Ergodic Theory

Ben Green, Oxford, Michaelmas Term 2015

Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG

E-mail address: ben.green@maths.ox.ac.uk

2000 *Mathematics Subject Classification.* Primary

Contents

Preface

These are notes for an advanced undergraduate course on ergodic theory. The first draft was written for a 24-hour course in Part III of the Mathematical Tripos in Cambridge in 2008. Those notes were then substantially revised in preparation for a 16-lecture advanced undergraduate course in Oxford in 2014.

The aim is to cover some topics in ergodic theory motivated by applications to number theory: normal numbers, continued fraction expansions, recurrence of polynomials, and Szemerédi's theorem on arithmetic progressions.

I wanted to get away, as far as possible, from the typical style of many texts on ergodic theory in which one first develops or recalls results from measure theory and functional analysis at length. I take the view that a student does not really need to be completely on top of measure theory to derive benefit from a course on ergodic theory. Indeed, such a course can help consolidate or refresh knowledge of measure theory, or act as motivation to go and learn about it. A particular point is that one really does not need to know very much about the *construction* of measures to benefit from this course.

The style of these notes, then, is to "recall" the measure theory we need as we go along. The first several chapters require only very basic notions, and it is only in Chapter **??** (which may well not make it into a typical 16-lecture course) that we require more serious material.

Similar remarks apply to topics in Fourier Analysis and Functional Analysis. Everything we need about the former is summarised (with proofs and further references) in Appendix B, but we would recommend the reader plunge straight into the course without first reading that appendix. Concerning the latter, for much of the course we have tried to minimise the use of anything other than very basic facts. Thus we do require some very simple properties of Hilbert space (projections to closed subspaces and existence of adjoints) but deeper topics such as the Riesz representation theorem are confined to optional parts of the course.

The focus of this book is on applications to number theory and simplicity of exposition. As such, it has been necessary to sacrifice generality. Very often we will make statements which are in fact valid in far greater generality than we state them. We will say nothing about actions of groups other than $\mathbb{Z}$, and we will restrict attention to relatively benign settings (measure-preserving systems on compact metric spaces) when it suit us.

Finally a word on the word *ergodic*, the etymology of which is less than obvious. Apparently the term was coined by Boltzmann and derives from the Greek words $\acute{\varepsilon}\rho\gamma o\nu$ (ergon: "work") and $o\delta\acute{o}\zeta$ (odos: "path" or "way").

CHAPTER 1

Introduction and some examples

Let X be a set, and let $T : X \to X$ be a map. This course is about what happens when the map T is applied repeatedly. If one takes a point x and applies T repeatedly, the resulting set $\{x, Tx, T^2x, \dots\}$ is called an orbit. In this course we will be concerned with variants of the following basic question:

Is the orbit $x, Tx, T^2x, \dots$ equidistributed in X?

It is not even obvious what this question means – how should we define *equidistributed*? We will turn to this question later, but let us begin by considering some examples of the sorts of systems we will be looking at.

Circle rotations. Let $X = \mathbb{R}/\mathbb{Z}$, and define the transformation $T : X \to X$ by $Tx = x + \alpha (\mathrm{mod}\, 1)$.

Torus rotations. A similar example works in higher dimensions: let $X = [0,1)^d$, write $\alpha = (\alpha_1, \dots, \alpha_d)$, and define $T : X \to X$ by $Tx = x + \alpha$.

Doubling map. Let $X = [0,1)$, and define the transformation $T : X \to X$ by $Tx = 2x (\mathrm{mod}\, 1)$. Similarly, if $k \in \mathbb{N}$, we may define the map $Tx = kx (\mathrm{mod}\, 1)$.

Bernouilli shift. Let X be the set of two sided 0-1 sequences $\vec{x} = (x_n)_{n \in \mathbb{Z}}$, and define $T : X \to X$ by $(T\vec{x})_n = x_{n+1}$. We may also consider the one-sided variant of this in which X is not the space of 0-1 sequences $(x_n)_{n=0}^{\infty}$ and we define T in the same way, thus $T(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$.

The Gauss map. Let $X = [0,1]$. If $x \in (0,1]$ then define $Tx := \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$, where $\lfloor t \rfloor$ denotes the integer part of t. Define $T(0) = 0$ (the inclusion of the point 0 is just to make the underlying space compact). The map T and its iterates map be used to compute the continued fraction expansion of $x \in (0,1)$. Indeed if $a_1, a_2, \dots$ are natural numbers and

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3} + \dots}} \quad \text{then} \quad Tx = \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4} + \dots}}$$

and so the nth partial quotient a_n can be recovered as $a_n = \lfloor 1/T^{n-1}x \rfloor$. Note that, in contrast to the earlier examples, the map T is not continuous with respect to the natural topology on $[0,1)$ (it has dramatic discontinuities at $x = \frac{1}{2}, \frac{1}{3}, \ldots$).

As a warm-up for the rest of the course, and to introduce some techniques we will use later, we begin by studying the first example, circle rotations, from first principles in a little more detail. In the case $X = \mathbb{R}/\mathbb{Z}$, there is a natural definition of what we should mean by saying that the orbit $(T^n x)_{n=0}^{\infty}$ is equidistributed.

Definition 1.1. We say that the orbit $(T^n x)_{n=0}^{\infty}$ is equidistributed in intervals if

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \in \{0, 1, \ldots, N-1\} : T^n x \in I\} \to \text{length}(I)$$

for all intervals $I \subset \mathbb{R}/\mathbb{Z}$.

Let us say for the sake of argument that our intervals are closed, but this is of no consequence later on. Here is the basic theorem about circle rotations and equidistribution in intervals.

Theorem 1.1. *Suppose that $Tx = x + \alpha (\mathrm{mod}\, 1)$ is a circle rotation. Then*

- *If $\alpha \in \mathbb{Q}$ then $(T^n x)_{n=0}^{\infty}$ is not equidistributed in intervals for any x;*
- *If $\alpha \notin \mathbb{Q}$ then $(T^n x)_{n=0}^{\infty}$ is equidistributed in intervals for every x.*

Proof. The first part is quite obvious: if $\alpha = \frac{a}{q}$ then $T^n x = x + \frac{an}{q}$ and so the orbit $(T^n x)_{n=0}^{\infty}$ consists of at most q points, separated by at least $1/q$. Thus there are intervals of length $1/2q$ (say) which are never visited by this orbit, which cannot therefore be equidistributed in intervals.

The second part is much less obvious. Suppose that $\alpha \notin \mathbb{Q}$. A slightly fancier way of writing what it means for $(T^n x)_{n=0}^{\infty}$ to be equidistributed in intervals is as follows: for all $x \in \mathbb{R}/\mathbb{Z}$ and for all functions f of the form $f = 1_I$ (defined by $f(x) = 1$ if $x \in I$ and 0 otherwise) we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \to \int_0^1 f(t)dt.$$

More succinctly,

$$S_N f(x) \to \int_0^1 f(t)dt, \tag{1.1}$$

where

$$S_N f(x) := \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x).$$

If this holds with the convergence of $S_N f(x)$ to $\int_0^1 f(t)dt$ being uniform in x we say that the function $f = 1_I$ has the "time averages = space averages" property (TASA[1]).

It turns out that the indicator functions $f = 1_I$ are not the easiest functions for proving TASA. It is much easier to prove that the exponential functions $f(x) = e^{2\pi ikx}$, $k \in \mathbb{Z}$, are TASA. The proof goes by summing a geometric progression, first noting that the case $k = 0$ is trivial. If $k \neq 0$ then we have

$$S_N f(x) = \frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi ik(x+n\alpha)} = \frac{e^{2\pi ikx}}{N}\frac{1-e^{2\pi ikN\alpha}}{1-e^{2\pi ik\alpha}}.$$

Here we used the fact that α is irrational: this guarantees that $k\alpha \notin \mathbb{Z}$ and hence that the denominator here is not zero. Therefore

$$|S_N f(x)| \leqslant \frac{2}{N}\frac{1}{|1-e^{2\pi ik\alpha}|},$$

a quantity which tends to 0 as $N \to \infty$ uniformly in x. Since

$$\int_0^1 f(t)dt = \int_0^1 e^{2\pi ikt}dt = 0,$$

we have $S_N f(x) \to \int_0^1 f(t)dt$ uniformly in x, and so the exponential functions have TASA.

[1]This is not standard terminology.

If f_1, f_2 have TASA then it is very easy to see that $c_1 f_1 + c_2 f_2$ does also, and so any finite sum of exponentials or *trigonometric polynomial* has the TASA property.

Next we prove that the set of functions with the TASA property is closed under uniform convergence: that is, if $f_1, f_2, f_3, \ldots$ is a sequence of functions with TASA for which $\sup_{x\in\mathbb{R}/\mathbb{Z}} |f_j(x) - f(x)| \to 0$ then f is also TASA.

To see this, let $\varepsilon > 0$. Then there is some i such that we have

$$\sup_t |f_i(t) - f(t)| < \varepsilon/3. \tag{1.2}$$

Since f_i has TASA, for N large enough we have

$$|S_N f_i(x) - \int_0^1 f_i(t)dt| < \varepsilon/3$$

for all $x \in \mathbb{R}/\mathbb{Z}$. However (1.2) and the triangle inequality easily implies that

$$|S_N f_i(x) - S_N f(x)| < \varepsilon/3$$

for all x, and it also implies that

$$|\int_0^1 f_i(t)dt - \int_0^1 f(t)dt| < \varepsilon/3.$$

Therefore by the triangle inequality we have

$$|S_N f(x) - \int_0^1 f(t)dt| < \varepsilon.$$

Since ε was arbitrary, it follows that $S_N f(x) \to \int_0^1 f(t)dt$, or in other words that f has TASA.

It is a well-known fact that the trigonometric polynomials are dense in the space $C(X)$ of continuous functions on X, with the topology induced by the uniform norm, or in other words that every continuous $f : X \to \mathbb{R}$ may be uniformly approximated by trigonometric polynomials. This follows from the Stone-Weierstrass theorem, or it may be proven directly: see Appendix B for a direct argument. Thus, by the above comments, every continuous function on $\mathbb{R}/\mathbb{Z}$ is TASA.

Unfortunately, the characteristic functions $f(t) = 1_I(t)$ are not continuous. However, for any $\varepsilon > 0$ we may find continuous functions $\chi_\varepsilon^-, \chi_\varepsilon^+$ such that $\chi_\varepsilon^-(t) \leqslant 1_I(t) \leqslant \chi_\varepsilon^+(t)$ pointwise for all $t \in \mathbb{R}/\mathbb{Z}$ and such that

$$\int_0^1 \chi_\varepsilon^-(t)dt > \text{length}(I) - \varepsilon$$

and

$$\int_0^1 \chi_\varepsilon^+(t)dt < \text{length}(I) + \varepsilon.$$

For example, we could take χ_ε^+ to look like this:

and χ_ε^- to be this function:

Hence, if N is large enough, for all $x \in \mathbb{R}/\mathbb{Z}$ we have

$$S_N f(x) \leqslant S_N \chi_\varepsilon^+(x) \leqslant \varepsilon + \int_0^1 \chi_\varepsilon^+(t)dt < 2\varepsilon + \text{length}(I)$$

and

$$S_N f(x) \geqslant S_N \chi_\varepsilon^-(x) \geqslant -\varepsilon + \int_0^1 \chi_\varepsilon^+(t)dt > -2\varepsilon + \text{length}(I).$$

Since $\varepsilon > 0$ was arbitrary, the result follows. □

Time averages $S_N f$. We wish to highlight a piece of notation introduced in this proof which will reappear throughout the course: namely the definition

$$S_N f(x) := \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$$

to describe the "time averages" of f along the orbit $(T^n x)_{n=0}^\infty$ starting from x.

To conclude this section we wish to highlight, for future reference, a key result established in the course of the proof of Theorem 1.1.

