.25, for every i,

$$d(\tau_j(t), \tau_{j-1})$$

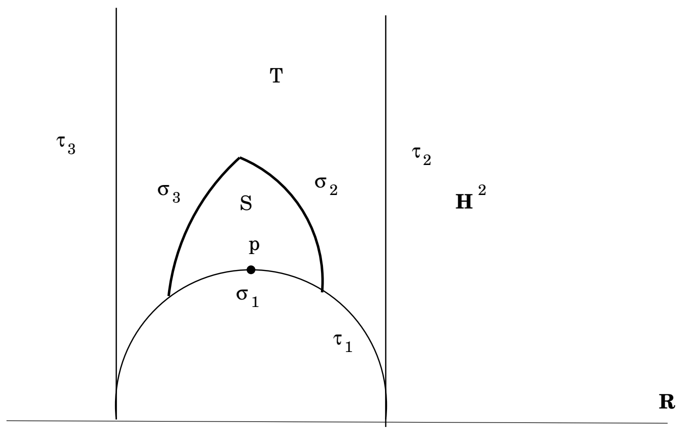


FIGURE 7.5. *E*nlarging hyperbolic triangle S.

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 7.13, we get $d(p, q) = \operatorname{arccosh}(\sqrt{2})$. Note that alternatively, one can get the formula for $d(p, q)$ from (7.7) by considering the right triangle $[p, q, -1]$ where the angle at p equals $\pi/4$. $\square$

As we will see in Section 8.1, the above propositions mean that *all hyperbolic triangles are uniformly thin.*

7.10. Existence-uniqueness theorem for triangles

Proof of Lemma 2.31. We will prove this result for the hyperbolic plane $\mathbb{H}^2$, this will imply lemma for all $\kappa < 0$ by rescaling the metric on $\mathbb{H}^2$. We leave the cases $\kappa \geq 0$ to the reader as the proof is similar. The proof below is goes back to Euclid (in the case of $\mathbb{R}^2$). Let c denote the largest of the numbers a, b, c. Draw a geodesic $\gamma \subset \mathbb{H}^2$ through points x, y so that $d(x, y) = c$. Then

$$\gamma = \gamma_x \cup [x, y] \cap \gamma_y,$$

where γ_x, γ_y are geodesic rays emanating from x and y respectively. Now, consider 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

$$S(x, b) \cap [x, y] = p, \quad S(y, a) \cap [x, y] = y.$$

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, for instance, from the hyperbolic cosine law. $\square$

7.11. Lattices

Recall that a *lattice* in a Lie group G is a discrete subgroup Γ such that the quotient $\Gamma \setminus G$ has finite volume. Here, the left-invariant volume form on G is defined by taking a Riemannian metric on G which is left-invariant under G and right-invariant under K, the maximal compact subgroup of G. Thus if $X := G/K$, then this quotient manifold has a Riemannian metric which is (left) invariant under G. Hence, Γ is a lattice iff Γ acts on X properly discontinuously so that $Vol(\Gamma \setminus X)$ is finite. Note that the action of Γ on X need not be free. Recall also that a lattice Γ is *uniform* if $\Gamma \setminus X$ is compact and Γ is *nonuniform* otherwise.

Each lattice is finitely-generated (this is clear for uniform lattices but is not at all obvious otherwise); in the case of the hyperbolic spaces finite generation follows from the thick-thin decomposition discussed below. Thus, if Γ is a lattice in a linear Lie group, then, by Selberg lemma 3.88, Γ contains a torsion-free subgroup of finite index. In particular, if Γ is a lattice in $PO(n,1)$ (which is isomorphic to the isometry group of the hyperbolic n-space) then Γ is virtually torsion-free. We also note that a finite-index subgroup in a lattice is again a lattice. Passing to a finite-index subgroup, of course, does not affect uniformity of a lattice.

EXAMPLE 7.44. 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 the fundamental group Γ. Therefore, Γ is a free group of finite rank.

EXERCISE 7.45. Show that groups Γ in the above example cannot be cyclic.

Recall that a *horoball* in $\mathbb{H}^n$ (in the unit ball model) is a domain bounded by a round Euclidean ball $B \subset \mathbb{H}^n$, whose boundary is tangent to the boundary of $\mathbb{H}^n$ in a single point (called the *center* or *footpoint* of the horoball). The boundary of a horoball in $\mathbb{H}^n$ is called a *horosphere*. In the upper half-space model, the horospheres with the footpoint ∞ are horizontal hyperplanes

$$\{(x_1, ..., x_{n-1}, t) : (x_1, ..., x_{n-1}) \in \mathbb{R}^{n-1}\},$$

where t is a positive constant.

LEMMA 7.46. *Suppose that $\Gamma < PO(n,1)$ is a torsion-free discrete group containing a parabolic element γ. Then Γ is a non-uniform lattice.*

PROOF. Recall that every parabolic isometry of $\mathbb{H}^n$ has unique fixed point in the ideal boundary sphere S^{n-1}. By conjugating Γ by an isometry of $\mathbb{H}^n$, we can assume that γ fixes the point ∞ in the upper half-space model $\mathbb{R}^n_+$ of $\mathbb{H}^n$. Therefore, γ acts on as a Euclidean isometry on $\mathbb{R}^n_+$. After conjugating γ by a Euclidean isometry, γ has the form

$$x \mapsto Ax + v,$$

where $v \in \mathbb{R}^{n-1} \setminus \{0\}$ and A is an orthogonal transformation fixing the vector v. Hence, γ preserves the Euclidean line $L \subset \mathbb{R}^{n-1}$ (spanned by v) and the restriction of γ to L is the translation $x \mapsto x + v$. Let H denote the hyperbolic plane in $\mathbb{H}^n$, which is the vertical Euclidean half-plane above the line L. Again, γ acts on H as the translation $x \mapsto x + v$. We introduce the coordinates (x, y) on H, where $x \in \mathbb{R}$ and $y > 0$. Then for every $z = (x, y) \in H$,

$$d(z, \gamma z) < \frac{|v|}{y}$$

where $|v|$ is the Euclidean norm of the vector v. Let c_z denote the projection of the geodesic $[z, \gamma z]$ to the hyperbolic manifold $M = \mathbb{H}^n/\Gamma$. By sending y to infinity, we conclude that the (nontrivial) free homotopy class $[\gamma]$ in $M = \mathbb{H}^n/\Gamma$ represented by $\gamma \in \Gamma$, contains loops c_z of arbitrarily short length. This is impossible if M were a compact Riemannian manifold. $\square$

The converse to the above lemma is much less trivial and follows from

THEOREM 7.47 (Thick-thin decomposition). *Suppose that Γ is a nonuniform lattice in $Isom(\mathbb{H}^n)$. Then there exists an (infinite) collection C of open horoballs $C := \{B_j, j \in J\}$, with pairwise disjoint closures, so that*

$$\Omega := \mathbb{H}^n \setminus \bigcup_{j \in J} B_j$$

is Γ-invariant and $M_c := \Omega/\Gamma$ is compact. Furthermore, 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. [**?**], [**Kap01**], [**Rat94**], [**Thu97**].

The quotient M_c is called the *thick part* of $M = \mathbb{H}^n/\Gamma$ and its (noncompact) complement in M is called the *thin part* of M. If Γ is torsion-free, then it acts freely on $\mathbb{H}^n$ and M has natural structure of a hyperbolic manifold of finite volume. If Γ is not torsion-free, then M is a *hyperbolic orbifold*. Clearly, when $\Gamma < PO(n, 1)$ is a lattice, the quotient $M = \mathbb{H}^n/\Gamma$ is compact if and only if $C = \emptyset$.

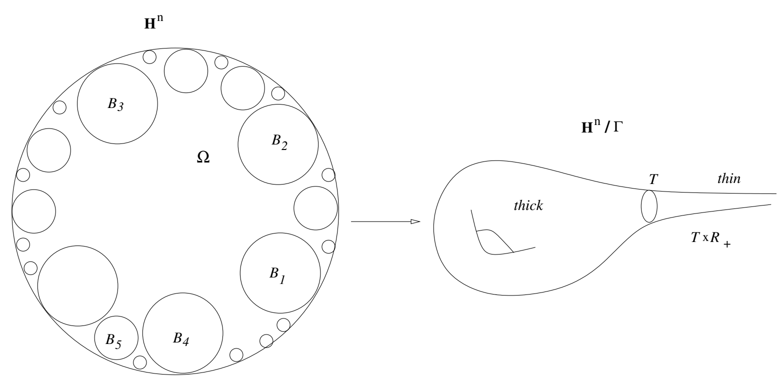


FIGURE 7.6. Truncated hyperbolic space and thick-thin decomposition.

The set Ω is called a *truncated hyperbolic space*. The boundary horospheres of Ω are called *peripheral horospheres*. Since each closed horoballs used to define Ω are pairwise disjoint, Ω is contractible. In particular, if Γ is torsion-free, then it has finite type. In general, Γ is of type $\mathbf{F}_\infty$.

Note that the stabilizer Γ_j of each horosphere ∂B_j acts on this horosphere cocompactly with the quotient $T_j := \partial B_j/\Gamma_j$. The quotient B_j/Γ_j is naturally homeomorphic to $T_j \times \mathbb{R}_+$, this product decomposition is inherited from the foliation of B_j by the horospheres with the common footpoint ξ_j and the geodesic rays asymptotic to ξ_j. If Γ is torsion-free, orientation preserving and $n = 3$, the quotients T_j are 2-tori.

Observe that a hyperbolic horoball cannot be stabilized by a hyperbolic isometry. Indeed, by working with the upper half-space model of $\mathbb{H}^n$, we can assume that the (open) horoball in question is given by

$$B = \{(x_1, \ldots, x_n) : x_n > 1\}.$$

Every hyperbolic isometry γ stabilizing B would have to fix ∞ and act and a Euclidean isometry on the boundary horosphere of B. Thus, γ is either elliptic or parabolic. In particular, stabilizers of the horoballs B_j in Theorem 7.47 contain no hyperbolic elements. Since we can assume that Γ is torsion-free, we obtain

COROLLARY 7.48. *A lattice in $PO(n,1)$ is uniform if and only if it does not contain parabolic elements.*

Arithmetic groups provide a general source for lattices in Lie groups. 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. Let G be a Lie group with finitely many components.

DEFINITION 7.49. An *arithmetic subgroup* in G is a subgroup of G commensurable to the 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.

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. [**?**, **?**].

Bianchi groups. We now describe a concrete class of non-uniform arithmetic lattices in the isometry group of hyperbolic 3-space, called *Bianchi groups*. Let D denote 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 the group 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 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, quotients $\mathbb{H}^3/\Gamma$ has finite volume. Furthermore, every arithmetic lattice in $SL(2,\mathbb{C})$ is *commensurable* to a Bianchi group. We refer the reader to [**?**] for the detailed discussion of these and other facts about Bianchi groups.

Commensurators of lattices.

Recall (see §3.4) 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. $|\Gamma : g\Gamma g^{-1} \cap \Gamma| < \infty$, $|g\Gamma g^{-1} : g\Gamma g^{-1} \cap \Gamma| < \infty$.

Below we consider commensurators in the situation when Γ is a lattice in a Lie group G.

EXERCISE 7.50. Let $\Gamma := SL(2, O_D) \subset G := SL(2, \mathbb{C})$ be a Bianchi group.

1. Show that $Comm_G(\Gamma) \subset SL(2, \mathbb{Q}(\omega))$. 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 [?], Chapter IX, Theorem B and Lemma 2.7; see also [?], 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.

Consider now the case when G is either a Lie group or a finitely-generated group and $\Gamma \leqslant G$ is a finitely-generated subgroup. We note that each element $g \in Comm_G(\Gamma)$ determines a quasi-isometry $f : \Gamma \to \Gamma$. Indeed, the Hausdorff distance between Γ and $g\Gamma g^{-1}$ is finite. Hence the quasi-isometry f is given by composing $g : \Gamma \to g\Gamma g^{-1}$ with the nearest-point projection to Γ.

The main goal of the remainder of the chapter is to prove the following

THEOREM 7.51 (R. Schwartz [?]). *Let $\Gamma \subset G = Isom(\mathbb{H}^n)$ be a nonuniform lattice, $n \geqslant 3$. Then:*

(a) For each quasi-isometry $f : \Gamma \to \Gamma$ there exists $\gamma \in Comm_G(\Gamma)$ which is within finite distance from f. The distance between these maps depends only on Γ and on the quasi-isometry constants of f.

(b) Suppose that Γ, Γ' are non-uniform lattices which are quasi-isometric to each other. Then there exists an isometry $g \in Isom(\mathbb{H}^n)$ such that the groups Γ' and $g\Gamma g^{-1}$ are commensurable.

(c) Suppose that Γ' is a finitely-generated group which is quasi-isometric to a nonuniform lattice Γ above. Then the groups Γ, Γ' are virtually isomorphic

Our proof will mostly follow [?].

Note that this 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 non-uniform lattice Γ acting on $\mathbb{H}^2$. In view of 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, so that for every nontrivial cyclic subgroup $C \subset F_n$ and every m, $\phi^m(C)$ is not conjugate to C, see e.g. [?]. Therefore, such ϕ cannot send parabolic subgroups of Γ to parabolic subgroups of Γ. Hence, the quasi-isometry of F_n given by ϕ cannot extend to a quasi-isometry $\mathbb{H}^2 \to \mathbb{H}^2$. It follows that (a) fails for $n = 2$. Similarly, one can show that (b) fails, since commensurability preserves arithmeticity and there are both arithmetic and non-arithmetic lattices in $Isom(\mathbb{H}^2)$. All these lattices are virtually free, hence, virtually isomorphic.

CHAPTER 8

Gromov-hyperbolic spaces and groups

The goal of this chapter is to define and 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 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 Mikhail Gromov on hyperbolic groups, although ideas of *combinatorial* curvature and (in retrospect) hyperbolic properties of finitely-generated groups are much older. They go back to work of Max Dehn (on word problem in groups), Martin Grindlinger (small cancelation theory), Alexandr Ol'shanskii (who used what we now would call *relative hyperbolicity* in order to construct finitely-generated groups with exotic properties) and many others.

8.1. 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 a *hyperbolic group* discussed later in this chapter. Rips notion of hyperbolicity is based on the thinness properties of hyperbolic triangles which are established in section 7.9.

Let (X, d) be a geodesic metric space. As in section 7.4, 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 Section 7.9 of Chapter 7.

DEFINITION 8.1. 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 8.2 (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. 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 few simple but important geometric features of δ-hyperbolic spaces.

First, not that general Rips–hyperbolic metric spaces X are by no means uniquely geodesics. Nevertheless, next lemma shows that geodesics in X between given pair of points are "almost unique":

LEMMA 8.3. *If X is δ–hyperbolic, then every pair of geodesics $[x,y],[x,z]$ with $d(y,z) \leqslant D$ are at Hausdorff distance at most $D+\delta$. In particular, if α, β are geodesic segments connecting points $x, y \in X$, then* $\mathrm{dist}_{Haus}(\alpha, \beta) \leqslant \delta$.

PROOF. Every point p on $[x,y]$ is, either at distance at distance at most δ from $[x,z]$, or at distance at most δ from $[y,z]$; in the latter case p is at distance at most $D+\delta$ from $[x,z]$. □

The next lemma, the *fellow-traveling property of hyperbolic geodesics* sharpens the conclusion of Lemma 8.3.

LEMMA 8.4. *Let $\alpha(t), \beta(t)$ be geodesics in a δ-hyperbolic space X, so 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 previous lemma, for every $t \in [0, t_0]$ there exists $s \in [0, t_0]$ so that

$$d(\beta(t), \alpha(s)) \leqslant c = \delta + D.$$

By 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* 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. 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 8.5. 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 8.6 (thin polygons). *If X is δ–hyperbolic then every geodesic n-gon in X is η_n-thin for*

$$\eta_n = 2\delta \log_2 n.$$

PROOF. We prove the estimate on thinness of n-gons by induction on m. For $n \leqslant 3$ the statement follows from δ-thinness of bigons and triangles. Suppose $n \geqslant 4$ and the inequality holds for all $m \leqslant n-1$. Consider a geodesic n-gon P which has edges $\tau_i = [A_i, A_{i+1}]$ and consider its edge $\tau = \tau_n$ of P. We will consider the case when n is odd, $n = 2k+1$, since the other case is similar. We subdivide P in two $k+1$-gons P', P'' and one triangle T by introducing the diagonals $[A_1, A_{k+1}]$ and $[A_{k+1}, A_n]$. By the induction hypothesis, P', P'' are η_{k+1}-thin, while the triangle T is δ-thin. Therefore, τ is within distance $\leqslant \eta_{k+1} + \delta$ from the union of the other sides of P. We leave it to the reader to check that

$$2\log_2(k+1) + 1 \leqslant 2\log(n) = 2\log_2(2k+1). \quad \square$$

We now give some examples of Rips–hyperbolic metric spaces.

EXAMPLE 8.7. (1) Proposition 7.42 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 section 2.1.8. 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 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 section 2.1.8. 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 edges $\tau_i, i = 1, 2, 3$ and a points $P_1 \in \tau_1$ 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 $i = 3$, so that $d(\tilde{P}_1, \tilde{P}_i) \leqslant \delta_\kappa$. Let $P_i \in \tau_i$ be the comparison point of $\tilde{P}_i$. Then, by the comparison inequality

$$d(P_1, P_i) \leqslant d(\tilde{P}_1, \tilde{P}_i) \leqslant \delta_\kappa.$$

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 2.36, 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 8.8. 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 8.9. For instance, let us verify that $\mathbb{R}^2$ 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 $c \to \infty$, show that $\mathbb{R}^2$ is not δ-hyperbolic for any $\delta > 0$. More generally, if a metric space X contains an isometrically embedded copy of $\mathbb{R}^2$, then X is not hyperbolic.

Here is an example of a metric space which is not hyperbolic, but does not contain a quasi-isometrically embedded copy of $\mathbb{R}^2$ either. Consider the wedge X of countably many circles C_i each given with path-metric of overall length $2\pi i$, $i \in \mathbb{N}$. We equip X with the path-metric so that each C_i is isometrically embedded. Exercise 8.8 shows that X is not hyperbolic.

EXERCISE 8.10. Show that X contains no quasi-isometrically embedded copy of $\mathbb{R}^2$. Hint: Use coarse topology.

More interesting examples of non-hyperbolic spaces containing no quasi–isometrically embedded copies of $\mathbb{R}^2$ are given by various solvable groups, e.g. the Sol_3 group and Cayley graph of the Baumslag–Solitar group $BS(n, 1)$, see [**?**].

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 the triangle, described in Definition 5.49, consequently it is related to the filling area of the triangle *via* a Besikovitch type inequality as described in Proposition 5.50.

DEFINITION 8.11. 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 7.41, this definition coincides with the radius of the largest circle inscribed in T. Clearly, $\Delta(T) \leqslant \delta(T)$ and

$$\Delta(T) \leqslant \text{minsize}(T) \leqslant 2\Delta(T) + 1\,.$$

It turns out that

$$\text{minsize}(T) \leqslant 2\delta. \tag{8.1}$$

Indeed, let τ_1, τ_2, τ_3 be the sides of T, we will assume that τ_1 is parameterized so that

$$\tau_1(0) \in Im(\tau_3), \tau_1(a_1) = Im(\tau_2),$$

where a_1 is the length of τ_1. Then by the intermediate value theorem, applied to the difference

$$d(\tau_1(t) - Im(\tau_2)) - d(\tau_1(t) - Im(\tau_3))$$

we conclude that there exists t_1 so that $d(\tau_1(t_1), Im(\tau_2)) = d(\tau_1(t_1), Im(\tau_3)) \leqslant \delta$. Taking $p_1 = \tau_1(t_1)$ and $p_i \in Im(\tau_i), i = 2, 3$, the points nearest to p_1, we get

$$d(p_1, p_2) \leqslant \delta, d(p_1, p_3) \leqslant \delta,$$

hence,

$$minsize(T) \leqslant 2\delta.$$

8.2. Geometry and topology of real trees

In this section we consider a special type of hyperbolic spaces, the *real trees*.

DEFINITION 8.12. A 0–hyperbolic (geodesic) metric space is called a *real tree*.

EXERCISE 8.13. 1. Show that every real tree is a $CAT(0)$ space.
2. Show that every real tree is a $CAT(\kappa)$ space for every κ.

It follows from Exercise 8.5 that every polygon in a real tree is 0-thin.

LEMMA 8.14. *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 continuous injective map (i.e., a topological arc) $x = \alpha(0), y = \alpha(1)$. Let $\alpha^* = [x, y], \alpha^* : [0, D] \to X$ be the geodesic connecting x to y. We claim that the image of α contains the image of α^*. Indeed, we can approximate α by piecewise-geodesic (nonembedded!) arcs

$$\alpha_n = [x_0, x_1] \cup ... \cup [x_{n-1}, x_n], \quad x_0 = x, x_n = y.$$

Since the $n+1$-gon P in X,which is the concatenation of α_n with $[y, x]$ is 0-thin, $\alpha^* \subset \alpha_n$. 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 8.15. Prove the converse to the above lemma.

DEFINITION 8.16. 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

$$\Re_p = \{r : [0, a) \to T \mid a > 0,\ r \text{ isometry},\ r(0) = p\}$$

and

$$r_1 \sim r_2 \iff \exists\, \varepsilon > 0 \text{ such that } r_1|_{[0,\varepsilon)} \equiv r_2|_{[0,\varepsilon)}.$$

Simplest examples of real trees are given by simplicial trees equipped with path-metrics. We will see, however, that other real trees also arise naturally in geometric group theory.

By Lemma 8.14, for every homeomorphism $c : [a, b] \to T$ the image $c([a, b])$ 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 $\Im_p / \sim$, where

$$\Im_p = \{c : [0, a) \to T \mid a > 0,\ c \text{ homeomorphism},\ c(0) = p\}$$

and

$$c_1 \sim c_2 \iff \exists\, \varepsilon_1 > 0,\ \varepsilon_2 > 0 \text{ such that } c_1([0, \varepsilon_1)) = c_2([0, \varepsilon_2)).$$

DEFINITION 8.17. Define *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 8.18. Show that $val(p)$ equals the number of connected components of $T \setminus \{p\}$.

DEFINITION 8.19. A real tree T is called *α–universal* if every real tree with valence at most α can be isometrically embedded into T.

See [?] for a study of universal trees. In particular, the following holds:

THEOREM 8.20 ([?]). *For every cardinal number $\alpha > 2$ there exists an α–universal tree, and it is unique up to isometry.*

Fixed-point properties.

Part 1 of Exercise 8.13 together with Corollary 2.43 implies:

COROLLARY 8.21. *If G is a finite group acting isometrically on a complete real tree T, then G fixes a point in T.*

DEFINITION 8.22. A group G is said to have *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 property FA.

8.3. 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 $|x|_p := \mathrm{dist}(x, p)$ and define the *Gromov product*

$$(x, y)_p := \frac{1}{2}\left(|x|_p + |y|_p - \mathrm{dist}(x, y)\right).$$

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 8.23. 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 8.24. Suppose that X is a metric tree. Then $(x, y)_p$ is the distance $\mathrm{dist}(p, \gamma)$ from p to the geodesic segment $\gamma = [xy]$.

In general a direct calculation shows that for each point $z \in X$

$$(p, x)_z + (p, y)_z \leqslant |z|_p - (x, y)_p$$

with equality

$$(p, x)_z + (p, y)_z = |z|_p - (x, y)_p. \tag{8.2}$$

if and only $d(x, z) + d(z, y) = d(x, y)$. Thus, for every $z \in \gamma = [x, y]$,

$$(x, y)_p = d(z, p) - (p, x)_z - (p, y)_z \leqslant d(z, p).$$

In particular, $(x, y)_p \leqslant \mathrm{dist}(p, \gamma)$.

LEMMA 8.25. *Suppose that X is δ–hyperbolic in the sense of Rips. Then the Gromov product in X is "comparable" to* $\mathrm{dist}(p, \gamma)$*: For every $x, y, p \in X$ and geodesic $\gamma = [x, y]$,*

$$(x, y)_p \leqslant \mathrm{dist}(p, \gamma) \leqslant (x, y)_p + 2\delta.$$

PROOF. The inequality $(x, y)_p \leqslant \mathrm{dist}(p, \gamma)$ was proved above; so we have to establish the other inequality. Note that since the triangle $\Delta(pxy)$ is δ–thin, for each point $z \in \gamma = [x, y]$ we have

$$\min\{(x, p)_z, (y, p)_z\} \leqslant \min\{\mathrm{dist}(z, [p, x]), \mathrm{dist}(z, [p, y])\} \leqslant \delta.$$

By continuity of the distance function, there exists a point $z \in \gamma$ such that $(x, p)_z, (y, p)_z \leqslant \delta$. By applying the equality (8.2) we get:

$$|z|_p - (x, y)_p = (p, x)_z + (p, y)_z \leqslant 2\delta.$$

Since $|z|_p \leqslant \mathrm{dist}(p, \gamma)$, we conclude that $\mathrm{dist}(p, \gamma) \leqslant (x, y)_p + 2\delta$. $\square$

Now, for a metric space X define a number $\delta_p = \delta_p(X) \in [0, \infty]$ as follows:

$$\delta_p := \sup\{\min((x, z)_p, (y, z)_p) - (x, y)_p\}$$

where the supremum is taken over all triples of points $x, y, z \in X$.

EXERCISE 8.26. If $\delta_p \leqslant \delta$ then $\delta_q \leqslant 2\delta$ for all $q \in X$.

DEFINITION 8.27. 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 8.28. The real line with the usual metric is 0-hyperbolic in the sense of Gromov.

EXERCISE 8.29. Gromov–hyperbolicity is invariant under $(1, A)$-quasi-isometries.

EXERCISE 8.30. Let X be a metric space and $N \subset X$ be an R-net. Show that the embedding $N \hookrightarrow X$ is an $(1, R)$-quasi-isometry. In particular, 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.

LEMMA 8.31. *Suppose that X is δ–hyperbolic in the sense of Rips. Then it is 3δ–hyperbolic in the sense of Gromov. In particular, a geodesic metric space is a real tree if and only if it is 0-hyperbolic in the sense of Gromov.*

PROOF. Consider points $x, y, z, p \in X$ and the geodesic triangle $T(xyz) \subset X$ with vertices x, y, z. Let $m \in [x, y]$ be the point nearest to p. Then, since the triangle $T(x, y, z)$ is δ–thin, there exists a point $n \in [x, z] \cup [y, z]$ so that $\text{dist}(n, m) \leq \delta$. Assume that $n \in [y, z]$. Then, by Lemma 8.25,

$$(y, z)_p \leqslant \text{dist}(p, [y, z]) \leqslant \text{dist}(p, [x, y]) + \delta.$$

On the other hand, by Lemma 8.25,

$$\text{dist}(p, [x, y]) \leqslant (x, y)_p - 2\delta.$$

By 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 the "converse" to the above lemma:

LEMMA 8.32. *Suppose that 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

$$\mathrm{dist}(y,p) - \mathrm{dist}(x,y) \geqslant \mathrm{dist}(z,p) - \mathrm{dist}(x,z) - 2\delta \iff$$
$$\mathrm{dist}(z,p) \leqslant 2\delta.$$

Thus $\mathrm{dist}(p,\alpha) \leqslant 2\delta$.

2. Consider now a geodesic triangle $[x,y,p] \subset X$ and let $z \in [x,y]$. Our goal is to show that z belongs to $\mathcal{N}_{4\delta}([p,x] \cup [p,y])$. 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 $\alpha := [p,y]$. 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 8.33 (M. Gromov, [**?**], section 6.3C.). *For geodesic metric spaces, Gromov–hyperbolicity is equivalent to Rips–hyperbolicity.*

The drawback is that in this generality, Gromov–hyperbolicity fails to be QI invariant:

EXAMPLE 8.34 (Gromov–hyperbolicity is not QI invariant). This example is taken from [**?**]. 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 quasi-isometry:

$$|x - x'| \leqslant d(f(x), f(x')) \leqslant \sqrt{2}|x - x'|.$$

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 not bounded from above as $t \to \infty$ and hence X is not δ-hyperbolic for any $\delta < \infty$. Thus 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 8.39), that in the context of geodesic spaces, Gromov–hyperbolicity is a QI invariant.

8.4. Ultralimits and stability of geodesics in Rips–hyperbolic spaces

In this section we will see that every hyperbolic geodesic metric spaces X *globally* resembles a tree. This property will be used to prove *Morse Lemma*, which establishes that quasi-geodesics in δ-hyperbolic spaces are uniformly close to geodesics.

LEMMA 8.35. *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 **??**, ultralimit of geodesic metric spaces is again a geodesic metric space. Thus, in view of Lemma 8.32, 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 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 ω-lim 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 ω-lim $\delta_i = 0$. Thus, X_ω is 0-hyperbolic. □

EXERCISE 8.36. Find a flaw in the following "proof" of this lemma: 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 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.

COROLLARY 8.37. *Every geodesic in the tree X_ω is a limit geodesic.*

The following fundamental theorem in the theory of hyperbolic spaces is called *Morse Lemma* or *stability of hyperbolic geodesics*.

THEOREM 8.38 (Morse Lemma). *There exists a function $\theta = \theta(L, A, \delta)$, so 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 the geodesic segment $[f(a), f(b)] \subset X$ is at most θ.*

PROOF. Set $c = d(f(a), f(b))$. Given quasi-geodesic f and geodesic $f^* : [0, c] \to X$ parameterizing $[f(a), f(b)]$, 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 $dist_{Haus}(Im(f), Im(f^*))$ is $\max(D_f, D_f^*)$. 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

$$|\text{dist}(f_n(t_n), [f(-n), f(n)]) - D_n| \leqslant 1.$$

As in the definition of asymptotic cones, consider two sequences of pointed metric spaces

$$\left(\frac{1}{D_n} X_n, f_n(t_n)\right), \quad \left(\frac{1}{D_n}[-n, n], t_n\right).$$

Note that $\omega\text{-lim}\,\frac{n}{D_n}$ could be infinite. Let

$$(X_\omega, x_\omega) = \omega\text{-lim}\left(\frac{1}{D_n}X_n, f_n(t_n)\right)$$

and

$$(Y, y) := \omega\text{-lim}\left(\frac{1}{D_n}[-n, n], t_n\right).$$

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 points $Im(f_n)$ to $Im(f_n^*)$ in the rescaled metric space $\frac{1}{D_n}X_n$ is at most $1 + 1/d_n$. Each map

$$f_n : Y_n \to \frac{1}{d_n}X_n$$

is an $(L, A/D_n)$–quasi-geodesic. Therefore the ultralimit

$$f_\omega = \omega\text{-lim}\, f_n : (Y, y) \to (X_\omega, x_\omega)$$

is an $(L, 0)$–quasi-isometric embedding, i.e. it is a L-bi-Lipschitz map. In particular this map is a continuous embedding. Therefore, the image of f_ω is a geodesic γ in X_ω, see Lemma 8.14.

On the other hand, the sequence of geodesic segments $[f_n(-n), f_n(n)] \subset \frac{1}{d_n}X_n$ 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 1-neighborhood of the geodesic γ^* and, at the same time, $\gamma \neq \gamma^*$ since $x_\omega \in \gamma \setminus \gamma^*$. This contradicts the fact that X_ω is a real tree. □

Historical Remark. Morse [**?**] proved a special case of this theorem in the case of $\mathbb{H}^2$ where the quasi-geodesics in question where geodesics in another Riemannian metric on $\mathbb{H}^2$, which admits a cocompact group of isometries. Busemann, [**?**], 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 quasi-geodesics is due to Mostow [**?**], in the context of negatively curved symmetric spaces, although his proof is general.

Corollary 8.39 (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)–quasi-isometry and $f' : X' \to X$ is its quasi-inverse. Pick a geodesic triangle $T \subset X$. Its image under f is a quasi-geodesic triangle S in X' whose sides are (L, A)–quasi-geodesic. Therefore each of the quasi-geodesic sides σ_i of S is within distance $\leqslant \theta = \theta(L, A, \delta')$ from a geodesic σ_i^* connecting the end-points of this side. See Figure 8.1. The geodesic triangle S^* formed by the segments $\sigma_1^*, \sigma_2^*, \sigma_3^*$ is δ'-thin. Therefore, the quasi-geodesic triangle $f'(S^*) \subset X$ is $\epsilon := L\delta' + A$–thin, i.e. each quasi-geodesic $\tau_i' := f'(\sigma_i^*)$ is within distance $\leqslant \epsilon$ from the union τ_{i-1}', τ_{i+1}'. However,

$$dist_{Haus}(\tau_i, \tau_i') \leqslant L\theta + 2A.$$

Putting this all together, we conclude that the triangle T is δ-thin with

$$\delta = 2(L\theta + 2A) + \epsilon = 2(L\theta + 2A) + L\delta' + A. \quad \square$$

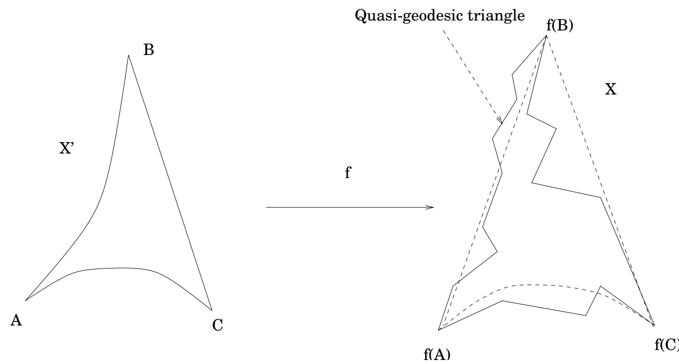


FIGURE 8.1. *I*mage of a geodesic triangle.

Note 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^* so that

$$d(f(t), f^*(s)) \leqslant \theta,$$

and

$$d(f^*(t), f(s^*)) \leqslant \theta.$$

Here $s = s(t), s^* = s^*(t)$. However, applying triangle inequalities one gets for $B = A + \theta$ the following estimates:

$$L^{-1}t - B \leqslant s \leqslant Lt + B \tag{8.3}$$

and

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B) \tag{8.4}$$

8.5. Quasi-convexity in hyperbolic spaces

The usual notion of convexity does not make much sense in the context of hyperbolic geodesic metric spaces. For instance, there is an example of a geodesic Gromov–hyperbolic metric space X where the convex hull of a finite subset is the entire X. The notion of convex hull is then replaces with

DEFINITION 8.40. 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 $[y_1, y_2] \subset X$, where $y_1, y_2 \in Y$.

Accordingly, a subset $Y \subset X$ is R-*quasiconvex* if $H(Y) \subset \mathcal{N}_R(Y)$. A subset Y is called *quasiconvex* if it is quasiconvex for some $R < \infty$.

EXAMPLE 8.41. Let X be a δ-hyperbolic geodesic metric space. Then thin triangle property immediately implies:

1. Every metric ball $B(x, R)$ in is δ-quasiconvex.
2. let $Y_i \subset X$ be 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.
3. Intersection of any family of R-quasiconvex sets is again R-quasiconvex.

An example of a non-quasiconvex subset is a horosphere in $\mathbb{H}^n$: Its quasiconvex hull is the horoball bounded by this horosphere.

The construction of quasiconvex hull could be iterated and, by applying the fact that quadrilaterals in X are 2δ-thin, we obtain:

LEMMA 8.42. *Let $Y \subset X$ be a subset. Then $H(Y)$ is 2δ-quasiconvex in X.*

The following results connect quasiconvexity and quasi-isometry for subsets of Gromov–hyperbolic geodesic metric spaces.

THEOREM 8.43. *Let X, Y be geodesic metric spaces, so that X is δ-hyperbolic geodesic metric space. Then for every quasi-isometric 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) quasi-isometric embedding, $\beta = f(\alpha)$ is an (L, A) quasi-geodesic in X. By Morse Lemma,

$$dist_{Haus}(\beta, \beta^*) \leqslant R = \theta(L, A, \delta),$$

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-quasi-convex. □

The map $f : Y \to f(Y)$ is a quasi-isometry, 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 quasi-convex, so applying the same arguments as in the proof of Theorem 8.39, we conclude that Y is also hyperbolic.

Conversely, let $Y \subset X$ be a coarsely connected subset, i.e., there exists a constant $c < \infty$ so that the complex $Rips_C(Y)$ 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 paths in $Rips_C(Y)$ connecting points of Y. The following is a converse to Theorem 8.43:

THEOREM 8.44. *Suppose that $Y \subset X$ is coarsely connected and Y is quasi-convex in X. Then the identity map $f : (Y, d_{Y,C}) \to (X, \mathrm{dist}_X)$ is a quasi-isometric 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 $dist_X(y, y') \leqslant C$ 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 in $n = [L]$ subintervals of unit intervals and an interval of the length $L - n$:

$$[z_0, z_1], ..., [z_{n-1}, z_n], [z_n, z_{n+1}],$$

where $z_0 = y, z_{n+1} = y'$. Since each z_i belongs to $\mathcal{N}_c(Y)$, there exist points $y_i \in Y$ so that $dist_X(y_i, z_i) \leqslant c$, where we take $y_0 = z_0, y_{n+1} = z_{n+1}$. Then

$$dist_X(z_i, z_{i+1}) \leqslant 2c + 1 \leqslant C$$

and, hence, z_i, z_{i+1} are connected by an edge (of length C) in $Rips_C(Y)$. Now it is clear that

$$d_{Y,C}(y, y') \leqslant C(n + 1) \leqslant C dist_X(y, y') + C. \quad \square$$

REMARK 8.45. It is proven in [?] that in the context of subsets of negatively pinched complete simply-connected Riemannian manifolds X, quasi-convex hulls $Hull(Y)$ are essentially the same as convex hulls:

There exists a function $L = L(C)$ so that for every C-quasiconvex subset $Y \subset X$,

$$H(Y) \subset Hull(Y) \subset \mathcal{N}_{L(C)}(Y).$$

8.6. 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 8.46. *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 quasi-geodesic subspace in a hyperbolic space is coarse Lipschitz:

LEMMA 8.47. *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 $x, y \in X$ so that $d(x,y) = D$. Let $x' = \pi(x), y' = \pi(y)$. Consider the quadrilateral formed by geodesic segments $[x,y] \cup [y,y'], [y',x'] \cup [x',x]$. 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 $[x,y] \cup [y,y]$.

Case 1. We first assume that there are points $x'' \in [x,x'], y'' \in [y,y]$ so 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

$$[x,x''] \cup [x'',q] \cup [q,q'], \quad [y,y''] \cup [y'',q] \cup [q,q']$$

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 $q'' \in [x,y]$ so 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, $d(x'.y') \leqslant 2d(x, y) + 2R + 9\delta$. $\square$

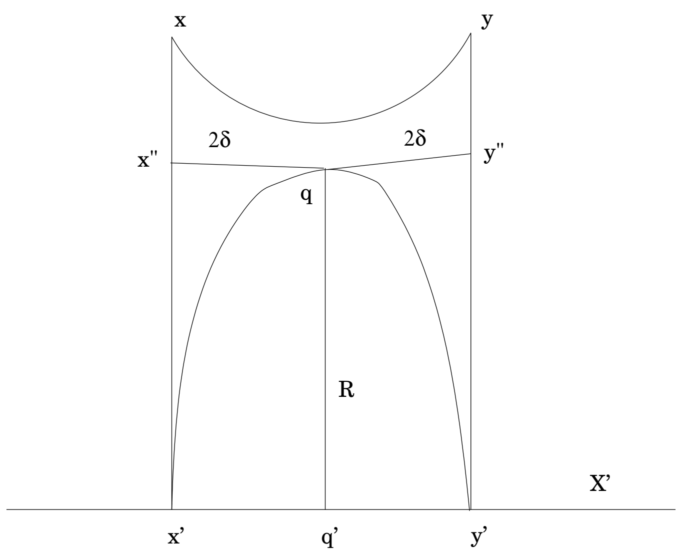


FIGURE 8.2. Projection to a quasiconvex subset.

8.7. 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 $(1, 14\delta)$-quasi-isometries. 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 is the union of three Euclidean line segments (called *legs* of the tripod) joined at a common vertex o, called the *centroid* of $\tilde{T}$. 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 $\mathcal{T} = \tilde{T} = T(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$.

REMARK 8.48. 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 2.33.

EXERCISE 8.49. Given 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.

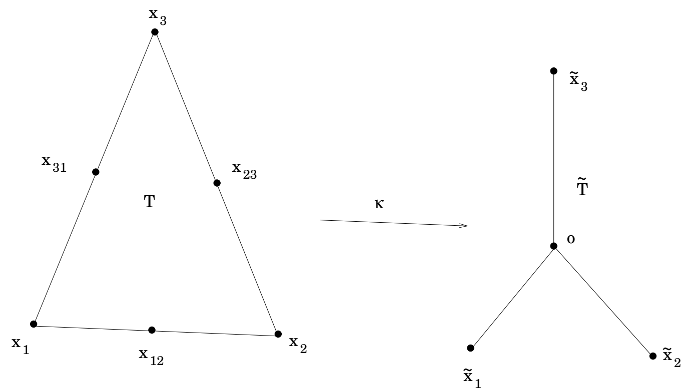


FIGURE 8.3. Collapsing map of triangle to a tripod.

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 is isometric 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 points $x, y \in T$ are *dual* to each other if $\kappa(x) = \kappa(y)$.

EXERCISE 8.50. 1. The collapsing map κ preserves the Gromov-products $(x_i, x_j)_{x_k}$.

2. κ is 1-Lipschitz.

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 8.51 (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 8.3 to obtain 2 and Lemma 8.4 to obtain 3. □

We thus obtain

PROPOSITION 8.52. *κ is a $(1, 14\delta)$-quasi-isometry.*

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 8.52 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 $m_{ij} \in [x_i, x_j]$ be points so 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 8.53. *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 $(1, 14\delta)$-quasi-isometry,

$$d(m_{ij}, x_{ij}) \leqslant d(\kappa(m_{ik}), \kappa(m_{jk})) + 14\delta \leqslant r + 14\delta.$$

DEFINITION 8.54. 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 8.55. *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. $\square$

Let $p_3 \in \gamma_{12} = [x_1, x_2]$ be a point closest to x_3. Taking $R = 2\delta$ and combining Lemma 8.25 with Lemma 8.46, we obtain:

COROLLARY 8.56. $d(p_3, x_{12}) \leqslant 2(2\delta + 2\delta) = 6\delta$.

We now can define a continuous quasi-inverse $\bar{\kappa}$ to κ as follows: We map $[\tilde{x}_1, \tilde{x}_2] \subset \tilde{T}]$ isometrically to a geodesic $[x_1, x_2]$. 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 8.57.

$$d(\bar{\kappa} \circ \kappa, Id) \leqslant 32\delta.$$

8.8. 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 8.6, as shown below. Our arguments are inspired by those in [**?**].

LEMMA 8.58. *Let X be a geodesic metric space, δ–hyperbolic in the sense of Rips' definition. If $[x, y]$ is a geodesic of length $2r$ and m is its midpoint then every path joining x, y outside the open ball $B(m, r)$ has length at least $2^{\frac{r-1}{2\delta}}$.*

PROOF. Consider such a path $\mathfrak{p}$, of length ℓ. Divide it first into two arcs of length $\frac{\ell}{2}$, then into four arcs of length $\frac{\ell}{4}$ etc, until we obtain k arcs of length $\frac{\ell}{2^k} \leqslant 1$. Consider the minimal k satisfying this, i.e. k is the integer part $\lfloor \log_2 \ell \rfloor$. Let $x_0 = x, x_1, ..., x_k = y$ be the consecutive points on $\mathfrak{p}$ obtained after this procedure. Lemma 8.6 applied to a geodesic polygon with vertices $x_0 = x, x_1, ..., x_k = y$ with $[x, y]$ as an edge, implies that m is contained in the $(2\delta k)$–tubular neighborhood of $\bigcup_{i=0}^{k-1} [x_i, x_{i+1}]$, hence in the $(2\delta k + 1)$–tubular neighborhood of $\mathfrak{p}$. However, we assumed that $\operatorname{dist}(m, \mathfrak{p}) \geqslant r$. Thus,

$$r \leqslant 2\delta k + 1 \leqslant 2 \log_2 \ell + 1 \Rightarrow \ell \geqslant 2^{\frac{r-1}{2\delta}}.$$

$\square$

LEMMA 8.59. *Let X be a geodesic metric space, δ–hyperbolic in the sense of Rips' definition, and let x and y be two points on the sphere $S(o, R)$ such that $\operatorname{dist}(x, y) = 2r$. Every path joining x and y outside $\overline{B}(o, R)$ has length at least $\psi(r) = 2^{\frac{r-1}{2\delta} - 3} - 12\delta$.*

PROOF. Let $m \in [x, y]$ be the midpoint. Since $d(o, x) = d(o, y)$, it follows that m is also one of the center-points of the triangle $\Delta(x, y, o)$ in the sense of Section 8.7. Then, by using Lemma 8.51 (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 $[x,x']$ and $[y',y]$ be subsegments of $[x,y]$ of length 6δ. Lemma 8.58 implies that the path $[x',x]\cup\mathfrak{p}\cup[y,y']$ has length at least

$$2^{\frac{r-6\delta-1}{2\delta}}$$

whence $\mathfrak{p}$ has length at least

$$2^{\frac{r-1}{\delta}-3}-12\delta.$$

□

LEMMA 8.60. *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 $[x,o]$ and $[y,o]$ 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 $[x,y]$, since $\Delta(x,y,o)$ is isosceles, m is one of the centroids of this triangle. Since $d(x',y')>14\delta$, they cannot be dual point on $\Delta(x,y,o)$ in the sense of Section 8.7. Let $x'',y''\in[x,y]$ be dual to x',y'. Thus (by Lemma 8.51 (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 $[x'',y'']\subset[x,y]$. Thus, by quasiconvexity of metric balls, see Section 8.5,

$$d(m,o)\leqslant r_1+14\delta+\delta=r_1+15\delta.$$

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 8.59 is increasing,

$$\psi(r_2-15\delta)\leqslant\psi(r).$$

Combining this with Lemma 8.59 (where we take $R=r_1+r_2$), we get the required inequality. □

For a more detailed treatment of divergence in metric spaces, see [**?**, **?**, **?**, **?**, **?**, **?**].

8.9. Ideal boundaries

We consider the general notion of ideal boundary defined in Section 2.1.10 of Chapter 1 in the special case when X is geodesic, δ–hyperbolic and locally compact (equivalently, proper).

LEMMA 8.61. *For each $p\in X$ and each element $\alpha\in\partial_\infty X$ there exists a geodesic ray ρ with initial point p and such that $\rho(\infty)=\alpha$.*

PROOF. Let ρ' be a geodesic ray from the equivalence class α, with 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 $D_n=d(p,\rho'(n))$. The δ-hyperbolicity of X implies that $Im(\gamma_n)$ is at Hausdorff distance at most $\delta+\mathrm{dist}(p,x_0)$ from $[x_0,x_n]$, where $[x_0,x_n]$ is the initial subsegment of ρ'.

Combining the properness of X with the Arzela-Ascoli theorem, we see that the geodesic maps γ_n subconverge to a geodesic ray ρ, $\rho(0) = p$. Clearly, $Im(\rho)$ is at Hausdorff distance at most $\delta + \text{dist}(p, x_0)$ from $Im(\rho))$. In particular, $\rho \sim \rho$. $\square$

Lemma 8.61 is very similar to the result in the case of X $CAT(0)$–space. The important difference with respect to that case is that the ray ρ may not be unique. Nevertheless we shall still use the notation $[p, \alpha)$ to designate a geodesic (one of the geodesics) with initial point x in the equivalence class α.

In view of this lemma, in order to understand $\partial_\infty X$ it suffices to restrict to the set $Ray_p(X)$ of geodesic rays in X emanating from $p \in X$.

It is convenient to extend the topology τ defined on $\partial_\infty X$ (i.e. the quotient topology of the compact-open topology on the set of rays) to a topology on $\bar{X} = X \cup \partial_\infty X$. Namely, we say that a sequence $x_n \in X$ converges to a point $\xi \in \partial_\infty X$ if a sequence of geodesics $[p, x_n]$ converges (uniformly on compacts) to a ray $[p, \xi)$. Then $\partial_\infty X \subset \bar{X}$ is a closed subset. Consider the set $Geo_p(X)$ consisting of geodesics in X (finite or half-infinite) emanating from p. We again quip $Geo_p(X)$ with the compact-open topology. There is a natural quotient map $Geo_p(X) \to \bar{X}$ which sends a finite geodesic or a geodesic ray emanating from p to its terminal point in $\bar{X}$.

COROLLARY 8.62. *If X is geodesic, hyperbolic and proper, then $\bar{X}$ is compact.*

PROOF. The space $Geo_p(X)$ is compact by Arzela-Ascoli theorem. Since a quotient of a compact is compact, the claim follows. $\square$

LEMMA 8.63 (Asymptotic rays are uniformly close). *Let ρ_1, ρ_2 be asymptotic geodesic rays in X such that $\rho_1(0) = \rho_2(0) = p$. 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 $T \succ t$. 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 $T \succ t$, it follows that there exists t' such that

$$\text{dist}(\rho_1(t), \rho_2(t')) \leqslant \delta.$$

By the triangle inequality, $|t - t'| \leqslant \delta$. It follows that $\text{dist}(\rho_1(t), \rho_2(t)) \leqslant 2\delta$. $\square$

COROLLARY 8.64. *$\partial_\infty X$ is Hausdorff.*

PROOF. Let ρ_n, ρ'_n be sequences of rays emanating from $p \in X$, so that $\rho_n \sim \rho'_n$ and

$$\lim_{n\to\infty} \rho_n = \rho, \quad \lim_{n\to\infty} \rho'_n = \rho'.$$

We claim that $\rho \sim \rho'$. Suppose not. Then there exists $a > 0$ so that $d(\rho(a), \rho'(a)) \geqslant 2\delta + 1$. For all sufficiently large n

$$d(\rho_n(a), \rho(a)) < 1/2, \quad d(\rho'_n(a), \rho'(a)) < 1/2,$$

while

$$d(\rho_n(a), \rho'_n(a)) \leqslant 2\delta.$$

Thus, $d(\rho(a), \rho'(a)) < 2\delta + 1$, contradicting our choice of a. $\square$

EXERCISE 8.65. Show that $\bar{X}$ is also Hausdorff.

Given a number $k > 2\delta$, define the topology τ_k on $Ray_p(X)/\sim$, where the basis of neighborhoods of a point $\rho(\infty)$ given by

$$U_{k,n}(\rho) := \{\rho' : \mathrm{dist}(\rho'(t), \rho(t)) < k, t \in [0, n]\}, n \in \mathbb{R}_+. \tag{8.5}$$

LEMMA 8.66. *Topologies τ and τ_k coincide.*

PROOF. 1. Suppose that ρ_j is a sequence of rays emanating from p such that $\rho_j \notin U_{k,n}(\rho)$ for some n. If $\lim_j \rho_j = \rho'$ then $\rho' \notin U_{k,n}$ and by Lemma 8.63, $\rho'(\infty) \neq \rho(\infty)$.

2. Conversely, if for each n, $\rho_j \in U_{k,n}(\rho)$ (provided that j is large enough), then the sequence ρ_j subconverges to a ray ρ' which belongs to each $U_{k,n}(\rho)$. Hence $\rho'(\infty) = \rho(\infty)$. ☐

LEMMA 8.67. *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 so that $d(\rho(t_i), \rho'(t_i)) \leqslant D$. Then, by Lemma 8.4, for every $t \leqslant t_i$,

$$d(\rho(t), \rho'(t)) \leqslant 2(D + \delta).$$

Since $\lim t_i = \infty$, it follows that $\rho \sim \rho'$. Contradiction. ☐

LEMMA 8.68. *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 $\xi \neq \eta$, by previous lemma, for each $R < \infty$ the set

$$K(R) := \{x \in X : \mathrm{dist}(x, \rho) \leqslant R, \mathrm{dist}(x, \rho') \leqslant R\}$$

is compact. Consider the sequences $x_n := \rho(n), x_n' := \rho'(n)$ on ρ, ρ' respectively. Since the triangles $[p, x_n, x_n']$ are δ–thin, each segment $\gamma_n := [x_n, x_n']$ contains a point within distance $\leqslant \delta$ from both $[p, x_n], [p, x_n']$, i.e. $\gamma_n \cap K(\delta) \neq \emptyset$. Therefore, by 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 η. ☐

EXERCISE 8.69. Suppose that X is δ-hyperbolic. Show that there are no complete geodesics γ in X so 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 8.70 (Ideal bigons are 2δ-thin). Suppose that α, β are geodesics in X which are both asymptotic to points $\xi, \eta \in \partial_\infty X$. Then $dist_{Haus}(\alpha, \beta) \leqslant 2\delta$. Hint: For $n \in \mathbb{N}$ define $z_n, w_n \in Im(\beta)$ 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 in X with the edges $[x_n, y_n], [y_n, w_n], [w_n, z_n], [z_n, x_n]$ is 2δ-thin.

We now compute two examples of ideal boundaries of hyperbolic spaces.

1. Suppose that $X = \mathbb{H}^n$ is the real-hyperbolic space. We claim that $\partial_\infty X$ is naturally homeomorphic to the sphere S^{n-1}, the boundary sphere of $\mathbb{H}^n$ in the unit ball model. Every ray $\rho \in Ray_o(X)$ (which is a Euclidean line segment $[o, \xi)$, $\xi \in S^{n-1}$) determines a unique point on the boundary sphere S^{n-1}, namely the point ξ. Furthermore, we claim that distinct rays $\rho_1, \rho_2 \in Ray_o(X)$ are never asymptotic. Indeed, consider the equilateral triangle $[o, \rho_1(t), \rho_2(t)]$ with the angle $\gamma > 0$ at o. Then the hyperbolic cosine law (7.4), implies that

$$\cosh(d(\rho_1(t), \rho_2(t))) = 1 + \sinh^2(t)(1 - \cos(\gamma)).$$

It is clear that this quantity diverges to ∞ as $t \to \infty$. We, thus, obtain a bijection

$$Ray_o(X) \to \partial_\infty(X).$$

We equip $Ray_p(X)$ with the topology given by the initial velocities $\rho'(0)$ of the geodesic rays $\rho \in Ray_o(X)$. Clearly, the map $Ray_o(X) \to S^{n-1}$, sending each ray $\rho = [o, \xi)$ to $\xi \in S^{n-1}$ is a homeomorphism. It is also clear that the above topology on $Ray_o(X)$ coincides with the compact-open topology on geodesic rays since the latter depend continuously on their initial velocities. Thus, the composition

$$S^{n-1} \to Ray_o(X) \to \partial_\infty X$$

is a homeomorphism.

2. Suppose that X is a simplicial tree of finite constant valence $val(X) \geqslant 3$, metrized so that every edge has unit length. As before, it suffices to restrict to rays in $Ray_p(X)$, where $p \in X$ is a fixed vertex. Note that $\rho, \rho' \in Ray_p(X)$ are equivalent if and only they are equal. We know that X is 0-hyperbolic. Our claim then is 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 fact 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 $\rho, \rho' \in Ray_p(X)$, there exist open, closed and disjoint neighborhoods $U_{n,k}(\rho), U_{n,k}(\rho')$ of the points ρ, ρ'. 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.

Gromov topology on $\bar{X} = X \cup \partial_\infty X$. The above definition of $\bar{X}$ was worked fine for geodesic hyperbolic metric spaces. Gromov extended this definition to the case when X is an arbitrary hyperbolic metric space. Pick a base-point $p \in X$. Gromov boundary $\partial_{Gromov} X$ of X consists of equivalence classes of sequences (x_n) in X so that $\lim d(p, x_n) = \infty$, where $(x_n) \sim (y_n)$ if

$$\lim_{n\to\infty} (x_n, y_n)_p = \infty.$$

One then defines the Gromov–product $(\xi,\eta)_p \in [0,\infty]$ for points ξ,η in Gromov-boundary of X by

$$(\xi,\eta)_p = \lim sup_{n\to\infty}(x_n,y_n)_p$$

where (x_n) and (y_n) are sequences representing ξ,η respectively. Then, Gromov topologizes $\bar{X} = X \cup \partial_{Gromov}X$ by:

$$\lim x_n = \xi, \xi \in \partial_{Gromov}X$$

if and only if

$$\lim_{n\to\infty}(x_n,\xi)_p = \infty.$$

It turns out that this topology is independent of the choice of p. In case when X is also a geodesic metric space, there is a natural map

$$X \cup \partial_\infty X \to X \cup \partial_{Gromov}X$$

which is the identity on X and which sends $\xi = [\rho]$ in $\partial_\infty X$ to the equivalence class of the sequence $(\rho(n))$. This map is a homeomorphism provided that X is proper.

Hyperbolic triangles with ideal vertices. We return to the case when X is a δ-hyperbolic proper geodesic metric space. We now generalize (geodesic) triangles in X to triangles where some vertices are in $\partial_\infty X$, similarly to the definitions made in section 7.3. Namely a (generalized) geodesic triangle in $\bar{X}$ is a concatenation of geodesics connecting (consecutively) 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 vertices of the triangle. As in the case of $\mathbb{H}^n$, *we do not allow two ideal vertices of a triangle T to be the same.* By abusing terminology, we will again refer to such generalized triangles as *hyperbolic triangles*.

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 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 a limit of the sides of the triangles T_i, we obtain limit geodesics connecting vertices A, B, C of T. The resulting triangle T', of course, need not be equal to T (since geodesics connecting points in $\bar{X}$ need not be unique), however, in view of Exercise 8.70, sides of T' are thin 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 5.25).

EXERCISE 8.71. Every (generalized) hyperbolic triangle T in X is 5δ-thin. In particular,

$$minsize(T) \leqslant 4\delta.$$

Hint: Use a sequence of finite triangles which coarsely converges to T and the fact that finite triangles are δ-thin.

This exercise allows one to define a *centroid* of a triangle T in X (with sides $\tau_i, i = 1,2,3$) to be a point $p \in X$ so that

$$d(p,\tau_i) \leqslant 5\delta, i = 1,2,3.$$

More generally, as in Definition 8.54, we say that a point $p \in X$ is an R-centroid T it p is within distance $\leqslant R$ from all three sides of T.

LEMMA 8.72. *Distance between any two R-centroids of a hyperbolic 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 8.55 applied to 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 $d(p, q) \leqslant 4R + 32\delta$. $\square$

We thus, define the correspondence

$$center : Trip(\partial_\infty X) \to X$$

which sends every triple of distinct points in $\partial_\infty X$ first to the set of ideal triangle T that they span and then to the set of centroid of these ideal triangles. Then Lemma 8.72 implies

COROLLARY 8.73. *For every $\xi \in Trip(\partial_\infty X)$,*

$$\operatorname{diam}(center(\xi)) \leqslant r(7\delta, \delta) = 60\delta.$$

EXERCISE 8.74. Suppose that γ_n are geodesics in X which limit to points $\zeta_n, eta_n \in \partial_\infty X$ and

$$\lim_n \zeta_n = \zeta, \lim_n \eta_n = \eta, \eta \neq \zeta.$$

Show that geodesics γ_n subconverge to a geodesic which is asymptotic to both ξ and η.

Use this exercise to conclude:

EXERCISE 8.75. If $K \subset Trip(\partial_\infty X)$ is a compact subset, then $center(K)$ is a bounded subset of X.

Conversely,

EXERCISE 8.76. Let $B \subset X$ be a bounded subset and $K \subset Trip(\partial_\infty X)$ is a subset such that $center(K) \subset B$. Show that K is relatively compact in $Trip(\partial_\infty X)$. Hint: For every $\xi \in K$, every ideal edge of a triangle spanned by ξ intersects 5δ-neighborhood of B. Now, use Arzela-Ascoli theorem.

Loosely speaking, the two exercises show that the correspondence *center* is coarsely continuous (image of a compact is bounded) and coarsely proper (preimage of a bounded subset is relatively compact).

Cone topology. Suppose that X is a proper geodesic hyperbolic metric space. Later on, it will be convenient to use another topology on $\bar{X}$, called *cone 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 8.77. We say that a sequence $x_n \in X$ converges to a point $\xi = \rho(\infty) \in \partial_\infty X$ in the **cone topology** if there is a constant C such that $x_n \in \mathcal{N}_C(\rho)$ and the geodesic segments $[x_1 x_n]$ converge to a geodesic ray asymptotic to ξ.

EXERCISE 8.78. If a sequence x_n 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 $X = \mathbb{H}^m$ in the upper half-space model, $\xi = 0 \in \mathbb{R}^{m-1}$, L is the vertical geodesic from the origin. Then a sequence $x_n \in X$ converges ξ in the cone topology if and only if all the points x_n belong to the Euclidean cone with the axis L and the Euclidean distance from x_n to 0 tends to zero. See Figure 8.4. This explains the name *cone topology*.

EXERCISE 8.79. 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 subsequence in the cone topology on $\bar{X}$.

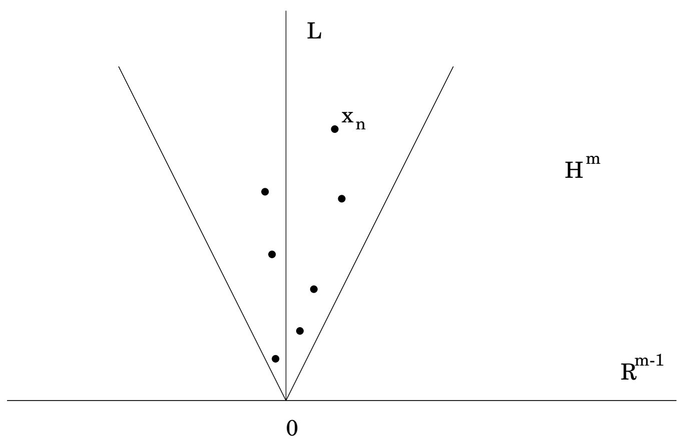


FIGURE 8.4. *C*onvergence in the cone topology.

8.10. Extension of quasi-isometries of hyperbolic spaces to the ideal boundary

The goal of this section is to explain how quasi-isometries of Rips–hyperbolic spaces extend to their ideal boundaries.

We first extend Morse lemma to the case of quasi-geodesic rays and complete geodesics.

LEMMA 8.80 (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)–quasi-geodesic. Then there is ρ^* which is either a geodesic ray or a complete geodesic in X so that the Hausdorff distance between $Im(\rho)$ and $Im(\rho^*)$ is $\leqslant \theta(L, A, \delta)$. Here θ is the function which appears in Morse lemma.*

Moreover, there are two functions $s = s(t), s^* = s^*(t)$ so that

$$L^{-1}t - B \leqslant s \leqslant Lt + B \tag{8.6}$$

and

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B) \tag{8.7}$$

and for every t, $d(\rho(t), \rho^*(s)) \leqslant \theta$, $d(\rho^*(t), \rho(s^*)) \leqslant \theta$. Here $B = A + \theta$.

PROOF. We will consider only the case of quasigeodesic rays $\rho : [0, \infty) \to X$ as the other case is similar. Let $\rho_i := \rho|[0, i]$, $i \in \mathbb{N}$. Consider the sequence of geodesic segments $\rho_i^* = [\rho(0)\rho(i)]$ as in Morse lemma. By Morse lemma,

$$dist_{Haus}(\rho_i, \rho_i^*) \leqslant \theta(L, A, \delta).$$

By properness, the geodesic segments ρ_i^* subconverge to a complete geodesic ray ρ^*. It is now clear that

$$dist_{Haus}(\rho, \rho^*) \leqslant \theta(L, A, \delta).$$

Estimates (8.6) and (8.7) follow from the estimates (8.3) and (8.4) in the case of finite geodesic segments. $\square$

COROLLARY 8.81. *If ρ is a quasi-geodesic ray as in the above lemma, there exists a point $\xi \in \partial_\infty X$ so that $\lim_{t\to\infty} \rho(t) = \xi$.*

PROOF. Take $\xi = \rho^*(\infty)$. Since $d(\rho(t), Im(\rho^*)) \leqslant \theta$, 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 quasi-geodesic ray which is 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 the nearest-point projection.

PROPOSITION 8.82 (Quasi-isometries commute with projections). *There exists $C = C(L, A, \delta)$ so that the following holds. Let X be a δ-hyperbolic geodesic metric space and let $f : X \to X$ be an (L, A)-quasi-isometry. Let α be a (finite or infinite) geodesic in X, and $\beta \subset X$ be a geodesic which is $\theta(L, A, \delta)$-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 a (finite or infinite) geodesic $\gamma \subset X$ consider the triangle $\Delta = \Delta_{x,\gamma}$ where one side is γ and x is a vertex: The other two sides are geodesics connecting x to the (finite or ideal) end-points of γ. Let $c = center(\Delta) \in \overline{\gamma}$ denote a centroid of Δ: The distance from c to each side of Δ is $\leqslant 6\delta$. By Corollary 8.56,

$$d(c, \pi_\gamma(x)) \leqslant 21\delta.$$

Applying f to the centroid $c(\Delta_{x,\alpha})$ we obtain a point $a \in X$ whose distance to each side of the quasi-geodesic triangle $f(\Delta_{x,\alpha})$ is $\leqslant 2\delta L + A$. Hence, the distance from a to each side of the geodesic triangle $\Delta_{y,\beta}, y = f(x)$ is at most $R := 2\delta L + A + D(L, A, \delta)$. Hence, a is an R-centroid of $\Delta_{y,\beta}$. By Lemma 8.72, it follows that

$$d(a, c(\Delta_{y,\beta})) \leqslant 8R + 32\delta.$$

Since $d(\pi_\beta(y), c(\Delta_{y,\beta})) \leqslant 21\delta$, we obtain:

$$d(f(\pi_\alpha(x)), \pi_\beta f(x)) \leqslant C := 21\delta + 8R + 27\delta + 21\delta L + A. \quad \square$$

Below is the main theorem of this section, which is a fundamental fact of the theory of hyperbolic spaces:

THEOREM 8.83 (Extension Theorem). *Suppose that X and X' are Rips–hyperbolic proper metric spaces. Let $f : X \to X'$ be a quasi-isometry. Then f admits a homeomorphic extension $f_\infty : \partial_\infty X \to \partial_\infty X'$. This extension is such that the map*

$f \cup f_\infty$ is continuous at each point $\eta \in \partial_\infty X$ with respect to the topology τ on $\bar{X}$. The extension satisfies the following functoriality properties:

1. For every pair of quasi-isometries $f_i : X_i \to X_{i+1}, i = 1, 2$, we have

$$(f_2 \circ f_1)_\infty = (f_2)_\infty \circ (f_1)_\infty.$$

2. For every pair of quasi-isometries $f_1, f_2 : X \to X'$ satisfying $\operatorname{dist}(f_1, f_2) < \infty$*, we have $(f_2)_\infty = (f_1)_\infty$.*

PROOF. First, we construct the extension f_∞. Let $\eta \in \partial_\infty X$, $\eta = \rho(\infty)$ where ρ is a geodesic ray in X. The image of this ray $f \circ \rho : \mathbb{R}_+ \to X'$ is a quasi-geodesic ray, hence we set $f_\infty(\eta) := f\rho(\infty)$. Observe that $f_\infty(\eta)$ does not depend on the choice of a geodesic ray asymptotic to η.

We will verify continuity for the map $f_\infty : \partial_\infty X \to \partial_\infty X$ and leave the case of $\bar{X}$ as an exercise to the reader. Let $\eta_n \in \partial_\infty X$ be a sequence which converges to η. Let ρ_n be a sequence of geodesic rays asymptotic to η_n with $\rho_n(0) = \rho(0) = x_0$. Then, by Lemma 8.66, for each $a \in \mathbb{R}_+$ there exists n_0 such that for all $n \geqslant n_0$ and $t \in [0, a]$ we have

$$d(\rho(t), \rho_n(t)) \leqslant 3\delta,$$

where δ is the hyperbolicity constant of X. Let $\rho_n' := (f \circ \rho_n)^*, \rho' := (f\rho)^*$ denote a geodesic rays given by Lemma 8.80. Thus, for all $t \in [0, a]$ there exist s and s_n,

$$L^{-1}t - A - \theta \leqslant \min(s, s_n),$$

so that

$$d(f\rho_n(t), \rho_n'(s_n)) \leqslant \theta,$$
$$d(f\rho(t), \rho'(s)) \leqslant \theta,$$

and for all $t \in [0, a]$,

$$d(f\rho_n(t), f\rho(t)) \leqslant 3\delta L + A.$$

Thus, by the triangle inequalities, for the above s, s_n we get

$$d(\rho_n'(s_n), \rho'(s)) \leqslant C = 3\delta L + A + 2\theta.$$

Since ρ_n', ρ' are geodesic, $|s - s_n| \leqslant C$. In particular, for $t = a$, and b the corresponding value of s, we obtain

$$d(\rho'(b), \rho_n'(b)) \leqslant 2C.$$

By the fellow-traveling property of hyperbolic geodesics, for all $u \in [0, b]$,

$$d(\rho'(u), \rho_n'(u)) \leqslant k := 2(2C + \delta).$$

Since $b \geqslant L^{-1}a - A - \theta$ and

$$\lim_{a \to \infty} (L^{-1}a - A - \theta) = \infty,$$

it follows that $\lim \rho_n'(\infty) = \rho'(\infty)$ in the topology τ_k. Since topologies τ and τ_k agree, it follows that $\lim_n f_\infty(\xi_n) = f_\infty(\xi)$. Hence, f_∞ is continuous.

Functoriality properties (1) and (2) of the extension are clear from the construction (in view of Morse Lemma). They also follow from continuity of the extension.

Let $\bar{f}$ be a quasi-inverse of $f : X \to X'$. Then, by the functoriality properties, $(\bar{f})_\infty$ is inverse of f_∞. Thus, extension of a quasi-isometry $X \to X'$ is a homeomorphism $\partial_\infty X \to \partial_\infty X'$. $\square$

EXERCISE 8.84. Suppose that f is merely a QI embedding $X \to X'$. Show that the continuous extension f_∞ given by this theorem is 1-1.

Remark 8.85. The above extension theorem was first proven by Efremovich and Tikhomirova in [?] for the real-hyperbolic space and, soon afterwards, reproved by Mostow [?]. We will see later on that the homeomorphisms f_∞ are *quasi-symmetric*, in particular, they enjoy certain regularity properties which are critical for proving QI rigidity theorems in the context of hyperbolic groups and spaces.

We thus obtained a functor from quasi-isometries between Rips–hyperbolic spaces to homeomorphisms between their boundaries.

The following lemma is a "converse" to the 2nd functoriality property in Theorem 8.83:

Lemma 8.86. *Let X and Y be proper geodesic δ–hyperbolic spaces. In addition we assume that centroids of ideal triangles in X form an R-net in X. Suppose that $f, f' : X \to Y$ are (L, A)–quasi-isometries such that $f_\infty = f'_\infty$ Then* $\operatorname{dist}(f, f') \leqslant$ *$D(L, A, R, \delta)$,*

Proof. Let $x \in X$ and $p \in X$ be a centroid of an ideal triangle T in X, so that $d(x, p) \leqslant R$. (Recall that p is a centroid of T if p is within distance $\leqslant 4\delta$ from all three sides of T). Then, by Lemma 8.80, $q = f(p), q' = f'(q')$ are C-centroids of the ideal geodesic triangle $S \subset Y$ whose ideal vertices are the images of the ideal vertices of T under f_∞. Here $C = 4\delta L + A + \theta(L, A, \delta)$. By Lemma 8.72, $d(q, q') \leqslant r(C, \delta)$. Therefore,

$$d(f(x), f'(x)) \leqslant D(L, A, R, \delta) = 2(LR + A) + r(C, \delta). \quad \square$$

Suppose that X is Gromov–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 quasi-homogeneous, then centroids of ideal triangles in X form a net. Thus, the above lemma applies to the real-hyperbolic space and, as we will sees soon, every non-elementary hyperbolic group.

Example 8.87. The line $X = \mathbb{R}$ is 0-hyperbolic, its ideal boundary consists of 2 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.

Corollary 8.88. *Let X be a Rips–hyperbolic space. Then the map $f \mapsto f_\infty$ (where $f : X \to X$ are quasi-isometries) descends to a homomorphism $QI(X) \to Homeo(X)$. Furthermore, under the hypothesis of Lemma 8.86, this homomorphism is injective.*

In Section **??** we will identify the image of this homomorphism in the case of real-hyperbolic space $\mathbb{H}^n$, it will be a subgroup of $Homeo(S^{n-1})$ consisting of *quasi-Moebius* homeomorphisms.

Boundary extension and quasi-actions. In view of Corollary 8.88, we have

Corollary 8.89. *Every quasi-action 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 8.90. *Suppose that X satisfies the hypothesis of Lemma 8.86 and the quasi-action $G \curvearrowright X$ is properly discontinuous. Then the kernel for the 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)-quasi-isometry of X, it follows from Lemma 8.86, that

$$\operatorname{dist}(\phi(g), id) \leqslant D(L, A, R, \delta).$$

Since ϕ was properly discontinuous, K is finite. □

Conical limit points of quasi-actions.

Suppose that ϕ is a quasi-action 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 quasi-action ϕ if there exists a sequence $g_i \in G$ so 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 $R < \infty$ so that:

- $\lim_{i\to\infty} \phi(g)(x) = \xi$.
- $d(\phi(g_i)(x), \gamma) \leqslant R$ for all i.

Lemma 8.91. *Suppose that $\psi : G \curvearrowright X$ is a cobounded quasi-action. 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 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 a ball $B = B_R(x)$ so that for every $x' \in X$ there exists $g \in G$ so that $d(x', \phi(g)(x)) \leqslant R$. Then, by coboundedness of the quasi-action ψ, there exists a sequence $g_i \in G$ so that

$$d(x_i, \phi(g_i)(x)) \leqslant R.$$

Thus, ξ is a conical limit point of the quasi-action. □

Corollary 8.92. *Suppose that G is a group and $f : X \to G$ is a quasi-isometry, $G \curvearrowright G$ is isometric action by left multiplication. Let $\psi : G \curvearrowright X$ be the quasi-action, obtained by conjugating $G \curvearrowright G$ via f. Then every point of $\partial_\infty X$ is a conical limit point for the quasi-action ψ.*

Proof. The action $G \curvearrowright G$ by left multiplication is cobounded, hence, the conjugate quasi-action $\psi : G \curvearrowright X$ is also cobounded. □

If ϕ_∞ is a topological action of a group G on $\partial_\infty X$ which is obtained by extension of a quasi-action ϕ of G on X, then we will say that *conical limit points* of the action $G \curvearrowright \partial_\infty X$ are the conical limit points for the quasi-action $G \curvearrowright X$.

8.11. Hyperbolic groups

We now come to the *raison d'être* for δ-hyperbolic spaces, namely, *hyperbolic groups.*

Definition 8.93. A finitely-generated group G is called *Gromov–hyperbolic* or *word-hyperbolic*, or simply *hyperbolic* if one of its Cayley graphs is hyperbolic.

Example 8.94. 1. Every finitely-generated free groups is hyperbolic: Taking Cayley graphs corresponding to a free generating set, we obtain a simplicial tree, which is 0-hyperbolic.

2. Finite groups are hyperbolic.

Many examples of hyperbolic groups can be constructed *via small cancelation* theory, see e.g. [**?**, **?**]. For instance, let G be a 1-relator group with the presentation

$$\langle x_1, \dots, x_n | w^m \rangle ,$$

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 [**?**, Theorem 3] before the notion of hyperbolic groups was introduced; Newman proved that for such groups G *Dehn's algorithm* applies, which is equivalent to hyperbolicity, see §8.13.)

Below is a combinatorial characterization of hyperbolic groups among Coxeter groups. Let Γ be a finite Coxeter graph and $G = C_\Gamma$ be the corresponding Coxeter group. A *parabolic subgroup* of Γ is the Coxeter subgroup defined by a subgraph Λ of Γ. It is clear that every parabolic subgroup of G admits a natural homomorphism to G; it turns out that such homomorphisms are always injective.

THEOREM 8.95 (G. Moussong [**?**]). *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 subgroup of rank 2.

PROBLEM 8.96. Is there a similar characterization of Gromov–hyperbolic groups among Shephard groups and generalized von Dyck groups?

Since changing generating set does not alter the quasi-isometry type of the Cayley graph and Rips–hyperbolicity is invariant under quasi-isometries (Corollary 8.39), we conclude that a group G is hyperbolic if and only if all its Cayley graphs are hyperbolic. Furthermore, if groups G, G' are quasi-isometric 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 Milnor-Schwarz lemma,

OBSERVATION 8.97. If G is a group acting geometrically on a Rips–hyperbolic metric space, then G is also hyperbolic.

DEFINITION 8.98. 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.

The following is an outstanding open problem in geometric group theory:

OPEN PROBLEM 8.99. Construct a hyperbolic group G which is not a $CAT(-1)$ or even a $CAT(0)$ group.

DEFINITION 8.100. A hyperbolic group is called *elementary* if it is virtually cyclic. A hyperbolic group is called non-elementary otherwise.

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 8.9).

2. A deeper fact is that if a group G contains a subgroup isomorphic to $\mathbb{Z}^2$ then G is not hyperbolic, see e.g. [**BH99**].

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. See e.g. [**BH99**].

5. Furthermore, if $Z \leqslant G$ is a central subgroup of a hyperbolic group, then either Z is finite, or G/Z is finite.

REMARK 8.101. There are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups, see Theorem 8.142. 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 8.44, H is quasi-isometrically 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 $g \in G$ to be

$$\|g\| := \lim_{n\to\infty} \frac{d(g^n, e)}{n}.$$

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**].

8.12. 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 8.83, that boundaries of different Cayley graphs are equivariantly homeomorphic. Here are two simple examples of computation of the ideal boundary.

Since $\partial_\infty \mathbb{H}^n = S^{n-1}$, we conclude that for the fundamental group G of a closed hyperbolic n-manifold, $\partial_\infty G \cong S^{n-1}$. Similarly, if $G = F_n$ is the free group of rank n, then free generating set S of G yields Cayley graph $X = \Gamma_{G,S}$ which is a simplicial tree of constant valence. Therefore, as we saw in Section 8.9, $\partial_\infty X$ is homeomorphic to the Cantor set. Thus, $\partial_\infty F_n$ is the Cantor set.

LEMMA 8.102. *Let G be a hyperbolic group and $Z = \partial_\infty G$. Then Z consists of 0, 2 or continuum of points, in which case it is perfect. In the first two cases G is elementary, otherwise G is non-elementary.*

PROOF. Let X be a Cayley graph of G. If G is finite, then X is bounded and, hence $Z =$. Thus, we assume that G is infinite. By Exercise 4.74, 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 quasi-isometric to $\mathbb{R}$ and, hence, G is 2-ended. Therefore, G is virtually cyclic by Part 3 of Theorem 6.8.

We assume, therefore, that $\mathrm{dist}_{Haus}(\gamma, X) = \infty$. Then there exists a sequence of vertices $x_n \in X$ so that $\lim \mathrm{dist}(x_n, \gamma) = \infty$. Let $y_n \in \gamma$ be a nearest vertex to x_n. Let $g_n \in G$ 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 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, so that for every $x \in \rho$, e is a nearest point on γ to x. Therefore, $\rho(\infty)$ is a point different from $\gamma(\pm\infty)$, so Z contains at least 3 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 described in Lemma 8.86. 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 8.86. It follows that K is a finite group. Since G is infinite, Z is also infinite.

Let $\xi \in Z$ and let ρ be a ray asymptotic to ξ. Then, there exists a sequence $g_n \in G$ so that $g_n(e) = x_n \in \rho$. 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 so that their images under the given sequence g_n are not all equal to ξ. Thus, ξ is an accumulation point of Z and Z is perfect. Since Z is infinite, it follows that it has cardinality continuum. □

Definition 8.103. Let Z be a compact and $G \subset Homeo(Z)$ be a subgroup. The group G is said to be a *convergence group* if G acts properly discontinuously on $Trip(Z)$, where $Trip(Z)$ is the set of triples of distinct elements of Z. A convergence group G is said to be a *uniform* if $Trip(Z)/G$ is compact.

Theorem 8.104 (P. Tukia, [**?**]). *Suppose that X is a proper δ-hyperbolic geodesic metric space with the ideal boundary $Z = \partial_\infty X$ which consists of at least 3 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 $center : Trip(Z) \to X$ sending each triple of distinct points in Z to the set of centroids of the corresponding ideal triangles. Furthermore, by Corollary 8.73, for every $\xi \in Trip(Z)$,

$$\mathrm{diam}(center(\xi)) \leqslant 60\delta.$$

Clearly, the correspondence *center* is G-equivariant. Moreover, the image of every compact K in $Trip(Z)$ under *center* is bounded (see Exercise 8.75).

Assume now that the action $G \curvearrowright X$ is geometric. Given a compact subset $K \subset Trip(Z)$, suppose 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 $g_n \in G$ so that $g_n(\xi_n) \in K$. Then the diameter of the set

$$E = \left(\bigcup_n center(\xi_n) \cup center(g_n(\xi_n)) \right) \subset X$$

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 $G \curvearrowright Trip(Z)$ is properly discontinuous.

Similarly, since $G \curvearrowright X$ is cobounded, the G-orbit of some metric ball $B(p, R)$ covers the entire X. Thus, using equivariance of *center*, for every $\xi \in Trip(Z)$, there exists $g \in G$ so that

$$center(g\xi) \subset B = B(x, R + 60\delta).$$

Since $center^{-1}(B)$ is relatively compact in $Trip(Z)$ (see Exercise 8.76), we conclude that G acts cocompactly on $Trip(Z)$. Thus, $G \subset Homeo(Z)$ is a uniform convergence group.

The proof of the converse is essentially the same argument run in the reverse. Let $K \subset Trip(Z)$ be a compact, so that G-orbit of K is the entire $Trip(Z)$. Then the set $center(K)$, which is the union of sets of centroids of points $\xi' \in K$, is a bounded subset $B \subset X$. Now, by equivariance of the correspondence *center*, it follows that G-orbit of B is the entire X. Hence, $G \curvearrowright X$ is cobounded. The argument for proper discontinuity of the action $G \curvearrowright Trip(Z)$ is similar, we just 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 $Trip(Z)$. Then proper discontinuity of the action $G \curvearrowright X$ follows from proper discontinuity of $G \curvearrowright Trip(Z)$. □

COROLLARY 8.105. *Every hyperbolic group G acts by homeomorphisms on $\partial_\infty G$ as a uniform convergence group.*

The converse to Theorem 8.104 is a deep theorem of B. Bowditch [**?**]:

THEOREM 8.106. *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 8.104 we did not really need the property that the action of G on itself was isometric, a geometric quasi-action (see Definition 5.59) suffices:

THEOREM 8.107. *Suppose that X is a δ-hyperbolic proper geodesic metric space. Assume that there exists R so that every point in X is within distance $\leqslant R$ from a centroid of an ideal triangle in X. Let $\phi : G \curvearrowright X$ be a geometric quasi-action. Then the extension $\phi_\infty : G \to Homeo(Z)$, $Z = \partial_\infty X$, of the quasi-action 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 8.104; the only difference is that ideal triangles $T \subset X$ are not mapped to ideal triangles by quasi-isometries $\phi(g), g \in G$. However, ideal quasi-geodesic triangles $\phi(g)(T)$ are uniformly close to ideal triangles which suffices for the proof. □

8.13. 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 X, the number of G-orbits of loops of length $\leqslant 10\delta$ in Γ is bounded. We attach a 2-cell along every such loop. Let X denote the resulting cell complex. Recall that for a loop γ in X, $\ell(\gamma)$ is the length of γ and $A(\gamma)$ is the least combinatorial area of a disk in X bounding γ, see Section 4.9.

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 2-dimensional cell complex whose 1-skeleton X^1 (metrized to have unit edges) is δ-hyperbolic (with $\delta \geqslant 2$ an integer) and so that for every loop γ of length $\leqslant 10\delta$ in X, $A(\gamma) \leqslant K < \infty$. Then:

The following theorem was first proven by Gromov in Section 2.3 of [**?**]:

THEOREM 8.108 (Hyperbolicity implies linear isoperimetric inequality). *Under the above assumptions, for every loop $\gamma \subset X$,*

$$A(\gamma) \leqslant K\ell(\gamma). \tag{8.8}$$

Since the argument in the proof of the theorem is by induction on the length of γ, the following result is the main tool.

PROPOSITION 8.109. *Every loop γ in $X^{(1)}$ of length larger than 10δ is a product of two loops, one of length $\leqslant 10\delta$ and another one of length $< \ell(\gamma)$.*

PROOF. We assume that γ is parameterized by its arc-length, and that it has length n.

Without loss of generality we may also assume that $\delta > 2$.