PROPOSITION 1.1. *Suppose that $X = \mathbb{R}/\mathbb{Z}$ and that $T : X \to X$ is the circle rotation given by $Tx = x + \alpha (\text{mod}\, 1)$. Suppose that $\alpha \notin \mathbb{Q}$. Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ be a continuous function. Then*

$$S_N f(x) \to \int_0^1 f(t)dt$$

as $N \to \infty$, this convergence being uniform in $x \in \mathbb{R}/\mathbb{Z}$ for each fixed f.

In the course of the proof of Theorem 1.1 this was the property that we described by saying that "every continuous f has TASA", but we will not use this nonstandard nomenclature in later chapters.

CHAPTER 2

Measure-preserving systems

2.1. Probability spaces

In the introduction we discussed the circle rotation system in which $X = \mathbb{R}/\mathbb{Z}$ and $T : X \to X$ is given by $Tx = x + \alpha (\bmod 1)$. We proved a result about when orbits $(T^n x)_{n=0}^{\infty}$ are equidistributed. We did not look in any detail at the other examples, and in some cases (for example the Bernouilli shift) it is less obvious what it should even mean for an orbit to be equidistributed.

One can easily guess a rough form for the definition.

Rough Definition. An orbit $(T^n x)_{n=0}^{\infty}$ is equidistributed if

$$\lim_{N \to \infty} \frac{1}{N} \#\{n \in \{0, 1, \ldots, N-1\} : T^n x \in E\} = \mathrm{vol}(E)$$

for all "nice" sets $E \subset X$.

In the circle rotation example, we took the nice sets to be intervals, and then it was obvious how to define their volume: the volume $\mathrm{vol}(E)$ of an interval E is $\mathrm{length}(E)$.

The right framework for understanding these concepts in much greater generality is *measure theory*. In this course, all the spaces X we will be studying will be *probability spaces*. A probability space is a set X together with a collection $\mathscr{B}$ of subsets of X which we know the volume, or *measure*, of, and this measure will always lie in $[0, 1]$. We write $\mu : \mathscr{B} \to [0, 1]$ for the function which assigns to a set $A \in \mathscr{B}$ its measure. The function μ is itself called a measure.

The collection $\mathscr{B}$ is required to be a σ-algebra, which means that it contains the empty set $\emptyset$ and X, and it is closed under complements, countable intersections and countable unions. To spell it out:

- We have $\emptyset, X \in \mathscr{B}$;
- If $A \in \mathscr{B}$ then $X \setminus A \in \mathscr{B}$;
- If $A_1, A_2, \dots \in \mathscr{B}$ then $\bigcap_{n=1}^{\infty} A_n \in \mathscr{B}$;
- If $A_1, A_2, \dots \in \mathscr{B}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathscr{B}$.

The measure μ is required to interact with $\mathscr{B}$ in a pleasant way. Specifically, for μ to qualify as a probability measure we must have, whenever $A, A', A_1, \ldots, A_n \in \mathscr{B}$:

- We have $\mu(\emptyset) = 0$ and $\mu(X) = 1$;

- (Additivity) If A, A' are disjoint then $\mu(A \cup A') = \mu(A) + \mu(A')$;
- (Limits) If $A_1 \subset A_2 \subset A_3 \subset \dots$ then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} A_n$ and if $A_1 \supset A_2 \supset A_3 \supset \dots$ then $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

In fact, each of the two limit conditions is easily seen to imply the other on taking complements. Often, the limit condition is phrased differently as a condition called "countable additivity": if $E_1, E_2, \dots$ are disjoint subsets of $\mathscr{B}$ then $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$. This corresponds exactly to the first limit condition, taking $A_1 = E_1$, $A_2 = E_1 \cup E_2$, $A_3 = E_1 \cup E_2 \cup E_3$, or in the other direction $E_1 = A_1$, $E_2 = A_2 \setminus A_1$, $E_3 = A_3 \setminus A_2$, and so on.

The standard convention, which we will follow in this course, is to write $(X, \mathscr{B}, \mu)$ for a probability space together with its associated σ-algebra $\mathscr{B}$ and measure $\mu : \mathscr{B} \to [0,1]$.

How does this fit with the discussion of the previous chapter? Let us consider first of all the circle rotation on $X = \mathbb{R}/\mathbb{Z}$, or the doubling map on the same space. Here $\mu((a,b)) = b - a$ for every open interval (a,b). What is the σ-algebra $\mathscr{B}$? It is not simply the set of all intervals (a,b), since this is not closed under taking countable complements or unions. To be closed under these operations, $\mathscr{B}$ must also contain all open subsets of $[0,1]$, and then all countable intersections of these sets (the so-called F_σ-sets), then all countable unions of these sets, and so on. For the purposes of this course, it is best not to think too explicitly about it.

Definition 2.1. The *Borel algebra* $\mathscr{B}$ on $\mathbb{R}/\mathbb{Z}$ is the smallest σ-algebra containing the open intervals, or in other words the σ-algebra *generated by* the open intervals. More generally, if X is a metric space then the Borel algebra is the smallest σ-algebra containing the open balls $\{y : d(x,y) < \varepsilon\}$.

The Borel σ-algebra on $\mathbb{R}/\mathbb{Z}$ (or on any metric space) does exist, because the intersection of any two σ-algebras is again a σ-algebra. Thus we may define it to be the intersection of all σ-algebras containing the open intervals.

In the introduction, we only considered the measure μ of intervals. However, a remarkable fact is that μ extends to a measure on all of $\mathscr{B}$. This measure is known as *Lebesgue measure*, and it can be defined by first defining the measure $\mu(U)$ of an open set U by setting

$$\mu(U) = \sum_{j=1}^{\infty} \mu(I_j),$$

where $\bigcup_{j=1}^{\infty} I_j$ is the decomposition of U as a disjoint union of countably many open intervals (you may wish to recall why there is such a decomposition, or see Appendix A). Then for an arbitrary Borel set E we define

$$\mu(E) = \inf_{U \supset E} \mu(U),$$

where the infimum is taken over all *open* sets U containing E. We will not really need to know any of the details of the proof that this definition works, which are given in the Part A course *Integration*. For the most part, in this course, we hardly need to know the definition. We *do* make repeated use of one important consequence of it, namely the following statement that "every Borel set is almost an open set".

LEMMA 2.1 (Regularity of Lebesgue measure). *Let E be a Borel set and let $\varepsilon > 0$. Then there is an open set $U \supset E$ such that $\mu(U \setminus E) < \varepsilon$.*

This allows one to approximate measurable sets E by "simple" sets such as finite unions of intervals in various ways. For a general statement of this type see Lemma A.2.

What about a different example, such as the one-sided Bernouilli shift? Here, we took $X = \{0,1\}^{\mathbb{N}}$. This space X is one that we are less used to dealing with, but it does seem intuitively reasonable that the measure $\mu(E)$ of any *cylinder set* E of the form

$$E = \{(x_1, x_2, \dots) : x_{i_1} = \varepsilon_1, \dots, x_{i_k} = \varepsilon_k\} \tag{2.1}$$

(that is to say, the set of sequences with k fixed digits) should be 2^{-k}.

The collection of cylinder sets is not a σ-algebra, but again there is some smallest σ-algebra $\mathscr{B}$ containing the cylinder sets. Once again, the measure μ extends to a measure on $\mathscr{B}$.

Both this fact and the existence of Lebesgue measure are consequences of a general result from measure theory called the *Carathéodory Extension Theorem*. We will state it in Appendix A, though for this course you do not really need to know either the statement or the proof. Note that there is some confusion in the literature about exactly what this theorem states – for example in the 2014 notes from Part A integration [**2**] this term is used for a more specific result about Lebesgue measure (though the general result is very closely related).

2.2. Measure-preserving transformations

We have talked in some generality about spaces X and measures on them. In the examples of the introduction we also had maps $T : X \to X$. What kind of maps were these? They were not always continuous with respect to some natural topology – for example the Gauss map is discontinuous at infinitely many points. These were examples of *measure-preserving* maps.

DEFINITION 2.2. Suppose that $(X_1, \mathscr{B}_1, \mu_1)$ and $(X_2, \mathscr{B}_2, \mu_2)$ are two probability spaces. If $T : X_1 \to X_2$ is a map, we say that T is *measurable* if $T^{-1}A \in \mathscr{B}_1$ for all $A \in \mathscr{B}_2$.

Here, $T^{-1}A$ denotes the set of all $x_1 \in X_1$ for which $Tx_1 \in A$. The definition of measurable should be compared to the definition of continuity in a topological space (note, however, that probability spaces do not automatically come with a topological structure and so it does not necessarily make sense to talk about continuous functions in this context). This notion was also covered in the Part A course on integration.

DEFINITION 2.3. Suppose that $(X_1, \mathscr{B}_1, \mu_1)$ and $(X_2, \mathscr{B}_2, \mu_2)$ are two probability spaces and that $T : X_1 \to X_2$ is a map. Then we say that T is *measure-preserving* if it is measurable and if $\mu_1(T^{-1}A) = \mu_2(A)$ for all $A \in \mathscr{B}_2$.

The main object of study in this course will be measure-preserving maps from a probability space $(X, \mathscr{B}, \mu)$ to itself. The quadruple $(X, \mathscr{B}, \mu, T)$ is called a *measure-preserving system.*

In the next section we revisit the examples from the introduction and see that they are indeed measure-preserving systems. In order to do this, we need to be able to check whether a map $T : X \to X$ is measure-preserving. Thankfully, to decide whether this is so we do not have to look at $T^{-1}(A)$ for an arbitrary $A \in \mathscr{B}$, as the following lemma shows.

LEMMA 2.2. *Suppose that $(X_1, \mathscr{B}_1, \mu_1)$, $(X_2, \mathscr{B}_2, \mu_2)$ are probability spaces and that $T : X_1 \to X_2$ is a map. Suppose that $\mathscr{B}_2$ is generated, as a σ-algebra, by a collection $\mathscr{A} \subset \mathscr{B}_2$. Suppose that for all $A \in \mathscr{A}$ we have $T^{-1}A \in \mathscr{B}_1$ and $\mu_1(T^{-1}A) = \mu_2(A)$. Then T is measure-preserving.*

PROOF. The set of all $E \subset \mathscr{B}_2$ for which $T^{-1}E \in \mathscr{B}_1$ and $\mu_1(T^{-1}E) = \mu_2(E)$ is closed under countable unions, intersections and complementation. Since we know this is true for all $E \in \mathscr{A}$, it also holds for all E in the σ-algebra generated by $\mathscr{A}$, that is to say $\mathscr{B}_2$. □

2.3. Examples

Let us revisit the examples of the introduction and explain why they are measure-preserving systems.

Circle rotations. Take $X = \mathbb{R}/\mathbb{Z}$, define $T : X \to X$ by $Tx = x + \alpha (\bmod 1)$, take $\mathscr{B}$ to be the σ-algebra of Borel sets and let μ be Lebesgue measure. Then $(X, \mu, \mathscr{B}, T)$ is a measure-preserving system. By Lemma 2.2, all we need check is that $T^{-1}((a, b))$ is measurable and has measure $b - a$; this, however, is obvious.

Doubling map. Let X, μ and $\mathscr{B}$ be as in the above example, but now take $Tx = 2x (\bmod 1)$. We have $T^{-1}((a, b)) = (\frac{a}{2}, \frac{b}{2}) \cup (\frac{1}{2} + \frac{a}{2}, \frac{1}{2} + \frac{b}{2})$, a union of two intervals with total measure $b - a$. (Here, the meaning of $\frac{a}{2}$ is ambiguous: we can take it to mean the unique element in $[0, \frac{1}{2})$ for which $2 \cdot \frac{a}{2} \equiv a (\mathrm{mod}\, 1)$.) Since the open intervals generate $\mathscr{B}$, it follows from Lemma 2.2 that T is measure-preserving.

Let us draw attention to an important point: in the case of the doubling map, we *do not* have $\mu(TE) = \mu(E)$ for all measurable E. Indeed, if $E = [0, \frac{1}{2}]$ then $TE = [0, 1]$ and so $\mu(TE) = 2\mu(E)$.

Bernouilli shift. Consider the one-sided Bernouilli shift, with $X = \{0, 1\}^{\mathbb{N}}$ and $(T\vec{x})_n = x_{n+1}$ where $\vec{x} = (x_n)_{n=1}^{\infty}$. Let $\mathscr{B}$ be the σ-algebra on X generated by the cylinder sets. If

$$E = \{(x_1, x_2, \dots) : x_{i_1} = \varepsilon_1, \dots, x_{i_k} = \varepsilon_k\}$$

is such a set then

$$T^{-1}E = \{(x_1, x_2, \dots) : x_{i_1+1} = \varepsilon_1, \dots, x_{i_k+1} = \varepsilon_k\}.$$

By definition of the measure μ on X, both of these sets have measure 2^{-k}. Thus $\mu(E) = \mu(T^{-1}E)$ for all cylinder sets E, and by Lemma 2.2 we see that T is measure-preserving.

Gauss map. Set $X = [0, 1]$. Recall that the Gauss map $T : X \to X$ is defined by $T(x) = \{1/x\}$ if $x \neq 0$ and $T(0) = 0$, where $\{t\}$ denotes the fractional part of t. The point 0 is only included to make the underlying space X compact, and plays no role in what follows. The Gauss map T is *not* measure-preserving with respect to the Lebesgue measure μ. Indeed,

$$T^{-1}([0, \frac{1}{2}]) = \bigcup_{n=1}^{\infty} [\frac{2}{2n+1}, \frac{1}{n}],$$

and so

$$\mu(T^{-1}([0, \frac{1}{2}]) = 2(\frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \dots) = 2 - 2\log 2 \neq \frac{1}{2}.$$

It turns out, however, that T is measure-preserving with respect to a different measure ν, the so-called Gauss measure. This is defined by[1]

$$\nu(E) := \frac{1}{\log 2} \int_E \frac{d\mu(x)}{1+x}.$$

LEMMA 2.3 (Gauss map preserves the Gauss measure). *Let $X = [0, 1]$, $\mathscr{B}$ be the Borel σ-algebra, $T : X \to X$ the Gauss map and ν the Gauss measure. Then $(X, \nu, \mathscr{B}, T)$ is a measure-preserving system.*

PROOF. By Lemma 2.2 it suffices to prove that $\nu(T^{-1}(a, b)) = \nu((a, b))$ for all $0 < a < b < 1$. This is just simple calculus, combined with a small amount of

[1]The integral here is a Lebesgue integral. We remind the reader what this means in the next chapter. In the case $E = (a, b)$ this is simply the Riemann integral $\int_a^b \frac{dx}{1+x}$.

thought. Indeed the inverse image $T^{-1}((a,b))$ is the union $\bigcup_{n=1}^{\infty}\big(\frac{1}{n+b},\frac{1}{n+a}\big)$ and so we have

$$\begin{aligned}\log 2\cdot\nu(T^{-1}(a,b)) &= \sum_{n=1}^{\infty}\int_{1/(n+b)}^{1/(n+a)}\frac{dx}{1+x}\\ &= \sum_{n=1}^{\infty}\Big(\log\big(\frac{1}{n+a}+1\big)-\log\big(\frac{1}{n+b}+1\big)\Big)\\ &= \lim_{N\to\infty}\sum_{n=1}^{N}\Big(\log(\frac{n+a+1}{n+a}-\log(\frac{n+b+1}{n+b})\Big)\\ &= \lim_{N\to\infty}\Big(\log(1+b)-\log(1+a)+\log(\frac{N+a+1}{N+b+1})\Big)\\ &= \log(1+b)-\log(1+a)\\ &= \log 2\cdot\nu((a,b)).\end{aligned}$$

□

2.4. Invariant measures

In what we have said so far, we have largely taken the view that the probability space $(X,\mu,\mathscr{B})$ has been fixed and we have been interested in transformations $T:X\to X$ that are measure-preserving, leading to the measure-preserving system $(X,\mu,\mathscr{B},T)$. It is very useful to consider an alternate perspective, in which X, T and the σ-algebra $\mathscr{B}$ are fixed and we are at liberty to choose the probability measure $\mu:\mathscr{B}\to[0,1]$ so that T is a measure-preserving transformation with respect to μ. We saw this point of view when we discussed the Gauss map $T:[0,1]\to[0,1]$, which was *not* measure-preserving for the Lebesgue measure, but was measure-preserving for a different measure (the Gauss measure).

When looking at things this way around, we say that μ is an *invariant probability measure* for T. If X is a metric space and $\mathscr{B}$ the σ-algebra of Borel sets, we say that μ is an invariant Borel probability measure.

There may well not be a unique choice of an invariant probability measure μ: some examples are presented on the first exercise sheet.

2.5. Poincáre recurrence theorem

We conclude this chapter by proving a basic theorem about measure-preserving systems in general, the *Poincaré recurrence theorem.*

We begin by with some remarks on the terms *almost all* and *measure zero.* A set of measure zero is a set $A\in\mathscr{B}$ for which $\mu(A)=0$. Note that this absolutely does *not* imply that A is empty. For example, if $X=[0,1]$ with Lebesgue measure then all singletons $\{x\}$ have measure zero, as do all countable sets and even some uncountable sets (such as the Cantor set). We say that some property P of points of X holds for almost every x if the set $\{x \in X : \neg P(x)\}$ is measurable and has measure zero.

THEOREM 2.1 (Poincáre Recurrence Theorem). *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. Let $A \in \mathscr{B}$ and let $A' \subset A$ be the set of $x \in A$ for which there are infinitely many $n \geqslant 1$ with $T^n x \in A$. Then $\mu(A \setminus A') = 0$, or in other words almost all points in A recur back to A.*

PROOF. Write

$$S_N := \bigcup_{n \geqslant N} T^{-n} A$$

for the set of $x \in X$ for which $T^n x \in A$ for some $n \geqslant N$, and

$$S := \bigcap_N S_N$$

for the set of x for which $T^n x \in A$ for infinitely many n. Note that S is measurable. Then $A' = A \cap S$. Now we have the nesting $S_0 \supset S_1 \supset \dots$, and by the measure-preserving nature of T we have $\mu(S_{N+1}) = \mu(T^{-1} S_N) = \mu(S_N)$. By the limit principle it follows that $\mu(S) = \lim_{N \to \infty} \mu(S_N) = \mu(S_0)$, and so $\mu(S_0 \setminus S) = 0$. But $A \subset S_0$, and therefore $\mu(A \setminus S)$ is zero as well. □

Another consequence of the Poincaré Recurrence Theorem is as follows: for any $A \in \mathscr{B}$,

$$\mu(A \setminus \bigcup_{n=1}^{\infty} T^{-n} A) = 0.$$

CHAPTER 3

Ergodic transformations

In the last chapter we introduced the notion of a measure-preserving system $(X, \mathscr{B}, \mu, T)$. In this chapter we will take a look a specific property that such a system may enjoy: that of *ergodicity*. Ergodic transformations are, roughly speaking, those for which the orbit $(T^n x)_{n=0}^{\infty}$ is almost always equidistributed on X.

3.1. The definition of ergodicity

To understand this more precisely, we need to know what "equidistributed" means. In the case $X = \mathbb{R}/\mathbb{Z}$, we said that this was so if

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \in \{0, 1, \dots, N-1\} : T^n x \in I\} \to \text{length}(I)$$

for all (closed) intervals $I \subset \mathbb{R}/\mathbb{Z}$.

In the more general setting of measure-preserving systems, there are no sets playing the distinguished role of intervals – we have only the sets E in $\mathscr{B}$.

Definition 3.1. Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system, and let $E \in \mathscr{B}$. Then we say that the orbit $(T^n x)_{n=0}^{\infty}$ equidistributes in E if

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \in \{0, 1, \dots, N-1\} : T^n x \in E\} \to \mu(E)$$

With this notion in hand, we may make the following further definition.

Definition 3.2. Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. Then we say that T is *equidistributing* if, for every $E \in \mathscr{B}$, the orbit $(T^n x)_{n=0}^{\infty}$ equidistributes in E for almost every $x \in X$.

Note that the order of the quantifiers here is important: we are not asserting that for almost every $x \in X$ it is the case that $(T^n x)_{n=0}^{\infty}$ equidistributes in E for all $E \in \mathscr{B}$. In fact, this could never be the case in any of the examples considered in Chapter 2: for any x, we can take $E = \{x, Tx, T^2x, \dots\}$, and if we assume that the orbit $(T^n x)_{n-0}^{\infty}$ equidistributes in E then $\mu(E) = 1$. However, in all the examples the measure of any point was 0 and so, by countable additivity, $\mu(E) = 0$.

The notion of a transformation T being equidistributing is not standard. Transformations with this property are usually called *ergodic*. The usual definition of ergodic, however, looks rather different.

DEFINITION 3.3 (Ergodicity). Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. By an *invariant set* we mean a set $E \in \mathscr{B}$ with the property that $T^{-1}E$ and E differ in a set of measure 0. We say that T is *ergodic* if all invariant sets have measure either 0 or 1.

Many texts say that T is ergodic all *strictly* invariant sets E, that is to say sets with $T^{-1}E = E$, have measure 0 or 1. Ostensibly this is a weaker notion, but in fact the two notions are equivalent. This is not quite trivial – see Examples 1, Ex. 9. In this course we will allow $T^{-1}E = E$ to differ in a set of measure zero, as all the tools we use work most naturally up to a.e. equivalence.

Another way to express the invariance of E is to say that $1_E(x) = 1_E(Tx)$ for a.e. x. It is an easy exercise to show that if E is invariant then, for every $n \geqslant 1$, the sets $T^{-n}E$ and E differ in a set of measure 0, and that $1_E(x) = 1_E(T^n x)$ for almost every x. Moreover, since a countable union of sets of measure zero has measure zero, we can assert that in fact for a.e. x we have $1_E(x) = 1_E(T^n x)$ for all n, and that we have

$$E = \bigcup_{n=1}^{\infty} T^{-n}E = \bigcap_{n=1}^{\infty} T^{-n}E \tag{3.1}$$

up to measure 0. (Note that these facts would be completely trivial if invariance were replaced by strict invariance.)

Here is a simple lemma about the relationship between ergodic and equidistributing transformations.

LEMMA 3.1. *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. If T is equidistributing, then it is also ergodic.*

PROOF. Suppose that T is equidistributing. Let $E \in \mathscr{B}$ be an invariant set, and suppose that $\mu(E) > 0$. Set

$$S_E := \{x \in X : (T^n x)_{n=0}^{\infty} \text{ equidistributes in } E\}.$$

By assumption, $\mu(S_E) = 1$. Since $\mu(E) > 0$, it follows from (3.1) that

$$\mu(\bigcap_{n=1}^{\infty} T^{-n}E) > 0,$$

and therefore S_E and $\bigcap_{n=1}^{\infty} T^{-n}E$ intersect. Let x be a point in the intersection. The fact that $x \in \bigcap_{n=1}^{\infty} T^{-n}E$ implies that $x, Tx, T^2x, T^3x, \cdots \in E$, and therefore

$$\frac{1}{N}\#\{n \in \{0, 1, \dots, N-1\} : T^n x \in E\} = 1.$$

On the other hand, since $x \in S_E$ we have

$$\frac{1}{N}\#\{n \in \{0, 1, \dots, N-1\} : T^n x \in E\} = \mu(E).$$

Therefore $\mu(E) = 1$. □

A way to think about this lemma is as follows: a relatively obvious obstruction to T being equidistributing is the existence of a nontrivial invariant set E, where nontrivial means that $\mu(E) \neq 0, 1$.

It is a remarkable fact that this is the *only* obstruction: that is, an ergodic transformation is also equidistributing. Thus being equidistributing and being ergodic are one and the same concept (which is why the term *equidistributing* is not standard). This is known as the Birkhoff ergodic theorem, and we will prove it in the next chapter.