Case 1. Assume that there exists a vertex $u = \gamma(t)$ such that the vertex $v = \gamma(t + 5\delta)$ satisfies $d(u, v) < 5\delta$. By a circular change of the parameterizations of γ we may assume that $t = 0$. Let p denote the geodesic $[v, u]$ in $X^{(1)}$. We then obtain two new loops

$$\gamma_1 = \gamma([0, 5\delta]) \cup p$$

and

$$\gamma_2 = (-p) \cup \gamma([5\delta, n]).$$

Here $-p$ is the geodesic p with the reversed orientation. Since $\ell(p) < \ell(\gamma([0, 5\delta]))$, we have $\ell(\gamma_1) \leqslant 10\delta$ and $\ell(\gamma_2) < \ell(\gamma_1)$.

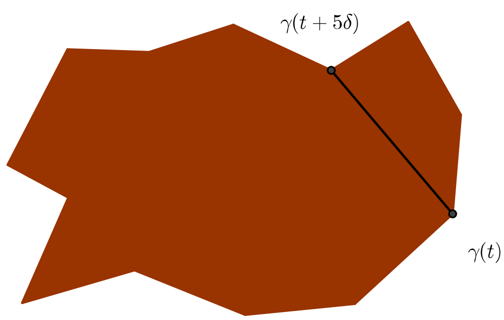


FIGURE 8.5. Case 1.

Case 2. Assume now that for every t, $d(\gamma(t), \gamma(t + 5\delta)) = 5\delta$, where $t + 5\delta$ is considered modulo n. In other words, every sub-arc of γ of length 5δ is a geodesic segment.

Let $v_0 = \gamma(0)$. Assume that $v = \gamma(t)$ is a vertex on γ whose distance to v_0 is the largest possible, in particular it is at least 5δ.

Consider the triangles $\Delta_\pm$ with the vertices $v_0, v = \gamma(t), v_\pm = \gamma(t \pm 5\delta)$. Each triangle in $X^{(1)}$ is δ-thin, therefore, $u_\pm = \gamma(t \pm (\delta + 1))$ is within distance $\leqslant \delta$ of a vertex on one of the sides $[v_0, v]$, $[v_0, v_\pm]$. If, say, u_+ is within $\leqslant \delta$ of some $w \in [v_0, v_+]$, then

$$d(v_0, v) \leqslant r + \delta + (\delta + 1) = r + 2\delta + 1,$$

$$d(v_0, v_+) = r + s \geqslant r + 3\delta - 1 > r + 2\delta + 1$$

where $r = d(v_0, w), s = d(w, v_+)$. Hence, $d(v_0, v_+) > d(v_0, v)$ which contradicts our choice of v as being farthest away from v_0. Therefore both $u_\pm$ are within distance $\leq \delta$ from *the same* point on the geodesic $[v_0, v]$ and, hence, $d(u_+, u_-) \leqslant 2\delta$. On the other hand, the distance between these vertices along the path γ is $2\delta + 2$. This contradicts our working hypothesis that every sub-arc of γ of length at most 5δ is a geodesic segment.

We have thus obtained that Case 2 is impossible. $\square$

Proof of Theorem 8.108.

The proof of the inequality (8.8) is by induction on the length of γ.

1. If $\ell(\gamma) \leqslant 10\delta$ then $A(\gamma) \leqslant K \leqslant K\ell(\gamma)$.

2. Suppose the inequality holds for $\ell(\gamma) \leqslant n$, $n \geqslant 10\delta$. If $\ell(\gamma) = n + 1$, then γ is the product of loops γ', γ'' as in Proposition 8.109: $\ell(\gamma') < \ell(\gamma)$, $\ell(\gamma'') \leqslant 10\delta$. Then, inductively,

$$A(\gamma') \leqslant K\ell(\gamma'), \quad A(\gamma'') \leqslant K,$$

and, thus,

$$A(\gamma) \leqslant A(\gamma') + A(\gamma'') \leqslant K\ell(\gamma') + K \leqslant K\ell(\gamma). \quad \square$$

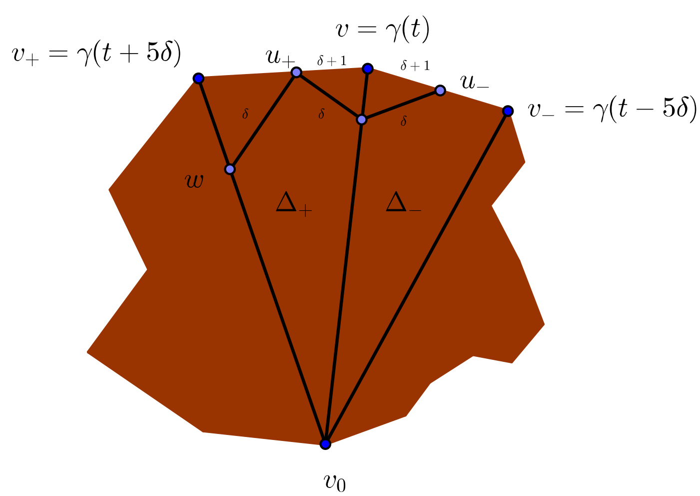


FIGURE 8.6. Case 2.

Below are two corollaries of Proposition 8.109, which was the key to the proof of the linear isoperimetric inequality.

COROLLARY 8.110 (M. Gromov, [?]). *Every hyperbolic group is finitely-presented.*

PROOF. Proposition 8.109 means that every loop in the Cayley graph of Γ is a product of loops of length $\leqslant 10\delta$. Attaching 2-cells to Γ along the G-images of these loops we obtain a simply-connected complex Y on which G acts geometrically. Thus, G is finitely-presented. $\square$

COROLLARY 8.111 (M. Gromov, [?], section 6.8N). *Let Y be a coarsely connected Rips–hyperbolic metric space. Then X satisfies linear isoperimetric inequality:*

$$Ar_{\mu}(\mathfrak{c}) \leqslant K\ell(\mathfrak{c})$$

for all sufficiently large μ and for appropriate $K = K(\mu)$.

PROOF. Quasi-isometry 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 8.109, every loop γ in Γ is a product of $\leqslant \ell(\gamma)$ loops of length $\leqslant 10\delta$, where Γ is δ-hyperbolic in the sense of Rips. Therefore, for any $\mu \geqslant 10\delta$, 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 relator. A word w is called *Dehn-reduced* if it does not contain 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' so that $u'u^{-1}$ is a relator so that $|u'| < |u|$. This, of course, does not change the element g of G represented by w. Since the length of w is decreasing on each step, eventually, we get a Dehn-reduced word v representing $g \in G$. Since $\langle X|R\rangle$ 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 word problems 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 *cyclic* Dehn reduction, where the reduction is applied to the (unbased) loop 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 8.112. *If G is δ-hyperbolic finitely-presented group then it 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 $\leqslant 10\delta$ 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 8.108. $\square$

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 $\leqslant 10\delta$) and, hence, is not *a priori* algorithmic. Nevertheless, the word problem in

δ-hyperbolic groups (with known δ) is solvable as we will see below, and, hence, a Dehn presentation is algorithmically computable.

The converse of Proposition 8.108 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 Section 8.17.

8.14. Central co-extensions of hyperbolic groups and quasi-isometries

We now consider a central co-extension

$$1 \to A \to \tilde{G} \xrightarrow{r} G \to 1 \tag{8.9}$$

with A a finitely-generated abelian group and G hyperbolic.

THEOREM 8.113. *$\tilde{G}$ is QI to $A \times G$.*

PROOF. In the case when $A \cong \mathbb{Z}$, the first published proof belongs to S. Gersten [**?**], although, it appears that D.B.A. Epstein and G. Mess also knew this result. Our proof follows the one in [**?**]. First of all, since an epimorphism with finite kernel is a quasi-isometry, 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 (8.9). We first consider the case when $A \cong \mathbb{Z}$. Each fiber $r^{-1}(g), g \in G$, is a copy of $\mathbb{Z}$ and, therefore, has a natural order denoted $\leq$. We let ι denote the embedding $\mathbb{Z} \cong A \to \tilde{G}$. We let $\mathcal{X}$ denote a symmetric generating set of $\tilde{G}$ and use the same name for its image under s. We let $\langle \mathcal{X} | \mathcal{R} \rangle$ be a finite presentation of G. Let $|w|$ denote the word length with respect to this generating set, for $w \in \mathcal{X}^*$, where $\mathcal{X}^*$ is the set of all words in $\mathcal{X}$, as in Section 4.2. Lastly, let $\tilde{w}$ and $\bar{w}$ denote the elements of $\tilde{G}$ and G represented by $w \in \mathcal{X}^*$.

LEMMA 8.114. *There exists $C \in \mathbb{N}$ so that for every $g \in G$ there exists*

$$r(g) := \max\{\tilde{w}\iota(-C|w|) : w \in X^*, \bar{w} = g\}.$$

Here the maximum is taken with respect to the natural order on $s^{-1}(g)$.

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 $R \in \mathcal{X}^*$ so 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 so that

$$\iota(T) = \max\{\tilde{R} : R^{\pm 1} \text{ is a defining relator of } G\}.$$

We then claim that for each $u \in \mathcal{X}^*$ representing the identity in G,

$$\iota(TArea(u)) \geqslant \tilde{u} \in A. \tag{8.10}$$

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 $w = h^{-1}Rh$,

$$\tilde{w} \leqslant \iota(T)$$

where R is a defining relator of G and $h \in \mathcal{X}^*$. The latter inequality follows from the fact that the multiplications by $\bar{h}$ and $\bar{h}^{-1}$ determine an order isomorphism and 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 (8.10),

$$\tilde{q} = \tilde{w}\tilde{u} \leqslant \iota(C|q|) = \iota(C|w|) + \iota(C|u|).$$

We now switch to the addition notation for $A \cong \mathbb{Z}$. Then,

$$w - v \leqslant \iota(C|w|) + \iota(C|v|),$$

and

$$w - \iota(C|w|) \leqslant v + \iota(C|v|).$$

Therefore, taking v to be a fixed word representing g, we conclude that all the differences $w - \iota(C|w|)$ are bounded from above. Hence their maximum exists. $\square$

Consider the section s (given by Lemma 8.114) of the exact sequence (8.9). A 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

$$\sigma(g_1, g_2) := s(g_1)s(g_2) - s(g_1 g_2)$$

where the difference is taking place in $r^{-1}(g_1 g_2)$. The next lemma does not use hyperbolicity of G, only the definition of s.

LEMMA 8.115. *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 $w_1 x$ also represents gx. Therefore, by the definition of s,

$$\widetilde{w_1 x}\iota(-C(|w_1| + 1)) \leqslant \tilde{w}_2\iota(-C|w_2|).$$

Hence, there exists $a \in A, a \geqslant 0$, so that

$$\widetilde{w_1}\iota(-C(|w_1|)\widetilde{x}\iota(-C)a = \tilde{w}_2\iota(-C|w_2|)$$

and, thus

$$s(g)\widetilde{x}\iota(-C)a = s(gx). \tag{8.11}$$

Since $w_2 x^{-1}$ represents g, we similarly obtain

$$s(gx)\widetilde{x}^{-1}\iota(-C)b = s(g), b \geqslant 0, b \in A. \tag{8.12}$$

By combining equations (8.11) and (8.12) and switching to the additive notation for the group operation in A we get

$$a + b = \iota(2C).$$

Since $a \geqslant 0, b \geqslant 0$, we conclude that $-\iota(C) \leqslant a - \iota(C) \leq \iota(C)$. Therefore, (8.11) 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. $\square$

REMARK 8.116. Actually, more is true: There exists a section $s' : G \to \tilde{G}$ so that $\sigma'(G, G)$ is a finite set. This follows from the fact that all (degree $\geqslant 2$) cohomology classes of hyperbolic groups are *bounded* (see [**?**]). However, the proof is more difficult and we will not need this fact.

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 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

$$\phi(\tilde{g}) = \tilde{g} - s \circ r(\tilde{g}). \tag{8.13}$$

It is immediate that ϕ is Lipschitz since s is Lipschitz.

We now extend the above 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 a central co-extension (8.9). 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

$$1 \to A_i \to \tilde{G}_i \xrightarrow{r_i} G \to 1. \tag{8.14}$$

Assuming that each central co-extension (8.14) has a Lipschitz section s_i, we obtain the corresponding Lipschitz projection $\phi_i : \tilde{G}_i \to A_i$ given by (8.13). This yields a Lipschitz projection

$$\Phi : \tilde{G} \to A, \Phi = (\phi_1 \circ \eta_1, ..., \phi_n \circ \eta_n).$$

We now set

$$s(r(\tilde{g})) := \tilde{g} - \Phi(\tilde{g}).$$

It is straightforward to verify that s is well-defined and that it is Lipschitz provided that each s_i is. We thus obtain

Corollary 8.117. *Given a finitely-generated free abelian group A and a hyperbolic group G, each central co-extension (8.9) admits a Lipschitz section $s : G \to \tilde{G}$ and a Lipschitz projection $\Phi : \tilde{G} \to A$ given by*

$$\Phi(\tilde{g}) = \tilde{g} - s(r(\tilde{g})).$$

We then 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 quasi-isometry. $\square$

Remark 8.118. The above proof easily generalizes to the case of an arbitrary finitely-generated group G and a central co-extension (8.9) given by a *bounded 2-nd cohomology class* (see e.g. [**?, ?, EF97a**] for the definition): One has to observe only that each cyclic central co-extension

$$1 \to A_i \to \tilde{G}_i \to G \to 1$$

is still given by a bounded cohomology class. We refer the reader to [**?**] for the details.

Example 8.119. 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 an co-extension is nilpotent but not virtually abelian. Hence, by Pansu's theorem (Theorem **??**), it is not quasi-isometric to $G \times A = \mathbb{Z}^3$.

One can ask if Theorem 8.113 generalizes to other normal co-extensions of hyperbolic groups G. We note that Theorem 8.113 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 *quasi-isometric 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 quasi-isometries of the product $G = G_1 \times ... \times G_n$ quasi-preserve the product structure:

THEOREM 8.120. *Let $\pi_j : G \to G_j, j = 1, ..., n$ be natural projections. Then for each (L, A)–quasi-isometry $f : G \to G$, there is $C = C(G, L, A) < \infty$, so that, up to a composition with a permutation of quasi-isometric factors G_k, the map f is within distance $\leqslant C$ from a product map $f_1 \times ... \times f_n$, where each $f_i : G_i \to G_i$ is a quasi-isometry.*

8.15. 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 8.121 (§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 ω-lim $\lambda_n = 0$, the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ 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 $\mathrm{Cone}_\omega(X, \boldsymbol{e}, (n))$ is a real tree.*

Then (X, dist) is hyperbolic.

The proof of Proposition 8.121 relies on the following lemma.

LEMMA 8.122. *Assume that a geodesic metric space (X, dist) satisfies either property (a) or property (b) in Proposition 8.121. Then there exists $M > 0$ such that for every geodesic triangle $\Delta(x, y, z)$ with $\mathrm{dist}(y, z) \geqslant 1$,the two edges with endpoint x are at Hausdorff distance at most $M\mathrm{dist}(y, z)$.*

PROOF. Suppose to the contrary that there exist sequences of triples of points x_n, y_n, z_n, such that $\mathrm{dist}(y_n, z_n) \geqslant 1$ and

$$\mathrm{dist}_{Haus}([x_n, y_n], [x_n, z_n]) = M_n \mathrm{dist}(y_n, z_n),$$

such that $M_n \to \infty$. Let a_n be a point on $[x_n, y_n]$ such that

$$\delta_n := \mathrm{dist}(a_n, [x_n, z_n]) = \mathrm{dist}_{Haus}([x_n, y_n], [x_n, z_n]).$$

Since $\delta_n \geqslant M_n$, it follows that $\delta_n \to \infty$.

Suppose condition **(a)** holds. Consider the sequence of base-points $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ and the sequence of scaling constants $\boldsymbol{\delta'} = (1/\delta_n)_{n\in\mathbb{N}}$. In the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\delta'})$, the limits of $[x_n, y_n]$ and $[x_n, z_n]$ are at Hausdorff distance 1.

The triangle inequalities imply that the limits

$$\omega\text{-lim} \frac{\mathrm{dist}(y_n, a_n)}{\delta_n} \text{ and } \omega\text{-lim} \frac{\mathrm{dist}(z_n, a_n)}{\delta_n}$$

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. All these cases are impossible in a real tree.

Suppose condition **(b)** holds. Let $\mathcal{S} = \{\lfloor \delta_n \rfloor \; ; \; n \in \mathbb{N}\}$, where $\lfloor \delta_n \rfloor$ is the integer part of δ_n. By Exercise **??**, there exists ω such that $\omega(\mathcal{S}) = 1$. Consider (x'_m), (y'_m), (z'_m) and (a'_m) defined as follows. For every m in the set $\mathcal{S}$ 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 $\mathcal{S}$ make an arbitrary choice for the entries of all four sequences.

In $\mathrm{Cone}_\omega(X, \boldsymbol{a}', (m))$ the limits $\omega\text{-lim}[x'_m, y'_m]$ and $\omega\text{-lim}[x'_m, z'_m]$ are as in one of the three cases discussed in the previous case, all cases being forbidden in a real tree. $\square$

Proof of Proposition 8.121. Suppose that the geodesic space X is not hyperbolic. For every triangle Δ in X and a point $a \in \Delta$ we define the quantity $d_\Delta(a)$, which is the distance from a to union of the two sides of Δ which do not contain a (if a lies on all three sides then we set $\epsilon(a) = 0$). Then for every $n \in \mathbb{N}$ there exists a geodesic triangle $\Delta_n = \Delta(x_n, y_n, z_n)$, and a point a_n on the edge $[x_n, y_n]$ such that

$$d_n = d_{\Delta_n}(a_n) \geqslant n.$$

For each Δ_n we then will choose the point a_n in Δ_n which maximizes the function d_{Δ_n}. After relabelling the vertices, we may assume that $a_n \in [x_n, z_n]$ and that $d_n = \mathrm{dist}(a_n, [y_n, z_n]) = \mathrm{dist}(a_n, b_n)$, where $b_n \in [y_n, z_n]$. Let δ_n be equal to $\mathrm{dist}(a_n, [x_n, z_n]) = \mathrm{dist}(a_n, c_n)$, for some $c_n \in [x_n, z_n]$. By hypothesis $\delta_n \geqslant d_n$.

Suppose condition **(a)** is satisfied. In the asymptotic cone $\mathbf{K} = \mathrm{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\lambda})$, where $\boldsymbol{a} = (a_n)$ and $\boldsymbol{\lambda} = (1/d_n)$ we look at the limit of Δ_n. There are two cases:

A) $\omega\text{-lim} \frac{\delta_n}{d_n} < +\infty$.

By Lemma 8.122, we have that $\mathrm{dist}_{Haus}([a_n, x_n], [c_n, x_n]) \leqslant M \cdot \delta_n$. 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 $a = \omega\text{-lim}\, a_n \in [x, y]$ is such that $\mathrm{dist}(a, [x, z] \cup [y, z]) \geqslant 1$, 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

$$\omega\text{-lim} \frac{\mathrm{dist}(a_n, x_n)}{d_n} = +\infty \text{ and } \omega\text{-lim} \frac{\mathrm{dist}(a_n, z_n)}{d_n} = +\infty.$$

By Lemma 8.122, we have $\mathrm{dist}_{Haus}([a_n, y_n], [b_n, y_n]) \leqslant M \cdot d_n$. Thus, the respective limits of the sequences of segments $[x_n, y_n]$ and $[y_n, z_n]$ are either two rays of origin $y = \omega\text{-lim}\, y_n$ or two complete geodesics asymptotic on one side. We denote them $\overline{xy}$ and $\overline{yz}$, respectively, with $y \in \mathbf{K} \cup \partial_\infty \mathbf{K}, x, z \in \partial_\infty \mathbf{K}$. The limit of $[x_n, z_n]$ 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 distance at most d_n from $[x_n, y_n] \cup [x_n, z_n]$. This implies that all points on the ray $\overline{bz}$ are at distance at most 1 from $\overline{xy}$. It follows that $\overline{xy}$ and $\overline{yz}$ are either asymptotic rays emanating from y or complete geodesics asymptotic on both sides, and they are at 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. Let $\mathcal{S} = \{\lfloor d_n \rfloor \, ; \, n \in \mathbb{N}\}$, and let ω be a non-principal ultrafilter such that $\omega(\mathcal{S}) = 1$ (see Exercise **??**). 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 $\mathrm{Cone}_\omega(X, \boldsymbol{a}', (m))$, with $\boldsymbol{a}' = (a'_m)$ we may consider the limit of the triangles (Δ'_m), argue as previously, and obtain a contradiction to the fact that the cone is a real tree. It follows that the condition **(b)** also implies the hyperbolicity of X. □

REMARK 8.123. An immediate consequence of Proposition 8.121 is an alternative proof of the quasi-isometric invariance of Rips-hyperbolicity among geodesic metric spaces: A quasi-isometry between two spaces induces a bi-Lipschitz map between asymptotic cones, and a topological space bi-Lipschitz equivalent to a real tree is a real tree.

As a special case, consider Proposition 8.121 in the context of hyperbolic groups: A finitely-generated group 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 [**?**], 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 8.124 (M.Kapovich, B.Kleiner [**?**]). *Let G be 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 of this theorem which we owe to Thomas Delzant. We will need the following

THEOREM 8.125 (B. Bowditch, [**?**], Theorem 8.1.2). *For every δ there exists δ' so 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 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 where each cell is isometric to a Euclidean simplex, satisfies this condition for every $m > 0$. We refer to [**?**, 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, so that one of its asymptotic cones is a 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 quasi-isometric to G. Suppose that ω is an ultrafilter, (λ_n) is a scaling sequence converging to zero, and X_ω is the asymptotic cone of X with respect to this sequence, so that X_ω is isometric to a tree. Consider the sequence of metric spaces $X_n = (X, \lambda_n \mathrm{dist})$. Then, since X_ω is a tree, by taking a diagonal sequence, there exists a pair of sequences r_n, δ_n with

$$\omega\text{-}\lim r_n = \infty, \quad \omega\text{-}\lim \delta_n = 0$$

so that for ω-all n, the 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 8.125, 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 Gromov-hyperbolic as well. □

We now continue discussion of properties of trees which appear as asymptotic cones of hyperbolic spaces.

PROPOSITION 8.126. *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-points with valence continuum.*

PROOF. STEP 1. By Theorem 5.29, the group G is finitely generated and hyperbolic and every Cayley graph Γ of G is quasi-isometric to X. It follows that there exists a bi-Lipschitz bijection between asymptotic cones

$$\Phi : \mathrm{Cone}_\omega(G, \mathbf{1}, \boldsymbol{\lambda}) \to \mathrm{Cone}_\omega(X, \boldsymbol{x}, \boldsymbol{\lambda}),$$

where x is an arbitrary base-point in X, and $\mathbf{1}, \boldsymbol{x}$ denote the constant sequences equal to $1 \in G$, respectively to $x \in X$. Moreover, $\Phi(\mathbf{1}_\omega) = \boldsymbol{x}_\omega$. The map Φ thus determines a bijection between the space of directions $\Sigma_{\mathbf{1}_\omega}$ in the cone of Γ and the space of directions $\Sigma_{\boldsymbol{x}_\omega}$ in the cone of X. It suffices therefore to prove that the set $\Sigma_{\mathbf{1}_\omega}$ has the cardinality of continuum. For simplicity, in what follows we denote the asymptotic cone $\mathrm{Cone}_\omega(G, \mathbf{1}, \boldsymbol{\lambda})$ by G_ω.

STEP 2. We show that the geodesic rays joining 1 to distinct points of $\partial_\infty \Gamma$ give distinct directions in $\mathbf{1}_\omega$ in the asymptotic cone.

Let $\rho_i : [0, \infty) \to \Gamma, i = 1, 2$ be geodesic rays, $\rho_i(0) = 1, i \in \{1, 2\}, \rho_1(\infty) = \alpha, \rho_2(\infty) = \beta$, where $\alpha \neq \beta$. For every t and s in $[0, \infty)$, we consider

$$a_t = \omega\text{-}\lim \rho_1(t/\lambda_n) \text{ and } b_s = \omega\text{-}\lim \rho_2(s/\lambda_n),\ a_t,\ b_s \in \Gamma_\omega\,.$$

We have

$$\mathrm{dist}(a_t, b_s) = \omega\text{-}\lim \lambda_n \mathrm{dist}(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n)) =$$

$$\omega\text{-}\lim\, [t + s - 2\lambda_n(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n))_1] = t + s,$$

because the sequence of Gromov products

$$(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n))_1$$

ω-converges to a constant. The two limit rays, ρ_1^ω and ρ_2^ω, of the rays ρ_1 and ρ_2, defined by $\rho_1^\omega(t) = a_t,\ \rho_2^\omega(s) = b_s$, have only the origin in common and give therefore distinct directions in $\mathbf{1}_\omega$.

We thus have found an injective map from $\partial_\infty \Gamma$ to $\mathbf{1}_\omega$

STEP 3. We argue that every direction of Γ_ω in $\mathbf{1}_\omega$ is determined by a sequence of geodesic rays emanating from 1 in Γ. This argument was suggested to us by P. Papasoglu.

An arbitrary direction of Γ_ω in $\mathbf{1}_\omega$ is the germ of a geodesic segment with one endpoint in $\mathbf{1}_\omega$, and this segment is the limit set of a sequence of geodesic segments of Γ with one endpoint in 1, with lengths growing linearly in $\frac{1}{\lambda_n}$.

LEMMA 8.127. *Every sufficiently long geodesic segment in a Cayley graph of a hyperbolic group is contained in the M-neighborhood of a geodesic ray, where M depends only on the Cayley graph.*

PROOF. According to [**?**] and to [**ECH$^+$92**, Chapter 3, §2], given a Cayley graph Γ of a hyperbolic group G, there exists a finite directed graph $\mathcal{G}$ with edges labeled by the generators of G such that every geodesic segment in Γ corresponds to a path in $\mathcal{G}$. If a geodesic segment is long enough, the corresponding path contains at least one loop in $\mathcal{G}$. The distance from the endpoint of the path to the last loop is bounded by a constant M which depends only on the graph $\mathcal{G}$. Let ρ be the geodesic ray obtained by going around this loop infinitely many times. The initial segment is contained in $\mathcal{N}_M(\rho)$. □

We conclude that every direction of Γ_ω in $\mathbf{1}_\omega$ is the germ of a limit ray. We then have a surjective map from the set of sequences in $\partial_\infty G$ to $\Sigma_{[\mathbf{1}_\omega]}$:

$$\{(\alpha_n)_{n\in\mathbb{N}}\,;\ \alpha_n \in \partial_\infty\Gamma\} = (\partial_\infty\Gamma)^{\mathbb{N}} \to \Sigma_{[\mathbf{1}_\omega]}.$$

Steps 2 and 3 imply that for a non-elementary hyperbolic group, the cardinality of $\Sigma_{[\mathbf{1}_\omega]}$ is continuum, . □

A. Dyubina–Erschler and I. Polterovich ([**?**], [**?**]) have shown a stronger result than Proposition 8.126:

THEOREM 8.128 ([**?**], [**?**]). *Let $\mathcal{A}$ be the $2^{\aleph_0}$–universal tree, as defined in Theorem 8.20.*

(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 sectional curvature at most $-k$, $k > 0$ a fixed constant, is isometric to $\mathcal{A}$.*

A consequence of Theorem 8.128 is that asymptotic cones of non-elementary hyperbolic groups and of complete, simply connected Riemannian manifold with strictly negative sectional curvature cannot be distinguished from one another.

8.16. Size of loops

The characterization of hyperbolicity with asymptotic cones allows one to define hyperbolicity in terms of size of its closed loops, in particular of the size of its geodesic triangles. Throughout this section X denotes a geodesic metric space. One parameter that measures the size of geodesic triangles is the minimal size introduced in Definition 5.49 for topological triangles. Only now, the three arcs that we consider are the three geodesic edges of the triangles. With this we can define the *minsize* function of a geodesic metric space X:

DEFINITION 8.129. The *m*inimal size function,

$$\mathrm{minsize} = \mathrm{minsize}_X : \mathbb{R}_+^* \to \mathbb{R}_+^*,$$

$$\mathrm{minsize}(\ell) = \sup\{\mathrm{minsize}(\Delta)\,;\ \Delta \text{ a geodesic triangle of perimeter } \leqslant \ell\}\,.$$

Note that according to (8.1), if X is δ–hyperbolic in the sense of Rips, the function minsize is bounded 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 [**?**, §6] that a 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 statement is straightforward:

PROPOSITION 8.130. *A geodesic metric space X is hyperbolic if and only if* $\mathrm{minsize}(\ell) = o(\ell)$.

PROOF. As noted above, the direct part follows from Lemma 8.51. Conversely, assume that $\mathrm{minsize}(\ell) = o(\ell)$. We begin by proving that in an arbitrary asymptotic cone of X every finite geodesic is a limit geodesic, in the sense of Definition **??**. More precisely:

LEMMA 8.131. *Let $\mathfrak{g} = [a_{\pm}, b_\omega]$ be a finite geodesic in $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ and assume that $a_\omega = (a_i), b_\omega = (b_i)$. Then for every geodesic $[a_i, b_i] \subset X$ connecting a_i to b_i, ω-lim$[a_i, b_i] = \mathfrak{g}$.*

PROOF. Let $c_\omega = (c_i)$ be an arbitrary point on $\mathfrak{g}$. Consider an arbitrary triangle $\Delta_i \subset X$ with vertices a_i, b_c, c_i. Let ℓ_i be the perimeter of Δ_i. Since $\omega\text{-lim}\, \lambda_i \ell_i < \infty$ and $minsize(\Delta_i) = o(\ell_i)$, we get

$$\omega\text{-lim}\, \lambda_i minsize(\Delta_i) = 0.$$

Taking the points x_i, y_i, z_i on the sides of Δ_i realizing the minsize of Δ_i, we conclude:

$$\omega\text{-lim}\, \lambda_i \,\mathrm{diam}(x_i, y_i, z_i) = 0.$$

Let $\{x_\omega\} = \omega\text{-lim}\{x_i, y_i, z_i\}$. Then

$$\mathrm{dist}(a_\omega, b_\omega) \leqslant \mathrm{dist}(a_\omega, x_\omega) + \mathrm{dist}(x_\omega, b_\omega) \leqslant$$

$$\mathrm{dist}(a_\omega, x_\omega) + \mathrm{dist}(x_\omega, b_\omega) + 2\mathrm{dist}(x_\omega, c_\omega) = \mathrm{dist}(a_\omega, c_\omega) + \mathrm{dist}(c_\omega, b_\omega)\,.$$

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 $c_\omega \in \omega\text{-lim}[a_i, b_i]$ and lemma follows. □

If one asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is not a real tree then it contains a geodesic triangle Δ which is not a tripod. Without loss of generality we may assume that the geodesic triangle is a simple loop. By the above lemma, the geodesic triangle is an ultralimit of a family of geodesic triangles $(\Delta_i)_{i \in I}$ with perimeters of the order $O\left(\frac{1}{\lambda_i}\right)$. The fact that $\mathrm{minsize}(\Delta_i) = o\left(\frac{1}{\lambda_i}\right)$ implies that the three edges of Δ have a common point, a contradiction. □

M. Gromov in [**?**, Proposition 6.6.F] proved the following version of Proposition 8.130:

THEOREM 8.132. *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*

$$\mathrm{minsize}(\Delta) \leqslant \varepsilon_0 \cdot \mathrm{perimeter}(\Delta)\,,$$

then X is hyperbolic.

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.

Let $\lambda \in \left(0, \frac{1}{2}\right)$. For a simple 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 $\lambda\ell$.

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 8.133 ([?], 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 $\ell > 1$,*

$$\mathrm{constr}_\lambda(\ell) \leqslant 2\delta\left[\log_2(\ell + 28\delta) + 6\right] + 2\,.$$

REMARK 8.134. One cannot obtain a better order than $O(\log \ell)$ for the general constriction function. This can be seen by considering, in the half-space model of $\mathbb{H}^3$, the horizontal circle of length ℓ.

PROOF. We begin by arguing that (2) implies (1). In what follows we define *limit triangles* in an asymptotic cone $Cone(X) = \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, to be the triangles in $Cone(X)$ whose edges are limit geodesics. Note that such triangles *a priori* need not be themselves limits of sequences of geodesic triangles in X.

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 it is a simple triangle. This triangle is the limit of a family of geodesic hexagons $(H_i)_{i\in I}$, with three edges of lengths of order $O\left(\frac{1}{\lambda_i}\right)$ alternating with three edges of lengths of 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 ω-$\lim H_i$ is not simple, a contradiction.

It remains to prove that every finite geodesic in every asymptotic cone is a limit geodesic. Let $\mathfrak{g}([a_\omega, b_\omega])$ be a geodesic in a cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, where $a_\omega = (a_i)$ and $b_\omega = (b_i)$; let $c_\omega = (c_i)$ be an arbitrary point on $\mathfrak{g}$. By the previous argument every limit geodesic triangle with vertices $a_\omega, b_\omega, c_\omega$ is a tripod. If c_ω does not coincide with the center of this tripod then this implies that

$$\mathrm{dist}(a_\omega, c_\omega) + \mathrm{dist}(c_\omega, b_\omega) \geqslant \mathrm{dist}(a_\omega, b_\omega),$$

a contradiction. Thus, $c_\omega \in \omega$-$\lim[a_i, b_i]$ and, hence, $\mathfrak{g} = \omega$-$\lim[a_i, b_i]$.

We thus proved that every geodesic triangle in every asymptotic cone of X is a tripod, hence every asymptotic cone is a real tree. Hence, 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 $\lambda = \frac{1}{4}$.

Consider an arbitrary simple 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. We denote $\mathrm{constr}_{\frac{1}{4}}(\mathfrak{c})$ simply by constr. Let x, y, z be three points on $\mathfrak{c}(\mathbb{S}^1)$ which are endpoints of arcs $\alpha_{xy}, \alpha_{yz}, \alpha_{zx}$ in $\mathfrak{c}(\mathbb{S}^1)$ so that the first two arcs have lenth $\frac{\ell}{4}$. Let $t \in \alpha_{zx}$ be the point minimizing the distance to y in X. Clearly,

$$R := \mathrm{dist}(y, t) \geqslant \mathrm{constr}, \quad R \leqslant d(x, y), \quad R \leqslant d(z, y).$$

The point t splits the arc $\alpha_{z,x}$ into two sub-arcs $\alpha_{z,t}, \alpha_{t,x}$. Without loss of generality, we can assume that length of $\alpha_{t,x}$ is $\geqslant \frac{\ell}{4}$. In partcicular, $d(x', t) = 2r \geqslant$ constr. Let $\alpha_{xx'}$ be the maximal subarc of α_{xy} disjoint from the interior of $B(y, r)$ (we allow $x = x'$). As $d(x', t) \geqslant$ constr, lemma 8.59 implies that

$$\ell \geqslant \ell(\alpha_{tx'}) \geqslant 2^{\frac{r-1}{2\delta}-3} - 12\delta,$$

and, thus,

$$\mathrm{constr} \leqslant 4\delta\left(\log_2(\ell + 12\delta) + 3\right) + 2$$

The inequality in (3) follows. □

8.17. Filling invariants

Recall that for every μ-simply connected geodesic metric space X we defined (in Section 5.4) the *filling area function* (or, *isoperimetric function*) $A(\ell) = A_X(\ell)$ (this function, technically speaking, depends on the choice of μ), which computes upper bound on the areas of disks bounding loops of lengths $\leqslant \ell$ in X. We also defined the *filling radius function* $r(\ell)$ which computes upper bounds on radii of such disks. The goal of this section is to relate both invariants to hyperbolicity of the sapce X. Recall also that hyperbolicity implies linearity of $A_X(\ell)$, see Corollary 8.111.

There is a stronger version of this (converse) statement. This version states that there is a gap between the quadratic filling order and the linear isoperimetric order: As soon as the isoperimetric inequality is less than quadratic, it has to be linear and the space has to be hyperbolic:

THEOREM 8.135 (Subquadratic filling, §2.3, §6.8, [**?**]). *If a coarsely simply-connected geodesic metric space X the isoperimetric function $A_X(\ell) = o(\ell^2)$, then the space is hyperbolic.*

Note that there is a second gap for the possible filling orders of groups.

REMARK 8.136 ([**?**], [**?**]). If a finitely presented group G has Dehn function $D(\ell) = o(\ell)$, then G is either free or finite.

Proofs of Theorem 8.135 can be found in [**?**], [**?**], [**?**] and [**?**]. B. Bowditch makes use of only two properties of the area function in his proof: The *quadrangle* (or *Besikovitch*) inequality (see Proposition 5.48) and a certain *theta–property.* In fact, as we will see below, only the quadrangle inequality or its triangle counterpart, the minsize inequality (see Proposition5.50) are needed. Also, we will see it suffices to have subquadratic isoperimetric function for geodesic triangles.

Proof of Theorem 8.135. Let X be a μ-simply-connected geodesic metric space and A_X be its isoperimetric function and $\mathrm{minsize}_X : \mathbb{R}_+ \to \mathbb{R}_+$ be the minsize function, see Definition 8.129. According to Proposition 5.50, for every $\delta \geqslant \mu$,

$$[\mathrm{minsize}_X(\ell)]^2 \leqslant \frac{\delta^2}{2\pi} A_X(\ell)\,,$$

whence $A_X(\ell) = o(\ell)$ implies $\mathrm{minsize}_X(\ell) = o(\ell)$. Proposition 8.130 then implies that X is hyperbolic. □

The strongest known version of the converse to Corollary 8.111 is:

THEOREM 8.137 (Strong subquadratic filling theorem,see §2.3, §6.8 of [?], and also [?], [?]). *Let X be a δ-simply connected geodesic metric space. If there exist sufficiently large N and L $\epsilon > 0$ sufficiently small, such that every loop $\mathfrak{c}$ in X with $N \leqslant \mathrm{Ar}_\delta(\mathfrak{c}) \leqslant LN$ satisfies*

$$\mathrm{Ar}_\delta(\mathfrak{c}) \leqslant \epsilon[length(\mathfrak{c})]^2\,,$$

then the space X is hyperbolic.

It seems impossible to prove this theorem using asymptotic cones.

In Theorem 8.137 it suffices to consider only geodesic triangles Δ instead of all closed curves, and to replace the condition $N \leqslant Ar_\delta(\Delta) \leqslant LN$ by *length* $(\Delta) \geqslant N$. This follows immediately from Theorem 8.132 and the minsize inequality in Proposition 5.50.

M. Coornaert, T. Delzant and A. Papadopoulos have shown that if X is a complete simply connected Riemannian manifold which is *reasonable* (see [?, 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 [?, Chapter 6, Theorem 2.1].

In terms of the multiplicative constant, a sharp inequality was proved by S. Wenger.

THEOREM 8.138 (S. Wenger [?]). *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 $\mathfrak{d} : D^2 \to X$ with*

$$Area(\mathfrak{d}) \leqslant \frac{1-\varepsilon}{4\pi}\,\mathrm{length}(\mathfrak{c})^2\,.$$

Then X is Gromov hyperbolic.

In the Euclidean space one has the classical isoperimetric inequality

$$Area(\mathfrak{d}) \leqslant \frac{1}{4\pi}\,\mathrm{length}(\mathfrak{c})^2\,,$$

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 8.138 is a generalization of the notion of the geometric area used in this book. If the Lipschitz map $\phi : D^2 \to X$ 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 Section 2.1.4. 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

$$Area(\phi) = \int_{D^2} |J_x\phi(x)|.$$

Another application of the results of Section 8.16 is a description of asymptotic behavior of the filling radius in hyperbolic spaces.

PROPOSITION 8.139 ([?], §6, [?], §3). *In a geodesic μ-simply connected metric space X the following statements are equivalent:*

(1) *X is hyperbolic;*

(2) *the filling radius $r(\ell) = o(\ell)$;*

(3) *the filling radius $r(\ell) = O(\log \ell)$.*

Furthermore, in (3) one can say that 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 $r(\mathfrak{d}) = O(\log \ell)$.

REMARK 8.140. The logarithmic order in (3) cannot be improved, as shown by the example of the horizontal circle in the half-space model of $\mathbb{H}^3$. We note that the previous result shows that, as in the case of the filling area, there is a gap between the linear order of the filling radius and the logarithmical one.

PROOF. In what follows, we let $\mathrm{Ar} = \mathrm{Ar}_\mu$ denote the μ-filling area function in the sense of Section 5.4, 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 8.111), there exists a constant K depending only on X such that

$$\mathrm{Ar}(\mathfrak{c}) \leqslant K \ell_X(\mathfrak{c}) \tag{8.15}$$

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-simpleces respectively in this complex. Thus, for a loop $\mathfrak{c}$ as above, we have

$$\ell_X(\mathfrak{c}) \leqslant \mu \; length(\mathcal{C}),$$

where $\mathcal{C}$ is the triangulation of the circle 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 denfine 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 Section 3.2.1) of 2-dimensional simplices σ in $\mathcal{D}$ so 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 D^2 is a union of several disjont 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 $\mathcal{D}'_j \subset \mathcal{D}$, so that $\mathcal{D}_j \subset \mathcal{D}'_j$. Let length$(\partial\mathcal{D}_j)$ and length(s_j) denote the number of edges of $\partial\mathcal{D}_j$ and of s_j respectively.

By 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 $\mathfrak{d}|_{\mathfrak{f}_j}$. 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 that (2) $\Rightarrow$ (1).

We first show that (2) implies that in an 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 seen as a loop $\mathfrak{c}_i : \mathbb{S}^1 \to \Delta_i$, and can be filled with a μ-disk $\mathfrak{d}_i : \mathcal{D}^{(1)} \to X$ of filling radius $r_i = r(\mathfrak{d}_i) = o\,(\mathrm{length}\,(\Delta_i))$. In particular, $\omega\text{-}\lim_i \lambda_i r_i = 0$.

Let $[x_i, y_i], [y_i, z_i]$ and $[z_i, x_i]$ 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 $[y_i, z_i]$. 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 $[y_i, x_i] \cup [x_i, z_i]$. 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 $[y_i, x_i] \cup [x_i, z_i]$. See Figure 8.7.

Consider an arbitrary vertex $\overline{u}$ on $\mathbb{S}^1$ between $\overline{y}_i, \overline{z}_i$ and its image $u \in [y_i, z_i]$. We have that $\mathfrak{p}_i \subset \overline{\mathcal{N}}_{r_i+\mu}([y_i, u]) \cup \overline{\mathcal{N}}_{r_i+\mu}([u, z_i])$, where $[y_i, u]$ and $[u, z_i]$ are sub-geodesics of $[y_i, z_i]$.

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 $\mathrm{dist}(u_1, u_2) \leqslant 2(r_i + \mu)$ it follow that, say, $\mathrm{dist}(u_1, u) \leqslant r_i + \mu$, whence $\mathrm{dist}(u, u') \leqslant 3(r_i + \mu)$. 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 $\overline{\mathcal{N}}_{4r_i+4\mu}\,([y_i, x_i] \cup [x_i, z_i])$.

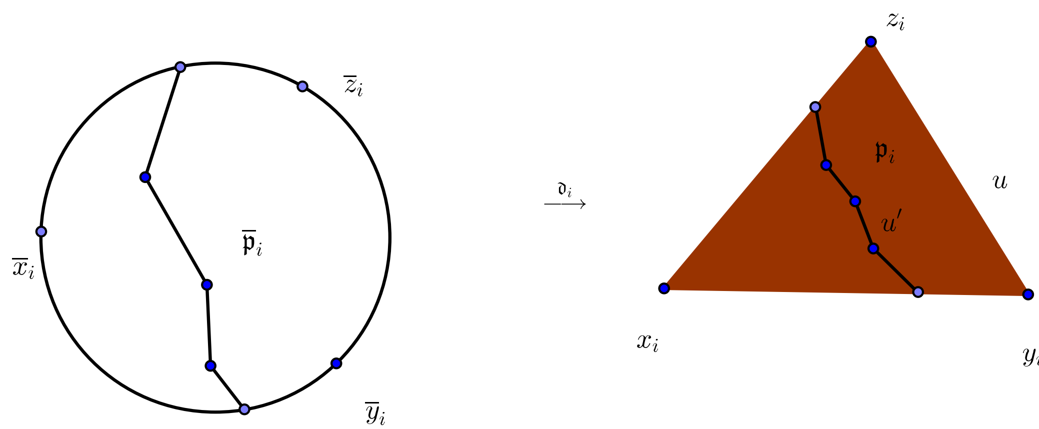


FIGURE 8.7. The path $\overline{\mathfrak{p}}_i$ and its image $\mathfrak{p}_i$.

This implies that in $\mathbf{\Delta}$ one edge is contained in the union of the other two. The same argument done for each edge implies that $\mathbf{\Delta}$ 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 $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ 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 Δ_i with vertices x_i, y_i, z_i yield a tripod $\Delta_\omega = \Delta(x_\omega, y_\omega, z_\omega)$ in the asymptotic cone, but since,

$$\mathrm{dist}(x_\omega, z_\omega) + \mathrm{dist}(z_\omega, y_\omega) = \mathrm{dist}(x_\omega, y_\omega),$$

it follows that the tripod must be degenerate. Thus $z_\omega \in \omega\text{-}\lim[x_i, y_i]$. $\square$

Like for the area, for the radius too there is a stronger version of the implication *sublinear radius* $\Longrightarrow$ *hyperbolicity*, similar to Theorem 8.137.

PROPOSITION 8.141 (M. Gromov; P. Papasoglou [?]). *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 [?], the best possible constant expected is not $\frac{1}{73}$, but $\frac{1}{8}$. Note that the proof of Proposition 8.141 cannot be extended from groups to metric spaces, because it relies on the bigon criterion for hyperbolicity [?], 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.

8.18. Rips construction

The goal of this section is to describe Rips construction which associates a hyperbolic group with to an arbitrary finite presentation.

THEOREM 8.142 (Rips Construction, I. Rips [**?**]). *Let Q be a group with a finite presentation $\langle A|R\rangle$. Then, with such presentation of Q one can associate 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. We will give here only a sketch of the argument. Let $A = \{a_1, ..., a_m\}$, $R = \{R_1, ..., R_n\}$. For $i = 1, ..., m, j = 1, 2$, pick even natural numbers $r_i < s_i$, $p_{ij} < q_{ij}$, $u_{ij} < v_{ij}$, so 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 R_k. Define the group G by the presentation P where generators are $a_1, ..., a_m, b_1, b_2$, and relators are:

$$R_i b_1 b_2^{r_2} b_1 b_2^{r_i+1} \cdots b_1 b_2^{s_i}, i = 1, ..., n \tag{8.16}$$

$$a_i^{-1} b_j a_i b_1 b_2^{u_{ij}} b_1 b_2^{u_{ij}+1} \cdots b_1 b_2^{v_{ij}}, i = 1, ..., m, j = 1, 2, \tag{8.17}$$

$$a_i b_j a_i^{-1} b_1 b_2^{p_{ij}} b_1 b_2^{p_{ij}+1} \cdots b_1 b_2^{q_{ij}}, i = 1, ..., m, j = 1, 2. \tag{8.18}$$

Now, define the map $\phi : G \to Q$, $\phi(a_i) = a_i, \phi(b_j) = 1, j = 1, 2$. Clearly, ϕ respects all the relators and, hence, it determines an epimomorphism $\phi : G \to Q$. 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 (8.17) and (8.18). Thus, indeed, b_1, b_2 generate K.

The reason that the group G is hyperbolic is that the presentation written above is Dehn: because of the choices of the numbers r_i etc., when we multiply conjugates of the relators of G, we cannot cancel more than half of one of the relators (8.16) — (8.18), namely, the product of generators b_1, b_2 appearing in the end of each relator. This argument is a typical example of application of the small cancelation theory, see [**LS77**]. Rips in his paper [**?**], did not use the language of hyperbolic groups, but the language of the small cancelation theory.

One then verifies that G has cohomological dimension 2 by showing that the presentation complex Z of the presentation P of the group G is aspherical, for this one can use, for instance, [**?**].

Now, R. Bieri proved in [**Bie76b**, Theorem B] that if G is a group of cohomological dimension 2 and $H \lhd G$ is a 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 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. Similar argument shows that K cannot be cyclic, so K is trivial and, hence, $b_1 = b_2 = 1$ in G. However, this clearly contradicts the fact that the presentation (8.16) — (8.18) is a Dehn presentation (since the words b_1, b_2 obviously do not contain more than half of the length of any relator). $\square$

In particular, there are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups. Furthermore,

COROLLARY 8.143. *Hyperbolic groups could 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 N if and only if g maps trivially to Q. Since Q has unsolvable word problem, the problem of membership of g in N is unsolvable as well. $\square$

On the other hand, the membership problem is solvable for quasiconvex subgroups, see Theorem 8.163.

8.19. 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:

$$d_\rho(x) := \max_k d(\rho(g_k)(x), x) \tag{8.19}$$

and set

$$d_\rho := \inf_{x \in X} d_\rho(x).$$

This function does not necessarily have minimum, so we choose $x_\rho \in X$ to be a point so that

$$d_\rho(x) - d_\rho \leqslant 1.$$

Such points x_ρ are called *min-max* points of ρ for obvious reason. 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 $\rho(G)$.

EXAMPLE 8.144. 1. Let $X = \mathbb{H}^n$, $G = \langle g \rangle$ be infinite cyclic group, $\rho(g) \in \mathrm{Isom}(X)$ is a hyperbolic translation along a geodesic $L \subset X$ by some amount $t > 1$, e.g. $\rho(g)(x) = e^t x$ in the upper half-space model. Then $d_\rho = t$ and we can take $x_\rho \in L$, since the set of points of minima of $d_\rho(x)$ is L.

2. Suppose that $X = \mathbb{H}^n = U^n$ and G are the same but $\rho(g)$ is a parabolic translation, e.g. $\rho(g)(x) = x + u$, where $u \in \mathbb{R}^{n-1}$ is a unit vector. Then d_ρ does not attain minimum, $d_\rho = 0$ and we can take as x_ρ any point $x \in U^n$ so that $x_n \geqslant 1$.

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 8.145. 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 subset topology.

EXERCISE 8.146. Show that 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 topology on $Hom(G, \mathrm{Isom}(X))$ to the Tychonoff topology on the infinite product.

Suppose now that the metric space X is proper. Pick a base-point $o \in X$. Then 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 8.147. Let $G = \langle g \rangle$ is infinite cyclic, $X = \mathbb{H}^n$. Show that trivial representation $\rho_0 : G \to 1 \in \mathrm{Isom}(X)$ and representation ρ_1 where $\rho_1(g)$ acts as a parabolic translation, project to points $[\rho_i]$ in $Rep(G, \mathrm{Isom}(X))$, so that every neighborhood of $[\rho_0]$ contains $[\rho_1]$.

EXERCISE 8.148. 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 $H \subset \mathrm{Isom}(X)$, one defines

$$Hom_D(G, H) = Hom_D(G, \mathrm{Isom}(X)) \cap Hom(G, H).$$

LEMMA 8.149. *Suppose that $H \subset \mathrm{Isom}(X)$ is a closed subgroup whose action on X is cobounded. Then for every $D \in \mathbb{R}_+$, $Rep_D(G, H) = Hom_D(G, H)/H$ 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 $\sigma \in H$ so 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 so that $x_\rho \in \bar{B}(o,R)$. 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. $\square$

In view of this lemma, even if X is not proper, we say that a sequence $\rho_i : G \to \mathrm{Isom}(X)$ *diverges* if

$$\lim_{i\to\infty} d_{\rho_i} = \infty.$$

DEFINITION 8.150. We say that an isometric action of a group on a real tree T is *nontrivial* if the group does not fix a point in T.

PROPOSITION 8.151 (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 a nontrivial 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 $Cone_\omega(X, \mathbf{P}, \lambda)$ of the space X; here $\mathbf{p} = (p_i)$. According to Lemma 8.35, the metric space $\mathbf{X}$ in this asymptotic cone is a real tree T. Furthermore, the sequence of group actions ρ_i converges to an isometric action $\rho_\omega : G \curvearrowright T$:

$$\rho_\omega(g)(x_\omega) = (\rho_i(x_i)),$$

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

$$|d_{\rho_i} - d(\rho_i(g)(p_i)| \leqslant 1$$

in X. Thus, the element $\rho_\omega(g)$ will move the point $\mathbf{p} \in T$ exactly by 1. Because p_i was a min-max point of ρ_i, it follows that

$$d_{\rho_\omega} = 1.$$

In particular, the action $\rho_\omega : G \curvearrowright T$ has no fixed point, i.e., is nontrivial. $\square$

One of the important applications of this proposition is

THEOREM 8.152 (F. Paulin, [**?**]). *Suppose that G is a finitely-generated group with 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. Then, by the above proposition, if $Hom(G,H)/H$ is noncompact, then G has a nontrivial 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, so for every generator g of G, the sequence $\rho_i(g)$ is eventually constant. Therefore, the entire sequence (ρ_i) consists of only finitely many representations. Contradiction. Thus, $Hom(G,H)/H$ is finite. $\square$

This theorem is one of many results of this type: Bounding number of homomorphisms from a group to a hyperbolic group. Having Property FA is a very strong restriction on the group, so, typically one improves Proposition 8.151 by making stronger assumptions on representations $G \to H$ and, accordingly, stronger conclusions about the action of G on the tree, for instance:

THEOREM 8.153. *Suppose that H is a hyperbolic group, X is its Cayley graph and all the representations $\rho_i : G \to H$ are faithful. Then the resulting nontrivial action of G on a real tree is* small*, i.e., stabilizer of every nontrivial geodesic segment is virtually cyclic.*

The key ingredient then is *Rips Theory* which converts small actions (satisfying some mild restrictions which will hold in the case of groups G which embed in hyperbolic groups) $G \curvearrowright T$, to decompositions of G as an amalgam $G_1 \star_{G_3} G_2$ or HNN-extension $G = G_1 \star_{G_3}$, where the subgroup G_3 is again virtually cyclic. Thus, one obtains:

THEOREM 8.154 (I. Rips, Z. Sela, [?]). *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 $Out(G) = Aut(G)/G$ 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 lack of finiteness is the fact that one can precompose homomorphisms $G \to H$ with automorphisms of G itself:

THEOREM 8.155 (I. Rips, Z. Sela, [?]). *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.

8.20. Further properties of hyperbolic groups

1. Hyperbolic groups are ubiquitous:

THEOREM 8.156 (See e.g. [?]). *Let G be a non-elementary δ-hyperbolic group. Then there exists N, so that for every collection $g_1, .., g_k \in G$ of elements of norm $\geq 1000\delta$, the following holds:*

i. The subgroup generated by the elements g_i^N and all their conjugates is free.

ii. Then the quotient group $G/\langle\langle g_1^n, ...g_k^n\rangle\rangle$ 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. Furthermore,

THEOREM 8.157 (A. Ol'shanskii, [?]). *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 8.158 (A. Ol'shanskii, [?], T. Delzant [?]). *Every non-elementary hyperbolic group G is SQ-universal, i.e., every countable group embeds in a quotient of G.*

"Most" groups are hyperbolic:

THEOREM 8.159 (A. Ol'shanskii [?]). *Fix $k \in \mathbb{N}$, $k \geq 2$ and let $A = \{a^{\pm 1}, a^{\pm 2}, ..., a_k^{\pm 1}\}$ be an alphabet. Fix $i \in \mathbb{N}$ and let $(n_1, ..., n_i)$ be a sequence of natural numbers. Let $N = N(k, i, n_1, ..., n_i)$ be the number of group presentations*

$$G = \langle a_1, ..., a_k | r_1, ..., r_i \rangle$$

such that $r_1, ..., r_i$ are reduced words in the alphabet A such that the length of r_j is n_j, $j = 1, 2, , i$. If N_h is the number of hyperbolic groups in this collection and if $n = min\{n_1, .., n_i\}$, 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, we refer the reader to [?], [?], [?], [?] for further discussion of random groups.

Theorems 8.160, 8.161, 8.162 below first appeared in Gromov's paper [?]; other proofs could be found for instance in [?], [**BH99**], [**ECH$^+$92**], [**ECH$^+$92**], [?].

2. Hyperbolic groups have finite type:

THEOREM 8.160. *Let G be δ-hyperbolic. Then there exists $D_0 = D_0(\delta)$ so that for all $D \geq D_0$ the Rips complex $\mathrm{Rips}_D(G)$ is contractible. In particular, G has type F_∞.*

3. Hyperbolic groups have controlled torsion:

THEOREM 8.161. *Let G be hyperbolic. Then G contains only finitely many conjugacy classes of finite subgroups.*

4. Hyperbolic groups have solvable algorithmic problems:

THEOREM 8.162. *Every δ-hyperbolic group has solvable word and conjugacy problems.*

Furthermore:

THEOREM 8.163 (I. Kapovich, [?]). *Membership problem is solvable 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 solvable.*

Isomorphism problem is solvable:

THEOREM 8.164 (Z. Sela, [?]; F. Dahmani and V. Guirardel [?]). *Given two δ-hyperbolic groups G_1, G_2, there is an algorithm to determine if G_1, G_2 are isomorphic.*

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.

5. Hyperbolic groups are hopfian:

THEOREM 8.165 (Z. Sela, [?]). *For every hyperbolic group G and every epimorphism $\phi : G \to G$, $Ker(\phi) = 1$.*

Note that every 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 linear group is residually finite, but there are nonlinear hyperbolic groups, see [?]. It is very likely that some (or even all) of the nonlinear hyperbolic groups described in [?] are not residually finite.

6. Hyperbolic groups tend to be co-Hopfian:

THEOREM 8.166 (Z. Sela, [?]). *For every 1-ended hyperbolic group G, every monomorphism $\phi : G \to G$ is surjective, i.e., such G is co-Hopf.*

7. All hyperbolic groups admit QI embeddings in the real-hyperbolic space $\mathbb{H}^n$:

THEOREM 8.167 (M. Bonk, O. Schramm [?]). *For every hyperbolic group G there exists n, such that G admits a quasi-isometric embedding in $\mathbb{H}^n$.*

CHAPTER 9

Tits' Alternative

In this chapter we will prove

THEOREM 9.1 (Tits' Alternative, [**?**]). *Let L be a Lie group with finitely many connected components and $\Gamma \subset L$ be a finitely generated subgroup. Then either Γ is virtually solvable or Γ contains a free nonabelian subgroup.*

REMARK 9.2. In the above one cannot replace 'virtually solvable' by 'solvable'. Indeed consider the Heisenberg group $H_3 \leqslant GL(3,\mathbb{R})$ and $A_5 \leqslant GL(5,\mathbb{R})$. The group $\Gamma = H_3 \times A_5 \leqslant GL(8,\mathbb{R})$ is not solvable (because A_5 is simple) and does not contain a free nonabelian subgroup (because it has polynomial growth).

COROLLARY 9.3. *Suppose that Γ is a finitely generated subgroup of $GL(n,\mathbb{R})$. Then Γ has either polynomial or exponential growth.*

PROOF. By 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 **??**. $\square$

9.1. 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 [**?**] and [**OV90**].

The proof of the following general lemma is straightforward, and left as an exercise to the reader.

LEMMA 9.4. *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 9.5. A commutative ring is called *noetherian* if it satisfies one (hence both) statements in Lemma 9.4.

Note that a field seen as a ring is always noetherian. Other examples of noetherian rings come from the following

THEOREM 9.6 (Hilbert's ideal basis theorem, see [**?**]). *If A is a noetherian ring then the ring of multivariable polynomials $A[X_1, ..., X_n]$ is also noetherian.*

From now on, we fix a field $\mathbb{K}$.

DEFINITION 9.7. 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 $\mathbb{K}$:

$$Z = \{(x_1, .., x_n) \in \mathbb{K}^n \; ; \; p_j(x_1, .., x_n) = 0, j \in J\}.$$

We will frequently say "algebraic subset" when referring to affine algebraic set.

For instance, the algebraic subsets in the affine line (1-dimensional vector space V) 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 $K[X_1, ..., X_n]$:

$$Z \mapsto I_Z = \{p \in K[X_1, ..., X_n] \; ; \; p|_Z \equiv 0\} .$$

Note that I_Z is the kernel of the homomorphism $p \mapsto p|_Z$ from $K[X_1, ..., X_n]$ to the ring of functions on Z. Thus, the ring $K[X_1, ..., X_n]/I_Z$ 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 9.6 and Lemma 9.4 imply the following.

LEMMA 9.8. (1) *Every algebraic set is defined by finitely many equations.*

(2) *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* $i \geqslant N$.

The pair $(Z, \mathbb{K}[Z])$ (a *ringed space*) is an *affine algebraic variety* or simply an *affine variety*, or, by abusing the terminology, just a *(sub)variety*. We will frequently conflate affine varieties and the corresponding algebraic subsets.

DEFINITION 9.9. A *morphism* between two affine varieties Y in $\mathbb{K}^n$ and Z in $\mathbb{K}^m$ is a map of the form $\varphi : Y \to Z$, $\varphi = (\varphi_1, ..., \varphi_m)$, such that φ_i is in $\mathbb{K}[Y]$ for every $i \in \{1, 2, ..., m\}$.

Note that every morphism is induced by a morphism $\tilde{\varphi} : \mathbb{K}^n \to \mathbb{K}^m$, $\tilde{\varphi} = (\tilde{\varphi}_1, ..., \tilde{\varphi}_m)$, with $\tilde{\varphi}_i : \mathbb{K}^n \to \mathbb{K}$ a polynomial function for every $i \in \{1, 2, ..., m\}$.

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 9.10. 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 $Z = Z_1 \cup ... \cup Z_\ell$ be a finite union of algebraic subsets, Z_i defined by a set of polynomials P_i, $i \in \{1, ..., \ell\}$. According to Lemma 9.8, (1), we may take each P_i to be finite. Define the new set of polynomials

$$P := \left\{ p = \prod_{i=1}^{\ell} p_i;\ \ p_i \in P_i \text{ for every } i \in \{1, ..., \ell\} \right\}.$$

The solution set of the system of equations $p = 0$, $p \in P$, is Z.

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 9.10, 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 9.8, Part (2), implies that the closed sets in Zariski topology satisfy the descending chain condition (DCC).

DEFINITION 9.11. A topological space such that the closed sets satisfy the DCC (or, equivalently, with the property that the open sets satisfy the ACC) is called *noetherian*.

LEMMA 9.12. *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:

$$Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_n \supseteq \ldots$$

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,

$$\overline{Z}_1 \supseteq \overline{Z}_2 \supseteq \cdots \supseteq \overline{Z}_n \supseteq \ldots$$

stabilizes, hence, so does the chain of the subsets Z_i. $\square$

PROPOSITION 9.13. *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 $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n}$; hence, there exists $Z_{i_{n+1}}$ with $i_{n+1} \in I$ such that $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n} \cap Z_{i_{n+1}}$ is a non-empty proper closed subset of $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n}$. $\square$

We now discuss a strong version of connectedness, relevant in the setting of noetherian spaces.

LEMMA 9.14. *For a topological space X the following properties are equivalent:*

(1) *every open non-empty subset of X is dense in X;*

(2) *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 9.15. A topological space is called *irreducible* if it is non-empty and one of (hence all) the properties in Lemma 9.14 is (are) satisfied. A subset of a topological space is *irreducible* if, when endowed with the subset topology, it is an irreducible space.

EXERCISE 9.16. (1) Prove that $\mathbb{K}^n$ with Zariski topology is irreducible.

(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 9.17. (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. Nevertheless, one has:

LEMMA 9.18. *Let V_1, V_2 be finite-dimensional vector spaces over $\mathbb{K}$ and $Z_i \subset V_i, i = 1, 2$, 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* projection to the second factor) and, hence, 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:

$$\{f_{i,z} = 0 : i = 1, \ldots, k, z \in Z_1\}.$$

Therefore, A_1 is a subvariety. This contradicts irreducibility of Z_2. $\square$

Lemma 9.19. *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 9.14, (2), imply the equivalence.

(2) Now let U, V be two open sets in A. Then $U = A \cap U_1$ and $V = A \cap V_1$, where U_1, V_1 are open in X. Since $U_1 \cap \overline{Y} \neq \emptyset$ and $V_1 \cap \overline{Y} \neq \emptyset$ it follows that both U_1 and V_1 have non-empty intersections with Y. Then irreducibility of Y implies that $U_1 \cap V_1 \cap Y$ is non-empty, whence $U \cap V \neq \emptyset$.

(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 9.14, (2).

It follows by Zorn's Lemma that $\mathcal{I}_Y$ has a maximal element.

(4) follows from (1) and (3). □

Theorem 9.20. *A noetherian topological space X is a union of finitely many distinct maximal irreducible subsets $X_1, X_2, ..., X_n$ 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 $X_1, X_2, ..., X_n$. 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 9.19, (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 $Y_1, Y_2, ..., Y_m$ 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 $j_i \in \{1, 2, ..., n\}$ such that $Y_i = X_{j_i}$. The map $i \mapsto j_i$ is injective, and if some $k \in \{1, 2, ..., n\}$ 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 9.21. The subsets X_i defined in Theorem 9.20 are called *the irreducible components* 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 9.22. 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 G.

EXAMPLE 9.23.

(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 to an algebraic subgroup of $SL(n+1, \mathbb{K})$ 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 9.24. *If Γ is a subgroup of $SL(V)$ then its Zariski closure $\bar{\Gamma}$ in $SL(V)$ is also a subgroup.*

PROOF. Consider the map $f : SL(V) \to SL(V)$ given by $f(\gamma) = \gamma^{-1}$. Then f is a polynomial isomorphism and, hence, $f(\bar{\Gamma})$ is Zariski closed in $SL(V)$. Since Γ is a subgroup, $f(\bar{\Gamma})$ contains Γ. Thus, $\bar{\Gamma} \cap f(\bar{\Gamma})$ is a Zariski closed set containing Γ. It follows that $\bar{\Gamma} = f(\bar{\Gamma})$ and hence $\bar{\Gamma}$ is stable under the inversion. The argument for the multiplication is similar. $\square$

If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, then $V = \mathbb{K}^n$ also has the *standard* or *classical* topology, given by the Euclidean metric 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 9.25 (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.*

PROPOSITION 9.26. *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.*

REMARK 9.27. Proposition 9.26, (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 $\{xy = 0\} \subset \mathbb{K}^2$.

PROOF. (1) Let $X_1, ..., X_k$ be irreducible components of G containing the identity. According to Lemma 9.18, 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 9.19, (3), and by Theorem 9.20 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. Therefore, $gG_0 = G_0$. 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. □

In what follows we list some useful properties of algebraic groups. We refer the reader to [**OV90**] for the details:

1. A complex or real algebraic group is a complex, respectively real, Lie group.

2. Every Lie group G (resp. algebraic group over a field $\mathbb{K}$), contains a *radical* Rad$G$, which is the largest connected (resp. irreducible) solvable normal Lie (resp. algebraic) subgroup of G. The radical is the same if the group is considered with its real or its complex Lie structure. A group with trivial radical is called *semisimple.*

3. The quotient of an algebraic group by its radical is an algebraic semisimple group.

4. The commutator subgroup of an irreducible algebraic group is an irreducible algebraic subgroup. An irreducible algebraic semisimple group coincides with its commutator subgroup.

5. One of the most remarkable properties of algebraic semisimple groups is the following: given such a group G and its representation as a linear group $G \hookrightarrow GL(V)$, the space V decomposes into a direct sum of G-invariant subspaces so that the restriction of the action of G to any of these subspaces is irreducible, i.e. there are no proper G-invariant subspaces.

6. From the classification of normal subgroups in a semisimple connected Lie group (see for instance [**OV90**, Theorem 4, Chapter 4, §3]) it follows that the image

of an algebraic irreducible semisimple group under an algebraic homomorphism is an algebraic irreducible semisimple group.

As an application of 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 9.28. *The subset of monomorphisms $F_2 \to SU(2)$ is dense in the variety $Hom(F_2, SU(2))$.*

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 4.38), 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 3.8, $SU(2)$ 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. $\square$

Since $SU(2)/\pm I$ is isomorphic to $SO(3)$, we obtain

LEMMA 9.29. *The subset of monomorphisms $F_2 \to SO(3)$ is dense in the variety $Hom(F_2, SO(3))$.*

9.2. Virtually solvable subgroups of $GL(n,\mathbb{C})$

This and the following section deal with virtually solvable subgroups of the general linear group and limits of sequences of such groups. This material (namely, Theorem 9.45 or the weaker Proposition 9.44 that will also suffice) will be needed in the proof of the Tits' Alternative.

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* $0 \subset V_1 \subset \ldots \subset V_n = V$ 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 9.30. *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. $\square$

For a vector space V over $\mathbb{K}$ we let $P(V)$ denote the corresponding projective space:

$$P(V) = (V \setminus \{0\})/\mathbb{K}^*.$$

LEMMA 9.31. *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 $\mathrm{Fix}(G)$. $\square$

In what follows, the topology on subgroups of $GL(V)$ is always the Zariski topology, in particular, connectedness always means Zariski–connectedness.

THEOREM 9.32 (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 9.30, 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 $\mathcal{E}(g)$ element-wise, 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 needed for the proof of Borel's theorem.

LEMMA 9.33. *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. $\square$

LEMMA 9.34. *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 $G < Z_{GL(V)}$ 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. $\square$

Similarly, we have:

LEMMA 9.35. *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. $\square$

We can now prove Theorem 9.32. Lemma 9.33 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 9.31, $\mathrm{Fix}(G') \subset P(V)$ 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 9.34. This concludes the proof of Theorem 9.32. □

The following is a converse to Theorem 9.32:

PROPOSITION 9.36. *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 9.37. Theorem 9.32 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 9.38. *There exist functions $\nu(n), \delta(n)$ so 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 9.20).

LEMMA 9.39. *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 (**??**). 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 9.40. *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 9.36. ☐

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 9.32 and Lemma 9.31). 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 9.41. 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 **??**. 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 **??**), we obtain a metabelian group $\tilde{A} < \Gamma$ whose center is contained in $Z_{GL(V)}$. Now the assertion follows from Lemma 9.35.

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 **??**, 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 9.36, $L = \mathrm{Ker}(\phi \times \psi)$ 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

$$\begin{array}{ccccccc} 1 \to & L & \longrightarrow & N & \overset{\phi\times\psi}{\longrightarrow} & M \times N_W & \longrightarrow \\ & || & & \uparrow \iota' & & \uparrow \iota & \\ 1 \to & L & \longrightarrow & \widehat{G} & \overset{\phi\times\psi}{\longrightarrow} & M \times \widehat{G_W} & \longrightarrow \end{array}$$

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 9.38 follows. $\square$

9.3. Limits of sequences of virtually solvable subgroups of $GL(n, \mathbb{C})$

Throughout this section, all vector spaces under consideration will be complex and finite-dimensional.

We say that a subgroup $G < GL(V)$, $V \cong \mathbb{C}^n$, is *virtually reducible* if G contains a finite index subgroup H which has reducible action on V. A subgroup which is not virtually reducible is called *virtually irreducible.* Recall that modules of dimension 1 are regarded as reducible by default.

REMARK 9.42. In order to distinguish this notion of irreducibility from the irreducibility in the context of algebraic groups, we will refer to the later as *Zariski–irreducibility.*

LEMMA 9.43. *Let $G \leqslant GL(V)$ be a subgroup which is not virtually solvable. Then G contains a finite index subgroup H which admits an H-module W, which is either a submodule or quotient module of $H \curvearrowright V$, such that $H \curvearrowright W$ is virtually irreducible.*

PROOF. The proof is by induction on the dimension of V. The statement is clear if V is 1-dimensional. Suppose it holds in all dimensions $< n$. If G itself is virtually irreducible, we are done. Otherwise, we take a finite index subgroup $G_1 < G$ so that the $G_1 \curvearrowright V$ is reducible. Let $W \subset V$ be a G_1-invariant subspace. If the images of G_1 in $GL(W)$ and $GL(V/W)$ are both virtually solvable, then G is itself virtually solvable. If one of these images is not virtually solvable, the statement follows from the induction hypothesis. $\square$

PROPOSITION 9.44. *Let $\Gamma \leqslant GL(n, \mathbb{C})$ be a finitely-generated virtually irreducible subgroup. Then there exists a neighborhood Ξ of id in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho \in \Xi$ has image which is not virtually solvable.*

PROOF. Suppose to the contrary that there exists a sequence

$$\rho_j \in \mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$$

converging to id, so that each $\Gamma_j := \rho_j(\Gamma)$ is virtually solvable. Since each Γ_j is virtually solvable, by Remark 9.41 it contains a reducible subgroup of index $\leqslant n!c(n)$. Let $\Phi < \Gamma$ denote the intersection of the preimages of these subgroups under ρ_j's. Clearly, $|\Gamma : \Phi| < \infty$. After passing to a subsequence, we may assume that each Γ_j preserves a proper projective subspace $P_j \subset \mathbb{CP}^{n-1}$ of a fixed dimension k. By passing to a further subsequence, we may assume that the subspaces P_j converge to a proper projective subspace $P \subset \mathbb{CP}^{n-1}$. Since each Γ_j preserves P_j, the group Φ also preserves P. Hence, $\Gamma \curvearrowright V$ is virtually reducible, contradicting our assumptions. $\square$

Although the above proposition will suffice for the proof of the Tits' Alternative, we will prove a slightly stronger assertion:

THEOREM 9.45. *Let $\Gamma \subset GL(n, \mathbb{C})$ be a finitely-generated subgroup which is not virtually solvable. Then there exists a neighborhood Σ of id in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho \in \Sigma$ has image which is not virtually solvable.*

PROOF. We argue analogously to the proof of Proposition 9.44. Suppose to the contrary that there exists a sequence $\rho_j \in \mathrm{Hom}(\Gamma, G)$ converging to id, so that each $\Gamma_j := \rho_j(\Gamma)$ is virtually solvable. By Theorem 9.38, for each j there exists a subgroup $\Lambda_j \leqslant \Gamma_j$ of index $\leqslant \nu(n)$ which is solvable of derived length $\leqslant d = \delta(n)$.

Let $\Lambda \leqslant \Gamma$ denote the intersection of $\rho_j^{-1}(\Lambda_j)$. Again, $|\Gamma : \Lambda| < \infty$. Each group Γ_j satisfies the law:

$$[\![g_1, ..., g_{2^d}]\!] = 1$$

where $[\![g_1, ..., g_{2^d}]\!]$ is the d-fold iterated commutator as in (**??**). Therefore, for every 2^d-tuple of elements γ_i of Λ we have

$$[\![\gamma_1, ..., \gamma_{2^d}]\!] = \lim_{j\to\infty} [\![\rho_j(\gamma_1), ..., \rho_j(\gamma_{2^d})]\!] = 1.$$

Hence, Λ is solvable of derived length $\leqslant d$. $\square$

9.4. Reduction to the case of linear subgroups

PROPOSITION 9.46. *It suffices to prove Theorem 9.1 for subgroups $\Gamma \leqslant GL(V)$, where V is a finite-dimensional real vector space, and the Zariski closure of Γ in $GL(V)$ is a Zariski–irreducible semisimple algebraic group, acting irreducibly on V.*

PROOF. The first step is to reduce the problem from subgroups in Lie groups with finitely many connected components to subgroups of some $GL(V)$.

Let L be a Lie group with finitely many components. The connected component of the identity $L_0 \subset L$ is then a finite index normal subgroup. Thus $\Gamma \cap L_0$ has finite index in Γ. Therefore, we can assume that L is connected.

LEMMA 9.47. *There exists a homomorphism $\phi : \Gamma \to GL_n(\mathbb{R})$, $n = \dim(G)$, whose kernel is contained in the center of Γ.*

PROOF. Since L is connected, kernel of the adjoint representation $Ad : L \to GL(T_e L)$ is contained in the center of L, see Lemma 3.10. Now, take $\phi := Ad|\Gamma$. $\square$

Observe that

1. Γ is virtually solvable if and only if $\phi(\Gamma)$ is virtually solvable.
2. Γ contains a free subgroup if and only if $\phi(\Gamma)$ contains a free subgroup.

Therefore, we can assume that Γ is a linear group, $\Gamma \subset GL(n, \mathbb{R})$.

Let G be the Zariski-closure of Γ in $GL(V)$. Although G need not be Zariski–irreducible, by Proposition 9.26 it has only finitely many irreducible components. Thus, after passing to a finite index subgroup in Γ, we may assume that G is Zariski–irreducible.

According to the results mentioned in the end of Section 9.1, G contains a normal algebraic Zariski–irreducible subgroup which is solvable, $\mathrm{Rad}(G)$, and the quotient $G/\mathrm{Rad}(G)$ is a semisimple algebraic Zariski–irreducible subgroup. Clearly the image of Γ by the algebraic projection $\pi : G \to G/\mathrm{Rad}(G)$ is Zariski dense in $G/\mathrm{Rad}(G)$, and it suffices to prove the alternative for $\pi(\Gamma)$. Thus we may assume that the Zariski closure G of Γ is Zariski–irreducible and semisimple.

If the action $G \curvearrowright V$ is reducible then we take the direct sum decomposition

$$V = \bigoplus_{i=1}^{s} V_i$$

in G-invariant subspaces, so that the action of G on each V_i is irreducible. If we denote by ρ_i the homomorphism $G \to GL(V_i)$ then it suffices to prove the Tits' Theorem for each $\rho_i(\Gamma)$. Indeed, if it is proved, then either all $\rho_i(\Gamma)$ are solvable, in which case Γ itself is solvable (see Exercise **??**), or some $\rho_i(\Gamma)$ contains a free non-abelian subgroup, in which case Γ itself does, as $\rho_i(\Gamma)$ is a quotient of Γ.

Note that when we replace in our problem Γ by $\rho_i(\Gamma)$, we have to replace G by the Zariski closure G_i of $\rho_i(\Gamma)$ in $GL(V_i)$. Note also that

$$\rho_i(\Gamma) \leqslant \rho_i(\overline{\Gamma}) = \rho_i(G) \leqslant \overline{\rho_i(\Gamma)} = G_i \leqslant \overline{\rho_i(G)}\,.$$

According to the considerations in the end of Section 9.1, $\rho_i(G)$ is an algebraic Zariski-irreducible semisimple group. In particular it coincides with its closure, hence $G_i = \rho_i(G)$. Thus G_i acts irreducibly on V_i because G does, and G_i is Zariski–irreducible and semisimple because $\rho_i(G)$ is. This concludes the proof of Proposition 9.46. $\square$

9.5. Tits' Alternative for unbounded subgroups of $SL(n)$

In this section we prove Tits' Alternative for subgroups Γ of $SL(n, \mathbb{K})$ that are unbounded with respect to the standard norm, where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. For technical reasons, one should also consider the case of other *local fields* $\mathbb{K}$. Recall that a local field is a field with a norm $|\cdot|$ which determines a locally compact topology on $\mathbb{K}$. The most relevant examples for us are when $\mathbb{K} = \mathbb{R}, \mathbb{K} = \mathbb{C}$, $\mathbb{K} = \mathbb{Q}_p$ and, more generally, $\mathbb{K}$ is the completion of a finite extension of $\mathbb{Q}$.

In what follows, V is an n-dimensional vector space over a local field $\mathbb{K}$, $n = \dim(V) > 1$. We fix a basis $e_1, \dots, e_n$ in V. Then the norm $|\cdot|$ on $\mathbb{K}$ determines the Euclidean norms $\|\cdot\|$ on V and on its exterior powers.

NOTATION 9.48. We will use the notation E^c to denote the complement $X \setminus E$ of a subset $E \subset X$.

We shall prove the following.

THEOREM 9.49. *Let $\Gamma \leqslant GL(V)$ be a finitely-generated group which is not relatively compact, and such that the Zariski closure of Γ in $GL(V)$ is a Zariski-irreducible semisimple algebraic group acting irreducibly on V. Then Γ contains a free non-abelian subgroup.*

PROOF. In the argument, the free subgroups will be constructed using the Ping-pong Lemma 4.37. The role of the space X in that lemma will be played by the projective space.

NOTATION 9.50. 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.

The ideal situation would be to find a pair of elements g, h in Γ with properties as in Chapter 4, Section 4.5. Since such elements may not exist in Γ in general, we try to 'approximate' the situation in Lemma 4.42.

Recall that, according to the Cartan decomposition (see Section 4.5), every element $g \in GL(V)$ can be written as $g = kdh$, where k and h are in the compact subgroup K of $GL(V)$ and d is a diagonal matrix with entries on the diagonal such that $|a_1| \geqslant |a_2| \geqslant \ldots \geqslant |a_n| > 0$.

DEFINITION 9.51. We call a sequence of elements (g_i) in $GL(V)$ a *diverging sequence* if their matrix norms diverge to infinity.

It is immediate from the compactness of K that the elements g_i of a diverging sequence have Cartan decomposition $g_i = k_i d_i h_i$ such that $|a_1(g_i)| \to \infty$ as $i \to \infty$.

For every diverging sequence, there exists a maximal $m \in \{1, \dots, n-1\}$ with the property that

$$\limsup_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} > 0\,.$$

By passing to a subsequence we may assume that

$$\lim_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} = 2\ell > 0$$

and also that k_i and h_i converge to some $k \in K$ and $h \in K$ respectively. We formalize these observations as follows:

DEFINITION 9.52. We call a sequence (g_i) *m-contracting*, for $m < \dim V$, if its elements have Cartan decompositions $g_i = k_i d_i h_i$ satisfying the following convergence properties:

(1) k_i and h_i converge to some k and h in K;

(2) d_i are diagonal matrices with diagonal entries $a_1(g_i), \dots, a_n(g_i)$ such that

$$|a_1(g_i)| \geqslant |a_2(g_i)| \geqslant \dots \geqslant |a_n(g_i)|, \quad |a_1(g_i)| \to \infty$$

and

$$\lim_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} > 0\,.$$

(3) The number m is maximal with the above properties.

Observe now that *since Γ is unbounded, it contains an m-contracting sequence (g_i), for some $1 \leqslant m < \dim V$.*

In what follows we analyze the dynamics of an m-contracting sequence $\sigma = (g_i)$. We use the following notation and terminology, consistent to that in Definition 9.52 and the notation used in §4.5:

NOTATION 9.53.

$$A(g_i) = k_i\,[\mathrm{Span}(e_1, \dots, e_m)] \quad \text{and } A(\sigma) = k\,[\mathrm{Span}(e_1, \dots, e_m)]\,.$$

$$E(g_i) = h_i^{-1}\,[\mathrm{Span}(e_{m+1}, \dots, e_n)] \quad \text{and } E(\sigma) = h^{-1}\,[\mathrm{Span}(e_{m+1}, \dots, e_n)]\,.$$

Here the bracket stands for the projection to $P(V)$. We call $A(\sigma)$ the *attracting subspace* of the sequence σ and $E(\sigma)$ the *repelling subspace* of the sequence σ.

When $m = 1$ we call $A(\sigma)$ the *attracting point* and $E(\sigma)$ (sometimes also denoted $H(\sigma)$) the *repelling hyperplane* of the sequence σ.

Note that since $k_i \to k$ and $h_i \to h$, they converge in the compact-open topology as transformations of $P(V)$; hence $A(g_i)$ converge to $A(\sigma)$, and $E(g_i)$ converge to $E(\sigma)$ with respect to the Hausdorff metric.

EXAMPLE 9.54. To make things more concrete, consider the case $\dim V = 2$ and $\mathbb{K} = \mathbb{R}$. Then $P(V) = \mathbb{P}^1$ is the circle on which the group $PSL(2, \mathbb{R})$ acts by linear-fractional transformations. Since $0 < m < 2$, it follows that $m = 1$ and, hence, every diverging sequence contains a 1-contracting subsequence. It is easy to see that, for a 1-contracting sequence, the sequence of inverses has to be 1-contracting as well. Moreover, the repelling hyperplanes in $P(V)$ are again points.

Thus, each diverging sequence $g_i \in PSL(2,\mathbb{R})$ contains a subsequence g_{i_n} for which there exists a pair of points A and H in $P(V)$ such that

$$\lim_{n\to\infty} g_{i_n}|_{P(V)\setminus\{H\}} = A \ \text{ and } \ \lim_{n\to\infty} g_{i_n}^{-1}|_{P(V)\setminus\{A\}} = H\,,$$

uniformly on compact sets. For instance, if $g_{i_n} = g^n$, and g is parabolic, then $A = H$ is the fixed point of g. If g is hyperbolic then A is the attractive and H is the repelling fixed point of g. Thus, in general (unlike in the diagonal case), $A(g_i)$ may belong to $E(g_i)$.