Whether or not a map $T : X \to X$ is ergodic is not an intrinsic property of T, but also depends on the measure μ and the σ-algebra $\mathscr{B}$ (see, for example, Examples 1, Ex. 8). Thus, in a sense, it is an abuse of nomenclature to say that T is ergodic. The word *ergodic* is also used for the whole system $(X, \mu, \mathscr{B}, T)$, this system being said to be ergodic if T is ergodic with respect to μ. As with the notion of *invariance* discussed in the last chapter, we often switch the emphasis from T to μ. Thus a measure μ is said to be ergodic (for a transformation $T : X \to X$) if μ is T-invariant and if the system $(X, \mu, \mathscr{B}, T)$ is ergodic.

3.2. Ergodicity and recurrence

Let $(X, \mu, \mathscr{B}, T)$ be a measure-preserving system. To get a feel for what it means for this system to be ergodic, we prove a lemma giving a property about recurrence which is equivalent to ergodicity.

LEMMA 3.2. *A system is ergodic if and only if the following is true: For every measurable set A with $\mu(A) > 0$, for almost every x there is some n such that $T^n x \in A$. In particular if $\mu(B) > 0$ then there is some $b \in B$ such that $T^n b \in A$ for some $n \geqslant 1$.*

PROOF. Suppose that the system is ergodic. We claim that the set $A_+ := \bigcup_{n=1}^{\infty} T^{-n} A$ of points x for which some $T^n x$ lies in A is T-invariant. Certainly $T^{-1} A_+ \subset A_+$, and the difference between these two sets, $T^{-1} A \setminus (T^{-2} A \cup T^{-3} A \cup \dots)$, has measure zero by the Poincaré Recurrence Theorem (in the form given after the proof of that result at the end of Chapter 2). Since T is ergodic, we have $\mu(A_+) = 0$ or 1. The first possibility is absurd since $A_+ \subset T^{-1} A$ and $\mu(T^{-1} A) = \mu(A) > 0$, and therefore $\mu(A_+) = 1$, that is to say almost every point of X lies in A_+.

Conversely, suppose that the recurrence condition holds and that E is T-invariant, thus $T^{-1} E = E$ up to measure zero. Then $E_+ = E$ up to measure zero. However, we are assuming that if if $\mu(E) > 0$ then $\mu(E_+) = 1$. Therefore $\mu(E) = 0$ or 1.

□

3.3. A very quick refresher on integration and L^1

From now on in the course we will use a little more measure theory, specifically some notions of integration. Familiarity with the *proofs* of the statements we require is not important.

If $(X, \mathscr{B}, \mu)$ is a probability space and if $f : X \to \mathbb{R}$ is a function then we say that f is *measurable* if $f^{-1}((a, b)) \in \mathscr{B}$ for every open interval $(a, b) \subset \mathbb{R}$. This notion may be extended to complex-valued functions by taking real and imaginary parts, and it has good closure properties: if f_1, f_2 are measurable then so are $c_1 f_1 + c_2 f_2$ and $|f_i|$, for example. If f is real-valued and non-negative we may define $\int_X f d\mu$ by approximating f by simple measurable functions, that is to say functions of the form $g = \sum_{i \in I} c_i 1_{E_i}$ with I finite and the E_i disjoint, in which case we define $\int_X g d\mu$ to be the "obvious" quantity $\sum_i c_i \mu(E_i)$. We discuss this a little more carefully in Appendix A.

The integral $\int_X f d\mu$ may take the value ∞, but we distinguish the space

$$L^1(X, \mathscr{B}, \mu) := \{f : X \to \mathbb{C} : f \text{ measurable}, \int_X |f| d\mu < \infty\}$$

of integrable functions. This space is called L^1 because it is a special case of a more general construction of L^p-spaces; we will see the case $p = 2$ later on. If $f \in L^1(X, \mathscr{B}, \mu)$ then we may define the integral $\int_X f d\mu$ by splitting $f = (\Re f)_+ - (\Re f)_- + i(\Im f)_+ - i(\Im f)_-$ and appealing to the definition of the integral in the non-negative case. This integral will always be finite if $f \in L^1(X, \mathscr{B}, \mu)$. The map $f \mapsto \int_X f d\mu$ is a linear map from $L^1(X, \mathscr{B}, \mu)$ to $\mathbb{C}$ and has various pleasant limit properties such as the monotone and dominated convergence theorems; we will not make explicit use of these in this course. If $X = \mathbb{R}/\mathbb{Z}$ or $[0, 1]$ and f is continuous then $\int_X f d\mu$ is equal to the Riemann integral of f.

Usually, functions in $L^1(X, \mathscr{B}, \mu)$ which agree outside of a set of measure 0 are regarded as the same. With this quotienting convention in force,

$$\|f\|_1 := \int_X |f| d\mu$$

defines a norm on $L^1(X, \mathscr{B}, \mu)$. (The point about quotienting is that for $\| \cdot \|_1$ to qualify as a norm we must have $\|f\|_1 = 0$ if and only if $f = 0$, but this is only true up to a set of measure 0.)

One sometimes calls the functions f in $L^1(X, \mathscr{B}, \mu)$ "integrable". A useful fact for us will be the following statement about the behaviour of integrals under measure-preserving transformations.

LEMMA 3.3. *Suppose that $(X, \mu, \mathscr{B}, T)$ is a measure-preserving system and that $f \in L^1(X, \mathscr{B}, \mu)$. Then*

$$\int_X f d\mu = \int_X (f \circ T) d\mu,$$

that is to say

$$\int_X f(x) d\mu(x) = \int_X f(Tx) d\mu(x).$$

PROOF. *Since every $f \in L^1(X, \mathscr{B}, \mu)$ can be approximated arbitrarily closely (in the $\|\cdot\|_1$ norm) by simple measurable functions, it suffices check this statement when f is simple-measurable. This quickly reduces to checking the case $f = 1_E$, E measurable. However

$$\int_X 1_E d\mu = \mu(E),$$

whilst

$$\int_X (1_E \circ T) d\mu = \int_X 1_E(Tx) d\mu(x) = \mu(T^{-1}E).$$

The result now follows from the fact that T is measure-preserving. □

We remark that this lemma is "the reason" why the notion of measure-preserving is defined using T^{-1} rather than T.

To conclude this section we formally record, in the language of this section, a fact that we already used in the proof of Theorem 1.1.

LEMMA 3.4. *The time averages map S_N is a contraction on $L^1(X, \mathscr{B}, \mu)$. That is, if $f \in L^1(X, \mathscr{B}, \mu)$ then $\|S_N f\|_1 \leqslant \|f\|_1$.*

PROOF. This is simply the triangle inequality:

$$\|S_N f\| = \frac{1}{N} \| \sum_{n=0}^{N-1} f \circ T^n \| \leqslant \frac{1}{N} \sum_{n=0}^{N-1} \|f \circ T^n\|_1 = \|f\|_1.$$

□

3.4. Irrational rotations are ergodic

We are going to give three proofs that irrational circle rotations are ergodic, that is to say of the following theorem.

PROPOSITION 3.1 (Irrational circle rotations are ergodic). *Let $X = \mathbb{R}/\mathbb{Z}$, $\mathscr{B}$ the Borel σ-algebra, μ the Lebesgue measure and $T : X \to X$ the rotation $Tx = x + \alpha (\mathrm{mod}\, 1)$. Suppose that α is irrational. Then T is ergodic.*

The first two proofs are somewhat similar and should be thought of as "L^1" proofs. The third proof is more "L^2" and will be given later. Both of the first two proofs rely on the following fact.

LEMMA 3.5. *Let $E \subset \mathbb{R}/\mathbb{Z}$ be measurable (with respect to the Borel σ-algebra). Suppose that $E \in \mathscr{B}$ is a measurable set. Then for every $\varepsilon > 0$ there is a continuous function $f : \mathbb{R}/\mathbb{Z} \to [0,1]$ such that $\|f - 1_E\|_1 \leqslant \varepsilon$.*

PROOF. By the regularity property of Lebesgue measure there is an open set U with $E \subset U$ and $\mu(U \setminus E) \leqslant \varepsilon/3$. Now U is a countable union of disjoint open intervals $\bigcup_{j=1}^{\infty} I_j$. By the limit property of measures we have $\lim_{J\to\infty} \bigcup_{j=1}^{J} I_j = \mu(U)$, and so there is a set $U' \subset U$ with U' a finite union of open intervals and $\mu(U \setminus U') \leqslant \varepsilon/3$. Note that $\mu(E \triangle U') \leqslant 2\varepsilon/3$, and so $\|1_E - 1_{U'}\|_1 \leqslant 2\varepsilon/3$. Finally, we can find a continuous function $f : \mathbb{R}/\mathbb{Z} \to [0,1]$ such that $\|f - 1_{U'}\|_1 \leqslant \varepsilon/3$ by adapting the contruction used in the proof of Theorem 1.1 to approximate characteristic functions of intervals by continuous functions. In fact, if $U' = \bigcup_{j=1}^{J} I_j$ then we can take, in the notation of that theorem, $f = \sum_{j=1}^{J} \chi^{+}_{I_j,\varepsilon/J}$. □

We remark that a small elaboration of this allows one to prove that the continuous functions are dense in $L^1(X, \mathscr{B}, \mu)$: see Appendix A for further comment.

First proof of Proposition 3.1. Suppose that E is an invariant set, that is to say $T^{-1}E = E$ up to measure 0. Then, as remarked above, $1_E(T^n x) = 1_E(x)$ for all n and for a.e. x. Let $\varepsilon > 0$, and choose some continuous function $f : \mathbb{R}/\mathbb{Z} \to [0,1]$ such that $\|f - 1_E\|_1 \leqslant \varepsilon$. We showed in Chapter 1 that all continuous functions have TASA, by which we meant that

$$S_N f(x) \to \int_0^1 f(t)dt$$

as $N \to \infty$, uniformly in x. In particular if N is big enough then

$$\|S_N f - \int_0^1 f(t)dt\|_1 \leqslant \varepsilon.$$

Since S_N is a contraction in L^1 we have

$$\|S_N f - S_N 1_E\|_1 \leqslant \|f - 1_E\|_1 \leqslant \varepsilon,$$

and therefore

$$\|S_N 1_E - \mu(E)\|_1 \leqslant 3\varepsilon.$$

However, the fact that $1_E(T^n x) = 1_E(x)$ for a.e. x implies that $S_N 1_E(x) = 1_E(x)$ for a.e. x, and thus

$$\|1_E - \mu(E)\|_1 \leqslant 3\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\|1_E - \mu(E)\|_1 = 0$, and so $1_E(x) = \mu(E)$ for a.e. x. This forces $\mu(E) = 0$ or 1.

Second proof of Proposition 3.1. This is quite similar to the first, but is uses less information. Rather than the fact that continuous functions are TASA, we just use the fact that the orbit $(\alpha n)_{n=1}^{\infty}$ is dense in $\mathbb{R}/\mathbb{Z}$. Whilst this is a straightforward consequence of the main result of Chapter 1, it may also be proven in a more elementary fashion. See Exercise sheet 1. We begin as before. Suppose that E is an invariant set, and note that $1_E(T^n x) = 1_E(x)$ for all n and for a.e. x. Let $\varepsilon > 0$, and choose a continuous function f such that $\|1_E - f\|_1 \leqslant \varepsilon$. Since T^n is measure-preserving we have

$$\|(1_E - f) \circ T^n\|_1 \leqslant \varepsilon.$$

Since $1_E = 1_E \circ T^n$ a.e., this implies that

$$\|1_E - f \circ T^n\|_1 \leqslant \varepsilon.$$

By the triangle inequality, it follows that

$$\|f - f \circ T^n\|_1 \leqslant 2\varepsilon,$$

or, written out in full,

$$\int_0^1 |f(x) - f(x + n\alpha)| dx \leqslant 2\varepsilon$$

for all $n \in \mathbb{N}$. Now since $(n\alpha)_{n=1}^{\infty}$ is dense in $\mathbb{R}/\mathbb{Z}$, for any $t \in \mathbb{R}/\mathbb{Z}$ we may choose a sequence n_i of integers with $n_i\alpha \to t (\mathrm{mod}\, 1)$. It follows from this and the continuity of f that in fact

$$\int_0^1 |f(x) - f(x + t)| dx \leqslant 2\varepsilon$$

for all $t \in \mathbb{R}/\mathbb{Z}$. From this it follows that

$$\begin{aligned} \|f - \int f\|_1 &= \int \big| f(x) - \int f(x+t)\, d\mu(t) \big|\, d\mu(x) \\ &\leqslant \int |f(x) - f(x+t)|\, d\mu(t) d\mu(x) \\ &\leqslant 2\varepsilon. \end{aligned}$$

Thus we have $\|1_E - \int f\|_1 \leqslant 3\varepsilon$. This implies that

$$|\mu(E) - \int f| = \big| \int (1_E - \int f)\, d\mu \big| \leqslant 3\varepsilon,$$

and hence by the triangle inequality $\|1_E - \mu(E)\|_1 \leqslant 6\varepsilon$. However $\varepsilon > 0$ was arbitrary, and so $\|1_E - \mu(E)\|_1 = 0$ which implies that $1_E(x) = \mu(E)$ a.e. It follows that $\mu(E) = 0$ or 1.