The following is a uniform version of Lemma 4.41 for m-contracting sequences:

LEMMA 9.55. *Let $\sigma = (g_i)$ be an m-contracting sequence. For each compact $K \subset E(\sigma)^c$ there exist L and i_0 so that g_i is L–Lipschitz on K, for every $i \geqslant i_0$.*

PROOF. Assume that g_i's satisfy (for all sufficiently large i) the following:

$$|a_1(g_i)| \geqslant |a_2(g_i)| \geqslant \ldots \geqslant |a_m(g_i)| \geqslant \ell |a_1(g_i)|\,,$$

where $\ell > 0$ is a constant independent of i.

By the assumption, hK is disjoint of $[\mathrm{Span}(e_{m+1}, \ldots, e_n)]$, so the Hausdorff distance between these two compact sets is $2\varepsilon > 0$. Since the sets $h_i K$ converge to hK in the Hausdorff metric, as $i \to \infty$, we may assume that for large i, the set $h_i K$ is contained in K_ε, where

$$K_\varepsilon = \overline{\mathcal{N}}_\varepsilon(hK) = \{p \in P(V) \mid \mathrm{dist}(p, hK) \leqslant \varepsilon\}\,.$$

Since k_i act as isometries on $P(V)$, it suffices to prove that d_i's are L-Lipschitz maps, for some uniform L and i large enough. In what follows, we consider an arbitrary diagonal matrix $d = d_i$ with eigenvalues $a_1, \ldots, a_n$.

Then every point $[u]$ of K_ε is at distance $\gg \varepsilon$ from $[\mathrm{Span}(e_{m+1}, \ldots, e_n)]$. Without loss of generality, we may assume that $u = (u_1, \ldots, u_n)$ is a unit vector. Set

$$u' = (u_1, \ldots, u_m, 0, \ldots, 0), \quad u'' = (0, \ldots, 0, u_{m+1}, \ldots, u_n)$$

Suppose that $0 < \delta \leqslant \frac{1}{2\sqrt{n}}$ and the vector u (as above) is such that

$$|u_i| \leqslant \delta, \forall i = 1, \ldots, m.$$

Then,

$$|u - u''|_{max} = |u'|_{max} \leqslant \delta.$$

Lemma 1.74 then implies that

$$|u \wedge u''| \leqslant 2n\delta,$$

while

$$|u''| \gg 1 - \sqrt{n}\delta \geqslant \frac{1}{2}.$$

Combining these inequalities,we obtain

$$d([u], [u'']) \leqslant 4n\delta.$$

Since, by assumption, $\varepsilon \leqslant d([u], [u''])$, we see that

$$\delta \geqslant \frac{\varepsilon}{4n}.$$

Therefore, for every unit vector u so that $[u] \in K_\varepsilon$,

$$\max_{k=1,\ldots,m} |u_k| \geqslant \delta = \delta(\varepsilon) = \min\left(\frac{\varepsilon}{4n}, \frac{1}{2\sqrt{n}}\right). \tag{9.1}$$

In particular, for such u, there exists $k \in \{1, \dots, m\}$, so that

$$|d(u)|^2 \geqslant |a_k|^2 |u_k|^2 \geqslant \ell^2 |a_1|^2 \delta^2 .$$

Let $[v]$ and $[w]$ be two points in K_ϵ. Then, in the archimedean case,

$$|d(v) \wedge d(w)|^2 = \sum_{p<q} |a_p v_p a_q w_q - a_q v_q a_p w_p|^2 = \sum_{p<q} |a_p a_q|^2 |v_p w_q - v_q w_p|^2 \leqslant$$

$$|a_1|^4 \sum_{p<q} |v_p w_q - v_q w_p|^2 = |a_1|^4 |v \wedge w|^2,$$

while in the nonarchimedean case we also get:

$$|d(v) \wedge d(w)| = \max_{p,q} |a_p v_p a_q w_q - a_q v_q a_p w_p| \leqslant |a_1|^2 |v \wedge w|.$$

By combining these inequalities, for unit vectors u, v satisfying $[u], [v] \in K_\epsilon$, we obtain

$$d(g(v), g(w)) = \frac{|g(v) \wedge g(w)|}{|g(v)| \cdot |g(w)|} \leqslant \frac{|v \wedge w|}{\ell^2 \delta} = \frac{d(v, w)}{\ell^2 \delta}. \quad \square$$

LEMMA 9.56. *Let g be an element in $GL(V)$ with Cartan decomposition $g = kdh$, where d is a diagonal matrix with entries $a_1, \dots, a_n$ on the diagonal such that $|a_1| > |a_2| \geqslant \dots \geqslant |a_n| > 0$. If $\frac{|a_2|}{|a_1|} < \varepsilon^2/\sqrt{n}$, then g maps the complement of the ε–neighborhood of the hyperplane $H = h^{-1}[\mathrm{Span}(e_2, \dots, e_n)]$ into the ball with center $k[e_1]$ and radius ε .*

PROOF. Since k and h are isometries of $P(V)$, it clearly suffices to prove the statement for $g = d$, $k = h = 1$. Let $[v]$ be a point in $P(V)$ such that $\mathrm{dist}([v], [\mathrm{Span}(e_2, ..., e_n)]) \geqslant \varepsilon$. Then, as in the proof of Lemma 4.42,

$$d([dv], [e_1]) = \frac{|dv \wedge e_1|}{|dv|} \leqslant \sqrt{n} \frac{|a_2|}{\varepsilon |a_1|} < \varepsilon .$$

$\square$

LEMMA 9.57. *If $\sigma = (g_i)$ is a 1-contracting sequence with attracting point $p = A(\sigma)$ and repelling hyperplane $H(\sigma)$, then for every closed ball $B \subseteq H(\sigma)^c$, the maps $g_i|_B$ converge uniformly to the constant function on B which maps everything to the point p.*

PROOF. Consider an arbitrary closed ball B in $H(\sigma)^c$. Then hB is a closed ball in the complement of $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. By compactness on $P(V)$, there exists $\varepsilon > 0$ so that the minimal distance from hB to $[\mathrm{Span}(e_{m+1}, ..., e_n)]$ is $\geqslant 2\varepsilon$. Consider

$$B_\varepsilon = \{x \in P(V) \mid \mathrm{dist}(x, B) \leqslant \varepsilon\} ,$$

which is also a closed ball, at minimal distance $\geqslant \varepsilon$ from $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. For all sufficiently large i, the ball $h_i B$ is contained in B_ε. Therefore, it suffices to prove that the maps $k_i d_i|_{B_\varepsilon}$ converge uniformly to the constant function on B_ε which maps everything to the point p

Consider $\delta = \varepsilon/2$. For all sufficiently large i, according to Lemma 9.56, $d_i(B_\delta)$ is contained in $B([e_1], \delta)$. On the other hand, for all large i, the point $k_i[e_1]$ belongs to the ball $B(p, \delta)$. Whence,

$$k_i d_i(B_\varepsilon) \subset B(p, \varepsilon).$$

$\square$

LEMMA 9.58. *Let (g_i) be a diverging sequence of elements in $GL(V)$.*

(1) *If there exists a closed ball B with non-empty interior and a point p such that $g_i|_B$ converge uniformly to the constant function on B which maps everything to the point p, then (g_i) contains a 1-contracting subsequence with attracting point p.*

(2) *If, moreover, there exists a hyperplane H such that for every closed ball $B \subseteq H^c$, $g_i|_B$ converge uniformly to the constant function on B which maps everything to the point p, then (g_i) contains a 1-contracting subsequence with the attracting point p and the repelling hyperplane H.*

PROOF. (1) Since (g_i) is diverging, it contains a subsequence σ (whose elements we again denote g_i) which is m-contracting for some m. By replacing B with a smaller ball, we may assume that B is in $E(\sigma)^c$.

Let $g_i = k_i d_i h_i$ denote the Cartan decomposition of g_i. By the above observations, for all sufficiently large i, the balls $h_i B$ are disjoint from $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. The sequence of closed metric balls $h_i B$ Hausdorff–converges to the closed metric ball hB. Therefore, there exists i_0 and a closed ball B' contained in the intersection

$$\bigcap_{i \geqslant i_0} h_i B.$$

By the hypothesis, the closed sets $k_i d_i(B')$ Hausdorff–converge to the point p.

For every point $[v] \in B'$ represented by a vector v, we have:

$$[d_i v] = \left[v_1 e_1 + \frac{a_2(g_i)}{a_1(g_i)} v_2 e_2 + \ldots + \frac{a_n(g_i)}{a_1(g_i)} v_n e_n \right].$$

After passing to a subsequence, we may assume that

$$\lim_{i\to\infty} \frac{a_k(g_i)}{a_1(g_i)} = \lambda_k, k = 1, \ldots, m.$$

Since our sequence is m-contracting,

$$|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_m| > 0.$$

If $m \geqslant 2$ then we may find two distinct points $[v], [v']$ in B' represented by two unit vectors $v = (v_1, \ldots, v_n), v' = (v_1', \ldots, v')$ so that

$$\lim_{i\to\infty} [d_i v] = [w], \quad \lim_{i\to\infty} [d_i v'] = [w'],$$

$[w] \neq [w'], \quad w = v_1 e_1 + \lambda_2 v_2 e_2 + \ldots + \lambda_m v_m e_m, w' = v_1' e_1 + \lambda_2 v_2' e_2 + \ldots + \lambda_m v_m' e_m$. Assume that $d([w], [w']) = \epsilon > 0$. As

$$[u] = \lim_{i\to\infty} [k_i d_i v] = \lim_{i\to\infty} [k_i w_i], \quad [u'] = \lim_{i\to\infty} [k_i d_i v'] = \lim_{i\to\infty} [k_i w_i'],$$

it follows that the $d([u], [u']) = \epsilon > 0$. This contradicts the assumption that the sequence of sets $k_i d_i(B')$ Hausdorff–converges to a point. It follows that $m = 1$, i.e., $\sigma = (g_i)$ is 1-contracting. If $A(\sigma) \neq p$ then a contradiction easily follows from Lemma 9.57.

(2) According to (1), the sequence (g_i) contains a subsequence σ which is 1-contracting, with $A(\sigma) = p$. We continue with the notation introduced in the proof of (1). If $H(\sigma) \neq H$ then at least one of the points $h^{-1}[e_2], \ldots, h^{-1}[e_n]$ is not in H. Assume that it is $h^{-1}[e_2]$, and that its distance to H is $2\epsilon > 0$. For sufficiently large all i's, the points $h_i^{-1}[e_2]$ belong to the ball $B(h^{-1}[e_2], \epsilon)$, disjoint

from H. It follows that the sequence $k_i d_i[e_2] = k_i[e_2]$ must converge to $p = k[e_1]$ by the assumption, and also to $k[e_2]$, since $\lim_{i\to\infty} k_i = k$. Contradiction. $\square$

The following lemma is an easy consequence of Lemma 9.58, and it is left as an exercise to the reader.

LEMMA 9.59. *Let (g_i) be a 1-contracting sequence in $PGL(V)$, and $f, h \in PGL(V)$. Then the sequence (fg_ih) contains a 1-contracting subsequence $\sigma = (g_i')$ such that*

$$A(\sigma) = f(A(\sigma)), \quad E(\sigma) = h^{-1}E(\sigma).$$

LEMMA 9.60. *Let (g_i) be a diverging sequence in $PGL(V)$. Then there exists a vector space W and an embedding $\rho : PGL(V) \hookrightarrow PGL(W)$ so that a subsequence in $(\rho(g_i))$ is 1-contracting in $PGL(W)$.*

PROOF. After passing to a subsequence, we may assume that the sequence $\sigma = (g_i)$ is m-contracting for some m, $0 < m < n$. We consider the m-th exterior power of V,

$$W := \Lambda^m V.$$

The action of $GL(V)$ on V extends naturally to its action on W we obtain the embedding $\rho : GL(V) \hookrightarrow GL(W)$. Clearly, for a matrix $g \in GL(V)$, the norms of the singular values of $\rho(g) \in GL(W)$ are the products

$$\prod_{j_1 < \ldots < j_m} |a_{j_1} \cdots a_{j_m}(g)|.$$

where $a_j(g)$ is the j-th singular value of g. Then, $|a_1(\rho(g_i))| = |a_1 \cdots a_m(g_i)|$ and it is immediate that

$$\lim_{i\to\infty} \frac{a_l(\rho(g_i))}{a_1(\rho(g_i))} = 0, \forall l > 1. \quad \square$$

We now return to the proof of the Tits alternative for the subgroup $\Gamma < GL(V)$. Recall that we are working under the assumption that the Zariski closure $G = \overline{\Gamma}$ of Γ in $GL(V)$ satisfies certain conditions, namely G is Zariski–irreducible, semisimple and it acts irreducibly on V.

After replacing V with W as above, since

$$\rho(\Gamma) \leqslant \rho(G) = \rho(\overline{\Gamma}) \leqslant \overline{\rho(\Gamma)} \leqslant \overline{\rho(G)}$$

and $\rho(G)$ is still an algebraic Zariski–irreducible semisimple subgroup (see the end of Section 9.1), it follows that $\overline{\rho(\Gamma)} = \rho(G)$. In what follows, we let Γ and G denote $\rho(\Gamma)$ and $\rho(G)$, and we denote the sequence $(\rho(g_i))$ by (g_i).

If the action $G \curvearrowright W$ is reducible, we take a direct sum decomposition

$$W = \bigoplus_{i=1}^{s} W_i$$

into G-invariant subspaces, so that the restriction of the G-action to each is irreducible. This defines homomorphisms $\rho_i : G \to GL(W_i)$, and all $G_i = \rho_i(G)$ are algebraic Zariski–irreducible semisimple subgroups. In particular, $G_i = [G_i, G_i]$, hence every G_i is, in fact, contained in $SL(W_i)$. In particular for the 1-dimensional spaces W_i, the group G_i is trivial. Without loss of generality, we can, therefore, assume that each subspace W_i has dimension > 1.

LEMMA 9.61. *For some s, the sequence $\sigma = (g_i)$ restricted to W_s is 1-contracting.*

Proof. Let $p = A(\sigma) \in P(W)$ and $H = H(\sigma) \subset P(W)$ be the attracting point and, respectively, the repelling hyperplane of the sequence $\sigma = (g_i)$. Since the subspaces W_t are G-invariant, for each t either $p \in P(W_t)$ or $P(W_t) \subset H$. Since H is a hyperplane in $P(W)$, it follows that $p \in P(W_s)$ for some s. The restriction of (g_i) to $P(W_s)$ converges to p away from $H \cap P(W_s)$. Since $\dim(W_s) > 1$, we are done. $\square$

Let ρ_s be the representation $G \to SL(W_s)$. Our goal will be to prove that $\rho_s(\Gamma)$ contains a free non-abelian group, whence it will follow that Γ contains such a group, which will conclude the proof. For simplicity of notation, in what follows, we denote $\rho_s(\Gamma)$ by Γ, its Zariski closure by G and the vector space W_s by V. As before, the Zariski closure of $\rho_s(\Gamma)$ is Zariski–irreducible and semisimple.

Theorem 9.62. *Let Γ be a subgroup in $SL(V)$ containing a 1-contracting sequence of elements, and such that the Zariski closure $\overline{\Gamma}$ of Γ is Zariski–irreducible and that Γ acts irreducibly on V. Then Γ contains a free non-abelian subgroup.*

Before beginning the proof, we note that the 1-contracting sequence that we now have at our disposal in the group Γ does not suffice yet, not even to construct one of the two elements in a ping-pong pair "modeled" after the one in Lemma 4.42. Indeed, for every $i \in \mathbb{N}$ the action of the element $g_i \in \Gamma$ on the projective space $P = P(V)$ is, as represented in Figure 9.1 (where we picture projective space as a sphere). According to Lemma 9.56, for every $\epsilon > 0$ and all sufficiently large i, the transformation g_i (with the Cartan decomposition $k_i d_i h_i$) maps the complement of the ϵ-neighborhood of $H(\sigma) = h_i^{-1}[\mathrm{Span}(e_2, \dots, e_n)]$ into the ϵ-neighborhood of the point $A(\sigma) = k_i[e_1]$, with the notation of 9.53.

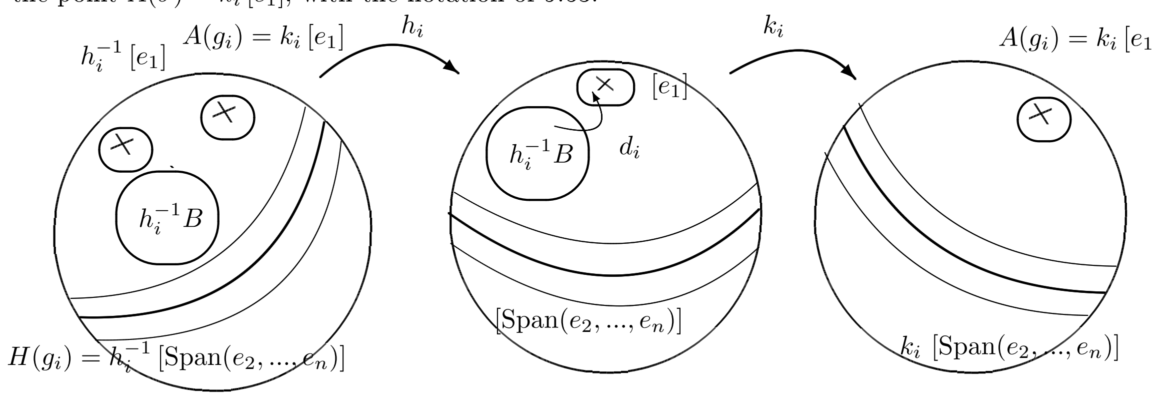


Figure 9.1. The action of g_i.

The first problem occurs when one iterates g_i, i.e. one considers g_i^2, g_i^3, etc. Nothing guarantees that g_i^2 would also map the complement of the ϵ-neighborhood of $H(g_i)$ into the ϵ-neighborhood of $A(g_i)$, for large i. This only happens when the ϵ-neighborhood of $A(g_i)$ is disjoint from the ϵ-neighborhood of $H(g_i)$. Our hypothesis does not ensure this, since no conditions can be imposed on h_i, k_i and their limits (see comments in Example 9.54). We will use Lemma 9.59 and the notion of a *separating set* developed in the sequel to circumvent this difficulty.

Separating sets.

DEFINITION 9.63. A subset $F \subset PGL(V)$ is called *m-separating* if for every choice of points $p_1, \dots, p_m \in P = P(V)$ and hyperplanes $H_1, \dots, H_m \subset P$, there exists $f \in F$ so that

$$f^{\pm 1}(p_i) \notin H_j, \forall i, j = 1, \dots, m.$$

It will now become apparent why we endeavored to ensure the two irreducibility properties (for the Zariski topology, and for the action) for the Zariski closure of Γ.

PROPOSITION 9.64. *Let $\Gamma \subset SL(V)$ be a subgroup with the property that its Zariski closure is Zariski–irreducible and it acts irreducibly on V. For every m, Γ contains a finite m-separating subset F.*

PROOF. Let G be the Zariski closure of Γ. Let $P^\vee$ denote the space of hyperplanes in P (i.e. the projective space of the dual of V). For each $g \in G$ let $M_g \subset P^m \times (P^\vee)^m$ denote the collection of $2m$-tuples

$$(p_1, \dots, p_m, H_1, \dots, H_m)$$

so that

$$g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j$$

for some $i, j = 1, \dots, m$.

LEMMA 9.65. *If Γ is as in Proposition 9.64 then*

$$\bigcap_{g \in \Gamma} M_g = \emptyset\,.$$

PROOF. Suppose to the contrary that the intersection is nonempty. Then there exists a $2m$-tuple $(p_1, \dots, p_m, H_1, \dots, H_m)$ so that for every $g \in \Gamma$,

$$\exists\, i, j \text{ so that } g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j\,. \tag{9.2}$$

The set of elements $g \in SL(V)$ such that (9.2) holds for the given $2m$-tuple is Zariski–closed, and G is the Zariski closure of Γ, hence all $g \in G$ also satisfy (9.2).

Let $G^{\pm}_{p_i, H_j}$ denote the set of $g \in G$ so that

$$g^{\pm 1}(p_i) \in H_j.$$

Clearly, these subsets are Zariski–closed and cover the group G. Since G Zariski–irreducible, it follows that one of these sets, say $G^{+}_{p_i, H_j}$, is the entire of G. Therefore, for every $g \in G$, $g(p_i) \in H_j$. Thus, projectivization of the vector subspace L spanned by the G–orbit (of lines) $G \cdot p_i$ is contained in H_j. The subspace L is proper and G-invariant. This contradicts the hypothesis that G acts irreducibly on V. $\square$

We now finish the proof of Proposition 9.64. Let M_g^c denote the complement of M_g in $P^m \times (P^\vee)^m$. This set is Zariski open. By Lemma 9.65, the sets M_g^c $(g \in \Gamma)$ cover the space $P^m \times (P^\vee)^m$. Since $\mathbb{K}$ is a local field, the product $P^m \times (P^\vee)^m$ is compact and, thus, the above open cover contains a finite subcover. Hence, there exists a finite set $F \subset \Gamma$ so that

$$\bigcup_{f \in F} M_f^c = P^m \times (P^\vee)^m.$$

This set satisfies the assertion of Proposition 9.64. $\square$

REMARK 9.66. The above proposition holds even if the field $\mathbb{K}$ is not local. Then the point is that by Hilbert's Nullstellensatz, there exists a finite subset $F \subset \Gamma$ so that

$$\bigcap_{f\in F} M_f = \bigcap_{g\in\Gamma} M_g = \emptyset.$$

With this modification, the above proof goes through.

Ping-pong sequences. We now begin the proof of Theorem 9.62, which will be split in several lemmas.

In what follows we fix a 4-separating finite subset $F \subset \Gamma \subset PGL(V)$. We will use the notation f for the elements of F.

LEMMA 9.67. *There exists $f \in F$ so that (after passing to a subsequence in (g_i)) both sequences $h_i := g_i f g_i^{-1}$ and $g_i f^{-1} g_i^{-1}$ are 1-contracting.*

PROOF. After passing to a subsequence $\sigma = (g_i)$, we can assume that the sequence $\sigma^- = (g_i^{-1})$ is m-contracting, with attracting subspace $A(\sigma^-)$ and repelling subspace $E(\sigma^-)$. Pick a point q in the complement of the subspace $E(\sigma^-)$. After passing to a subsequence in (g_i) again, we can assume that $\lim_i g_i^{-1}(q) = u \in A(\sigma^-)$. Let $A(\sigma)$ and $H(\sigma)$ be the attracting point and the repelling hyperplane of the sequence σ.

Since F is a separating subset, there exists $f \in F$ so that $f^{\pm 1}(u) \notin H(\sigma)$.

Take a small closed ball $B(q,\epsilon) \subset P$ centered at q and disjoint from $E(\sigma^-)$. According to Lemma 9.55, $g_i^{-1}(B(q,\epsilon)) \subset B\left(g_i^{-1}(q), L\epsilon\right)$ for al large i and L independent of i. It follows that for all large i

$$g_i^{-1}(B(q,\epsilon)) \subset B(u, 2L\epsilon).$$

By Lemma 4.41, $fg_i^{-1}(B(q,\epsilon)) \subset B(f(u), L'\epsilon)$ for all large i and L' independent of i. Note that if we reduce ϵ, the constants L and L' will not change. We take ϵ small enough so that the sets $B(f(u), L'\epsilon)$ and $\mathcal{N}_\epsilon(H(\sigma))$ are disjoint. Since the sequence (g_i) restricted to the complement of $\mathcal{N}_\epsilon(H(\sigma))$ converges uniformly to the point $A(\sigma)$ it follows that the sequence $g_i f g_i^{-1}|_{B(q,\epsilon)}$ converges uniformly to the point $A(\sigma)$. Lemma 9.58, (1), now implies that (g_i) contains a 1-contracting subsequence.

The same argument for f^{-1} concludes the proof. $\square$

Thus, we have found a 1-contracting sequence $\tau = (h_i)$ in Γ such that the sequence $\tau^- = (h_i^{-1})$ is also 1-contracting.

LEMMA 9.68. *There exists $f \in F$ such that, for a subsequence $\eta = (y_i)$ of the sequence (fh_i), both η and $\eta^- = \left(y_i^{-1}\right)$ are 1-contracting. Moreover,*

$$A(\eta) \notin H(\eta) \quad \text{and} \quad A(\eta^-) \notin H(\eta^-). \tag{9.3}$$

PROOF. By Lemma 9.59, for any choice $f \in F$, the sequence (fh_i) contains a 1-contracting subsequence $\eta = (y_i)$, with $\eta^- = \left(y_i^{-1}\right)$ likewise 1-contracting, and

$$A(\eta) = f(A(\tau)),\ H(\eta) = H(\tau),$$

$$A\left(\eta^-\right) = A\left(\tau^-\right),\ H\left(\eta^-\right) = fH\left(\tau^-\right).$$

Now, the assertion follows from the fact that F is a 4-separating set. $\square$

DEFINITION 9.69. [Ping-pong pair] A pair of sequences $\eta = (y_i)$ and $\zeta = (z_i)$ is called a *ping-pong pair* if both sequences are as in Lemma 9.68 and, furthermore, $A(\eta^{\pm}) \notin H(\zeta^{\pm})$ and $A(\zeta^{\pm}) \notin H(\eta^{\pm})$.

Let $\eta = (y_i)$ be the sequence from Lemma 9.68.

LEMMA 9.70. *There exists $f \in F$ so that the sequences $(y_i), (z_i) = (f y_i f^{-1})$ contain subsequences that form a ping-pong pair.*

PROOF. By Lemma 9.59, after replacing $\eta = (y_i)$ with a subsequence, we may assume that $\zeta = (z_i)$ and $\zeta^- = \left(z_i^{-1}\right)$ are 1-contracting and $A\left(\zeta^{\pm 1}\right) = fA\left(\eta^{\pm 1}\right)$, while $H\left(\zeta^{\pm 1}\right) = fH\left(\eta^{\pm 1}\right)$. Now, the assertion follows from the fact that F is 4-separating. □

End of proof of Theorem 9.62. Lemma 9.70 implies that Γ contains a ping-pong pair of sequences $\eta = (y_i)$, $\zeta = (z_i)$. For every small ϵ and all large i, we have:

$$\mathcal{N}_\epsilon (H(\eta))^c \xrightarrow{y_i} B(A(\eta), \epsilon)$$

$$\mathcal{N}_\epsilon \left(H\left(\eta^-\right)\right)^c \xrightarrow{y_i^{-1}} B\left(A\left(\eta^-\right), \epsilon\right)$$

$$\mathcal{N}_\epsilon (H(\zeta))^c \xrightarrow{z_i} B(A(\zeta), \epsilon)$$

$$\mathcal{N}_\epsilon \left(H\left(\zeta^-\right)\right)^c \xrightarrow{z_i^{-1}} B\left(A\left(\zeta^-\right), \epsilon\right)$$

Moreover, for ϵ sufficiently small, the balls on the right-hand side are contained in the complements of tubular neighborhoods on the left-hand side. Therefore, the above statements also hold with transformations $y_i, y_i^{-1}, z_i, z_i^{-1}$ replaced by their k-th iterations for all $k > 0$.

We choose ϵ small enough so that

$$B(A(\eta), \epsilon) \cap \mathcal{N}_\epsilon \left(H(\eta) \cup H(\zeta) \cup H\left(\zeta^-\right)\right) = \emptyset,$$

$$B\left(A\left(\eta^-\right), \epsilon\right) \cap \mathcal{N}_\epsilon \left(H\left(\eta^-\right) \cup H(\zeta) \cup H\left(\zeta^-\right)\right) = \emptyset,$$

$$B(A(\zeta), \epsilon) \cap \mathcal{N}_\epsilon \left(H(\zeta) \cup H(\eta) \cup H\left(\eta^-\right)\right) = \emptyset,$$

$$B\left(A\left(\zeta^-\right), \epsilon\right) \cap \mathcal{N}_\epsilon \left(H\left(\zeta^-\right) \cup H(\eta) \cup H\left(\eta^-\right)\right) = \emptyset.$$

For ϵ small as above, we consider the sets

$$\widetilde{A} = B(A(\eta), \epsilon) \cup B\left(A\left(\eta^-\right), \epsilon\right)$$

and

$$\widetilde{B} = B(A(\zeta), \epsilon) \cup B\left(A\left(\zeta^-\right), \epsilon\right).$$

Since $A(\eta) \in H\left(\eta^-\right)$, $A\left(\eta^-\right) \in H(\eta)$ and $A(\zeta) \in H\left(\zeta^-\right)$, $A\left(\zeta^-\right) \in H(\zeta)$, our hypotheses imply that $\widetilde{A} \cap \widetilde{B} = \emptyset$. Moreover for all large i, for every $k \in \mathbb{Z} \setminus \{0\}$,

$$y_i^k\left(\widetilde{B}\right) \subseteq \widetilde{A} \text{ and } z_i^k\left(\widetilde{A}\right) \subseteq \widetilde{B}.$$

Lemma 4.37 now implies that for all large i, the group $\langle y_i, z_i\rangle$ is a free group of rank 2. □

9.6. Free subgroups in compact Lie groups

The compact case is more complicated. Let Γ be a relatively compact finitely-generated subgroup of $G = SL(n, \mathbb{C})$. According to Proposition 9.46, we may assume that the Zariski closure of Γ in $SL(n, \mathbb{C})$ is Zariski–irreducible, semisimple, and that it acts irreducibly, i.e., it does not preserve a proper subspace of $\mathbb{C}^n$. Note that in this section, unlike in the previous one, G denotes $SL(n, \mathbb{C})$, not the Zariski closure of Γ.

Let $\gamma_1, \ldots, \gamma_m$ denote generators of Γ and consider the subfield F in $\mathbb{C}$ generated by the matrix entries of these matrices.

Reduction to a number field case. Consider the *representation variety* $R(\Gamma, G) = \mathrm{Hom}(\Gamma, G)$. This space can be described as follows. Let

$$\langle \gamma_1, \ldots, \gamma_m | r_1, \ldots \rangle$$

be a presentation of Γ (the number of relators could be infinite). Each homomorphism $\rho : \Gamma \to G$ is determined by the images of the generators of Γ. Hence $R(\Gamma, G)$ is a subset of G^m. A map $\rho : \gamma_i \mapsto G, i = 1, \ldots, m$ extends to a homomorphism of Γ if and only if

$$\forall j, \rho(r_j) = 1. \tag{9.4}$$

Since the relators r_j are words in $\gamma_1^{\pm 1}, \ldots, \gamma_m^{\pm 1}$, the equations (9.4) amount to polynomial equations on G^m. Hence, $R(\Gamma, G)$ is given by a system of polynomial equations and has a natural structure of an affine algebraic variety. Since the formula for the inverse in $SL(n)$ involves only integer linear combinations of products of matrix entries, it follows that the above equations have integer (in particular, rational) coefficients. In other words, the representation variety $R(\Gamma, G)$ is *defined over* $\mathbb{Q}$.

PROPOSITION 9.71. *Let Z be an affine variety in $\mathbb{C}^N$ defined by polynomial equations with rational coefficients and let $\overline{\mathbb{Q}}$ be the field of algebraic numbers, the algebraic closure of $\mathbb{Q}$. Then the set $Z \cap \overline{\mathbb{Q}}^N$ is dense in Z with respect to the classical topology on $\mathbb{C}^N$.*

PROOF. The proof is by induction on N. The assertion is clear for $N = 1$. Indeed, in this case either $Z = \mathbb{C}$ or Z is a finite set of roots of a polynomial with rational coefficients: These roots are algebraic numbers. Suppose the assertion holds for subvarieties in $\mathbb{C}^{N-1}$. Pick a point $x = (x_1, \ldots, x_N) \in Z$ and let q_i be a sequence of rational numbers converging to the first coordinate x_1. For each rational number q_i, the intersection $Z \cap \{x_1 = q_i\}$ is again an affine variety defined over $\mathbb{Q}$ which sits inside $\mathbb{C}^{N-1}$. Now the claim follows from the induction hypothesis by taking a diagonal sequence. $\square$

COROLLARY 9.72. *Algebraic points are dense in $R = R(\Gamma, G)$ with respect to the classical topology. In other words, for every homomorphism $\rho : \Gamma \to G$, there exists a sequence of homomorphisms $\rho_j : \Gamma \to G$ converging to ρ so that the matrix entries of the images of generators $\rho_j(\gamma_i)$ are in $\overline{\mathbb{Q}}$.*

We now let $\rho_i \in R(\Gamma, G)$ be a sequence which converges to the identity representation $\rho : \Gamma \to \Gamma \subset G$. Recall that in section 9.3, we proved that for every finitely-generated subgroup $\Gamma \subset GL(n, \mathbb{C})$ which is not virtually solvable, there exists a neighborhood Σ of $\rho = id$ in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho' \in \Xi$ has image which is not virtually solvable. Therefore, without loss of generality, we may assume that each $\rho_j(\Gamma)$ constructed above is not virtually solvable.

LEMMA 9.73. *If $\Gamma_j := \rho_j(\Gamma)$ contains a free subgroup Λ_j of rank 2 then so does Γ.*

PROOF. Let $g_1, g_2 \in \Gamma$ be the elements which map to the free generators h_1, h_2 of Λ_j under ρ_j. Let Λ be the subgroup of Γ generated by g_1, g_2. We claim that Λ is free of rank 2. Indeed, since Λ_j is free of rank 2, there exists a homomorphism $\phi_j : \Lambda_j \to \Lambda$ sending h_k to g_k, $k = 1, 2$. The composition $\phi_j \circ \rho_j$ is the identity since it sends each h_k to itself. Hence, $\phi_j : \Lambda_j \to \Lambda$ is an isomorphism. □

Thus, it suffices to consider the case when the field F (generated by matrix entries of generators of Γ) is a number field, i.e., is contained in $\overline{\mathbb{Q}}$. The absolute Galois group $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on F and hence on $SL(n, F)$:

$$\forall \sigma \in Gal(\overline{\mathbb{Q}}/\mathbb{Q}), \quad A = (a_{ij}) \in SL(n, F), \quad \sigma(A) := (a_{ij}^{\sigma}).$$

Every $\sigma \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ extends (by identity) to the set of transcendental numbers and, hence, extends to an automorphism σ of $\mathbb{C}$. Therefore, σ determines an automorphism σ of $SL(n, \mathbb{C})$ (which, typically, it discontinuous in the classical topology). Therefore, σ will send the subgroup $\Gamma \subset SL(n, F)$ to $\sigma(\Gamma) \subset G(\sigma(F)) \subset SL(n, \mathbb{C})$. The homomorphism $\sigma : \Gamma \to \Gamma' := \sigma(\Gamma)$ is $1-1$ and, therefore, if for some σ the group $SL(n, \sigma(F))$ happens to be a non-relatively compact subgroup of $SL(n, \mathbb{C})$ we are done by Theorem 9.49.

However, it could happen that for each σ the group $G(\sigma(F))$ is relatively compact and, thus, we seemingly gained nothing. Nevertheless, there is a remarkable construction which saves the proof.

Adeles. (See [?, Chapter 6].) The ring of adeles was introduced by A. Weil in 1936. For a number field F consider various norms $|\cdot| : F \to \mathbb{R}_+$, see §1.7.

Suppose that F is a finitely–generated number field. Then F is a finite extension of $\mathbb{Q}$. Let $\mathrm{Nor}(F)$ denote the set of all norms on F which restrict to either the absolute value or to one of the p-adic norms on $\mathbb{Q} \subset F$. We will use the notation F_ν, $\mathbb{Q}_\nu$ to denote the completion of F with respect to the norm ν, we let $O_\nu \subset F_\nu$ denote the ring of integers:

$$O_\nu = \{x \in F_\nu : \nu(x) \leqslant 1\}.$$

Note that for each $x \in \mathbb{Q}$, $x \in O_p$ for all but finitely many p's, since x has only finitely many primes in its denominator. The same is true for elements of F: For all but finitely many $\nu \in \mathrm{Nor}(F)$, $\nu(x) \leqslant 1$. We will use the notation ν_p for the p-adic norm on $\mathbb{Q}$.

Product formula: For each $x \in \mathbb{Q} \setminus \{0\}$

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

Indeed, 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 $x \neq 0$. A similar product formula is true for an arbitrary algebraic number field F:

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

where $N_\nu = [F_\nu : \mathbb{Q}_\nu]$, see [**?**, Chapter 6].

DEFINITION 9.74. The ring of *adeles* is the *restricted product*

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

i.e. the subset of the direct product

$$\prod_{\nu \in \mathrm{Nor}(F)} F_\nu \tag{9.5}$$

which consists of sequences whose projection to F_ν belongs to O_ν for all but finitely many ν's.

We topologize $\mathbb{A}(F)$ *via* the subset topology induced from the product (9.5), which, in turn, is equipped with the product topology. Note that the ring operations are continuous with respect to this topology.

For instance, if $F = \mathbb{Q}$ then $\mathbb{A}(\mathbb{Q})$ is the restricted product

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

REMARK 9.75. Actually, it suffices to use the ring of adeles $\mathbb{A}(\mathbb{Q})$. This is done *via* the following procedure called *the restriction of scalars*: The field F is an m-dimensional vector space over $\mathbb{Q}$. This determines an embedding

$$\Gamma \subset GL(n, F) \hookrightarrow \prod_{i=1}^{m} GL(n, \mathbb{Q}) \subset GL(n+m, \mathbb{Q})$$

and reduces our discussion to the case $\Gamma \subset GL(n+m, \mathbb{Q})$.

Now, a miracle happens:

THEOREM 9.76 (See e.g. Chapter 6, Theorem 1 of [**?**]). *The image of the diagonal embedding $F \hookrightarrow \mathbb{A}(F)$ is a discrete subset in $\mathbb{A}(F)$.*

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

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

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

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

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

In order to appreciate this theorem, the reader should consider the case $F = \mathbb{Q}$ which is dense in the completion of $\mathbb{Q}$ with respect to every norm.

Recall that Γ is a subgroup in $SL(n, F)$. The diagonal embedding above defines an embedding

$$\Gamma \subset SL(n, F) \hookrightarrow SL(n, \mathbb{A}(F)) \subset \prod_{\nu \in \mathrm{Nor}(F)} SL(n, F_\nu)$$

with discrete image.

For each norm $\nu \in \mathrm{Nor}(F)$ we have the projection $p_\nu : \Gamma \to SL(n, F_\nu)$. If the image $p_\nu(\Gamma)$ is relatively compact for each ν then Γ is relatively compact in $\prod_{\nu \in \mathrm{Nor}(F)} SL(n, F_\nu)$, by Tychonoff's Theorem. As Γ is also discrete, this implies that Γ is finite, a contradiction.

Thus, there exists a norm $\nu \in \mathrm{Nor}(F)$ such that the image of Γ in $SL(n, F_\nu)$ is not relatively compact. If ν happens to be archimedean we are done as before. The more interesting case occurs if ν is nonarchimedean. Then the field $F_\nu = \mathbf{k}$ is a local field (just like the p-adic completion of the rational numbers) and we appeal to Theorem 9.49 to conclude that Γ contains a free subgroup in this case as well. This concludes the proof of the Tits' Alternative (Theorem 9.1). □

REMARK 9.77. 1. The above proof works only if Γ is finitely generated. The general case will be treated below.

2. Tits' proof also works for algebraic groups over fields of positive characteristic, see [**?**]. However, in the case of infinitely-generated groups one has to modify the assertion, since $GL(n, F)$, where F is an infinite algebraic extension of a finite field, provides a counter-example otherwise.

3. The arguments in the above proof mostly follow the ones of Breuillard and Gelander in [**?**].

Note that a consequence of the previous arguments is the following.

THEOREM 9.78. *Let Γ be a finitely generated group that does not contain a free non-abelian subgroup. Then:*

(1) *If Γ is a subgroup of an algebraic group L then its Zariski closure G is virtually solvable.*
(2) *If Γ is a subgroup of a Lie group L with finitely many connected components, then the closure G of Γ in the Lie group L is virtually solvable.*

Furthermore, in both cases above, the solvable subgroup S of G has derived length at most $\delta = \delta(L)$ and the index $|G : S|$ is at most $\nu = \nu(L)$.

PROOF. The arguments in the proof of Theorem 9.1 imply the statement (1). The statement (2) follows in a similar manner. Indeed, as in Section 9.4, using the adjoint representation one can reduce the problem to the setting of linear subgroups, and there the closure in the standard topology is contained in the Zariski closure. □

Tits Alternative without finite generation assumption.

We will need

LEMMA 9.79. *Every countable field F of zero characteristic embeds in $\mathbb{C}$.*

PROOF. Since F has characteristic zero, its prime subfield P is isomorphic to $\mathbb{Q}$. Then F is an extension of the form

$$P \subset E \subset F,$$

where $P \subset E$ is an algebraic extension and $E \subset F$ is a purely transcendental extension (see [Chapter VI.1][**?**]). The algebraic number field E embeds in $\bar{Q} \subset \mathbb{C}$. Since F is countable, F/E has countable dimension and, therefore,

$$F = E(u_1, \ldots, u_m)$$

or

$$F = E(u_1, \ldots, u_m, \ldots).$$

Sending variables u_j to independent transcendental numbers $z_j \in \mathbb{C}$, we then obtain an embedding $F \hookrightarrow \mathbb{C}$. $\square$

THEOREM 9.80 (Tits Alternative). *Let F be a field of zero characteristic and Γ be a subgroup of $GL(n, F)$. Then either Γ is virtually solvable or Γ contains a free nonabelian subgroup.*

PROOF. The group Γ is the direct limit of the direct system of its finitely-generated subgroups Γ_i. Let $F_i \subset F$ denote the subfield generated by the matrix entries of the generators of Γ_i. Then $\Gamma_i \leqslant GL(n, F_i)$. Since F (and, hence, every F_i) has zero characteristic, the field F_i embeds in $\mathbb{C}$ (see Lemma 9.79).

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 linear groups), each Γ_i is virtually solvable. For $\nu = \nu(GL(n, \mathbb{C}))$ and $\delta = \delta(GL(n, \mathbb{C}))$, every i there exists a subgroup $\Lambda_i \leqslant \Gamma_i$ of index $\leqslant \delta$, so that Λ_i has derived length $\leqslant \delta$ (see Theorem 9.38). In view of Exercise **??**, the group Γ is also virtually solvable. $\square$

CHAPTER 10

The Banach-Tarski paradox

10.1. Paradoxical decompositions

DEFINITION 10.1. Two subsets A, B in a metric space (X, dist) are *congruent* if there exists an isometry $\phi : X \to X$ such that $\phi(A) = B$.