Remark: we did not really use much about integration and L^1 in either of these proofs. All of the functions we were integrating in these two arguments were linear combinations of continuous functions and characteristic functions of sets.

3.5. A very quick refresher on L^2

The space $L^2(X, \mathscr{B}, \mu)$ is defined by

$$L^2(X, \mathscr{B}, \mu) = \{f : X \to \mathbb{C} : f \text{ measurable}, \int_X |f|^2 d\mu < \infty\}.$$

As with L^1, we identify functions which differ only in a set of measure zero, and then $\|f\|_2 := (\int_X |f|^2 d\mu)^{1/2}$ becomes a norm. The most important fact about L^2 is that it is a Hilbert space, that is to say a complete inner product space. We will need very little of the general theory of Hilbert spaces in this course: what we do need is summarised in Section 4.1

THEOREM 3.1. *$L^2(X, \mathscr{B}, \mu)$ is a Hilbert space, that is to say a complete inner product space when endowed with the inner product $\langle f, g\rangle = \int_X f\overline{g} d\mu$.*

PROOF. * That $L^2(X, \mathscr{B}, \mu)$ is a complex vector space, and that $\langle f, g\rangle$ is well-defined if $f, g \in L^2(X, \mathscr{B}, \mu)$, are both consequences of the Cauchy-Schwarz inequality. The deep part of this theorem is the assertion that $L^2(X, \mathscr{B}, \mu)$ is *complete*, which is sometimes called the Riesz-Fischer theorem. This will have been discussed in Part A *Integration* in the case $X = [0, 1]$. The proof in the general case – involving the monotone convergence theorem – is essentially the same. □

It follows from Lemma 3.3 that if $f \in L^2(X, \mathscr{B}, \mu)$, and if $T : X \to X$ is measure-preserving, then $f \circ T \in L^2(X, \mathscr{B}, \mu)$. We introduce the special notation U_T for the map which sends f to $f \circ T$: that is to say, $U_T f(x) = f(Tx)$. We only use this notation when working in L^2. This map[1] U_T is an *isometry* that is, $\|U_T f\|_2 = \|f\|_2$ for all f, and more generally

$$\langle U_T f, U_T g\rangle = \int_X f(Tx)\overline{g(Tx)} d\mu(x) = \int_X f(x)\overline{g(x)} d\mu(x) = \langle f, g\rangle.$$

Finally, we note that if X is a probability space then, by the Cauchy-Schwarz inequality,

$$\int_X |f| \leqslant (\int_X 1)^{1/2} (\int_X |f|^2)^{1/2} = (\int_X |f|^2)^{1/2},$$

or in other words $\|f\|_1 \leqslant \|f\|_2$. In particular,

$$L^2(X, \mathscr{B}, \mu) \subset L^1(X, \mathscr{B}, \mu).$$

[1]The letter U stands for *unitary*: a unitary map on a Hilbert space is an *invertible* isometry. Note that U_T will not always be unitary if T is not invertible. For example, if $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is the doubling map $Tx = 2x (\text{mod } 1)$ then U_T maps $L^2(\mathbb{R}/\mathbb{Z})$ into the space of functions satisfying $f(x) = f(x + \frac{1}{2})$, which is certainly a proper subspace of $L^2(\mathbb{R}/\mathbb{Z})$.

3.6. Irrational rotations are ergodic: an L^2 proof

We now give our third proof of Proposition 3.1, which may be thought of as a "Hilbert space" or L^2 proof of the proposition, though we use rather little about Hilbert spaces, and most of what we do use is embedded in the proofs of facts about Fourier series, for which see Appendix B. We use the notation U_T for the isometry on $L^2(X, \mathscr{B}, \mu)$ induced by T, introduced in the last section.

Given $f \in L^2$ and $r \in \mathbb{Z}$, we define the Fourier coefficient

$$\widehat{f}(r) := \langle f, e_r \rangle = \int_0^1 f(x)\overline{e_r(x)}\, dx,$$

where $e_r(x) := e^{2\pi i r x}$. This is well-defined since if $f \in L^2$ then certainly $f \in L^1$, as remarked at the end of the last section. Now we have

$$\widehat{U_T f}(r) = \int_0^1 f(x+\alpha)e^{-2\pi i r x}dx = e^{2\pi i r \alpha}\int_0^1 f(x+\alpha)e^{-2\pi i r(x+\alpha)}dx = e^{2\pi i r\alpha}\hat{f}(r).$$

Suppose that E is T-invariant. Then $1_E = U_T 1_E$ a.e., and so $\widehat{1_E}(r) = \widehat{U_T 1_E}(r)$ for all $r \in \mathbb{Z}$. By the preceding computation, this implies that

$$\widehat{1_E}(r) = e^{2\pi i r\alpha}\widehat{1_E}(r)$$

for all $r \in \mathbb{Z}$. Since $\alpha \notin \mathbb{Q}$, we have $e^{2\pi i r \alpha} \neq 1$ when $r \neq 0$, and so $\widehat{1_E}(r) = 0$ whenever $r \neq 0$. It follows that $f := 1_E - \mu(E)$ has $\hat{f}(r) = 0$ for *all* integers r. By a standard fact from Fourier analysis (uniqueness of Fourier coefficients, see Appendix B) this implies that $f = 0$ a.e., and hence $1_E(x) = \mu(E)$ a.e. This implies that $\mu(E) = 0$ or 1, and so T is indeed ergodic.

3.7. The doubling map is ergodic

We have given three proofs that irrational circle rotations are ergodic, but in a sense we have learned little more than we knew already in the first chapter. However, we have acquired some techniques, and we now use these to prove that the doubling map is ergodic and then, in the next section, to show that the Gauss map is ergodic.

PROPOSITION 3.2 (Doubling map is ergodic). *The doubling map $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is ergodic with respect to the Lebesgue measure μ.*

We give two proofs, an "L^1-style proof" and an "L^2-style proof".

L^1 proof. Let $D_{a,n}$ be a "standard dyadic interval at scale n", that is to say an interval of type $(\frac{a}{2^n}, \frac{a+1}{2^n})$, $a \in \mathbb{Z}$. If E is any measurable set then it is not hard to check that $\mu(T^{-n}E \cap D_{a,n}) = 2^{-n}\mu(E)$. Thus if E is T-invariant, so that $T^{-n}E = E$ for all n, then $\mu(E \cap D_{a,n}) = 2^{-n}\mu(E)$ (or in other words the relative

density of E on $D_{a,n}$ is $\mu(E)$). It follows that if $\tilde{U}$ is any finite union of standard dyadic intervals then $\mu(E \cap \tilde{U}) = \mu(E)\mu(\tilde{U})$. Now by the regularity of Lebesgue measure, for any $\varepsilon > 0$ there is an open set U with $E \subset U$ and $\mu(U \setminus E) \leqslant \varepsilon$. For each n write U_n for the union of all the standard dyadic intervals at scale n which are contained in U. We clearly have $U_1 \subset U_2 \subset \dots$, and since U is open we have $\bigcup_n U_n = U$. By the limit principle for measures we therefore have

$$\mu(E) = \mu(E \cap U) = \lim_{n\to\infty} \mu(E \cap U_n) = \mu(E) \lim_{n\to\infty} \mu(U_n) = \mu(E)\mu(U).$$

Thus if $\mu(E) \neq 0$ then $\mu(U) = 1$, and hence $\mu(E) \geqslant \mu(U) - \varepsilon = 1 - \varepsilon$. Since ε was arbitrary we must have $\mu(E) = 1$.

One could also have appealed to Lemma A.2, taking $U_{n,i} = [\frac{i-1}{2^n}, \frac{i}{2^n}]$.

L^2 proof. Suppose that $E \subset \mathbb{R}/\mathbb{Z}$ is measurable and T-invariant, thus $T^{-1}E = E$ up to measure 0. Equivalently, $U_T 1_E(x) = 1_E(x)$ for a.e x. Using Fourier analysis, with the same notation as before, we have for any function $f \in L^2(X)$ that

$$\widehat{f}(r) = \int_0^1 f(x)\overline{e_r(x)}dx = \int_0^1 f(2x)\overline{e_r(2x)}dx = \widehat{U_T f}(2r)$$

for all integers r. Taking $f = 1_E$, it follows that $\widehat{1_E}(r) = \widehat{1_E}(2r)$ for all r. By the Riemann-Lebesgue lemma (see Appendix B), which states that $\lim_{r\to\infty} \hat{f}(r) = 0$ for all $f \in L^2(\mathbb{R}/\mathbb{Z})$, we are forced to conclude that $\widehat{1_E}(r) = 0$ whenever $r \neq 0$. As before, this implies that $\mu(E) = 0$ or 1.

3.8. The Gauss map is ergodic

This is a little harder than the previous two arguments.

PROPOSITION 3.3. *The Gauss map T is ergodic with respect to the Gauss measure ν.*

PROOF. In a sense, this proof is similar in structure to the proof that the $\times 2$ map is ergodic. The details are, however, rather more difficult. Rather than work with the Gauss measure ν, we work with the Lebesgue measure μ. We have $\mu(E) = 0$ if and only if $\nu(E) = 0$, so the notion of T-invariant set does not depend on which of these measures we use. Recall, however, that we do not in general have $\mu(T^{-1}A) = \mu(A)$; this will not be a problem in the argument that follows.

For any choice of integers $k_1, \dots, k_n \geqslant 1$ define the map $\psi_{k_n,\dots,k_1} : (0,1] \to (0,1]$ by

$$\psi_{k_n,\dots,k_1}(x) := \frac{1}{k_n + \frac{1}{k_{n-1} + \dots + \frac{1}{k_1 + x}}},$$

that is to say

$$\psi_{k_n,\dots,k_1} = \phi_{k_n} \circ \dots \circ \phi_{k_1}$$

where $\phi_k(x) := \frac{1}{k+x}$. Note that $T^n \circ \psi_{k_n,\dots,k_1}$ is the identity map, and indeed every preimage of x under T^n is of the form $\psi_{k_n,\dots,k_1}(x)$ for some choice of the k_i.

Write $U_{k_n,\dots,k_1} := \psi_{k_n,\dots,k_1}((0,1)) = \phi_{k_n} \circ \dots \circ \phi_{k_1}((0,1))$. $U_{k_n,\dots,k_1}$ is easily seen to be an interval. Furthermore, for any $0 < a < b \leqslant 1$ and for any $k \in \mathbb{N}$ we have

$$\phi_k(a) - \phi_k(b) = \frac{b-a}{(k+a)(k+b)} \leqslant b - a,$$

and so

$$1 \geqslant \mu(U_{k_1}) \geqslant \mu(U_{k_2,k_1}) \geqslant \mu(U_{k_3,k_2,k_1}) \geqslant \dots$$

We claim that, for any choice of $k_1, k_2, \dots,$

$$\mu(U_{k_{n+2},k_{n+1},k_n,\dots,k_1}) \leqslant \frac{1}{2}\mu(U_{k_n,k_{n-1},\dots,k_1}). \tag{3.2}$$

To prove this set $k := k_{n+1}, k' := k_{n+2}$ and suppose that $U_{k_n,k_{n-1},\dots,k_1} = (a,b)$, $U_{k_{n+1},\dots,k_1} = (a',b')$ and $U_{k_{n+2},\dots,k_1} = (a'',b'')$. Then

$$|b' - a'| = |\phi_k(a) - \phi_k(b)| = |\frac{b-a}{(k+a)(k+b)}|.$$

This is $\leqslant \frac{1}{2}(b-a)$ if $k \geqslant 2$, and so the claim follows in this case. If $k = 1$ then $a' = \frac{1}{1+b}$, $b' = \frac{1}{1+a}$ and so $a', b' \geqslant \frac{1}{2}$. But then

$$|b'' - a''| = |\phi_{k'}(a') - \phi_{k'}(b')| = |\frac{b'-a'}{(k'+a')(k'+b')}| \leqslant |b'-a'|\frac{1}{(k'+\frac{1}{2})^2} \leqslant \frac{1}{2}|b'-a'|.$$

This establishes the claim (3.2).

It follows from (3.2) and a simple induction that

$$\mu(U_{k_n,\dots,k_1}) \leqslant 2^{-\lfloor n/2 \rfloor}. \tag{3.3}$$

If $[a,b]$ and $F : [a,b] \to \mathbb{R}$ is a function which never vanishes, we define the *variation* of $\operatorname{var}_{[a,b]} F$ on $[a,b]$ to be $\sup_{x,y\in[a,b]} |\frac{F(x)}{F(y)}|$. Our next claim is that the variation of the derivative $\psi'_{k_n,\dots,k_1}$ is bounded by an absolute constant, independently of n and the choice of $k_n, \dots, k_1$. This means that the map $\psi_{k_n,\dots,k_1}$ from X to $U_{k_n,\dots,k_1}$ is "almost affine" in a weak sense.

To see this, we use the chain rule to conclude that

$$\operatorname{var}_{[0,1]} \psi'_{k_n,\dots,k_1} \leqslant \operatorname{var}_{[0,1]} \psi'_{k_{n-1},\dots,k_1} \cdot \operatorname{var}_{\psi_{k_{n-1},\dots,k_1}((0,1])} \phi'_{k_n}.$$

By (3.3) the right hand quantity is bounded by $\max_I \operatorname{var}_I \phi'_{k_n}$, the maximum being taken over all intervals I of length at most $2^{-\lfloor n/2 \rfloor}$. However

$$\operatorname{var}_{[a,b]} \phi'_k = |\frac{k+b}{k+a}|^2 \leqslant |\frac{1+b}{1+a}|^2 \leqslant |1+b-a|^2 \leqslant 1 + 3(b-a).$$

It follows that

$$\operatorname{var}_{[0,1]} \psi'_{k_n,\dots,k_1} \leqslant \prod_{n=1}^{\infty}(1+3\cdot 2^{-\lfloor n/2\rfloor}) \leqslant e^{3\sum_{n=1}^{\infty} 2^{-\lfloor n/2\rfloor}} < \infty.$$

We call any interval $U_{k_n,\dots,k_1}$ a *standard interval at level* n (compare with the notion of a standard dyadic interval at scale n that we encountered in our discussion of the doubling map). Let $E \subset [0,1]$ be measurable. Fix $k_1,\dots,k_n$ and set $\psi = \psi_{k_n,k_{n-1},\dots,k_1}$. Then the map $\psi : E \to \psi(E)$ is a bijection. Hence by change of variables we have

$$\mu(\psi(E)) = \int 1_{\psi(E)}(x)\, d\mu(x) = \int 1_E(y)\psi'(y)\, d\mu(y).$$

By the fact that $\operatorname{var}_{[0,1]} \psi'$ is bounded this lies between $M_1\mu(E)$ and $M_2\mu(E)$ where the ratio M_2/M_1 is bounded independently of E and of $k_1,\dots,k_n$. Noting that when $E=[0,1]$ we have $\psi(E) = U_{k_n,\dots,k_1}$, we see that

$$\mu(\psi(E)) \geqslant c\mu(E)\mu(U_{k_n,\dots,k_1})$$

for some absolute constant $c>0$ (there is a similar upper bound, but we do not require it). Now $\psi(E) = T^{-n}E \cap U_{k_n,k_{n-1},\dots,k_1}$, and thus

$$\mu(T^{-n}E \cap U_{k_n,\dots,k_1}) \geqslant c\mu(E)\mu(U_{k_n,\dots,k_1}).$$

At last, we reach the main argument. Suppose that E is T-invariant. Then $T^{-n}E = E$ up to measure zero, and so we see that

$$\mu(E\cap \tilde{U}) \geqslant c\mu(E)\mu(\tilde{U})$$

for all sets $\tilde{U}$ which are finite unions of standard intervals at level n. Call such sets $\tilde{U}$ *standard at level* n.

Now the union of all the standard intervals at level n consists of all of $(0,1)$ except for some numbers with finite continued fraction expansion (numbers of the form $\psi_{k_n,\dots,k_1}(0)$ or $\psi_{k_n,\dots,k_1}(1)$). All of these are rational, and so the union of all the standard intervals at level n has measure 1. We showed above that the length of any standard interval of level n is at most $2^{-\lfloor n/2\rfloor}$, which of course tends to 0 with n. We may therefore apply Lemma A.2 to conclude that every measurable set can be approximated arbitrarily well by standard sets at some level. It follows that

$$\mu(E\cap E') \geqslant c\mu(E)\mu(E')$$

for all measurable $E' \in \mathscr{B}$.

Taking $E' = [0,1]\setminus E$ we immediately obtain $\mu(E) = 0$ or 1. This concludes the proof that the Gauss map is ergodic. □

CHAPTER 4

The mean ergodic theorem

In this chapter we will develop some aspects of the "L^2-theory" of ergodic transformations, and in particular prove our first ergodic theorem, the mean ergodic theorem. It is stated in Section 4.3 below.

An ergodic theorem is a theorem in the opposite direction to the rather trivial Lemma 3.1, asserting that ergodic transformations have equidistributing properties. More accurately, an ergodic theorem is any result stating that the time averages $S_N f := \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$ converge to the space average $\int f\, d\mu$. There are various notions of convergence of functions we might consider, some stronger than others. In this chapter we will consider the case of convergence in L^2. In the next chapter we will consider the rather harder question of pointwise convergence.

4.1. $L^2(X)$ as a Hilbert space

If $(X, \mu, \mathscr{B}, T)$ is a measure-preserving system we will make much use of the Hilbert space $L^2(X) = L^2(X, \mathscr{B}, \mu)$ and of the associated isometry $U_T : L^2(X) \rightarrow L^2(X)$ given by $U_T f(x) = f(Tx)$. We introduced these objects in the last chapter, but made no real use of the fact that $L^2(X)$ is a Hilbert space.

We do not wish to assume that students have taken a course in Hilbert spaces, although many will have done. Here we will use two basic facts: the existence of projection operators to closed subspaces, and the existence of adjoints of bounded linear operators on Hilbert space. Students do not need to know the proofs of these results, though they are not difficult. Here is a very brief refresher on their statements.

Definition 4.1. A (complex) Hilbert space H is a real or complex vector space which is endowed with an inner product $\langle , \rangle : H \times H \rightarrow \mathbb{C}$ with the property that H is complete with respect to the norm $\|x\| := \sqrt{\langle x, x\rangle}$ induced by this inner product.

For the rest of this chapter

Lemma 4.1 (Projections). *Let H be a Hilbert space. Suppose that $V \subset H$ is a closed subspace of H. Then we have a direct sum decomposition $H = V \oplus V^{\perp}$. Hence there is a unique linear projection operator*

$$\pi : H \rightarrow V$$

such that

(i) $\pi|_V$ *is the identity on* V*;*
(ii) $\pi|_{V^\perp}$ *is identically zero;*
(iii) *For all* $x \in H$*,* $x - \pi(x) \in V^\perp$*, that is to say* $\langle x, v\rangle = \langle \pi(x), v\rangle$ *for all* $v \in V$*.*

If H is a Hilbert space, then a *bounded linear operator* is a linear map $\phi : H \to H$ such that $\|\phi(x)\| \leqslant C\|x\|$ for all $x \in H$ and for some constant C. The infimum of all such C is called the *norm* of ϕ and is written $\|\phi\|$.

LEMMA 4.2 (Adjoints). *Let* H *be a Hilbert space and suppose that* $\phi : H \to H$ *is a bounded linear operator. Then* ϕ *has an* adjoint ϕ^**, which is a bounded linear map* $\phi^* : H \to H$ *with* $\|\phi\| = \|\phi^*\|$ *and* $\langle \phi(x), y\rangle = \langle x, \phi^*(y)\rangle$ *for all* $x, y \in H$*.*

4.2. The space of invariant functions.

Suppose that (X, μ, T) is a measure-preserving system, not necessarily ergodic. Then we consider[1] the space

$$I_T := \{f \in L^2(X) : f(x) = f(Tx) \quad \text{a.e.}\}$$

of invariant functions on $L^2(X)$. Alternatively, recalling that $U_T : L^2(X) \to L^2(X)$ is the isometry given by $U_T f(x) = f(Tx)$, we have

$$I_T = \{f \in L^2(X) : f = U_T f\} = \ker(\mathrm{id} - U_T).$$

It is easy to see that I_T is a closed subspace of $L^2(X)$.

The following fact is fairly straightforward.

LEMMA 4.3. *The transformation* T *is ergodic if and only if the space* I_T *of invariant functions consists of just the functions which are constant a.e.*

PROOF. The "if" direction is trivial: suppose that $E \subset \mathscr{B}$ is T-invariant, so $T^{-1}E = E$ up to measure zero. Alternatively, $1_E = U_T 1_E$ a.e., and so $1_E \in I_T$. Hence, by assumption, 1_E is constant a.e., and this implies that $\mu(E) = 0$ or 1.

The "only if" direction is not much harder. Suppose that T is ergodic. Suppose that $f \in I_T$, that is to say $f = U_T f$. In other words, $f(x) = f(Tx)$ a.e. Then the level sets

$$E_{a,b} := \{x \in X : a \leqslant |f(x)| < b\}$$

are T-invariant, that is to say $x \in E_{a,b}$ if and only if $Tx \in E_{a,b}$, outside a set of measure 0. All these sets are measurable and hence must have measure 0 or 1, and this implies that f is constant a.e. You are asked to supply a detailed proof in Example Sheet 2, Q2. □

[1]Note that the space of *strictly* invariant functions, those with $f(x) = f(Tx)$ for all x, is not really well-defined since functions in $L^2(X)$ are only defined almost everywhere.

Since I_T is a closed linear subspace of a Hilbert space, it follows from Lemma 4.1 that there is a projection operator

$$\pi_T : L^2(X) \to I_T$$

to this space of invariant functions.

Let us note a few simple but useful facts about how π_T interacts with U_T.

LEMMA 4.4. *We have the following.*

(i) *If $f \in I_T$ then $U_T^* f = f$.*

(ii) *If T is ergodic then $\pi_T f = \int f d\mu$.*

PROOF. (i) Let $y \in H$ be arbitrary, and let $\phi : H \to H$ be any isometry. Then we have

$$\begin{aligned}\langle y, x - \phi^*(x)\rangle &= \langle y, x\rangle - \langle y, \phi^*(x)\rangle \\ &= \langle \phi(y), \phi(x)\rangle - \langle \phi(y), x\rangle \\ &= \langle \phi(y), \phi(x) - x\rangle = 0.\end{aligned}$$

Taking $y = x - \phi^*(x)$ tells us that $\|x - \phi^*(x)\| = 0$, and the result follows. (We remark that the assumption that ϕ is an isometry is not really necessary, and can be replaced with the weaker condition that $\|\phi\| \leqslant 1$, but it slightly simplifies the proof.)