DEFINITION 10.2. Two sets A, B in a metric space X are *piecewise congruent* (or *equidecomposable*) if, for some $k \in \mathbb{N}$, they admit partitions $A = A_1 \sqcup ... \sqcup A_k$, $B = B_1 \sqcup ... \sqcup B_k$ such that for each $i \in \{1, ..., k\}$, the sets A_i and B_i are congruent.

Two subsets A, B in a metric space X are *countably piecewise congruent* (or *countably equidecomposable*) if they admit partitions $A = \bigsqcup_{n \in \mathbb{N}} A_n$, $B = \bigsqcup_{n \in \mathbb{N}} B_n$ such that for every $n \in \mathbb{N}$, the sets A_n and B_n are congruent.

REMARK 10.3. Thus, by using empty sets for some A_n, B_n, we see that piecewise congruence as a stronger form of countably piecewise congruence.

EXERCISE 10.4. Prove that (countably) piecewise congruence is an equivalence relation.

DEFINITION 10.5. A set E in a metric space X is *paradoxical* if there exists a partition

$$E = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

and isometries $\varphi_1, ..., \varphi_k, \psi_1, ..., \psi_m$ of X, so that

$$\varphi_1(X_1) \sqcup ... \sqcup \varphi_k(X_k) = E$$

and

$$\psi_1(Y_1) \sqcup ... \sqcup \psi_m(Y_m) = E\ .$$

A set E in a metric space X is *countably paradoxical* if there exists a partition

$$E = \bigsqcup_{n \in \mathbb{N}} X_n \sqcup \bigsqcup_{m \in \mathbb{N}} Y_m$$

and two sequences of isometries $(\varphi_n)_{n \in \mathbb{N}}, (\psi_m)_{m \in \mathbb{N}}$ of X, so that

$$\bigsqcup_{n \in \mathbb{N}} \varphi_n(X_n) = E, \text{ and } \bigsqcup_{m \in \mathbb{N}} \psi_m(Y_m) = E.$$

EXERCISE 10.6. 1. If $E, E' \subset X$ are piecewise-congruent and E is paradoxical, then so is E'.

2. If $E, E' \subset X$ are countably piecewise-congruent and E is countably paradoxical, then so is E'.

Using earlier work of Vitali and Hausdorff, Banach and Tarski proved the following:

THEOREM 10.7 (Banach-Tarski paradox [**?**]). (1) *Any two bounded subsets with non-empty interior in $\mathbb{R}^n$ (for $n \gg 3$) are piecewise congruent.*

(2) *Any two bounded subsets with non-empty interior in $\mathbb{R}^n$ (for $n \in \{1, 2\}$) are countably piecewise congruent.*

COROLLARY 10.8. (1) *Every Euclidean ball is paradoxical in $\mathbb{R}^n, n \geqslant 3$, and countably paradoxical in $\mathbb{R}^n, n \in \{1, 2\}$.*

(2) *For every $n \geqslant 3$ and every $m \in \mathbb{N}$, every ball in $\mathbb{R}^n$ is piecewise congruent to m copies of this ball (one can "double" the ball).*

(3) *A pea and the sun are piecewise congruent (any two Euclidean n-balls are piecewise-congruent for $n \geqslant 3$).*

REMARK 10.9. The Banach-Tarski paradox emphasizes that it is impossible to find a finitely-additive measure defined on *all* subsets of the Euclidean space of dimension at least 3 that is invariant with respect to isometries and takes the value one on a unit cube. The main point in their theorem is that the congruent pieces A_i, B_i are not *Lebesgue measurable.*

REMARK 10.10 (Banach-Tarski paradox and axiom of choice). The Banach-Tarski paradox is neither provable nor disprovable with Zermelo-Fraenkel axioms (ZF) only: It is impossible to prove that a unit ball in $\mathbb{R}^3$ 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 [**?**] and J. Pawlikowski [**?**] 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 **??**).

10.2. Step 1 of the proof of the Banach–Tarski theorem

We will prove only Corollary 10.8, Parts 1 and 2 and only in the case $n \leqslant 3$. The general statement of Theorem 10.7 for two bounded subset with non-empty interiors is derived from the doubling of a ball by using the Banach–Bernstein-Schroeder theorem (see [**?**]). The general statement in $\mathbb{R}^n, n \geqslant 3$, can be easily either derived from the statement for $n = 3$, or proved directly by adapting the proof in dimension 3.

The first step in the proof is common to all dimensions.

Step 1: The unit sphere $\mathbb{S}^n$ is piecewise congruent to $\mathbb{S}^n \setminus C$, where C is any countable set, and $n \geq 2$.

We first prove that there exists a rotation ρ around the origin such that for any integer $n \geqslant 1$, $\rho^n(C) \cap C = \emptyset$. This is obvious in the plane (only a countable set of rotations do not satisfy this).

In the space we first select a line ℓ through the origin such that its intersection with $\mathbb{S}^2$ is not in C. Such a line exists because the set of lines through the origin containing points in C is countable. Then we look for a rotation ρ_θ of angle θ around ℓ such that for any integer $n \geqslant 1$, $\rho_\theta^n(C) \cap C = \emptyset$. Indeed take A the set of angles α such that the rotation of angle α around ℓ sends a point in C to another point in C. There are countably many such angles, therefore the set $A' = \bigcup_{n \geqslant 1} \frac{1}{n} A$ is also countable. Thus, we may choose an angle $\theta \notin A'$.

Take $\mathcal{O} = \bigcup_{n \geqslant 0} \rho_\theta^n(C)$ and decompose $\mathbb{S}^2$ as $\mathbb{S}^2 = \mathcal{O} \sqcup (\mathbb{S}^2 \setminus \mathcal{O})$. Then $(\mathcal{O} \setminus C) \sqcup (\mathbb{S}^2 \setminus \mathcal{O}) = \mathbb{S}^2 \setminus C$. We, thus, have a piecewise congruence of $\mathbb{S}^2$ to $\mathbb{S}^2 \setminus C$ which sends $\mathcal{O}$ to $\mathcal{O} \setminus C$ by ρ_θ and is the identity on $\mathbb{S}^2 \setminus \mathcal{O}$.

Step 2 (using the axiom of choice): The unit circle $\mathbb{S}^1$ is countably paradoxical.

Let α be an irrational number and let $R \in SO(2)$ be the counter-clockwise rotation of angle $2\pi\alpha$. Then the map $m \mapsto R^m$ is an injective homomorphism $\mathbb{Z} \to SO(2)$. *Via* this homomorphism, $\mathbb{Z}$ acts on the unit circle $\mathbb{S}^1$. *According to the axiom of choice* there exists a subset $D \subset \mathbb{S}^1$ which intersects every $\mathbb{Z}$-orbit in exactly one point.

Since $\mathbb{Z}$ decomposes as $2\mathbb{Z} \sqcup (2\mathbb{Z}+1)$, the unit circle decomposes as

$$2\mathbb{Z} \cdot D \sqcup (2\mathbb{Z}+1) \cdot D.$$

Now, for each $X_n = R^{2n} \cdot D$ consider the isometry $\varphi_n = R^{-n}$, and for each $Y_n = R^{2n+1} \cdot D$ consider the isometry $\psi_n = R^{-n-1}$. Clearly $\mathbb{S}^1 = \bigsqcup_{n\in\mathbb{Z}} \varphi_n(X_n)$ and $\mathbb{S}^1 = \bigsqcup_{n\in\mathbb{Z}} \psi_n(Y_n)$.

Step 3: The unit disk $\mathbb{D}^2$ is countably paradoxical.

Let $\mathbb{D}^2$ be the closed unit disk in $\mathbb{R}^2$ centered at a point O. Step 1 and the fact that $\mathbb{D}^2 \setminus \{O\}$ can be written as the set

$$\{\lambda x \,;\, \lambda \in (0,1]\,,\, x \in \mathbb{S}^1\},$$

imply that $\mathbb{D}^2\setminus\{O\}$ is countably paradoxical. Thus, it suffices to prove that $\mathbb{D}^2\setminus\{O\}$ is piecewise congruent to $\mathbb{D}^2$. Take $\mathbb{S}^1\left(\left(\frac{1}{2},0\right), \frac{1}{2}\right)$, the unit circle with center $\left(\frac{1}{2},0\right)$ and radius $\frac{1}{2}$. For simplicity, we denote this circle $\mathbb{S}_{1/2}$. Then

$$\mathbb{D}^2 \setminus \{O\} = \mathbb{D}^2 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} \setminus \{O\}.$$

According to Step 1, $\mathbb{S}_{1/2} \setminus \{O\}$ is piecewise congruent to $\mathbb{S}_{1/2}$, hence $\mathbb{D}^2 \setminus \{O\}$ is piecewise congruent to

$$\mathbb{D}^2 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} = \mathbb{D}^2. \quad \square$$

Remark 10.11 (Stronger result). Instead of the splitting $\mathbb{Z} = 2\mathbb{Z} \sqcup (2\mathbb{Z}+1)$ of $\mathbb{Z}$ into two 'copies' of itself, we might consider a splitting of $\mathbb{Z}$ into infinitely countably many 'copies' of itself. Indeed the subsets $\mathbb{Z}^{(k)} = 2^k\mathbb{Z} + 2^{k-1}\,,\, k \in \mathbb{N}\,,$ form a partition of $\mathbb{Z}$. This allows to prove, following the same proof as above, that a unit disk is countably piecewise congruent to countably many copies of itself.

Proof. As in Step 2, we write $\mathbb{S}^1 = \mathbb{Z}D = \bigsqcup_{k\in\mathbb{N}} \mathbb{Z}^{(k)} D$. The idea is to move by isometries the copies of D in $\mathbb{Z}^{(k)}D$ so as to form the k-th copy of the unit circle. Indeed, if for the set $X_{k,m} = R^{2^k m + 2^{k-1}} D$ we consider the isometry

$$\phi_{k,m} = T_{(2k,0)} \circ R^{-2^k m - 2^{k-1} + m},$$

then

$$\bigsqcup_{m\in\mathbb{Z}} \phi_{k,m}(X_{k,m})$$

is equal to $\mathbb{S}^1((2k,0),1)$.

Thus, $\mathbb{S}^1$ is countably piecewise congruent to

$$\bigsqcup_{k\in\mathbb{N}} \mathbb{S}^1((2k,0),1).$$

This extends to the corresponding disks with their centers removed. In Step 3 we proved that a punctured disk is piecewise congruent to the full disk. This allows to finish the argument. $\square$

10.4. Proof of the Banach–Tarski theorem in the space

We now explain prove Banach–Tarski theorem for A, the unit ball in $\mathbb{R}^3$ and B, the disjoint union of two unit balls in $\mathbb{R}^3$.

Step 2: a paradoxical decomposition for the free group of rank 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

$$F_2 = \{1\} \sqcup \mathcal{W}_a \sqcup \mathcal{W}_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{10.1}$$

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 $\mathcal{W}_{a^{-1}}$:

$$\mathcal{W}'_a = \mathcal{W}_a \setminus \{a^n \; ; \; n \geqslant 1\} \text{ and } \mathcal{W}'_{a^{-1}} = \mathcal{W}_{a^{-1}} \sqcup \{a^n \; ; \; n \geqslant 0\} .$$

Then

$$F_2 = \mathcal{W}'_a \sqcup \mathcal{W}'_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{10.2}$$

and

$$F_2 = L_a\mathcal{W}'_{a^{-1}} \cup \mathcal{W}'_a .$$

Step 3: A paradoxical decomposition for the unit sphere (using the axiom of choice).

According to the Tits Alternative (see also Example 9.29), the free group F_2 embeds as a subgroup in the orthogonal group $SO(3)$. For every $w \in F_2$ we denote by R_w the rotation of $\mathbb{R}^3$ given by this embedding.

Let C be the (countable) set of intersections of $\mathbb{S}^2$ with the union of axes of the rotations R_w, $w \in F_2 \setminus \{1\}$. Since C is countable, by Step 1, $\mathbb{S}^2$ is piecewise congruent to $\mathbb{S}^2 \setminus C$. The set $\mathbb{S}^2 \setminus C$ is a disjoint union of orbits of F_2. *According to the axiom of choice* there exists a subset $D \subset \mathbb{S}^2 \setminus C$ which intersects every F_2-orbit in $\mathbb{S}^2 \setminus C$ exactly once. (The removal of the set C ensures that the action of F_2 is free, i.e., no nontrivial element of F_2 fixes a point, that is all orbits are copies of F_2.)

By Step 2,

$$F_2 = \mathcal{W}'_a \sqcup \mathcal{W}'_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}}.$$

This defines a decomposition

$$\mathbb{S}^2 \setminus C = F_2 \cdot D = \mathcal{W}'_a \cdot D \sqcup \mathcal{W}'_{a^{-1}} \cdot D \sqcup \mathcal{W}_b \cdot D \sqcup \mathcal{W}_{b^{-1}} \cdot D . \tag{10.3}$$

The fact that the subsets in the union (10.3) are pairwise disjoint reflects the fact that the action of F_2 on $\mathbb{S}^2 \setminus C$ is free. Since F_2 admits a paradoxical decomposition, so does $\mathbb{S}^2 \setminus C$. Since the latter is piecewise-congruent to $\mathbb{S}^2$, it follows that $\mathbb{S}^2$ also admits a paradoxical decomposition.

We will now show that $\mathbb{S}^2$ is piecewise congruent to a disjoint union of two copies of $\mathbb{S}^2$. Let v denote the vector $(3,0,0)$ in $\mathbb{R}^3$ and let T_v denote the isometry of $\mathbb{R}^3$ which is the translation by v.

In view of the decomposition (10.3), the set $\mathbb{S}^2 \setminus C$ is piecewise congruent to

$$\mathcal{W}'_a \cdot D \sqcup R_a \mathcal{W}'_{a^{-1}} \cdot D \sqcup T_v\left(\mathcal{W}_b D\right) \sqcup T_v \circ R_b\left(\mathcal{W}_{b^{-1}} D\right) = \mathbb{S}^2 \setminus C \sqcup T_v\left(\mathbb{S}^2 \setminus C\right) .$$

This and Step 1 imply that $\mathbb{S}^2$ is piecewise congruent to $\mathbb{S}^2 \sqcup T_v \mathbb{S}^2$, i.e., one can "double" the ball. Part 2 of Corollary 10.8 now follows by induction.

Step 4: A paradoxical decomposition for the unit ball.

The argument is very similar to the last step in the 2-dimensional case. Let $\mathbb{B}^3$ denote the closed unit ball in $\mathbb{R}^3$ centered at O. Step 3 and the fact that the unit ball $\mathbb{B}^3 \setminus \{O\}$ can be written as the set

$$\{\lambda x\ ;\ \lambda \in (0,1]\,,\, x \in \mathbb{S}^2\},$$

imply that $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to

$$\mathbb{B}^3 \setminus \{O\} \sqcup T_v\left(\mathbb{B}^3 \setminus \{O\}\right).$$

Thus, it remains to prove that $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to $\mathbb{B}^3$. We denote by $\mathbb{S}_{1/2}$ the sphere with the center $\left(\frac{1}{2},0,0\right)$ and radius $\frac{1}{2}$. Then

$$\mathbb{B}^3 \setminus \{O\} = \mathbb{B}^3 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} \setminus \{O\}.$$

According to Step 1, $\mathbb{S}_{1/2} \setminus \{O\}$ is piecewise congruent to $\mathbb{S}_{1/2}$; hence, $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to $\mathbb{B}^3 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} = \mathbb{B}^3$.

This concludes the proof of Corollary 10.8, Parts 1 and 2, for $n \leqslant 3$. □

Remark 10.12. Banach and Tarski's proof relies on the Hausdorff's paradox, discovered several years prior to their proof. Inspired by the Hausdorff's argument, R. M. Robinson, answering a question of von Neumann, proved in [**?**] that five is the minimal number of pieces in a paradoxical decomposition of the unit 3-dimensional ball. See Proposition 11.90 for a proof of this statement, and Section 11.7 for a discussion on the minimal number of pieces in a paradoxical decomposition.

Remark 10.13.

(1) The free group F_2 of rank 2 contains a free subgroup of countably infinite rank, see Proposition 4.47. This and a proof similar to the one of Theorem 10.7 yields that the unit sphere $\mathbb{S}^{n-1}$ is countably piecewise congruent to countably many copies of $\mathbb{S}^{n-1}$.

(2) It can be proved that the unit 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 [**?**]).

CHAPTER 11

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 the graph version of amenability, then we will turn to the case of groups, and then to the opposite property of being *paradoxical*.

11.1. Amenable graphs

DEFINITION 11.1. A graph $\mathcal{G}$ is called *amenable* if its Cheeger constant, as described in Definition **??**, is zero. Equivalently, there exists a sequence F_n of finite subsets of V such that

$$\lim_{n\to\infty} \frac{|E(F_n, F_n^c)|}{|F_n|} = 0\,.$$

Such sequence F_n is called a *Følner sequence for the graph* $\mathcal{G}$.

A graph $\mathcal{G}$ is *non-amenable* if its Cheeger constant is strictly positive.

It is immediate from the definition that every finite graph is amenable (take $F_n = V$).

We describe in what follows various metric properties equivalent to non-amenability. Our arguments are adapted from [**?**]. 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 11.2 (Hall-Rado [**?**], §$III.2$). *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 edge-boundary $E(A, A^c)$ contains at least $k|A|$ elements.*
- *For every finite subset B in Z, its edge-boundary $E(B, B^c)$ 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 $\mathrm{dist}(y, z) \leqslant C$. (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 graph with the vertex-set V we will use the notation $\overline{\mathcal{N}}_C(F)$ and $\mathcal{N}_C(F)$ to denote the "closed" and "open" C-neighborhood of F in V:

$$\overline{\mathcal{N}}_C(F) = \{v \in V : \text{dist}(v, F) \leqslant C\}, \quad \mathcal{N}_C(F) = \{v \in V : \text{dist}(v, F) < C\}.$$

THEOREM 11.3. *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 $F \subset V$, the set $\overline{\mathcal{N}}_C(F) \subset V$ contains at least twice as many vertices as F.*

(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 5.10) such that for every $v \in V$ the preimage $f^{-1}(v)$ contains exactly two elements.*

(e) (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.*

REMARK 11.4. Property (b) can be replaced by the property (b') that for some (equivalently, every) $\beta > 1$ there exists $C > 0$ such that $\overline{\mathcal{N}}_C(F) \cap V$ has cardinality at least β times the cardinality of F. Indeed, it suffices to observe that for every $\alpha > 1$, $C > 0$,

$$\forall F, |\overline{\mathcal{N}}_C(F)| \geqslant \alpha |F| \Rightarrow \forall k \in \mathbb{N}, \quad |\overline{\mathcal{N}}_{kC}(F)| \geqslant \alpha^k |F|.$$

PROOF. We will now prove Theorem 11.3. Let $m \geq 1$ denote the valence of $\mathcal{G}$.
(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, $|E(F, F^c)| \geqslant \eta |F|$. This implies that $\overline{\mathcal{N}}_1(F)$ contains at least $(1 + \frac{\eta}{m})|F|$ vertices, which, according to Remark 11.4, implies property (b).

(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 11.2 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 $\text{dist}_{\mathcal{G}}(z, f(z)) \leqslant C$. Hence, $f \in \mathcal{B}(V)$ and $|f^{-1}(y)| = 2$ for every $y \in V$.

The implication (d) $\Rightarrow$ (e) is obvious. We show that (e) $\Rightarrow$ (b). According to (e), there exists a constant $M > 0$ and a map $f : V \to V$ such that for every $x \in V$, $\text{dist}(x, f(x)) \leqslant M$, and $|f^{-1}(y)| \geq 2$ for every $y \in V$. For every finite nonempty set $F \subset V$, $f^{-1}(F)$ is contained in $\mathcal{N}_M(F)$ and it has at least twice as many elements. Thus, (b) is satisfied.

Thus, we proved that the properties (b) through (e) 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 $F \subset V$, $|\overline{\mathcal{N}}_C(F) \cap V| \geqslant 2|F|$. Without loss of generality, we may assume that C is a positive integer. Recall that $\partial_V F$ is the vertex–boundary of the subset $F \subset V$. Since $\overline{\mathcal{N}}_C(F) = F \cup \mathcal{N}_C(\partial_V F)$, it follows that $|\mathcal{N}_C(\partial_V F) \setminus F| \geqslant |F|$.

Recall that the graph $\mathcal{G}$ has finite valence $m \geqslant 1$. Therefore,

$$|\overline{\mathcal{N}}_C(\partial_V F)| \leqslant m^C |\partial_V F| \,.$$

We have, thus, obtained that for every finite nonempty set $F \subset V$,

$$|E(F, F^c)| \geqslant |\partial_V F| \geqslant \frac{1}{m^C} |\mathcal{N}_C(\partial_V F)| \geqslant \frac{1}{m^C} |F|.$$

Therefore, the Cheeger constant of $\mathcal{G}$ is at least $\frac{1}{m^C} > 0$, and the graph is non-amenable. $\square$

EXERCISE 11.5. Show that a sequence $F_n \subset V$ is Følner if and only if for every $C \in \mathbb{R}_+$

$$\lim_{n \to \infty} \frac{|\mathcal{N}_C(F_n)|}{|F_n|} = 1.$$

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 [?].

PROPOSITION 11.6. *A graph $\mathcal{G}$ of bounded geometry and sub-exponential growth (in the sense of Definition ??) 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. $\square$

For the sake of completeness we mention without proof two more properties equivalent to those in Theorem 11.3.

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 11.77).

THEOREM 11.7 (Theorem 1.3 in [**?**]). *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.

An amenability criterion with random walks. 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 [**?, ?, ?**] 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 11.8 (J. Dodziuk, [**?**]). *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 [**?**].

COROLLARY 11.9. *In a non-amenable graph of bounded geometry, the simple random walk is transient, that is, for every $x, y \in V$,*

$$\sum_{n=1}^{\infty} p_n(x,y) < \infty.$$

11.2. Amenability and quasi-isometry

THEOREM 11.10 (Graph amenability is QI invariant). *Suppose that $\mathcal{G}$ and $\mathcal{G}'$ are quasi-isometric 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 quasi-isometry invariant. We will assume that both $\mathcal{G}$ and $\mathcal{G}'$ are infinite, otherwise the assertion is clear. Note that according to Theorem 11.3, 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 $\bar{\mathcal{N}}_C(F)$ contains at least $2|F|$ 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 quasi-inverse 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

$$F \xrightarrow{f} F' = f(F) \xrightarrow{g} F'' = g(F').$$

Since F'' is at Hausdorff distance $\leqslant A$ from F, it follows that $|F| \leqslant b|F''|$, where $b = m^L$. In particular,

$$|f(F)| \geqslant b^{-1}|F|.$$

Likewise, for every finite set F' in V' we obtain

$$|g(F')| \geqslant b^{-1}|F'| \ .$$

Remark 11.4 implies that 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, 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 $F \subset V$ and set $F' := f(F), F'' = gf(F)$. Then $|F'| \geq b^{-1}|F|$ and, therefore,

$$|g\left(\bar{\mathcal{N}}_C(F')\right)| \geqslant \frac{\alpha}{b^2}|F|.$$

Since g is L–Lipschitz,

$$g\left(\bar{\mathcal{N}}_C(F')\right) \subset \bar{\mathcal{N}}_{LC}(F'') \subset \bar{\mathcal{N}}_{LC+A}(F).$$

We conclude that

$$|\bar{\mathcal{N}}_{LC+A}(F)| \geqslant \frac{\alpha}{b^2}|F|.$$

Setting $C' := LC + A$, and $\beta := \frac{\alpha}{b^2} > 1$, we conclude that $\mathcal{G}$ satisfies the expansion property (b') in Theorem 11.3. Hence, $\mathcal{G}$ is also non-amenable. $\square$

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, quasi-isometric 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 2.20). This may be seen as a version within the setting of amenability/isoperimetric problem of the Milnor–Efremovich–Schwartz Theorem **??** stating that the growth functions of M and $\mathcal{G}$ are equivalent.

In what follows we use the terminology in Definitions 2.56 and 2.60 for the bounded geometry of a Riemannian manifold, respectively of a simplicial graph.

THEOREM 11.11. *Let M be a complete connected n-dimensional Riemannian manifold and $\mathcal{G}$ a simplicial graph, both of bounded geometry. Assume that M is quasi-isometric to $\mathcal{G}$. Then the Cheeger constant of M is strictly positive if and only if the graph $\mathcal{G}$ is non-amenable.*

REMARKS 11.12. (1) Theorem 11.11 was proved by R. Brooks [**Bro82a**], [**Bro81**] 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 11.11 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 [**?**] and Proposition 11 in [**?**].

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 2.24, 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 2.58, (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 5.41, the manifold M is quasi-isometric to $\mathcal{G}_\varepsilon$. Furthermore, $\mathcal{G}_\varepsilon$ has bounded geometry as well. This and Theorem 11.10 imply that $\mathcal{G}_\varepsilon$ has strictly positive Cheeger constant if and only if $\mathcal{G}$ has. Thus, it suffices to prove the equivalence in Theorem 11.11 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 1.11),

$$\text{card}\, E(F, F^c) \geqslant B\, Area(\partial\Omega) \;\text{ and }\; C Vol(\Omega) \geqslant \text{card}\, F\,. \tag{11.1}$$

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 2.8, such that all the functions φ_x are L–Lipschitz for some constant $L > 0$ independent of x, see Lemma 2.23. Let $F \subset N$ be a finite subset. Consider the smooth function $\Phi = \sum_{x\in F} \varphi_x$. By hypothesis and since $\mathcal{U}$ has multiplicity at most m, the function Φ is Lm–Lipschitz. Furthermore, 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

$$\Omega_t = \Phi^{-1}((t,\infty)) \subset M$$

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 $\Phi^{-1}(t)$, 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 2.24 we get

$$Vol(\Omega_t) \geqslant \sum_{x\in F} Vol\left(x, \frac{\varepsilon}{2}\right) \geqslant \operatorname{card} F \cdot V_b\left(\varepsilon/2\right).$$

Therefore, for every $t \notin \Theta$, the domain Ω_t satisfies the second inequality in (11.1) 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 (11.1) 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,

$$U \subset \left(\bigcup_{x\in F'} \overline{B(x,\varepsilon)} \right) \cap \left(\bigcup_{z\in N\setminus F} \overline{B(z,\varepsilon)} \right). \tag{11.2}$$

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 2.16, with $g \equiv 1$, $f = \Phi$ and $U = \Phi^{-1}(0,\eta)$, we obtain:

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{dt} = \int_U |\nabla\Phi| \mathrm{d}V \leqslant mLVol(U) \leqslant mL \sum_{x\in F'} Vol(B(x,\varepsilon)).$$

The last inequality follows from the inclusion (11.2). At the same time, by applying Theorem 2.24, we obtain that for every $x \in M$

$$V_a(\varepsilon) \geqslant Vol(B(x,\varepsilon)).$$

By combining these inequalities, we obtain

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{dt} \leqslant mLV_a(\varepsilon)\,|F'| \leqslant mLV_a(\varepsilon)\,|E\left(F, F^c\right)|.$$

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 (11.1) 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 F_k in N such that for two constants P and Q independent of Ω, we have

$$Area(\partial\Omega) \geqslant P\,|E(F_k, F_k^c)| \;\text{ and }\; |F_k| \geqslant QVol(\Omega)\,. \tag{11.3}$$

This would imply positivity of Cheeger constant of M. Note that, since the graph $\mathcal{G}$ has finite valence, in the first inequality of (11.3) we may replace the edge boundary $E(F_k, F_k^c)$ by the vertex boundary $\partial_V F_k$ (see Definition 1.11).

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 into two parts:

$$F_1 = \left\{ x \in F \;:\; Vol[\Omega \cap B(x, \delta)] > \frac{1}{2} Vol[B(x, \delta)] \right\} \tag{11.4}$$

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 2.24, this inequality implies that

$$\frac{1}{2} Vol(\Omega) \leqslant \sum_{x\in F_1} Vol\,(B(x, \delta)) \leqslant |F_1|\, V_a(\delta)\,. \tag{11.5}$$

This gives the second inequality in (11.3). A point x in $\partial_V F_1$ is then a point in N satisfying (11.4), such that within distance 8δ of x there exists a point $y \in N$ satisfying the inequality opposite to (11.4). 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

$$V_b\,(\delta/2)\ |\partial_V F_1| \leqslant Vol\left[\bigsqcup_{x\in\partial_V F_1} B\left(x, \frac{\delta}{2}\right)\right] \leqslant$$

$$\sum_{y\in H_\delta} Vol\,[B(y, 10\delta)] \leqslant V_b(10\delta)\,|H_\delta|. \tag{11.6}$$

Since H_δ extends to a maximal δ–separated subset H' of M, Lemma 2.58, (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,

$$\lambda Area(\partial\Omega \cap B(y,\delta)) \geqslant Vol\,[\Omega \cap B(y,\delta)] = \frac{1}{2} Vol[B(y,\delta)]\,.$$

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 (11.6), 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 (11.3) 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

$$mArea(\partial\Omega) \geqslant \sum_{y\in F_2} Area\,(\partial\Omega \cap B(y,\delta)) \geqslant \frac{1}{\lambda} \sum_{y\in F_2} Vol\,[\Omega \cap B(y,\delta)] \geqslant \frac{1}{2\lambda} Vol(\Omega)\,.$$

Thus, in the Case 2 we obtain the required lower bound on $Area(\partial\Omega)$ directly. □

COROLLARY 11.13. *Let M and M' be two complete connected Riemann manifolds of bounded geometry which are quasi-isometric 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 quasi-isometric to M and M' respectively. Then $\mathcal{G}, \mathcal{G}'$ are also quasi-isometric to each other. The result now follows by combining Theorem 11.11 with Theorem 11.10. □

An interesting consequence of Corollary 11.13 is quasi-isometric invariance of a certain property of the Laplace-Beltrami operator for Riemannian manifolds of bounded geometry. Cheeger constant for Riemannian manifold M is closely connected to the bottom of the spectrum of the Laplace-Beltrami operator Δ_M on $L^2(M) \cap C^\infty(M)$. Let M be a complete connected Riemannian manifold of infinite volume, let $\lambda_0(M)$ denote the lowest eigenvalue of Δ_M. Then $\lambda_0(M)$ can be computed as

$$\inf\left\{\frac{\int_M |\nabla f|^2}{\int_M f^2} \mid f : M \to \mathbb{R} \text{ smooth with compact support}\right\}$$

(see [**?**] or [**SY94**], Chapter I). J. Cheeger proved in [**Che70**] that

$$\lambda_0(M) \geqslant \frac{1}{4} h^2(M)\,,$$

where $h(M)$ is the Cheeger constant of M. Even though Cheeger's original result was formulated for compact manifolds, his argument works for all complete manifolds, 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 that $h(M) = 0 \iff \lambda_0(M) = 0$.

COROLLARY 11.14. *Let M and M' be two complete connected Riemann manifolds of bounded geometry which are quasi-isometric to each other. Then $\lambda_0(M) = 0 \iff \lambda_0(M') = 0$.*

We finish the section by noting a remarkable property of quasi-isometries between non-amenable graphs.

THEOREM 11.15 (K. Whyte [**?**]). *Let $\mathcal{G}_i, i = 1, 2,$ be two non-amenable graphs of bounded geometry. Then every quasi-isometry $\mathcal{G}_1 \to \mathcal{G}_2$ is at bounded distance from a bi-Lipschitz map.*

Note that this theorem was also implicit in [**?**].

11.3. Amenability for groups

We now discuss the concept of amenability for groups. We introduce various versions of amenability and non-amenability, formulated in terms of actions and inspired by the Banach-Tarski paradox. We then show that in the case of finitely generated groups one of the notions of amenability is equivalent to the metric amenability for (arbitrarily chosen) Cayley graphs, as formulated in Definition 11.1.

Let G be a group acting on a set X. We assume that the action is on the left (for an action on the right a similar discussion can be carried out). We denote the action by $\mu(g, x) = g(x) = g \cdot x$.

We say that two subsets $A, B \subset X$ are *G–congruent* if there exists $g \in G$ such that $g \cdot A = B$.

We say subsets $A, B \subset X$ are *G–piecewise congruent* (or A and B are *G–equidecomposable*) if, for some $k \in \mathbb{N}$, there exist partitions $A = A_1 \sqcup \ldots \sqcup A_k$, $B = B_1 \sqcup \ldots \sqcup B_k$ such that A_i and B_i are G–congruent for every $i \in \{1, \ldots, k\}$.

The subsets A, B are *G–countably piecewise congruent* (or *G–countably equidecomposable*) if they admit countable partitions $A = \bigsqcup_{n\in\mathbb{N}} A_n$, $B = \bigsqcup_{n\in\mathbb{N}} B_n$ such that A_n and B_n are G–congruent for every $n \in \mathbb{N}$.

EXERCISE 11.16. Verify that piecewise congruence and countably piecewise congruence are equivalence relations.

DEFINITIONS 11.17. (1) A *G–paradoxical subset of* X is a subset E that admits a *G-paradoxical decomposition*, i.e., a finite partition

$$E = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

such that for some elements $g_1, \ldots, g_k, h_1, \ldots, h_m$ of G,

$$g_1(X_1) \sqcup \ldots \sqcup g_k(X_k) = E \ \text{ and } \ h_1(Y_1) \sqcup \ldots \sqcup h_m(Y_m) = E\,.$$

(2) A *G–countably paradoxical subset of* X is a subset F admitting a countable partition

$$F = \bigsqcup_{n\in\mathbb{N}} X_n \sqcup \bigsqcup_{m\in\mathbb{N}} Y_m$$

such that for two sequences $(g_n)_{n\in\mathbb{N}}$ and $(h_m)_{m\in\mathbb{N}}$ in G,

$$\bigsqcup_{n\in\mathbb{N}} g_n(X_n) = F \ \text{ and } \ \bigsqcup_{m\in\mathbb{N}} \psi_m(Y_m) = F\,.$$

John von Neumann [?] studied properties of group actions that make paradoxical decompositions possible (like for the action of the group of isometries of $\mathbb{R}^n$ for $n \geqslant 3$) or, on the contrary forbid them (like for the action of the group of isometries of $\mathbb{R}^2$). He defined the notion of *amenable group*, based on the existence of a mean/finitely additive measure invariant under the action of the group, and equivalent to the *nonexistence* of paradoxical decompositions 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 *super-amenable groups*. In what follows we discuss all these variants of amenability and paradoxical behavior.

To clarify the setting, we recall the definition of a finitely additive (probability) measure.

DEFINITION 11.18. An *algebra of subsets* of a set X is a non-empty collection $\mathcal{A}$ of subsets of X such that:

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

DEFINITION 11.19. (1) 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 $A, B \in \mathcal{A}$.

(2) If moreover $\mu(X) = 1$ then μ is called a *finitely additive probability (f.a.p.) measure*.

(3) Let G be a group acting on X preserving $\mathcal{A}$, i.e., $gA \in \mathcal{A}$ for every $A \in \mathcal{A}$ and $g \in G$. If μ is a finitely additive measure on $\mathcal{A}$, so that $\mu(gA) = \mu(A)$ for all $g \in G$ and $A \in \mathcal{R}$, then μ is called *G–invariant*.

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

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

REMARK 11.20. 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.

We recall without proof a strong result relating the existence of a finitely additive measure to the non-existence of paradoxical decompositions. It is due to Tarski ([**?**], [**?**, pp. 599–643]), see also [**?**, Corollary 9.2].

THEOREM 11.21 (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 $\mu(E) = 1$.*

11.4. Super-amenability, weakly paradoxical actions, elementary amenability

DEFINITION 11.22. (1) A group action $G \curvearrowright X$ is *weakly paradoxical* if there exists a G-paradoxical subset in X. An action $G \curvearrowright X$ is *super-amenable* if it is not weakly paradoxical.

(2) An action $G \curvearrowright X$ is *paradoxical* if the entire set X is G-paradoxical.

(3) A group G is *(weakly) paradoxical* if the action $G \curvearrowright G$ by left multiplications is (weakly) paradoxical.

(4) Likewise, a group G is called *super-amenable* if the action $G \curvearrowright G$ by left multiplications is super-amenable.

Note that, by using the inversion map $x \mapsto x^{-1}$, one easily sees that in Definition 11.22, (3) and (4), it does not matter if one considers left or right multiplication.

PROPOSITION 11.23. (1) *A group is super-amenable if and only if every action of it is super-amenable.*

(2) *A group is weakly paradoxical if and only if it has at least one weakly paradoxical action.*

PROOF. (1) and (2) are equivalent, therefore it suffices to prove (1). The 'if' part of the statement is obvious. We prove the 'only if' part.

Consider an arbitrary action $G \curvearrowright X$ and an arbitrary non-empty subset E of X. Without loss of generality we may assume that the action is $G \curvearrowright X$ is to the left.

Let x be a point in E and let G_E be the set of $g \in G$ such that $gx \in E$. By hypothesis, G is super-amenable, therefore G_E is not paradoxical with respect to the left-action $G \curvearrowright G$. Theorem 11.21 implies that there exists a G–left-invariant finitely additive measure $\mu_G : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(G_E) = 1$.

We define a G–invariant finitely additive measure $\mu : \mathcal{P}(X) \to [0, \infty]$ by

$$\mu(A) = \mu_G(\{g \in G \mid gx \in A\}) .$$

This measure satisfies $\mu(E) = 1$, hence, E cannot be G–paradoxical. $\square$

PROPOSITION 11.24. *A weakly paradoxical group has exponential growth.*

PROOF. Let G be weakly paradoxical and let E be a G-paradoxical subset of G. Then

$$E = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

and there exist elements $g_1, \ldots, g_k, h_1, \ldots, h_m$ in G such that

$$g_1X_1 \sqcup ... \sqcup g_kX_k = E \;\text{ and }\; h_1Y_1 \sqcup ... \sqcup h_mY_m = E .$$

We define two piecewise left translations $\bar{g} : E \to E$ and $\bar{h} : E \to E$ as follows: The restriction of $\bar{g}$ to g_iX_i coincides with the left translation by g_i^{-1}, for every $i \in \{1, \ldots, k\}$; the restriction of $\bar{h}$ to h_jY_j coincides with the left translation by h_j^{-1}, for every $j \in \{1, \ldots, m\}$. Both maps are injective. Indeed if a, b are two distinct elements of E, either they are in the same subset g_iX_i in which case the injectivity follows from the injectivity of left translations, or $a \in g_iX_i$ and $b \in g_jX_j$, for some $i \neq j$. In the latter case, $\bar{g}(a) \in X_i$ and $\bar{g}(b) \in X_j$ ad since $X_i \cap X_j = \emptyset$, the two images are distinct. A similar argument shows the injectivity of $\bar{h}$.

Given an alphabet of two letters $\{x, y\}$ we denote by W_n the set of words of length n. For $w \in W_n$ we denote by $w(\bar{g}, \bar{h})$ the map $E \to E$ obtained by replacing x with $\bar{g}$, y with $\bar{h}$ and considering the composition of the finite sequence of maps thus obtained.

We prove by induction on $n \geqslant 1$ that the subsets $w(\bar{g}, \bar{h})(E)$, $w \in W_n$, are pairwise disjoint. For $n = 1$ this means that $\bar{g}(E)$ and $\bar{h}(E)$ are disjoint, which is obvious.

Assume that the statement is true for n. Let u and v be two distinct words of length $n + 1$. Assume that they both begin with the same letter, say $u = xu'$ and $v = xv'$, where u' and v' are distinct words of length n (the case when the letter is y is similar).

Then $u(\bar{g}, \bar{h})(E) = \bar{g}u'(\bar{g}, \bar{h})(E)$ and $v(\bar{g}, \bar{h})(E) = \bar{g}v'(\bar{g}, \bar{h})(E)$. The induction hypothesis implies that the sets $u'(\bar{g}, \bar{h})(E)$ and $v'(\bar{g}, \bar{h})(E)$ are disjoint, and since $\bar{g}$ is injective, the same is true for the two initial sets.

If $u = xu'$ and $v = yv'$ then

$$u(\bar{g}, \bar{h})(E) \subset \bar{g}(E) \subset X_1 \sqcup \ldots \sqcup X_k$$

while

$$v(\bar{g}, \bar{h})(E) \subset \bar{h}(E) \subset Y_1 \sqcup \ldots \sqcup Y_m .$$

Thus, $u(\bar{g}, \bar{h})(E)$ and $v(\bar{g}, \bar{h})(E)$ are disjoint in this case too, which concludes the induction step, and the proof.

It follows from the statement just proved, that for every $n \geqslant 1$, given an arbitrary $a \in E$, the set $w(\bar{g}, \bar{h})(a)$, $w \in W_n$, contains as many elements as W_n, that is 2^n. By the definition of $\bar{g}$ and $\bar{h}$, for every $w \in W_n$, $w(\bar{g}, \bar{h})(a) = g_wa$, where g_w is an element in G obtained by replacing in w every occurrence of the letter x by one of the elements $g_1, \ldots, g_k$, every occurrence of the letter y by one of the elements $h_1, \ldots, h_m$, and taking the product in G. Since g_wa, $w \in W_n$, are pairwise distinct, the elements g_w, $w \in W_n$, are pairwise distinct. With respect to a generating set S containing $g_1, \dots, g_k, h_1, \dots, h_m$ we have $|g_w|_S \leqslant n$, whence, $\mathfrak{G}_S(n) \geqslant 2^n$. □

COROLLARY 11.25. *Every group with sub-exponential growth is super-amenable.*

Corollary 11.25 is a strengthening of Proposition 11.6 in the group-theoretic setting, in view of the discussion in Section 11.3.

COROLLARY 11.26. *Virtually nilpotent groups and finite extensions of Grigorchuk groups are super-amenable.*

EXERCISE 11.27. Given a finite group G and a non-empty subset $E \subset G$, construct a G–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(E) = 1$.

It is not known if the converse of Proposition 11.24 is true or if on the contrary there exist super-amenable groups with exponential growth.

A weaker version of the converse of Proposition 11.24 is known though, and it runs as follows.

PROPOSITION 11.28. *A free two-generated sub-semigroup S of a group G is always G–paradoxical, where the action $G \curvearrowright G$ is either by left of by right multiplication.*

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 $b^{-1}S_b = S$. □

REMARK 11.29. The converse of Proposition 11.28, on the other hand is not true: a weakly paradoxical group does not necessarily contain a nonabelian free subsemigroup. There exist torsion groups that are paradoxical (see the discussion following Remark 11.81).