(ii) We know that I_T consists of just the constant functions when T is ergodic. Now it is clear that

$$\langle \int f d\mu, 1\rangle = \langle f, 1\rangle \qquad (= \int f d\mu),$$

i.e.

$$\langle f - \int f d\mu, 1\rangle = 0.$$

Therefore the map sending f to $\int f d\mu$ is precisely the orthogonal projection to the space I_T. □

4.3. Von Neumann's mean ergodic theorem

THEOREM 4.1. *Suppose that (X, μ, T) is an ergodic m.p.s. Then for any $f \in L^2(X)$ we have $S_N f \to \int f d\mu$ in L^2.*

There is, in fact, a version of the theorem for measure-preserving transformations which are not necessarily ergodic.

THEOREM 4.2. *Suppose that (X, μ, T) is a m.p.s. Then for any $f \in L^2(X)$ we have $\|S_N f - \pi_T(f)\|_2 \to 0$ as $N \to \infty$.*

It is immediate that Theorem 4.1 follows from 4.2 using Lemma 4.4 (ii), so we will prove only Theorem 4.2.

PROOF. The key idea of the proof is to identify the orthogonal complement $I_T^\perp$ of I_T, the space of T-invariant functions, as the closed subspace spanned by *cocycles*. If $g \in L^2(X)$ we write $\partial g = g - U_T g$: this is called a cocycle. Let M be the closed subspace of $L^2(X)$ spanned by all cocycles ∂g. It is clear that $I_T \subset M^\perp$, since (by Lemma 4.4 (i)) if f is U_T-invariant then it is also U_T^*-invariant and we have

$$\langle \partial g, f\rangle = \langle g - U_T g, f\rangle = \langle g, f\rangle - \langle g, U_T^* f\rangle = 0.$$

Suppose, conversely, that $f \in L^2(X)$ is orthogonal to all cocycles. Then in particular we have $\langle f, \partial f\rangle = 0$. It follows that

$$\begin{aligned}\|f - U_T f\|^2 &= \langle f, f - U_T f\rangle + \langle f - U_T f, f\rangle - \|f\|^2 + \|U_T f\|^2 \\ &= -\|f\|^2 + \|U_T f\|^2 \\ &= 0.\end{aligned}$$

Thus $f = U_T f$ a.e., that is to say $f \in I_T$.

Now by the standard facts about Hilbert spaces mentioned in Section **??** we have

$$L^2(X) = I_T \oplus I_T^\perp.$$

Now that we know that $I_T^\perp$ is the closure of the space of cocycles, so we have the following statement. We state it as a separate lemma as we will use it in the next chapter.

LEMMA 4.5. *Let $f \in L^2(X)$ be arbitrary. Then for any $\varepsilon > 0$ we may write*

$$f = \pi_T(f) + \partial g + h$$

where $\|h\| \leqslant \varepsilon$. In particular if T is ergodic we have

$$f = \int_X f + \partial g + h$$

where $\|h\| \leqslant \varepsilon$.

Taking time averages, and using the fact that $S_N(\pi_T f) = \pi_T f$ since $\pi_T f \in I_T$, we thus have

$$\|S_N f - \pi_T(f)\| \leqslant \|S_N(\partial g)\| + \|S_N h\|. \tag{4.1}$$

Now $S_N : L^2(X) \to L^2(X)$ is easily seen to be a contraction (the proof is the same as for $L^1(X)$) and so

$$\|S_N h\| \leqslant \varepsilon.$$

Now by telescoping the sum we see that

$$S_N(\partial g) = \frac{1}{N}(g - U_T^N g),$$

and so

$$\|S_N(\partial g)\| \leqslant \frac{2}{N}\|g\|.$$

Comparing with (4.1) we see that

$$\|S_N f - \pi_T(f)\| \leqslant \frac{2}{N}\|g\| + \varepsilon,$$

which is less than 2ε if N is big enough. Since $\varepsilon > 0$ was arbitrary, the result follows. □

CHAPTER 5

The pointwise ergodic theorem

Our aim in this chapter is to prove the following result, asserting that "time averages converge to space averages" in a rather strong sense.

THEOREM 5.1 (Birkhoff's almost-everywhere ergodic theorem). *Suppose that $(X, \mu, \mathscr{B}, T)$ is an ergodic measure-preserving system, and suppose that $f \in L^1(X)$. Then $S_N f \to \int_X f d\mu$ pointwise almost everywhere. That is, for all $x \in X$ outside of a set of measure zero,*

$$\lim_{N\to\infty} S_N f(x) \to \int_X f d\mu.$$

The proof of this is somewhat tricky. An important ingredient of it is a kind of special case known (essentially) as the maximal ergodic theorem. This is an assertion to the effect that if the "space average" $\|f\|_1 = \int_X |f(x)| d\mu(x)$ is small then so are many of the time averages.

5.1. The maximal ergodic theorem

Before stating and proving the maximal ergodic theorem, we isolate a simple lemma (a special case of the Vitali covering lemma) from its proof.

LEMMA 5.1. *Let $\{I_m\}_{m\in S}$ be a collection of intervals of form $I_m = [m, m + \ell(m)) \subset \mathbb{Z}$. Then there is a disjoint subcollection $\{I_m\}_{m\in S'}$ whose union has length at least $\frac{1}{2}|S|$.*

PROOF. First of all pass to a minimal subcollection $\{I_m : m \in S_*\}$ with the property that $\bigcup_{m\in S_*} I_m = \bigcup_{m\in S} I_m$. By a simple inspection this subcollection has the property that no point y lies in three of the I_m. Writing $S_* = \{m_1 < m_2 < \cdots < m_k\}$, it may now be seen that the two collections $I_{m_1} \cup I_{m_3} \cup \dots$ and $I_{m_2} \cup I_{m_4} \cup \dots$ consist of disjoint intervals. Furthermore the union of these two collections is $\bigcup_{m\in S} I_m$ and hence contains S. Thus at least one of these two collections has union of size at least $\frac{1}{2}|S|$, and the lemma follows. □

PROPOSITION 5.1 (Maximal ergodic theorem). *Suppose that $(X, \mu, \mathscr{B}, T)$ is a measure-preserving system. Suppose that $f \in L^1(X)$ is a function with $\|f\|_1 = \int_X |f(x)| d\mu(x) \leqslant \varepsilon$. Let $E \subset X$ be the set of all $x \in X$ for which some time average $S_N f(x)$, $N = 1, 2, 3, \dots$ has magnitude at least δ. Then $\mu(E) \leqslant 2\varepsilon/\delta$.*

Note that we do not need to assume that T is ergodic here.

PROOF. By replacing f by $|f|$, we may assume that $f \geqslant 0$ everywhere. For integer N_0, Write $E(N_0) \subset X$ for the set of all $x \in X$ for which some time average $S_N f(x)$ with $N \leqslant N_0$ has magnitude at least δ. Then $E(N_0) \nearrow E$ as $N_0 \to \infty$. By the limit principle it is therefore enough to show that $\mu(E(N_0)) \leqslant 2\varepsilon/\delta$. For notational brevity, let us replace $E(N_0)$ by E in what follows.

If $x \in E$ then, by definition, there is some $N(x) \leqslant N_0$ such that

$$|S_{N(x)} f(x)| \geqslant \delta,$$

or in other words

$$\mathbb{E}_{0 \leqslant n < N(x)} f(T^n x) \geqslant \delta. \tag{5.1}$$

Here we have used the convenient notation $\mathbb{E}_{x \in X} = \frac{1}{|X|} \sum_{x \in X}$ to denote averaging over a set X. Thus, since E is large, many averages such as (5.1) are large. We wish to use this fact to show that $\|f\|_1$ is also large. The first trick is to shift (5.1) around to get many further large averages. Indeed by an obvious change of variables we have, for any positive integer m,

$$\mathbb{E}_{m \leqslant n < m+N(x)} f(T^{n-m} x) \geqslant \delta \tag{5.2}$$

and so

$$\mathbb{E}_{m \leqslant n < m+N(T^m x')} f(T^n x') \geqslant \delta \tag{5.3}$$

for all $x' \in T^{-m} E$.

Let N_1 be a quantity to be specified later, much larger than N_0. For $x \in X$, let $A(x)$ be the set of all return times of x to E before time N_1, that is to say the set of all $m \in [0, N_1)$ such that $x \in T^{-m} E$. Thus if $x \in X$ and if $m \in A(x)$ then we have

$$\mathbb{E}_{m \leqslant n < m+N(T^m x)} f(T^n x) \geqslant \delta, \tag{5.4}$$

Note that $N(T^m x) \in [0, N_0)$.

Now we have

$$\int_X |A(x)| d\mu(x) = \sum_{0 \leqslant m < N_1} \mu(T^{-m} E) = \mu(E) N_1, \tag{5.5}$$

since T is measure-preserving.

For each $x \in X$, consider the collection of intervals $\{I_m\}_{m \in A(x)}$, where $I_m = [m, m + N(T^m x))$. In general these intervals will overlap, but by Lemma 5.1 there is a disjoint subcollection $\{I_m\}_{m\in A'(x)}$ with

$$\sum_{m\in A'(x)} |I_m| \geqslant \frac{1}{2}|A(x)|.$$

Since all these intervals are contained in $[0, N_0 + N_1)$ we have

$$\begin{aligned}\sum_{0\leqslant n<N_0+N_1} f(T^n x) &\geqslant \sum_{m\in A'(x)} \sum_{n\in I_m} f(T^n x)\\ &\geqslant \delta \sum_{m\in A'(x)} |I_m|\\ &\geqslant \frac{1}{2}\delta|A(x)|.\end{aligned}$$

Now we integrate both sides over all $x \in X$. Since T is measure-preserving, we have $\int_X f \circ T^n d\mu = \int_X f d\mu$ for all n. Using (5.5), it follows that

$$\begin{aligned}(N_0 + N_1)\|f\|_1 &= (N_0 + N_1) \int_X f d\mu\\ &= \int_X \big(\sum_{0\leqslant n<N_0+N_1} f(T^n x)\big) d\mu(x)\\ &\geqslant \frac{1}{2}\delta \int_X |A(x)| d\mu(x)\\ &= \frac{1}{2}\delta\mu(E)N_1.\end{aligned}$$

It follows that

$$\mu(E) \leqslant \frac{2}{\delta}\frac{N_0 + N_1}{N_1}\|f\|_1.$$

Letting $N_1 \to \infty$, we obtain

$$\mu(E) \leqslant \frac{2}{\delta}\|f\|_1 \leqslant \frac{2\varepsilon}{\delta}.$$

This completes the proof. □

5.2. Time averages, space averages and limits

Suppose that $(X, \mu, \mathscr{B}, T)$ is an ergodic measure-preserving system. In Chapter 1 we talked about continuous functions having the time averages - space averages property (TASA). As we remarked, this nomenclature was slightly nonstandard. Here, we require a very similar notion for functions in $L^1(X)$.

Given a function $f \in L^1(X)$, we say that f has TASA if

$$\lim_{N\to\infty} S_N f(x) = \int_X f d\mu \tag{5.6}$$

for a.e. x. Note that now we do *not* require the convergence to be uniform in x, so this notion of TASA is not the same as the one in Chapter 1. We will only be using this terminology during the proof of the pointwise ergodic theorem.

In this section we use the maximal ergodic theorem to establish a pleasant closure property of the set of all $f \in L^1(X)$ with TASA.

LEMMA 5.2. *Suppose that a sequence of functions $f_1, f_2, \dots \in L^1(X)$ all have TASA, and that $f_j \to f$ in $L^1(X)$ (that is, $\lim_{j\to\infty} \|f_j - f\|_1 = 0$). Then f also has TASA.*

PROOF. Let $\varepsilon > 0$, and suppose that $0 < \delta < \frac{1}{3}$. Choose j so large that $\|f_j - f\|_1 \leqslant \frac{1}{12}\delta\varepsilon$, thus

$$|\int_X f_j d\mu - \int_X f d\mu| \leqslant \frac{1}{12}\delta\varepsilon.$$

Write A_M for the set of all $x \in X$ for which $|S_N f_j(x) - \int_X f_j d\mu| \geqslant \frac{1}{3}\varepsilon$ for some $N \geqslant M$. Then $\bigcap_{M=1}^{\infty} A_M$ is contained in the set of x for which $S_N f_j(x) \nrightarrow \int_X f_j d\mu$, and hence has measure 0. Noting that we have the nesting $A_1 \supset A_2 \supset \dots$ it follows from the limit property of μ that $\mu(A_M) \searrow 0$, and so there is some M_* such that $\mu(A_{M_*}) \leqslant \frac{1}{2}\delta$. Thus

$$|S_N f_j(x) - \int_X f_j d\mu| \leqslant \frac{1}{3}\varepsilon$$

for all $N \geqslant M_*$ and all $x \notin A_{M_*}$.