PROPOSITION 11.30.
(1) *A subgroup of a super-amenable group is super-amenable.*
(2) *A finite extension of a super-amenable group is super-amenable.*
(3) *A quotient of a super-amenable group is super-amenable.*
(4) *A direct limit of a directed system of super-amenable groups is super-amenable.*

REMARKS 11.31. The list of group constructions under which super-amenability is stable cannot be completed with:

- if a normal subgroup N in a group G is super-amenable and G/N is super-amenable then G is super-amenable;
- a direct product of super-amenable groups is super-amenable.

It is simply not known if the second property is true, while the first property is known to be false. Otherwise, this property and Corollary 11.25 would imply that all solvable groups are super-amenable. On the other hand, solvable groups that are not virtually nilpotent contain a nonabelian free subsemigroup [**?**].

PROOF. (1) Let $H \leqslant G$ with G super-amenable and let E be a non-empty subset of H. By Theorem 11.21, there exists a G–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(E) = 1$. Theorem 11.21 applied to μ restricted to $\mathcal{P}(H)$ imply that E cannot be H-paradoxical either.

(2) Let $H \leqslant G$ with H super-amenable and $G = \bigsqcup_{i=1}^m Hx_i$. Let E be a non-empty subset of G.

The group H acts on G, whence by Proposition 11.23, (1), and Theorem 11.21, there exists an H–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu\left(\cup_{i=1}^m x_i E\right) = 1$.

Define the measure $\nu : \mathcal{P}(G) \to [0, \infty]$ by

$$\nu(A) = \frac{\sum_{i=1}^m \mu(x_i A)}{\sum_{i=1}^m \mu(x_i E)}\,.$$

It is clearly finitely additive and satisfies $\nu(E) = 1$.

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 $\varphi : \{1, \dots, m\} \to \{1, \dots, m\}$ dependent on g such that $Hx_i g = Hx_{\varphi(i)}$.

We may then rewrite the denominator in the expression of $\nu(gA)$ as

$$\sum_{i=1}^m \mu(x_i g A) = \sum_{i=1}^m \mu(h_i x_{\varphi(i)} A) = \sum_{i=1}^m \mu(x_{\varphi(i)} A) = \sum_{j=1}^m \mu(x_j A)\,.$$

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. Theorem 11.21 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) 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 Section 1.1.

Consider a non-empty subset E of G. Without loss of generality we may assume that all $h_i(H_i)$ intersect E: there exists i_0 such that for every $i \geqslant i_0$, $h_i(H_i)$ intersects E, and we can 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 (see Remark **??**, **??**).

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. measures μ on $\mathcal{P}(G)$ such that $\mu(E) = 1$. Since H_i is super-amenable, Proposition 11.23, (1), and Theorem 11.21 imply that the set $\mathcal{M}_i$ is non-empty.

Let us prove that $\mathcal{M}_i$ is closed in $\prod_{\mathcal{P}(G)}[0, \infty]$. Let $f : \mathcal{P}(G) \to [0, \infty]$ be an element of $\prod_{\mathcal{P}(G)}[0, \infty]$ in the closure of $\mathcal{M}_i$. Then, for every finite collection $A_1, \dots, A_n$ of subsets of X 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 $\epsilon > 0$, $|f(E) - 1| \leqslant \epsilon$,

$$|f(A \sqcup B) - f(A) - f(B)| \leqslant 3\epsilon$$

and

$$|f(gA) - f(A)| \leqslant 2\epsilon,$$

$\forall A, B \in \mathcal{P}(X)$ 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. measure such that $\mu(E) = 1$, and μ is also G–left-invariant because

$$G = \bigcup_{i \in I} h_i(H_i).$$

□

In view of Corollary 11.26, Proposition 11.30 and Remarks 11.31 it is natural to consider the class of groups that contains all finite and abelian groups, that is stable with respect to the operations described in Proposition 11.30, plus the one of extension:

DEFINITION 11.32. 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 sums, finite-index extensions, direct limits, subgroups, quotients and extensions

$$1 \to G_1 \to G_2 \to G_3 \to 1,$$

where both G_1, G_3 are elementary amenable.

Neither of the two classes of super-amenable and of elementary amenable groups contains the other:

- solvable groups are all elementary amenable, while they are super-amenable only if they are virtually nilpotent;
- there exist Grigorchuk groups of intermediate growth that are not elementary amenable

The following result due to C. Chou (and proved previously for the smaller class of solvable groups by Rosenblatt [**?**]) describes, within the setting of finitely generated groups, the intersection between the two classes, and brings information on the set of elementary amenable groups that are not super-amenable.

THEOREM 11.33 ([**?**]). *A finitely generated elementary amenable group either is virtually nilpotent or it contains a free non-abelian subsemigroup.*

11.5. Amenability and paradoxical actions

In this section we define amenable actions and amenable groups, and prove that paradoxical behavior is equivalent to non-amenability.

DEFINITION 11.34.

(1) A group action $G \curvearrowright X$ is *amenable* if there exists a G–invariant f.a.p. measure μ on $\mathcal{P}(X)$, the set of all subsets of X.
(2) A group is *amenable* if the action of G on itself *via* left multiplication is amenable.

LEMMA 11.35. *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 assumption that $\mu(X) = 1$. $\square$

REMARK 11.36. We will prove in Corollary 11.63 that a finitely-generated group is amenable if and only if it is non-paradoxical.

EXAMPLE 11.37. 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 G define $\mu : \mathcal{P}(X) \to [0, 1]$ by $\mu(A) = \frac{|A|}{|X|}$, where $|\cdot|$ denotes cardinality of a subset.

EXAMPLE 11.38. The free group of rank two F_2 is non-amenable since F_2 is paradoxical, as explained in Chapter 10, Section 10.4.

Yet another equivalent definition for amenability can be formulated using the concept of an *invariant mean*, which is responsible for the terminology 'amenable':

DEFINITION 11.39.

(1) A *mean* on a set X is a linear functional $m : \ell^\infty(X) \to \mathbb{C}$ defined on the set $\ell^\infty(X)$ of bounded functions on X, with the following properties:
(M1) if f takes values in $[0, \infty)$ then $m(f) \geqslant 0$;
(M2) $m(\mathbf{1}_X) = 1$.

Assume, moreover, that X is endowed with the action of a group G, $G \times X \to X, (g, x) \mapsto g \cdot x$. This induces an action of G on the set $\ell^\infty(X)$ of bounded complex-valued functions on X defined by $g \cdot f(x) = f(g^{-1} \cdot x)$.

A mean is called *left-invariant* if $m(g \cdot f) = m(f)$ for every $f \in \ell^\infty(X)$ and $g \in G$.

A special case of the above is when $G = X$ and G acts on itself by left translations.

PROPOSITION 11.40. *A group action $G \curvearrowright X$ is amenable (in the sense of Definition 11.34) if and only if it admits a left-invariant mean.*

PROOF. Given a f.a.p. measure μ on X one can apply the standard construction of integrals (see [?, Chapter 1] or [?, Chapter 11]) and define, for any function $f : X \to \mathbb{C}$, $m(f) = \int f\,d\mu$. Since $\mu(X) = 1$, for every bounded function f, $m(f) \in \mathbb{C}$. Thus, we obtain a linear functional $m : \ell^\infty(X) \to \mathbb{C}$. If the measure μ is G-invariant then m is also G-invariant.

Conversely, given a G-invariant mean m on X, one defines an invariant f.a.p. measure μ on X by $\mu(A) = m(\mathbf{1}_A)$.

EXERCISE 11.41. Prove that μ thus defined is a f.a.p. measure and that G-invariance of m implies G-invariance of μ.

$\square$

REMARK 11.42. Suppose that in Proposition 11.40, $X = G$ and $G \curvearrowright X$ is the action by left multiplication. Then:

(a) In the above proposition, left-invariance can be replaced by right-invariance.

(b) Moreover, both can be replaced by bi-invariance.

PROOF. (a) It suffices to define $\mu_r(A) = \mu(A^{-1})$ and $m_r(f) = m(f_1)$, where $f_1(x) = f(x^{-1})$.

(b) Let μ be a left-invariant f.a.p. measure and μ_r the right-invariant measure in (a). Then for every $A \subseteq X$ define

$$\nu(A) = \int \mu(Ag^{-1})d\mu_r(g)\,.$$

$\square$

QUESTION 11.43. Suppose that G is a group which admits a mean $m : \ell^\infty(G) \to \mathbb{R}$ that is quasi-invariant, i.e., there exists a constant κ such that

$$|m(f \circ g) - m(f)| \leqslant \kappa$$

for all functions $f \in \ell^\infty(G)$ and all group elements g. Is it true that G is amenable?

LEMMA 11.44. *Every action $G \curvearrowright X$ of an amenable group G is also amenable.*

PROOF. Choose a point $x \in X$ and define $\nu : \mathcal{P}(X) \to [0, 1]$ by

$$\nu(A) = \mu(\{g \in G\ ;\ gx \in A\}).$$

We leave it to the reader to verify that ν is again a G-invariant f.a.p. measure. $\square$

COROLLARY 11.45. *If G is a group which admits a paradoxical action, then G is non-amenable. In other words, if an amenable group G acts on a space X, then X cannot be G–paradoxical.*

This corollary and the fact that the sphere $\mathbb{S}^2$ is $O(3)$–paradoxical imply that the group $O(3)$ is not amenable (as an abstract group). More generally, in view of Tits' Alternative, if G is a connected Lie group then either G is solvable or non-amenable (since every non-solvable connected Lie group contains a free nonabelian subgroup).

The converse to Lemma 11.44 is false: The action of any group on a one-point set is clearly amenable, see, however, Proposition 11.58. On the other hand, Glasner and Monod [?] proved that every countable group admits an amenable faithful action on a set X.

A natural question to ask is whether on an amenable group there exists only one invariant finitely additive probability measure. It turns out that this is far from being true:

THEOREM 11.46 (J. Rosenblatt [?]). *Let G be a non-discrete σ-compact locally compact metric group. If G is amenable as a discrete group, then there are $2^{\aleph_0}$ mutually singular G-invariant means on $L^\infty(G)$.*

REMARK 11.47. Theorem 11.21 and the Banach-Tarski paradox prove that there exists no $\mathrm{Isom}(\mathbb{R}^3)$–left-invariant finitely additive measure $\mu : \mathcal{P}(\mathbb{R}^3) \to [0, \infty]$ such that the unit ball has positive measure.

PROPOSITION 11.48. (1) *A 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 11.44, 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 a G–left-invariant probability measure.

(3) The proof is along the same lines as that of Proposition 11.30, (4). The only difference is that the compact $\prod_{\mathcal{P}(G)}[0, \infty]$ is replaced in this argument by

$$\{f : \mathcal{P}(G) \to [0, 1]\} = \prod_{\mathcal{P}(G)} [0, 1] .$$

□

COROLLARY 11.49. *Let G_1 and G_2 be two groups that are co-embeddable in the sense of Definition 3.40. Then G_1 is amenable if and only if G_2 is amenable.*

COROLLARY 11.50. *Any group containing a free nonabelian subgroup is non-amenable.*

PROOF. Note that every non-abelian free group contains a subgroup isomorphic to F_2, free group of rank 2. Now, the statement follows from Proposition 11.48, Part (1), and Example 11.38. □

COROLLARY 11.51. *A semidirect product $N \rtimes H$ is amenable if and only if both N and H are amenable.*

PROOF. The statement follows immediately from Part (2) of the above proposition. $\square$

COROLLARY 11.52. *1. If G_i, $i = 1, \ldots, n$, are amenable groups, then the Cartesian product $G = G_1 \times \ldots \times G_n$ is also amenable.*

2. 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 11.51.
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. $\square$

COROLLARY 11.53. *A group G is amenable if and only if all finitely generated subgroups of G are amenable.*

PROOF. The direct part follows from (1). The converse part follows from (3), where, given the group G, we let I be the set of all finite subsets in G, and for any $i \in I$, 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 Proposition 11.48. $\square$

COROLLARY 11.54. *Every abelian group G is amenable.*

PROOF. Since every abelian group is a direct limit of finitely-generated abelian subgroups, by Part (3) of the above proposition, it suffices to prove the corollary for finitely-generated abelian groups. Amenability of such groups will be proven in Proposition 11.69 as an application of the Følner criterion for amenability. $\square$

REMARK 11.55. Even for the infinite cyclic group $\mathbb{Z}$, amenability is nontrivial, it depends on a form of Axiom of Choice, e.g., ultrafilter lemma: One can show that Zermelo–Fraenkel axioms are insufficient for proving amenability of $\mathbb{Z}$.

COROLLARY 11.56. *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 11.54.

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 Proposition 11.48, (2), imply that G is amenable. $\square$

In view of the above results, every elementary amenable group is amenable. On the other hand, all finitely generated groups of intermediate growth are amenable but not elementary amenable.

EXAMPLE 11.57 (Infinite direct products of amenable groups need not be amenable). Let $F = F_2$ be free group of rank 2. Recall, Corollary 3.86, that F is residually finite, hence, for every $g \in F \setminus \{1\}$ there exists a homomorphism $\varphi_g : F \to \Phi_g$ so 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.

The following is a generalization of Proposition 11.48, (2); this proposition also completes the result in Lemma 11.44.

PROPOSITION 11.58. *Let G be a group acting on a set X. The group G is amenable if and only if $G \curvearrowright X$ is amenable and for every $p \in X$ the stabilizer* $\mathrm{Stab}(p)$ *of the point p is amenable.*

PROOF. The direct implication follows from Lemma 11.44 and from Proposition 11.48, (1).

Assume now that for every $p \in X$ its G-stabilizer S_p is amenable and that $m_X : \ell^\infty(X) \to \mathbb{C}$ is a G-invariant mean. By proposition 11.40, for every $p \in X$ there exists a left-invariant mean $m_p : \ell^\infty(S_p) \to \mathbb{C}$.

We define a left-invariant mean on $\ell^\infty(G)$ using a construction in the spirit of Fubini's Theorem.

Let $F \in \ell^\infty(G)$. We split X into G–orbits $X = \bigsqcup_{p\in\Re} Gp$.

For every $p \in \Re$ we define a function F_p on the orbit Gp by $F_p(gp) = m_p\left(F|_{gS_p}\right)$. Then we define a function F_X on X which coincides with F_p on each orbit Gp.

The fact that F is bounded implies that F_X is bounded. 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 two properties (M1) and (M2) in Definition 11.39 are easily checked for the mean m. We now check that m is left-invariant. Let h be an arbitrary element of G, and let $h \cdot F$ be defined by $h \cdot F(x) = F(h^{-1} \cdot x)$, for every $x \in G$.

Then

$$(h\cdot F)_p(gp) = m_p\left((h\cdot F)|_{gS_p}\right) = m_p\left(F|_{h^{-1}gS_p}\right) = F_p(h^{-1}gp).$$

We deduce from this that $(h \cdot F)_X = F_X \circ h^{-1} = h \cdot F_X$, whence

$$m(h\cdot F) = m_X\left((h\cdot F)_X\right) = m_X\left(h\cdot F_X\right) = m_X\left(F_X\right) = m(F).$$

□

COROLLARY 11.59. *Amenability is preserved by virtual isomorphisms of groups.*

PROOF. The only nontrivial part of this statement is: If H is an amenable subgroup of finite index in a group G, then G is also amenable. Consider the action of G on $X = G/H$ by left multiplications. Stabilizers of points under this action are conjugates of the group H in G, hence, they are amenable. The set X is finite and, hence, the action $G \curvearrowright X$ is amenable. Thus, G is amenable by Proposition 11.58. □

For topological groups and topological group actions one can refine the notion of amenability as follows:

DEFINITION 11.60 (Amenability for topological group actions). 1. Let $G\times X \to X$ be a topological action of a topological group G. Then this action is *topologically amenable* if there exists a continuous G-invariant linear functional m defined on the space of all Borel measurable bounded functions $X \to \mathbb{C}$, such that:

- $m(f) \geqslant 0$ when $f \geqslant 0$;

- $m(\mathbf{1}_X) = 1$;

Such a linear functional is called an *invariant mean.*

2. A topological group G is said to be *amenable* if the action of G on itself *via* left multiplication is amenable. The corresponding linear functional m is called a *left-invariant mean.*

REMARK 11.61. With this notion, for instance, every compact group is topologically amenable (we can take m to be the integral with respect to a left Haar measure). In particular, the group $SO(3)$ is topologically amenable. On the other hand, as we saw, $SO(3)$ is not amenable as an abstract group. More generally, if $\mathcal{H}$ is a separable Hilbert space and $G = U(\mathcal{H})$ is the group of unitary operators on $\mathcal{H}$ endowed with the weak operator topology, then G is topologically amenable, see [**?**]. We refer to [**?**] for further details on topological amenability.

11.6. Equivalent definitions of amenability for finitely generated groups

In view of Corollary 11.53, amenability in the case of finitely generated groups is particularly significant. In this case, one can relate the group amenability to the metric amenability for Cayley graphs.

THEOREM 11.62. *Let G be a finitely-generated group. The following are equivalent:*

(1) *G is amenable in the sense of Definition 11.34;*

(2) *one (every) Cayley graph of G is amenable in the sense of Definition 11.1.*

PROOF. According to Theorem 11.10, if one Cayley graph of 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 $\mathcal{G} = \text{Cayley}(G, S)$, and word metric, and we assume that the statement (2) refers to $\mathcal{G}$.

(2)$\Rightarrow$ (1). We first illustrate the proof in the case $G = \mathbb{Z}$ 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. Puck a non-principal ultrafilter ω on $\mathbb{N}$. For a subset $A \subset \mathbb{Z}$ we define a f.a.p. measure μ by

$$\mu(A) := \omega\text{-lim} \frac{|A \cap \Omega_n|}{2n+1}.$$

Let us show that μ is invariant under the unit 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.

We now consider the general case. Since $\mathcal{G}$ is amenable, there exists a Følner sequence of subsets $(\Omega_n) \subset G$ (since G is the vertex set of $\mathcal{G}$). We use the sets Ω_i to construct a G-invariant f.a.p. measure on $\mathcal{P}(G)$. Following Remark 11.42, we can and will use the action to the right of G on itself in this discussion.

Let ω be a non-principal ultrafilter on $\mathbb{N}$. For every $A \subset G$ define

$$\mu(A) = \omega\text{-lim} \frac{|A \cap \Omega_n|}{|\Omega_n|}\,.$$

We leave it to the reader to check that μ is a f.a.p. measure on G. Now consider an arbitrary generator $g \in S$. We have

$$|\mu(Ag) - \mu(A)| = \omega\text{-lim} \frac{||Ag \cap \Omega_n| - |A \cap \Omega_n||}{|\Omega_n|} = \omega\text{-lim} \frac{\big||A \cap \Omega_n g^{-1}| - |A \cap \Omega_n|\big|}{|\Omega_n|}\,.$$

Now $A \cap \Omega_n g^{-1} = \left(A \cap \Omega_n g^{-1} \cap \Omega_n\right) \sqcup \left(A \cap \Omega_n g^{-1} \setminus \Omega_n\right)$. Likewise,

$$A \cap \Omega_n = \left(A \cap \Omega_n \cap \Omega_n g^{-1}\right) \sqcup \left(A \cap \Omega_n \setminus \Omega_n g^{-1}\right).$$

Therefore, the ultralimit above can be rewritten as

$$\omega\text{-lim} \frac{\big||A \cap (\Omega_n g^{-1} \setminus \Omega_n)| - |A \cap (\Omega_n \setminus \Omega_n g^{-1})|\big|}{|\Omega_n|} \leqslant$$
$$\omega\text{-lim} \frac{|A \cap (\Omega_n g^{-1} \setminus \Omega_n)| + |A \cap (\Omega_n \setminus \Omega_n g^{-1})|}{|\Omega_n|}$$
$$= \omega\text{-lim} \frac{|A \cap (\Omega_n g^{-1} \setminus \Omega_n)| + |Ag \cap (\Omega_n g \setminus \Omega_n)|}{|\Omega_n|} \leqslant \omega\text{-lim} \frac{2|E(\Omega_n, \Omega_n^c)|}{|\Omega_n|} = 0\,.$$

The last equality follows from amenability of the graph $\mathcal{G}$. Therefore, $\mu(Ag) = \mu(A)$ for every $g \in S$. Since S is a generating set of G, we obtain the equality $\mu(Ag) = \mu(A)$ for every $g \in G$.

(1)$\Rightarrow$ (2). We prove this implication by proving the contrapositive, that is $\neg(2) \Rightarrow \neg(1)$. We shall, in fact, prove that $\neg(2)$ implies that G is paradoxical.

Assume that $\mathcal{G}$ is non-amenable. According to Theorem 11.3, this implies that there exists a map $f : G \to G$ which is at finite distance from the identity map, such that $|f^{-1}(y)| = 2$ for every $y \in G$. Lemma 5.27 implies that there exists a finite set $\{h_1, ..., h_n\}$ and a decomposition $G = T_1 \sqcup ... \sqcup T_n$ such that f restricted to T_i coincides with the multiplication on the right R_{h_i}.

For every $y \in G$ we have that $f^{-1}(y)$ consists of two elements, which we label as $\{y_1, y_2\}$. This gives a decomposition of G 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 $A_1 h_1 \sqcup ... \sqcup A_n h_n = G$ and $B_1 h_1 \sqcup ... \sqcup B_n h_n = G$. We have, thus, proved that G is paradoxical. $\square$

The equivalence in Theorem 11.62 allows to give 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 reduced word. This map satisfies Gromov's condition in Theorem 11.3. Hence, the Cayley graph of F_2 is non-amenable; thus, F_2 is non-amenable as well.

Another consequence of the proof of Theorem 11.62 is the following weaker version of the Tarski's Alternative Theorem 11.21:

COROLLARY 11.63. *A finitely generated group is either paradoxical or amenable.*

PROOF. Indeed, in the proof of Theorem 11.62, we proved that Cayley graph $\mathcal{G}$ of G is amenable if and only if the group G is, and that if $\mathcal{G}$ is non-amenable then G is paradoxical. Thus, we have that group amenability is equivalent not only to the Cayley graph amenability but also to non-paradoxical behavior. $\square$

Note that the above proof 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. In particular, for any group G containing an element of infinite order, ZF are not enough to conclude that G admits an invariant mean.

QUESTION 11.64. Is there a finitely-generated infinite group which admits an invariant mean under the ZF axioms in set theory?

COROLLARY 11.65. *Every super-amenable group is amenable.*

LEMMA 11.66. *Let (Ω_n) be a sequence of subsets of a finitely-generated group G. The following are equivalent:*

(1) *(Ω_n) is a Følner sequence for one of the Cayley graphs of G.*

(2) *For every $g \in G$*

$$\lim_{n\to\infty} \frac{|\Omega_n g \,\Delta\, \Omega_n|}{|\Omega_n|} = 0\,. \tag{11.7}$$

(3) *For every element g of a generating set S of G,*

$$\lim_{n\to\infty} \frac{|\Omega_n g \,\Delta\, \Omega_n|}{|\Omega_n|} = 0\,. \tag{11.8}$$

PROOF. Let S be a finite generating set that determines a Cayley graph $\mathcal{G}$ of G, we will assume that $1 \notin S$. Let $\Omega \subset G$, i.e., Ω is a subset of the vertex set of $\mathcal{G}$. Then the vertex boundary of Ω in $\mathcal{G}$ is

$$\partial_V \Omega = \bigcup_{s\in S} \Omega \setminus \Omega s^{-1}\,.$$

Thus, for a sequence (Ω_n) the equality

$$\lim_{n\to\infty} \frac{|E(\Omega_n, \Omega_n^c)|}{|\Omega_n|} = 0\,.$$

is equivalent to the set of equalities

$$\lim_{n\to\infty} \frac{|\Omega_n \setminus \Omega_n s^{-1}|}{|\Omega_n|} = 0 \text{ for every } s \in S\,,$$

which in its turn is equivalent to (11.8) for every $g \in S^{-1}$. Thus, (1) is equivalent to (3).

It remains to show that (1) implies that (11.7) holds for an arbitrary $g \in G$. In view of Exercise 11.5, if Ω_n is a Følner sequence for one finite generating set of G, the sequence Ω_n is also Følner for every finite generating set of G. By taking a finite generating set of G which contains given $g \in G$, we obtain the desired conclusion. □

DEFINITION 11.67. If G is a group, then a sequence of subsets $\Omega_n \subset G$ satisfying property (11.7) in Lemma 11.66, is called a *Følner sequence for the group G*. Note that the form of this definition makes sense even if G is not finitely-generated.

EXERCISE 11.68. Prove that the subsets $\Omega_n = \mathbb{Z}^k \cap [-n, n]^k$ form a Følner sequence for $\mathbb{Z}^k$.

PROPOSITION 11.69. (1) *If (Ω_n) is a Følner sequence in a countable group G and ω is a non-principal ultrafilter on $\mathbb{N}$ then a left-invariant mean $m : \ell^\infty(G) \to \mathbb{C}$ may be defined by*

$$m(f) = \omega\text{-}lim \frac{1}{|\Omega_n|} \sum_{x \in \Omega_n} f(x)$$

.

(2) *For any $k \in \mathbb{N}$ the group $\mathbb{Z}^k$ has an invariant mean $m : \ell^\infty\left(\mathbb{Z}^k\right) \to \mathbb{C}$ is defined by*

$$m(f) = \omega\text{-}lim \frac{1}{(2n+1)^k} \sum_{x \in \mathbb{Z}^k \cap [-n,n]^k} f(x) .$$

PROOF. (1) It suffices to note that $\mu(A) = m(\mathbf{1}_A)$ is the left invariant f.a.p. measure defined in the proof of (2)$\Rightarrow$ (1) above.

(2) is a consequence of (1) and Exercise 11.68. □

We are now able to relate amenable groups to the Banach–Tarski paradox.

PROPOSITION 11.70. (1) *The group of isometries* $\mathrm{Isom}(\mathbb{R}^n)$ *with $n = 1, 2$ is amenable.*

(2) *The group of isometries* $\mathrm{Isom}(\mathbb{R}^n)$ *with $n \geqslant 3$ is non-amenable.*

PROOF. (1) The group $\mathrm{Isom}(\mathbb{R}^n)$ is the semidirect product of $O(n)$ and $\mathbb{R}^n$. The group $\mathbb{R}^n$ is abelian and, hence, amenable, by Corollary 11.54. Furthermore, $O(1) \cong \mathbb{Z}_2$ is finite and, hence, amenable. The group $O(2)$ contains the abelian subgroup $SO(2)$ of index 2. Hence, $O(2)$ is also amenable. Thus, $\mathrm{Isom}(\mathbb{R}^n)$ $(n \leqslant 2)$ is amenable as a semidirect product of two amenable groups, see Corollary 11.51.

(2) This follows from Corollary 11.45 and Banach-Tarski paradox. □

In many textbooks one finds the following property (first introduced by Følner in [**?**]) as an alternative characterization of amenability. Though it is close to the one provided by Lemma 11.66, we briefly discuss it here, for the sake of completeness.

DEFINITION 11.71. A group G is said to have *the Følner Property* if for every finite subset K of G and every $\epsilon > 0$ there exists a finite non-empty subset F such that for all $g \in K$

$$\frac{|Fg \,\Delta\, F|}{|F|} \leqslant \epsilon . \tag{11.9}$$

REMARK 11.72. The relation (11.9) can be rewritten as

$$\frac{|FK \,\Delta\, F|}{|F|} \leqslant \epsilon , \tag{11.10}$$

where $FK = \{fk \,:\, f \in F , k \in K\}$.

EXERCISE 11.73. Verify that a group has Følner property if and only if it contains a Følner sequence in the sense of Definition 11.67.

LEMMA 11.74. (1) *In both Definition 11.71 and in the characterization of the Følner Property provided by Lemma 11.66, one can take the action of G on the left, i.e. $\frac{|gF \,\Delta\, F|}{|F|} \leqslant \epsilon$ in (11.9) etc.*

(2) *When G is finitely generated, it suffices to check Definition 11.71 for a finite generating set.*

PROOF. (1) One formulation is equivalent to the other *via* the anti-automorphism $G \to G$ given by the inversion $g \mapsto g^{-1}$.

In Definition 11.71, for every finite subset K and every $\epsilon > 0$ it suffices to apply the property with multiplication on the left to the set $K^{-1} = \{g^{-1} ; \, g \in K\}$, obtain the set F, then take $F' = F^{-1}$. This set will verify $\frac{|F'K \triangle F'|}{|F'|} \leqslant \epsilon$. The proof for Lemma 11.66 is similar.

(2) Let S be an arbitrary finite generating set of G. The general statement implies the one for $K = S$. Conversely, assume that the condition holds for $K = S$. In other words, there exists a sequence F_n of finite subsets of G, so that for every $s \in S$,

$$\lim_n \frac{|F_n s \triangle F_n|}{|F|} = 0.$$

In view of Lemma 11.66, for every $g \in G$

$$\lim_n \frac{|F_n g \triangle F_n|}{|F|} = 0.$$

Thus, there exists a sequence of finite subsets F_n so that for every $g \in G$ there exists $N = N_g$ so that

$$\forall n \geq N, \quad \frac{|F_n g \triangle F_n|}{|F|} < \epsilon.$$

Taking $N = \max\{N_g : g \in K\}$, we obtain the required statement with $F = F_N$. $\square$

COROLLARY 11.75. *A finitely-generated group is amenable if and only if it has Følner property.*

We already know that subgroups of amenable groups are again amenable, below we show how to construct Følner sequences for subgroups directly.

PROPOSITION 11.76. *Let H be a subgroup of an amenable group G, 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 intersection $g_n^{-1}\Omega_n \cap H = F_n$ 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

$$\alpha_n = \frac{|\Omega_n K \triangle \Omega_n|}{|\Omega_n|} \tag{11.11}$$

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

$$\Omega_n K \triangle \Omega_n = \left(\Omega_n^{(1)} K \triangle \Omega_n^{(1)}\right) \sqcup \cdots \sqcup \left(\Omega_n^{(k_m)} K \triangle \Omega_n^{(k_n)}\right).$$

The inequality

$$\frac{|\Omega_n K \triangle \Omega_n|}{|\Omega_n|} \leqslant \alpha_n$$

implies that there exists $i \in \{1, 2, \dots, k_n\}$ such that

$$\frac{|\Omega_n^{(i)} K \bigtriangleup \Omega_n^{(i)}|}{|\Omega_n^{(i)}|} \leqslant \alpha_n.$$

In particular, $g_i^{-1}\Omega_n^{(i)} = F_n$, with $F_n \subseteq H$, and we obtain that

$$\frac{|F_n K \bigtriangleup F_n|}{|F_n|} \leqslant \alpha_n.$$

□

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, under the name of *measurable groups* [**?**]. It is this observation that has led to the following question:

QUESTION 11.77 (the von Neumann-Day problem). Does every non-amenable group contain a free non-abelian subgroup?

The question is implicit in [**?**], and it was formulated explicitly by Day [**?**, §4].

When restricted to the class of subgroup of Lie groups with finitely many components (in particular, subgroups of $GL(n, \mathbb{R})$), Question 11.77 has an affirmative answer, since, in view of the Tits' alternative (see Chapter 9, Theorem 9.1), every such group without either contains a free non-abelian subgroup or is virtually solvable. Since all virtually solvable groups are amenable, the claim follows.

Note that more can be said about finitely generated amenable subgroups Γ of a Lie group L with finitely many connected components than just the fact that Γ is virtually solvable. To begin with, according to Theorem 9.78, Γ contains a solvable subgroup Σ of derived length $\leqslant \delta(L)$ so that $|\Gamma : \Sigma| \leqslant \nu(L)$.

THEOREM 11.78 (Mostow–Tits). *A discrete amenable subgroup Γ of a Lie group L with finitely many components, contains a polycyclic group of index at most $\eta(L)$.*

PROOF. We will prove this theorem for subgroups of $GL(n, \mathbb{C})$ as the general case is obtained *via* the adjoint representation of L. Let G denote the Zariski closure of Γ in $GL(n, \mathbb{C})$. Then, by Part 1 of Theorem 9.78, G contains a connected solvable subgroup S of derived length at most $\delta = \delta(n)$ and $|G : S| \leqslant \nu = \nu(n)$. Note that, up to conjugation, S is a subgroup of the group B of upper-triangular matrices in $GL(n, \mathbb{C})$, see Proposition 9.36. The intersection $\Lambda := \Gamma \cap B$ is a discrete subgroup of a connected solvable Lie group. Mostow proved in [**?**] that such a group is necessarily polycyclic. Furthermore, he established an upper bound on ranks of quotients $\Lambda^{(k)}/\Lambda^{(k+1)}$. □

When the subgroup $\Gamma < GL(n, \mathbb{C})$ is not discrete, not much is known. We provide below a few examples to illustrate that when one removes the hypothesis of discreteness, the variety of subgroups that may occur is much larger. Since this already occurs in $SL(2, \mathbb{R})$, it is natural to ask the following.

QUESTION 11.79. 1. What are the possible solvable subgroups of $SL(2,\mathbb{R})$? Equivalently, what are the possible subgroups of the group of affine transformations of the real line?

2. What are the possible solvable subgroups of $SL(2,\mathbb{C})$?

EXAMPLES 11.80. 1. We first note that for all integers $m, n \geqslant 1$, the wreath product $\mathbb{Z}^m \wr \mathbb{Z}^n$ is a subgroup of $SL(2,\mathbb{R})$. Indeed, consider $\mathcal{O}_K$, the ring of integers of a totally-real algebraic extension K of $\mathbb{Q}$ of degree m. This ring is a free $\mathbb{Z}$-module with a basis $\omega_1, ..., \omega_m$. Let $t_1, ..., t_n$ be transcendental numbers that are independent over $\mathbb{Q}$, i.e. for every $i \in \{1, ..., n\}$, t_i is transcendental over $\mathbb{Q}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)$.

Then the subgroup G of $SL(2,\mathbb{R})$ generated by the following matrices

$$s_1 = \begin{pmatrix} t_1 & 0 \\ 0 & 1 \end{pmatrix}, \dots, s_n = \begin{pmatrix} t_n & 0 \\ 0 & 1 \end{pmatrix},$$

$$u_1 = \begin{pmatrix} 1 & \omega_1 \\ 0 & 1 \end{pmatrix}, \dots, u_m = \begin{pmatrix} 1 & \omega_m \\ 0 & 1 \end{pmatrix}$$

is isomorphic to $\mathbb{Z}^m \wr \mathbb{Z}^n$.

Indeed, G is a semidirect product of its unipotent subgroup consisting of matrices

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \text{ with } x \in O_K(t_1, ..., t_n),$$

isomorphic to the direct sum $\bigoplus_{z \in \mathbb{Z}^n} O_K$, and of its abelian subgroup consisting of matrices

$$\begin{pmatrix} t_1^{k_1} \cdots t_n^{k_n} & 0 \\ 0 & 1 \end{pmatrix} \text{ with } (k_1, ..., k_n) \in \mathbb{Z}^n.$$

2. Every free metabelian group (see Definition **??**) is a subgroup of $SL(2,\mathbb{R})$. This follows from the fact that a free metabelian group with m generators appears as a subgroup of $\mathbb{Z}^m \wr \mathbb{Z}^m$, using the Magnus embedding (Theorem **??**).

3. All the examples above can be covered by the following general statements. Given an arbitrary free solvable group S with derived length $k \geqslant 1$, we have:

- S is a subgroup of $SL\left(2^{k-1}, \mathbb{R}\right)$;
- for every $m \in \mathbb{N}$ the wreath product $\mathbb{Z}^m \wr S$ is a subgroup of $SL\left(2^k, \mathbb{R}\right)$.

Indeed, one can construct by induction on k the necessary injective homomorphisms. The initial step for both statements above is represented by the examples 1 and 2. We assume that the second statement is true for k and we deduce that the first statement is true for $k+1$. This implication and the Magnus embedding described in Theorem **??** suffice to finish the inductive argument.

Consider the free solvable group $S_{n,k}$ of derived length k with n generators $s_1, ..., s_n$. According to the hypothesis, $S_{n,k}$ embeds as a subgroup of $SL\left(2^{k-1}, \mathbb{R}\right)$; thus, we will regard $s_1, ..., s_n$ as $2^{k-1} \times 2^{k-1}$ real matrices. Let $\mathcal{O}_K$ be the ring of integers of a totally-real algebraic extension of degree m, and let $\{\omega_1, ..., \omega_m\}$ be a basis of $\mathcal{O}_K$ as a free $\mathbb{Z}$-module.

We consider the subgroup G of $SL\left(2^k, \mathbb{R}\right)$ generated by the following matrices (described by square $2^{k-1} \times 2^{k-1}$ blocks; in particular the notations I and 0 below signify the identity respectively the zero square $2^{k-1} \times 2^{k-1}$ matrices):

$$\sigma_1 = \begin{pmatrix} s_1 & 0 \\ 0 & I \end{pmatrix}, \dots, \sigma_n = \begin{pmatrix} s_n & 0 \\ 0 & I \end{pmatrix},$$

$$u_1 = \begin{pmatrix} I & \omega_1 I \\ 0 & I \end{pmatrix}, \dots, u_m = \begin{pmatrix} I & \omega_m I \\ 0 & I \end{pmatrix}.$$

The group G is isomorphic to $\mathbb{Z}^m \wr S_{n,k}$: It is a semidirect product of the unipotent subgroup consisting of matrices

$$\begin{pmatrix} I & x \\ 0 & I \end{pmatrix}, \text{ with } x \text{ in the group ring } O_K S_{n,k} \simeq \bigoplus_{S_{n,k}} \mathbb{Z}^m,$$

and the subgroup isomorphic to $S_{n,k}$ consisting of matrices

$$\begin{pmatrix} g & 0 \\ 0 & I \end{pmatrix} \text{ with } g \in S_{n,k}.$$

Remark 11.81. Other classes of groups satisfying the Tits' alternative are:

(1) finitely generated subgroups of $GL(n, \mathbb{K})$ for some integer $n \geqslant 1$ and some field $\mathbb{K}$ of finite characteristic [**?**];

(2) subgroups of Gromov hyperbolic groups ([**?**, §8.2.F], [**?**, Chapter 8]);

(3) subgroups of the mapping class group, see [**?**];

(4) subgroups of $Out(F_n)$, see [**?, ?, ?**];

(5) fundamental groups of compact manifolds of nonpositive curvature, see [**Bal95**].

Hence, for all such groups Question 11.77 has positive answer.

The first examples of finitely-generated non-amenable groups with no (non-abelian) free subgroups were given in [**?**]. In [**?**] 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 finitely presented examples of non-amenable groups with no (non-abelian) free subgroups were given in [**?**].

Still, metric versions of the von Neumann-Day Question 11.77 have positive answers. One of these versions is Whyte's Theorem 11.7 (a graph of bounded geometry is non-amenable if and only if it admits a free action of a free non-Abelian group by bi-Lipschitz maps at finite distance from the identity).

Another metric version of the von Neumann-Day Question was established by Benjamini and Schramm in [**?**]. They proved that:

- *An infinite locally finite simplicial graph $\mathcal{G}$ with positive Cheeger constant contains a tree with positive Cheeger constant.*

  Note that in the result above uniform bound on the valency is not assumed. The definition of the Cheeger constant is considered with the edge boundary.

- *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 [**?**]:

OPEN QUESTION 11.82. Is it true every Cayley graph of every finitely generated group with exponential growth contains a tree with positive Cheeger constant?

Note that the open case is that of amenable non-linear groups with exponential growth.

11.7. Quantitative approaches to non-amenability

One can measure "how paradoxical" a group or a group action is *via* the *Tarski numbers.* In what follows, groups are not required to be finitely generated.

DEFINITION 11.83. (1) Given an action of a group G on a set X, and a subset $E \subset X$, which admits a G–paradoxical decomposition in the sense of Definition 11.17, the *Tarski number of the paradoxical decomposition* is the number $k+m$ of elements of that decomposition.

(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 $\mathrm{Tar}_G(X)$ simply by $\mathrm{Tar}(G)$ and we call it the *Tarski number of* G.

Note that G–invariance of the subset E is not required in Definition 11.83.
It is easily seen that $\mathrm{tar}(G) \leqslant \mathrm{Tar}(G)$ for every group G.
Of course, in view of the notion of countably paradoxical sets, one could refine the discussion further and use other cardinal numbers besides the finite ones. We do not follow this direction here.

PROPOSITION 11.84. *Let G be a group, $G \curvearrowright X$ be an action and $E \subset X$ be a nonempty subset.*

(1) *If H is a subgroup of G then* $\mathrm{Tar}_G(X,E) \leqslant \mathrm{Tar}_H(X,E)$.

(2) *The lower Tarski number* $\mathrm{tar}(G)$ *of a group is at least two.*
   Moreover, $\mathrm{tar}(G) = 2$ *if and only if G contains a free two-generated sub-semigroup S.*

PROOF. (1) If the subset E does not admit a paradoxical decomposition with respect to the action of H on X then there is nothing to prove. Consider an H-paradoxical decomposition

$$E = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$E = h_1 X_1 \sqcup ... \sqcup h_k X_k = h'_1 Y_1 \sqcup ... \sqcup h'_m Y_m,$$

and $k + m = \mathrm{Tar}_H(X, E)$. The above decomposition is paradoxical for the action of G on X as well, hence $\mathrm{Tar}_G(X, E) \leqslant \mathrm{Tar}_H(X, E)$.

(2) The fact that every $\mathrm{Tar}_G(X, E)$ is at least two is immediate.

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 of X with a decomposition $E = A \sqcup B$ and two elements $g, h \in G$ such that $gA = E$ and $hB = E$. Set $g' := g^{-1}, h' := h^{-1}$. We claim that g' and h' generate a free subsemigroup in G. Indeed every non-trivial word w in g', h' cannot equal the identity because, depending on whether its first letter is g' or h', it will have the property that $wE \subseteq A$ or $wE \subseteq B$.

Two non-trivial words w and u in g', h' cannot be equal either. Indeed, without loss of generality we may assume that the first letter in w is g', while the first letter in u is h'. Then $wE \subseteq A$ and $uE \subseteq B$, whence $w \neq u$.

We now prove the converse part of the equivalence. Let x, y be two elements in G generating the free sub-semigroup S, let S_x be the set of words beginning in x and S_y be the set of words beginning in y. Then $S = S_x \sqcup S_y$, with $x^{-1} S_x = S$ and $y^{-1} S_y = S$. □

R. Grigorchuk constructed in [?] examples of finitely-generated amenable torsion groups G which are weakly paradoxical, thus answering Rosenblatt's conjecture [?, Question 12.9.b]. Thus, every such amenable group G satisfies

$$3 \leqslant \mathrm{tar}(G) < \infty.$$

QUESTION 11.85. What are the possible values of $\mathrm{tar}(G)$ for an amenable 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 $\mathrm{Tar}(G)$, that is for G–paradoxical sets that are moreover G–invariant.

PROPOSITION 11.86. *Let G be a group, and let $G \curvearrowright X$ be an action.*

(1) $\mathrm{Tar}_G(X) \geq 4$.

(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 in every paradoxical decomposition of X one must have $k \geqslant 2$ and $m \geqslant 2$, the Tarski number is always at least 4.

(2) The proof of this statement is identical to the one appearing in Chapter 10, Section 10.4, Step 3, for $E = \mathbb{S}^2 \setminus C$. □

Proposition 11.86, (2), has a strong converse, appearing as a first statement in the following proposition.

Proposition 11.87. *1. If* $\mathrm{Tar}_G(X) = 4$*, then* G *contains a non-abelian free subgroup.*

2. If X *admits a* G*-paradoxical decomposition*

$$X = X_1 \sqcup X_2 \sqcup Y_1 \sqcup \ldots \sqcup Y_m,$$

then G *contains an element of infinite order. In particular, if* G *is a torsion group then for every* G*-action on a set* X*,* $\mathrm{Tar}_G(X) \geq 6$*.*

Proof. 1. By hypothesis, there exists a decomposition

$$X = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2$$

and elements $g_1, g_2, h_1, h_2 \in G$, such that

$$g_1 X_1 \sqcup g_2 X_2 = h_1 Y_1 \sqcup h_2 Y_2 = X\,.$$

Set $g := g_1^{-1} g_2$ and $h := h_1^{-1} h_2$; then

$$X_1 \sqcup g X_2 = X, Y_1 \sqcup h Y_2 = X. \tag{11.12}$$

This implies that

$$g X_1 \sqcup g Y_1 \sqcup g Y_2 = X \setminus g(X_2) = X_1$$

and, similarly,

$$h X_1 \sqcup h X_2 \sqcup h Y_1 = Y_1.$$

In particular, $gX_1 \subset X_1$, $hY_1 \subset Y_1$. It follows that for every $n \in \mathbb{N}$,

$$g^n X_1 \subseteq X_1, \quad \text{ and } h^n Y_1 \subseteq Y_1.$$

It also follows that for every $n \in \mathbb{N}$,

$$g^n (Y_1 \sqcup Y_2) \subseteq g^{n-1}(X_1) \subseteq X_1$$

and that

$$h^n (X_1 \sqcup X_2) \subseteq h^{n-1}(Y_1) \subseteq Y_1.$$

Equations (11.12) also imply that

$$X = g^{-1} X_1 \sqcup X_2 = h^{-1} Y_1 \sqcup Y_2.$$

Furthermore, for every $n \in \mathbb{N}$,

$$g^{-n}(X_2) \subseteq X_2 \quad \text{and} \quad h^{-n}(Y_2) \subseteq Y_2$$

and

$$g^{-n}(Y_1 \sqcup Y_2) \subseteq X_2 \quad \text{and} \quad h^{-n}(X_1 \sqcup X_2) \subseteq Y_2.$$

Now we can apply Lemma 4.37 with $A := Y_1 \sqcup Y_2$ and $B := X_1 \sqcup X_2$; it follows that bijections g and h of X generate a free subgroup F_2.

2. Let $g_1, g_2 \in G$ be such that

$$g_1 X_1 \sqcup g_2 X_2 = X\,.$$

Again, set $g := g_1^{-1} g_2$. The same arguments as in the proof of Part 1 show that for every $n > 0$,

$$g^n (Y_1 \sqcup \ldots \sqcup Y_m) \subseteq X_1.$$

Therefore, $g^n \neq 1$ for all $n > 0$. $\square$

S. Wagon (Theorems 4.5 and 4.8 in [?]) proved a stronger form of Proposition 11.87 and Proposition 11.86, part (2):

THEOREM 11.88 (S. Wagon). *Let G be a group acting on a set X. The Tarski number* $\mathrm{Tar}_G(X)$ *is four if and only if G contains a free non-abelian subgroup F such that the stabilizer in F of each point in X is abelian.*

As an immediate consequence of Proposition 11.86 is the following

COROLLARY 11.89. *The Tarski number for the action of $SO(n)$ on the $(n-1)$–dimensional sphere $\mathbb{S}^{n-1}$ is 4, for every $n \geqslant 3$.*

The result on the paradoxical decomposition of Euclidean balls can also be refined, and the Tarski number computed. We begin by noting that the Euclidean unit ball $\mathbb{B}$ in $\mathbb{R}^n$ centered in the origin 0 is never paradoxical with respect to the action of the orthogonal group $O(n)$. Indeed, assume that there exists a decomposition

$$\mathbb{B} = X_1 \sqcup \cdots \sqcup X_n \sqcup Y_1 \sqcup \cdots \sqcup Y_m$$

such that

$$\mathbb{B} = g_1 X_1 \sqcup \cdots \sqcup g_n X_n = h_1 Y_1 \sqcup \cdots \sqcup h_m Y_m$$

with

$$g_1, \ldots, g_n, h_1, \ldots, h_m \in O(n).$$

Then the origin 0 is contained in only one of the sets of the initial partition, say, in X_1. It follows that none of the sets Y_j contains 0; hence, neither does the union

$$h_1 Y_1 \sqcup \cdots \sqcup h_m Y_m$$

which contradicts the fact that this union equals $\mathbb{B}$.

The following result was first proved by R. M. Robinson in [?].

PROPOSITION 11.90. *The Tarski number for the unit ball $\mathbb{B}$ in $\mathbb{R}^n$ with respect to the action of the group of isometries G of $\mathbb{R}^n$ is 5.*

PROOF. We first prove that the Tarski number cannot be 4. Assume to the contrary that there exists a decomposition

$$\mathbb{B} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2$$

and $g_1, g_2, h_1, h_2 \in G = \mathrm{Isom}(\mathbb{R}^n)$, such that

$$\mathbb{B} = g_1 X_1 \sqcup g_2 X_2 = h_1 Y_1 \sqcup h_2 Y_2.$$

By Proposition 3.56, the elements g_i and h_j are compositions of linear isometries and translations. Since, as we observed above, elements g_i, h_j cannot all belong to $O(n)$, it follows that, say, g_1 has a non-trivial translation component:

$$g_1(x) = U_1 x + T_1, \quad U_1 \in O(n), T_1 \neq 0.$$

We claim that $g_2 \in O(n)$ and that X_2 contains a closed hemisphere of the unit sphere $\mathbb{S} = \partial\mathbb{B}$.

Indeed, $g_1 X_1 \subset T_1 \mathbb{B}$. As T_1 is non-trivial, $T_1 \mathbb{B} \neq \mathbb{B}$; hence, $T_1 \mathbb{S}$ contains no subsets of the form $\{x, -x\}$, where x is a unit vector. Therefore, $T_1 \mathbb{B} \cap \mathbb{S}$ is contained in an open hemisphere of the unit sphere $\mathbb{S}$. Since the union $g_1(X_1) \cup g_2(X_2)$ contains the sphere $\mathbb{S}$, it follows that $g_2 X_2$ contains a closed hemisphere in $\mathbb{S}$, and, hence, so does $g_2 \mathbb{B}$. Since $g_2 \mathbb{B} \subset \mathbb{B}$, it follows that $g_2 \mathbb{B} = \mathbb{B}$, hence, $g_2(0) = 0$ and, thus, X_2 contains a closed hemisphere of $\mathbb{S}$.

This claim implies that $(Y_1 \sqcup Y_2) \cap \mathbb{S}$ is contained in an open hemisphere of $\mathbb{S}$. By applying the above arguments to the isometries h_1, h_2, we see that both h_1, h_2 belong to $O(n)$. We then have that

$$\mathbb{S} = h_1(Y_1 \cap \mathbb{S}) \sqcup h_2(Y_2 \cap \mathbb{S}).$$

On the other hand, both Y_1, Y_2 and, hence, $h_1(Y_1), h_2(Y_2)$ are contained in open hemispheres of $\mathbb{S}$. Union of two open hemispheres in $\mathbb{S}$ cannot be the entire $\mathbb{S}$. Contradiction. Thus, $\mathrm{Tar}_G(\mathbb{B}) \geq 5$.

We now show that there exists a paradoxical decomposition of $\mathbb{B}$ with five elements. Corollary 11.89 implies that there exist g_1, g_2, h_1, h_2 in $SO(n)$ such that

$$\mathbb{S} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2 = g_1 X_1 \sqcup g_2 X_2 = h_1 Y_1 \sqcup h_2 Y_2$$

As in the proof of Proposition 11.87, we take $g := g_1^{-1} g_2, h := h_1^{-1} h_2$ and obtain

$$\mathbb{S} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2 = X_1 \sqcup g X_2 = Y_1 \sqcup h Y_2.$$

It follows that for every $\lambda > 0$ the sphere $\lambda\mathbb{S}$ (of radius λ) has the paradoxical decomposition

$$\lambda\mathbb{S} = \lambda X_1 \sqcup \lambda X_2 \sqcup \lambda Y_1 \sqcup \lambda Y_2 = \lambda X_1 \sqcup g\, \lambda X_2 = \lambda Y_1 \sqcup h\, \lambda Y_2$$

The group $\Gamma := \langle g, h \rangle$ generated by g and h contains countably many nontrivial orthogonal transformations; the fixed-point set of every such transformation is a proper linear subspace in $\mathbb{R}^n$. Therefore, there exists a point $P \in \mathbb{S}$ not fixed by any nontrivial element of Γ. Let Ω denote the Γ-orbit of P. Since the action of Γ on Ω is free, the map

$$\gamma \mapsto \gamma P$$

is a bijection $\Gamma \to \Omega$. The group Γ is a free group of rank two with free generators g, h, hence as in equation (10.1) of Section 10.4, we have the following paradoxical decomposition of the group Γ:

$$\langle g, h \rangle = \{1\} \sqcup \mathcal{W}_g \sqcup \mathcal{W}_{g^{-1}} \sqcup \mathcal{W}_h \sqcup \mathcal{W}_{h^{-1}}\,,$$

where

$$\Gamma = \mathcal{W}_g \sqcup g\mathcal{W}_{g^{-1}}, \quad \Gamma = \mathcal{W}_h \sqcup h\mathcal{W}_{h^{-1}}.$$

We now replace the original paradoxical decomposition of $\mathbb{S}$ by

$$\mathbb{S} = X_1' \sqcup X_2' \sqcup Y_1' \sqcup Y_2' \sqcup \{P\}$$

where

$$\begin{aligned} X_1' &= (X_1 \setminus \Omega) \sqcup \mathcal{W}_g P, \\ X_2' &= (X_2 \setminus \Omega) \sqcup \mathcal{W}_{g^{-1}} P, \\ Y_1' &= (X_1 \setminus \Omega) \sqcup \mathcal{W}_h P, \\ Y_2' &= (Y_2 \setminus \Omega) \sqcup \mathcal{W}_{h^{-1}} P\,. \end{aligned}$$

Clearly, $X_1' \sqcup g X_2' = Y_1' \sqcup h Y_2' = \mathbb{S}$.

We now consider the decomposition

$$\mathbb{B} = U_1 \sqcup U_2 \sqcup V_1 \sqcup V_2 \sqcup \{P\}\,,$$

where

$$U_1 = \{O\} \sqcup \bigsqcup_{0<\lambda<1} \lambda X_1 \sqcup X_1',$$

$$U_2 = \bigsqcup_{0<\lambda<1} \lambda X_2 \sqcup X_2',$$

$$V_1 = \bigsqcup_{0<\lambda<1} \lambda Y_1 \sqcup Y_1',$$

and

$$V_2 = \bigsqcup_{0<\lambda<1} \lambda Y_2 \sqcup Y_2'.$$

Then $U_1 \sqcup gU_2 = \mathbb{B}$, while $V_1 \sqcup hV_2 \sqcup \{T(P)\} = \mathbb{B}$, where T is the translation sending the point P to the origin O. $\square$

Below we describe the behavior of the Tarski number of groups with respect to certain group operations.

PROPOSITION 11.91. (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) *If N is a normal subgroup of G then* $\mathrm{Tar}(G) \leqslant \mathrm{Tar}(G/N)$.

PROOF. (1) If H is amenable then there is nothing to prove. Consider a decomposition

$$H = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$H = h_1 X_1 \sqcup ... \sqcup h_k X_k = h_1' Y_1 \sqcup ... \sqcup h_m' Y_m$$

and $k + m = \mathrm{Tar}(H)$.

Let $\mathcal{R}$ be the set of representatives of right H–cosets inside G. Then $\widetilde{X}_i = X_i\mathcal{R}$, $i \in \{1, 2, ..., k\}$ and $\widetilde{Y}_j = Y_j\mathcal{R}$, $j \in \{1, 2, ..., m\}$ form a paradoxical decomposition for G.

(2) Given a decomposition

$$G = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$G = g_1 X_1 \sqcup ... \sqcup g_k X_k = h_1 Y_1 \sqcup ... \sqcup h_m Y_m$$

and $k + m = \mathrm{Tar}(G)$, consider the subgroup H generated by $g_1, ..., g_k, h_1, ..., h_m$. Thus $\mathrm{Tar}(H) \leqslant \mathrm{Tar}(G)$; since the converse inequality is also true, the equality holds.

(3) Set $\overline{Q} := G/N$. As before, we may assume, without loss of generality, that $\overline{Q}$ is paradoxical. Let

$$\overline{Q} = \overline{X}_1 \sqcup ... \sqcup \overline{X}_k \sqcup \overline{Y}_1 \sqcup ... \sqcup \overline{Y}_m$$

be a decomposition such that

$$\overline{Q} = g_1 \overline{X}_1 \sqcup ... \sqcup g_k \overline{X}_k = h_1 \overline{Y}_1 \sqcup ... \sqcup h_m \overline{Y}_m$$

and $k + m = \mathrm{Tar}\left(\overline{Q}\right)$.

Consider an (injective) section $\sigma : \overline{Q} \to G$, for the projection $G \to \overline{Q}$; set $Q := s(\overline{Q})$. Then $G = QN$ and the sets $X_i = \sigma\left(\overline{X}_i\right) N$, $i \in \{1, 2, ..., k\}$ and $Y_j = \sigma\left(\overline{Y}_j\right) N$, $j \in \{1, 2, ..., m\}$ form a paradoxical decomposition for G. $\square$

Proposition 11.91, (1), allows to formulate the following quantitative version of Corollary 11.49.

COROLLARY 11.92. *If two groups are co-embeddable then they have the same Tarski number.*

It is proven in [?], [**Ady79**, Theorem VI.3.7] that, for every odd $m \geqslant 665$, two free Burnside groups $B(n;m)$ and $B(k;m)$ of exponent m and with $n \geqslant 2$ and $k \geqslant 2$, are co-embeddable. Thus:

COROLLARY 11.93. *For every odd $m \geqslant 665$, and $n \geqslant 2$, the Tarski number of a free Burnside groups $B(n;m)$ of exponent m is independent of the number of generators n.*

COROLLARY 11.94. *A group has the Tarski number 4 if and only if it contains a non-abelian free subgroup.*

PROOF. If a group G contains a non-abelian free subgroup then the result follows by Proposition 11.86, (1), (2), and Proposition 11.91, (1). If a group G has $\mathrm{Tar}(G) = 4$ then the claim follows from Proposition 11.87. □

Thus, the Tarski number helps to classify the groups that are non-amenable and do not contain a copy of F_2. This class of groups is not very well understood and, as noted in the end of Section 11.6, its only known members are "infinite monsters". For torsion groups G as we proved above $\mathrm{Tar}(G) \geqslant 6$. On the other hand, Ceccherini, Grigorchuk and de la Harpe proved:

THEOREM 11.95 (Theorem 2, [?]). *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.*

Natural questions, in view of Corollary 11.93, are the following:

QUESTION 11.96. How does the Tarski number of a free Burnside group $B(n;m)$ depend on the exponent m? What are its possible values?

QUESTION 11.97 (Question 22 [?], [?]). What are the possible values for the Tarski numbers of groups? Do they include 5 or 6? Are there groups which have arbitrarily large Tarski numbers?

It would also be interesting to understand how much of the Tarski number is encoded in the large scale geometry of a group. In particular:

QUESTION 11.98. 1. Is the Tarski number of a group G equal to that of its direct product $G \times F$ with an arbitrary finite group F?
2. Is the same true when F is an arbitrary amenable group?
3. Is the Tarski number invariant under virtual isomorphisms?

Note that the answers to Questions 11.98 are positive for the Tarski number equal to ∞ or 4.

QUESTION 11.99. 1. Is the Tarski number of groups a quasi-isometry invariant?
2. Is it at least true that the existence of an (L, C)-quasi-isometry between groups implies that their Tarski number differ at most by a constant $K = K(L, A)$?

The answer to Question 11.98 (Part 1) is, of course, positive for $\mathrm{Tar}(G) = \infty$, but, already, for $\mathrm{Tar}(G) = 4$ this question is equivalent to a well-known open problem below. A group G is called *small* if it contains no free nonabelian subgroups. Thus, G is small iff $\mathrm{Tar}(G) > 4$.

QUESTION 11.100. Is smallness invariant under quasi-isometries of finitely generated groups?

11.8. Uniform amenability and ultrapowers

In this section we discuss a *uniform version of amenability* and its relation to ultrapowers of groups.

Recall (Definition 11.71) that a (discrete) group G is *amenable* (has *the Følner Property*) iff for every finite subset K of G and every $\epsilon \in (0,1)$ there exists a finite non-empty subset $F \subset G$ satisfying:

$$|KF \bigtriangleup F| < \epsilon |F|.$$

DEFINITION 11.101. 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

$$|F| \leqslant \phi(\epsilon, |K|)\,.$$

EXAMPLES 11.102.
(1) Nilpotent groups have the uniform Følner property, [?].
(2) A subgroup of a group with the uniform Følner Property also has this property, [?].
(3) Let N be a normal subgroup of G. The group G has the uniform Følner Property if and only if N and G/N have this property, [?].
(4) There is an example of a countable (but infinitely generated) group that is amenable but does not satisfy the uniform Følner Property, see [?, §IV].

THEOREM 11.103 (G. Keller [?], [?]).
(1) *If for some non-principal ultrafilter ω the ultrapower G^ω has the Følner Property, then G also has the uniform Følner Property.*
(2) *If G has the uniform Følner property, then for every non-principal ultrafilter ω, the ultrapower G^ω also has the uniform Følner property.*

PROOF. (1) The group G can be identified with the "diagonal" subgroup $\widehat{G}$ of G^ω, represented by constant sequences in G. It follows by Proposition 11.76 that G has the Følner property. Assume that it 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 **??**, K has cardinality k. Since G^ω is amenable it follows that there exists a finite subset $U \in G^\omega$ such that $|KU \bigtriangleup U| < \epsilon|U|$. Let c be the cardinality of U. According to Lemma **??**, (**??**), $U = (U_n)^\omega$, 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.

(2) Let $k \in \mathbb{N}$ and $\epsilon > 0$; define $m := \phi(\epsilon, k)$ where ϕ is the function in the uniform Følner property of G. Let K be a subset of cardinality k in G^ω. Lemma **??** 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 **??**, (**??**), implies that

$$KF \bigtriangleup F = (K_n\Omega_n \bigtriangleup \Omega_n)^\omega,$$

whence $|KF \bigtriangleup F| < \epsilon|F|$. Since $|F| \leqslant m$ according to Lemma **??**, (**??**), the claim follows. $\square$

PROPOSITION 11.104 (G. Keller, [**?**], Corollary 5.9). *Every group with the uniform Følner property satisfies a law.*

PROOF. Indeed, by Theorem 11.103, (2), if G has the uniform Følner Property then any ultrapower G^ω has the uniform Følner Property. Assume that G does not satisfy any law, i.e., in view of Lemma **??**, the group G^ω contains a subgroup isomorphic to the free group F_2. By Proposition 11.76 it would then follow that F_2 has the Følner Property, a contradiction. $\square$

11.9. Quantitative approaches to amenability

One quantitative approach to amenability is due to A.M. Vershik, who introduced in [**?**] the *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 $\mathrm{F}_o^{\mathcal{G}}(x)$ is the minimal cardinality of a finite non-empty set F of vertices satisfying the inequality

$$|E(F, F^c)| \leqslant \frac{1}{x}|F|\,.$$

According to the inequality (1.1) relating the cardinalities of the vertex and edge boundary, if one replaces in the above $E(F,F^c)$ by the vertex boundary $\partial_V F$ of F, one obtains a Følner function asymptotically equal to the first, in the sense of Definition 1.7.

The following is a quantitative version of Theorem 11.10.

PROPOSITION 11.105. *If two graphs of bounded geometry are quasi-isometric then they are either both non-amenable or both amenable and 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)–quasi-isometries 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 either $\mathcal{G}$ or $\mathcal{G}'$.

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 $|A''| \leqslant |A'| \leqslant |A|$. By hypothesis, the Hausdorff distance between A'' and A is C, therefore $|A| \leqslant \alpha|A''|$. Thus we have the inequalities

$$\frac{1}{\alpha}|A| \leqslant |f(A)| \leqslant |A|\,, \tag{11.13}$$

and similar inequalities for finite subsets in $V(\mathcal{G}')$ and their images by g.

The first part of the statement follows from Theorem 11.10.

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 $x \in (0, \infty)$, and let A be a finite subset in $V(\mathcal{G})$ such that $|A| = \mathrm{F}_o^{\mathcal{G}}(x)$ and

$$|\partial_V(A)| \leqslant \frac{1}{x}|A|\,.$$

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 $\partial_V(B)$ is composed 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 (11.13),

$$|\partial_V(B)| \leqslant \alpha\left|g\left(\partial_V(B)\right)\right| \leqslant \alpha\lambda\left|\partial_V(A)\right| \leqslant \alpha\lambda\frac{1}{x}|A| \leqslant$$

$$\alpha^2\lambda\frac{1}{x}|A'| \leqslant \alpha^2\lambda\frac{1}{x}|B|\,.$$

We have thus obtained that, for $\kappa = \alpha^2\lambda$ and every $x > 0$, the value $\mathrm{F}_o^{\mathcal{G}'}\left(\frac{x}{\kappa}\right)$ is at most $|B| \leqslant m^R|A'| \leqslant m^R|A| = m^R\,\mathrm{F}_o^{\mathcal{G}}(x)$. 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 similarly. □

Proposition 11.105 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 11.106. (1) We say that a sequence (F_n) of finite subsets in a group *realizes the Følner function* of that group if for some generating set S, $\mathrm{card}\, F_n = \mathrm{F}_o^{\mathcal{G}}(n)$, where $\mathcal{G}$ is the Cayley graph of G with respect to S, and

$$|E(F_n, F_n^c)| \leqslant \frac{1}{n}\,|F_n|\,.$$

(2) We say that a sequence (A_n) of finite subsets in a group *quasi-realizes the Følner function* of that group if $\mathrm{card}\, A_n \asymp \mathrm{F}_o^{G}(n)$ 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 11.107. *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 X of H. We shall prove that for the Følner functions defined with respect to these generating sets, we can write $\mathrm{F}_o^H(x) \leqslant \mathrm{F}_o^G(x)$ for every $x > 0$. Let F be a finite subset in G such that $|F| = \mathrm{F}_o^G(x)$ and $|\partial_V F| \leqslant \frac{1}{x}|F|$.

The set F intersects finitely many cosets of H, $g_1 H, \dots, g_k H$. In particular $F = \bigsqcup_{i=1}^k F_i$, where $F_i = F \cap g_i H$. We denote by $\partial_V^i F_i$ the set of vertices in $\partial_V F_i$ joined to vertices in F_i by edges with labels in X. The sets $\partial_V^i F_i$ are contained in $g_i H$ for every $i \in \{1, 2, \dots, k\}$, hence they are pairwise disjoint subsets of $\partial_V F$. We can thus write

$$\sum_{i=1}^k \left|\partial_V^i F_i\right| \leqslant |\partial_V F| \leqslant \frac{1}{x}|F| = \frac{1}{x}\sum_{i=1}^k |F_i|\,.$$

It follows that there exists $i \in \{1, 2, \dots, k\}$ such that $\left|\partial_V^i F_i\right| \leqslant \frac{1}{x}|F_i|$. By construction, $F_i = g_i K_i$ with K_i a subset of H, and the previous inequality is equivalent to $|\partial_V K_i| \leqslant \frac{1}{x}|K_i|$, where the vertex-boundary $\partial_V K_i$ is considered in the Cayley graph of H with respect to the generating set X. We then have that $\mathrm{F}_o^H(x) \leqslant |K_i| \leqslant |F| = \mathrm{F}_o^G(x)$. □

One may ask how do the Følner functions relate to the growth functions, and when do the sequences of balls of fixed centre quasi-realize the Følner function, especially under the extra hypothesis of subexponential growth, see Proposition 11.6.

THEOREM 11.108. *Let G be an infinite finitely generated group.*

(1) $\mathrm{F}_o^G(n) \succeq \mathfrak{G}_G(n)$.

(2) *If the sequence of balls $B(1, n)$ quasi-realizes the Følner function of G then G is virtually nilpotent.*

PROOF. (1) Consider a sequence (F_n) of finite subsets in G that realizes the Følner function of that group, for some generating set S. In particular

$$|E(F_n, F_n^c)| \leqslant \frac{1}{n}\,|F_n|\,.$$

We let $\mathfrak{G}$ denote the growth function of G with respect to the generating set S.

Inequality (??) in Proposition ?? implies that

$$\frac{|F_n|}{2dk_n} \leqslant \frac{1}{n}\,|F_n|\,,$$

where $d = |S|$ and k_n is such that $\mathfrak{G}(k_n - 1) \leqslant 2|F_n| < \mathfrak{G}(k_n - 1)$.

This implies that

$$k_n - 1 \geqslant \frac{n}{2d} - 1 \geqslant \frac{n}{4d}\,,$$

whence,

$$2\,\mathrm{F}_o^G(n) \geqslant \mathfrak{G}\left(k_n - 1\right) \geqslant \mathfrak{G}\left(\frac{n}{4d}\right)\,.$$

(2) The inequality in (2) implies that for some $a > 0$,

$$|S(1, n+1)| \leqslant \frac{a}{n}\,|B(1, n)|\,.$$

In terms of the growth function, this inequality can be re-written as

$$\frac{\mathfrak{G}(n+1)-\mathfrak{G}(n)}{\mathfrak{G}(n)} \leqslant \frac{a}{n}\,. \tag{11.14}$$

Let $f(x)$ 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 (11.14) means that for all $x \notin \mathbb{N}$,

$$\frac{f'(x)}{f(x)} \leqslant \frac{a}{x}\,.$$

which, by integration, implies that $\ln|f(x)| \leqslant a \ln|x| + b$. In particular, it follows that $\mathfrak{G}(n)$ is bounded by a polynomial in n, whence, G is virtually nilpotent. $\square$

In view of Theorem 11.108, (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. The particular case of the wreath products already shows that there is no upper bound for the Følner functions, and that consequently they can differ a lot from the growth function.

THEOREM 11.109 (A. Erschler, [**?**]). *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 $aF(x) < F(bx)$ for every $x > 0$.*

Then the Følner function of the wreath product $A \wr B$ is asymptotically equal to $[\mathrm{F}_o^B(x)]^{\mathrm{F}_o^A(x)}$.

A. Erschler proved in [**?**] that 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.

11.10. Amenable hierarchy

We conclude this chapter with the following diagram illustrating hierarchy of amenable groups:

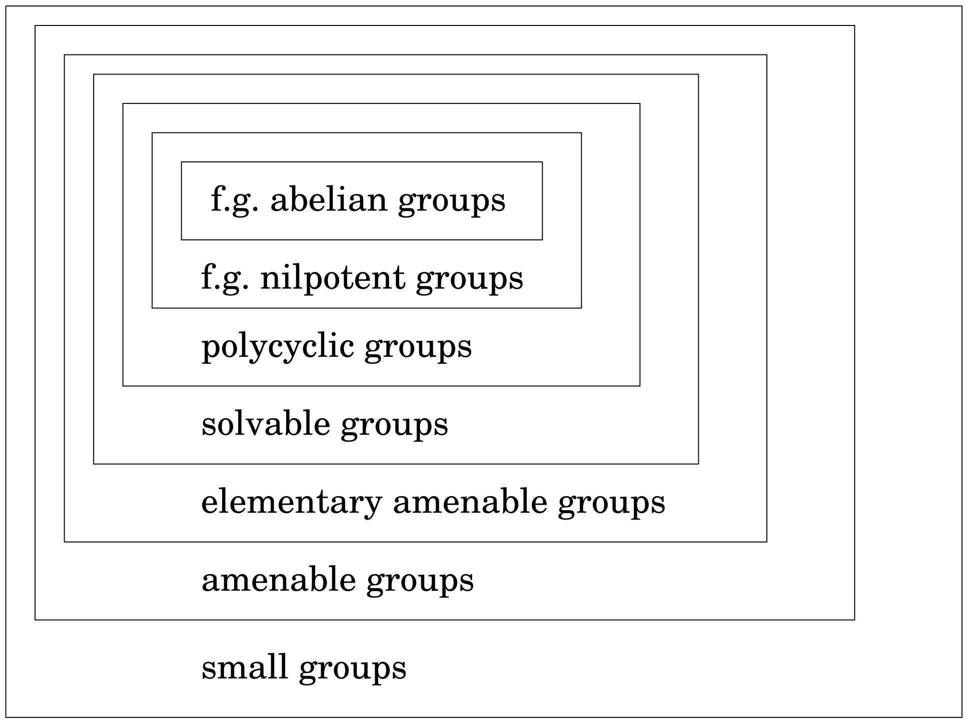


FIGURE 11.1. Hierarchy of amenable groups

Bibliography

[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.

[And05] J. W. Anderson, *Hyperbolic geometry*, Springer, 2005.

[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.

[Bal95] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Seminar, Band 25, Birkhäuser, 1995.

[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.

[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), 339–356.

[BdlHV08] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, Cambridge, 2008.

[Bea83] A.F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer, 1983.

[Ber68] G. Bergman, *On groups acting on locally finite graphs*, Annals of Math. **88** (1968), 335–340.

[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.

[BGS85] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of non-positive curvature*, Progress in Math., vol. 61, Birkhauser, 1985.

[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.

[Big01] S. Bigelow, *Braid groups are linear*, J. Amer. Math. Soc. **14** (2001), no. 2, 471–486 (electronic).

[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.

[BK02] ———, *Rectifying separated nets*, Geom. Funct. Anal. **12** (2002), 80–92.

[Bol79] B. Bollobás, *Graph theory, an introductory course*, Graduate Texts in Mathematics, vol. 63, Springer, 1979.

[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.

[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.

[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.

[Bow91] 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.

[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.

[Bow98] ———, *A topological characterisation of hyperbolic groups*, J. Amer. Math. Soc. **11** (1998), 643–667.

[Boż80] M. Bożejko, *Uniformly amenable discrete groups*, Math. Ann. **251** (1980), 1–6.

[BP92] R. Benedetti and C. Pertonio, *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.

[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.

[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.S. Brown, *Cohomology of Groups*, Graduate Texts in Mathematics, vol. 87, Springer, 1982.

[BS97] 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.

[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.

[BW11] M. Bridson and H. Wilton, *On the difficulty of presenting finitely presentable groups*, Groups Geom. Dyn. **5** (2011), no. 2, 301–325.

[Cal08] D. Calegari, *Length and stable length*, Geom. Funct. Anal. **18** (2008), no. 1, 50–76.

[Can84] J. Cannon, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata **16** (1984), no. 2, 123–148.

[CDP90] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et théorie des groupes*, 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.

[Cho80] C. Chou, *Elementary amenable groups*, Illinois J. Math. **24** (1980), no. 3, 396–407.

[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.

[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.

[Dav08] M. Davis, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series, vol. 32, Princeton University Press, Princeton, NJ, 2008.

[Day57] M.M. Day, *Amenable semigroups*, Illinois J. Math. **1** (1957), 509–544.

[dC92] M. P. do Carmo, *Riemannian geometry*, Birkhäuser Boston Inc., Boston, MA, 1992.

[DD89] W. Dicks and M. J. Dunwoody, *Groups acting on graphs*, Cambridge University Press, Cambridge-New York, 1989.

[Del96] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), no. 3, 661–682.

[DG] T. Delzant and M. Gromov, *Groupes de Burnside et géométrie hyperbolique*, preprint.

[DG11] F. Dahmani and V. Guirardel, *The isomorphism problem for all hyperbolic groups*, Geom. Funct. Anal. **21** (2011), 223–300.

[dlH73] P. de la Harpe, *Moyennabilité de quelques groupes topologiques de dimension infinie*, C. R. Acad. Sci. Paris Sér. I Math. **277** (1973), no. 14, 1037–1040.

[dlH00] ______, *Topics in geometric group theory*, Chicago Lectures in Mathematics, University of Chicago Press, 2000.

[dlHGCS99] P. de la Harpe, R. I. Grigorchuk, and T. Chekerini-Silʹberstaĭn, *Amenability and paradoxical decompositions for pseudogroups and discrete metric spaces*, Proc. Steklov Inst. Math. **224** (1999), 57–97.

[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.

[DR09] M. Duchin and K. Rafi, *Divergence of geodesics in Teichmüller space and the Mapping Class group*, GAFA **19** (2009), 722–742.

[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.

[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.

[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.
[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.
[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.
[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.
[Fed69] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.
[Fø55] E. Følner, *On groups with full banach mean value*, Math. Scand. **3** (1955), 243–254.
[Fox53] R. H. Fox, *Free differential calculus. I. Derivation in the free group ring*, Ann. of Math. (2) **57** (1953), 547–560.
[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.
[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.
[G\"60] P. Günther, *Einige Sätze über das Volumenelement eines Riemannschen Raumes*, Publ. Math. Debrecen **7** (1960), 78–93.
[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 hyperbolic d'apres Mikhael Gromov*, Progress in Mathematics, vol. 83, Birkhäuser, 1990.
[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.
[Ger93] ———, *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.
[Ger94] ———, *Quadratic divergence of geodesics in CAT(0)-spaces*, Geom. Funct. Anal. **4** (1994), no. 1, 37–51.
[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.
[GM07] Y. Glasner and N. Monod, *Amenable actions, free products and fixed-point property*, Bull. London Math. Society **39** (2007), 138–150.
[GP10] V. Guillemin and A. Pollack, *Differential topology*, AMS Chelsea Publishing, Providence, RI, 2010, Reprint of the 1974 original.
[Gri87] R. I. Grigorchuk, *Superamenability and the problem of occurence of free semigroups*, Functional Analysis and its Applications **21** (1987), 64–66.

[Gro81] M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. IHES **53** (1981), 53–73.

[Gro82] ______, *Volume and bounded cohomology*, Publ. Math. IHES (1982), no. 56, 5–99 (1983).

[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.

[GS90] S. Gersten and H. Short, *Small cancellation theory and automatic groups*, Invent. Math. **102** (1990), 305–334.

[Hat02] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[Hei01] J. Heinonen, *Lectures on analysis on metric spaces*, Universitext, Springer-Verlag, New York, 2001.

[Hel01] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Graduate Studies in Mathematics, vol. 34, Amer. Math. Soc., 2001.

[Hig40] G. Higman, *The units of group-rings*, Proc. London Math. Soc. (2) **46** (1940), 231–248.

[Hir76] M. Hirsch, *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer, 1976.

[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)

[Hum75] J. E. Humphreys, *Linear algebraic groups*, Springer-Verlag, New York, 1975, Graduate Texts in Mathematics, No. 21.

[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.

[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.

[Iva92] N. Ivanov, *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.

[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.

[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.

[KK05] M. Kapovich and B. Kleiner, *Coarse Alexander duality and duality groups*, Journal of Diff. Geometry **69** (2005), 279–352.

[KKL98] M. Kapovich, B. Kleiner, and B. Leeb, *Quasi-isometries and the de Rham decomposition*, Topology **37** (1998), 1193–1212.

[KL98a] M. Kapovich and B. Leeb, *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.

[KL09] ———, *Induced quasi-actions: A remark*, Proc. Amer. Math. Soc. **137** (2009), no. 5, 1561–1567.

[KS08] I. Kapovich and P. Schupp, *On group-theoretic models of randomness and genericity*, Groups Geom. Dyn. **2** (2008), no. 3, 383–404.

[Lan64] S. Lang, *Algebraic numbers*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-Palo Alto-London, 1964.

[Li04] P. Li, *Harmonic functions and applications to complete manifolds*, Preprint, 2004.

[LS77] R.C. Lyndon and P.E. Schupp, *Combinatorial group theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 89, Springer, 1977.

[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.

[Mal40] A. I. Mal'cev, *On isomorphic matrix representations of infinite groups*, Mat. Sb. **8** (1940), 405–422.

[Mar91] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag, Berlin, 1991.

[Mas91] W.S. Massey, *A basic course in algebraic topology*, Graduate Texts in Mathematics, vol. 127, Springer, 1991.

[Mil68] J. Milnor, *A note on curvature and fundamental group*, J. Diff. Geom. **2** (1968), 1–7.

[Min01] I. Mineyev, *Straightening and bounded cohomology of hyperbolic groups*, Geom. Funct. Anal. **11** (2001), no. 4, 807–839.

[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.

[Mor24] M. Morse, *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of AMS **26** (1924), 25–60.

[Mos57] G. D. Mostow, *On the fundamental group of a homogeneous space*, Ann. of Math. **66** (1957), 249–255.

[Mos73] ———, *Strong rigidity of locally symmetric spaces*, Annals of mathematical studies, vol. 78, Princeton University Press, Princeton, 1973.

[Mou88] G. Moussong, *Hyperbolic Coxeter groups*, Ph.D. thesis, The Ohio State University, 1988.

[MR03] C. Maclachlan and A. Reid, *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics, vol. 219, Springer-Verlag, New York, 2003.

[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.

[Ń64] A. Néron, *Modèles minimaux des variétes abèliennes sur les corps locaux et globaux*, Math. Publ. of IHES **21** (1964), 5–128.

[Nag83] J. Nagata, *Modern dimension theory*, revised ed., Sigma Series in Pure Mathematics, vol. 2, Heldermann Verlag, Berlin, 1983.

[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.

[Nic] B. Nica, *Linear groups: Mal'cev's theorem and Selberg's lemma*, Preprint, http://www.uni-math.gwdg.de/nica/linear_groups.pdf.

[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.

[NR97] W. Neumann and L. Reeves, *Central extensions of word hyperbolic groups*, Ann. of Math. **145** (1997), 183–192.

[Ol'80] A. Yu. Ol'shanskii, *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.
[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. **14** (2004), no. 3, 595–679.
[OOS09] A. Yu. Ol'shanskii, 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.
[OS02] A. Yu. Ol'shanskii and M. V. Sapir, *Non-amenable finitely presented torsion-by-cyclic groups*, Publ. Math. IHES **96** (2002), 43–169.
[OV90] A. Onishchik and E. Vinberg, *Lie groups and algebraic groups*, Springer, 1990.
[Pan89] P. Pansu, *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.
[Pap] P. Papasoglu, *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.
[Pap95a] ———, *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.
[Pap95b] ———, *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.
[Par08] J. Parker, *Hyperbolic spaces*, vol. 2, Jyväskylä Lectures in Mathematics, 2008.
[Pau91] 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.
[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.
[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.
[Rat94] J.G. Ratcliffe, *Foundations of hyperbolic manifolds*, Springer, 1994.
[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.

[Ros76] ———, *Invariant means for the bounded measurable functions on a non-discrete locally compact group*, Math. Ann. **220** (1976), no. 3, 219–228.

[Roy68] H. Royden, *Real Analysis*, Macmillan, New York, 1968.

[RS94] E. Rips and Z. Sela, *Structure and rigidity in hyperbolic groups. I*, Geom. Funct. Anal. **4** (1994), no. 3, 337–371.

[Rud87] W. Rudin, *Real and complex analysis*, McGraw-Hill International editions, 1987.

[Sau06] R. Sauer, *Homological invariants and quasi-isometry*, Geom. Funct. Anal. **16** (2006), no. 2, 476–515.

[Sch38] I.J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.

[Sch96] R. Schwartz, *The quasi-isometry classification of rank one lattices*, Publ. Math. IHES **82** (1996), 133–168.

[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.

[Sel95] Z. Sela, *The isomorphism problem for hyperbolic groups. I*, Ann. of Math. (2) **141** (1995), no. 2, 217–283.

[Sel97] ———, *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.

[Sel99] ———, *Endomorphisms of hyperbolic groups. I. The Hopf property*, Topology **38** (1999), no. 2, 301–321.

[Ser80] J. P. Serre, *Trees*, Springer, New York, 1980.

[Sha04] Y. Shalom, *Harmonic analysis, cohomology, and the large-scale geometry of amenable groups*, Acta Math. **192** (2004), no. 2, 119–185.

[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.

[Sta68] J. Stallings, *On torsion-free groups with infinitely many ends*, Ann. of Math. **88** (1968), 312–334.

[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.

[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.

[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.

[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.

[Tuk94] P. Tukia, *Convergence groups and Gromov's metric hyperbolic spaces*, New Zealand J. Math. **23** (1994), no. 2, 157–187.

[Väi05] J. Väisälä, *Gromov hyperbolic spaces*, Expo. Math. **23** (2005), no. 3, 187–231.

[Vav] N. Vavilov, *Concrete group theory*, http://www.math.spbu.ru/user/valgebra/groubook.pdf.

[Ver82] A. Vershik, *Amenability and approximation of infinite groups*, Selecta Math. Soviet. **2** (1982), no. 4, 311–330, Selected translations.

[Ver11] R. Vershynin, *Lectures in geometric functional analysis*, Preprint, University of Michigan, 2011.

[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.
[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.
[Wen08] S. Wenger, *Gromov hyperbolic spaces and the sharp isoperimetric constant*, Invent. Math. **171** (2008), no. 1, 227–255.
[Why99] K. Whyte, *Amenability, bilipschitz equivalence, and the von Neumann conjecture*, Duke Math. J. **99** (1999), 93–112.
[Woe00] W. Woess, *Random walks on infinite graphs and groups*, Cambridge University Press, 2000.
[Wys88] J. Wysoczánski, *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.
[Yom83] Y. Yomdin, *The geometry of critical and near-critical values of differentiable mappings*, Math. Ann. **264** (1983), no. 4, 495–515.
[Zim84] R. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Math, vol. 81, Birkhauser, 1984.