Also if B is the set of all $x \in X$ for which we do not have

$$|S_N f_j(x) - S_N f(x)| \leqslant \frac{1}{3}\varepsilon$$

for all $N = 1, 2, 3, \dots$ then, by the maximal ergodic theorem, $\mu(B) \leqslant \frac{1}{2}\delta$. If $x \notin A_{M_*} \cup B$, a set of measure at most δ, then by combining the three displayed equations and using the triangle inequality we have

$$|S_N f(x) - \int_X f d\mu| \leqslant \varepsilon$$

for all $N \geqslant M_*$, that is to say

$$\limsup_{N\to\infty} |S_N f(x) - \int_X f d\mu| \leqslant \varepsilon.$$

This is true for all x outside a set of measure at most δ, but δ was arbitrary: therefore it is true for a.e. x. Take a sequence $\varepsilon_1 > \varepsilon_2 > \varepsilon_3 > \dots$ with $\varepsilon_i \to 0$. For each i and for a.e. x we have

$$\limsup_{N\to\infty} |S_N f(x) - \int_X f d\mu| \leqslant \varepsilon_i.$$

Since a countable union of sets of measure zero has measure zero, this is true, for a.e. x, for all i simultaneously. For these x we of course have

$$\limsup_{N\to\infty} |S_N f(x) - \int_X f d\mu| = 0,$$

and this concludes the proof. $\square$

5.3. Proof of the pointwise ergodic theorem

Proof of the pointwise ergodic theorem. We now wish to prove the pointwise ergodic theorem, namely that in an ergodic measure-preserving system $(X, \mu, \mathscr{B}, T)$ *every* function $f \in L^1(X)$ has the TASA property in the sense defined in the last section.

Motivated by the proof of the mean ergodic theorem in the last chapter, let us begin by noting that there is a class of functions for which the property is quite easy to prove, namely the "constants plus bounded cocycles". Indeed suppose that $f = c + \partial g$ for some constant c and some $g \in L^\infty(X)$, where here $\partial g = g(x) - g(Tx)$ and $L^\infty(X)$ denotes the space of measurable functions for which $|g(x)| \leqslant M$ for a.e. x, for some finite M. Then by telescoping the sum we have

$$S_N f(x) = c + \frac{1}{N}(g(x) - g(T^{n-1}x)).$$

Since

$$|\frac{1}{N}(g(x) - g(T^{n-1}x))| \leqslant \frac{2}{N} M \to 0$$

for a.e. x., it follows that $S_N f(x) \to c$ for a.e. x. On the other hand,

$$\int_X \partial g d\mu = \int_X f d\mu - \int_X f \circ T = 0,$$

using the fact that T is measure-preserving.

By this observation and the main result of the last section, it is therefore enough to prove the following lemma.

LEMMA 5.3. *Suppose that $(X, \mu, \mathscr{B}, T)$ is an ergodic measure-preserving system. Then the constant-plus-bounded-cocycle functions are dense in $L^1(X)$.*

PROOF. In the proof of the von Neumann ergodic theorem we proved a superficially similar result, Lemma 4.5. We will use this fact to establish Lemma 5.3. The appeal to this lemma is in fact the only place in this section that we need the assumption that T is ergodic. We will also make two appeals to Lemma A.4, which states that the simple measurable functions $f = \sum_{i \in I} c_i 1_{E_i}$ are dense in $L^1(X)$.

Let f be such a simple measurable function. Then f certainly lies in $L^2(X)$. By Lemma 4.5, there is some $g \in L^2(X)$ with

$$\|f - \int_X f d\mu - \partial g\|_2 \leqslant \frac{1}{9}\varepsilon^2$$

and hence, by the Cauchy-Schwarz inequality,

$$\|f - \int_X f d\mu - \partial g\|_1 \leqslant \frac{1}{3}\varepsilon.$$

This is almost what we want, except that there is no guarantee that g is bounded. However, since g lies in $L^2(X)$ it also lies in $L^1(X)$, and so by Lemma A.4 it may be approximated arbitrarily closely by simple measurable functions, all of which lie in $L^\infty(X)$. In particular we may find $\tilde{g} \in L^\infty(X)$ with $\|g - \tilde{g}\|_1 \leqslant \frac{1}{3}\varepsilon$. By the triangle inequality we then have $\|\partial\tilde{g} - \partial g\|_1 \leqslant \frac{2}{3}\varepsilon$ and so

$$\|f - \int_X f d\mu - \partial\tilde{g}\|_1 \leqslant \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that f may be approximated arbitrarily well in L^1 by constant-plus-bounded-cocycle functions.

However, the simple measurable functions like f are dense in $L^1(X)$, and the result follows. □

5.4. Normal numbers.

Let $x \in [0, 1]$, and write $x = 0.a_1a_2a_3\ldots$ in base k (thus each a_i lies in the set $\{0, 1, \ldots, k-1\}$). For each sequence $b_1 \ldots b_j$ of digits, look at those n for which $a_{n+1} = b_1, \ldots, a_{n+j} = b_j$. We say that x is *normal* in base k if

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leqslant N : a_{n+1} = b_1, \ldots, a_{n+j} = b_j\} = k^{-j} \tag{5.7}$$

for all j and for all choices of $b_1, \ldots, b_j$, that is to say the number of occurrences of the pattern $b_1 \ldots b_j$ amongst the base k digits of x is what one expects it to be.

The normality of a number may be interpreted in terms of the $\times k$ maps T_k on the circle $\mathbb{R}/\mathbb{Z}$. Indeed $a_{n+1} = b_1, \ldots, a_{n+j} = b_j$ if, and only if, $T_k^n x$ lies in the interval

$$I := \frac{b_1}{k} + \cdots + \frac{b_j}{k^j} + [0, \frac{1}{k^j}).$$

Since the maps T_k are ergodic (we only proved this for $k = 2$, but the proof for general k is the same) it follows from the pointwise ergodic theorem that

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leqslant N : a_{n+1} = b_1, \ldots, a_{n+j} = b_j\} = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_I(T_k^n x) = \mu(I) = k^{-j}$$

for a.e. x. There are only countable many choices for j and $b_1, \ldots, b_j$, so in fact the above holds for all such choices for a.e. x, that is to say almost all $x \in [0, 1]$ are normal in base k.

It follows immediately that almost all $x \in [0, 1]$ are normal to all bases; such numbers are called *absolutely normal*.

5.5. The continued fraction expansion of a typical number

Let us derive a corollary about the partial quotients of a "typical" number in $(0,1)$. In this section we will write $X = [0,1]$, $T : X \to X$ will be the Gauss map, and $\nu(E) = \frac{1}{\log 2}\int_X \frac{d\mu(x)}{1+x}$ will be Gauss measure. As we showed in Chapter 3, T is ergodic with respect to ν.

Recall our observation that if

$$x = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

then $a_n = \lfloor 1/T^{n-1}x \rfloor$. Writing $f : (0,1] \to \mathbb{R}$ for the function defined by $f(t) := \log(\lfloor 1/t \rfloor)$ it is not hard to check that $f \in L^1([0,1], \mathscr{B}, \nu)$. It follows from the pointwise ergodic theorem that for ν-a.e. x (and hence for μ-a.e. x) we have

$$\frac{1}{N}\sum_{n=1}^{N} \log a_n = S_N f(x) \to \int f \, d\nu = \frac{1}{\log 2}\int_0^1 \frac{\log(\lfloor 1/t \rfloor)}{1+t}\, dt.$$

Splitting the integral into the ranges $(1/(k+1), 1/k)$ and making the substitution $u = 1/t - k$, we see that the right hand side is

$$\frac{1}{\log 2}\sum_{k=1}^{\infty} \log k \int_0^1 \frac{1}{(u+k)(u+k+1)}\, du.$$

The integration is easily accomplished using the obvious partial fraction expansion, and we see that the above is

$$\frac{1}{\log 2}\sum_{k=1}^{\infty} \log k \cdot \log \frac{(k+1)^2}{k(k+2)}.$$

Thus for almost all x the partial quotients satisfy

$$\lim_{N\to\infty} (a_1 \dots a_N)^{1/N} = \prod_{k=1}^{\infty} \left(\frac{(k+1)^2}{k(k+2)} \right)^{\log k/\log 2}.$$

The constant here is called Khintchine's constant, and its numerical value is approximately 2.685452...

More applications of ergodic theory to the study of continued fractions may be found on the third example sheet.

CHAPTER 6

Combinatorial number theory and the correspondence principle

6.1. Szemerédi's theorem

A beautiful application of ergodic theory, due to Furstenberg, is to give a proof of a famous theorem of Szemerédi in combinatorial number theory. Here, and in the rest of the chapter, $[N] := \{1, \ldots, N\}$.

THEOREM 6.1. *Let $k \in \mathbb{N}$ be an integer and suppose that $\delta > 0$. Let $A \subset [N]$ be a set with density $|A|/N$ at least δ. Then, provided N is large enough, A contains k distinct elements in arithmetic progression.*

To prove this theorem, we relate it to a result about multiple recurrence on certain types of system. The systems we will consider to begin with are called Cantor systems (this is not standard terminology).

6.2. Cantor systems

DEFINITION 6.1 (Cantor system). A Cantor system is a quadruple $(X, \mathscr{R}, \mu, T)$ where

- $X = \{0,1\}^{\mathbb{Z}}$ is the space of doubly-infinite sequences $\vec{x} = (x_n)_{n \in \mathbb{Z}}$;
- $\mathscr{R}$ is the *ring of clopen sets* [1] which consists of all finite unions of cylinder sets, that is to say sets obtained by fixing some finite number of coordinates of x (for example the set $R \subset X$ given by $R = \{\vec{x} \in X : x_{-3} = 0, x_2 = 1, x_7 = 0\}$;
- $T : X \to X$ is the right shift map defined by $(T\vec{x})_n = x_{n+1}$;
- $\mu : \mathscr{R} \to [0,1]$ is a probability measure which is invariant with respect to the shift: $\mu(T^{-1}R) = \mu(R)$ for all $R \in \mathscr{R}$.

A Cantor system is not *a priori* a measure-preserving system, because $\mathscr{R}$ is only a ring of sets, and not a σ-algebra (it is not closed under countable unions).

[1] clopen stands for "closed and open". We never need the fact that all clopen sets are finite unions of cylinder sets, though this is true and not hard to prove.

The definition of probability measure on $\mathscr{R}$ is almost identical to that of a probability measure on a σ-algebra $\mathscr{B}$, but with one slight change: the countable additivity axiom now becomes

- If $A_i \in \mathscr{R}$, if $A_1 \subset A_2 \subset \dots$ and if $\bigcup_{n=1}^{\infty} A_n \in \mathscr{R}$ then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

The only difference here is that it is not automatic that $\bigcup_{n=1}^{\infty} A_n \in \mathscr{R}$.

In the specific case we are discussing here, the ring of clopen sets on the Cantor space, the countable additivity axiom follows from the weaker *finite additivity* axiom $\mu(A \cup A') = \mu(A) + \mu(A')$ if $A, A' \in \mathscr{R}$ are disjoint. This is because $X = \{0,1\}^{\mathbb{Z}}$ (with the natural topology – see Examples Sheet 1, Q10) is compact and all elements of $\mathscr{R}$ are closed (and hence compact), so if $\bigcup_{n=1}^{\infty} A_n \in \mathscr{R}$ then in fact $\bigcup_{n=1}^{k} A_n = \bigcup_{n=1}^{\infty} A_n$ for some finite k.

6.3. The correspondence principle

In this section we show that Szemerédi's theorem is a consequence of the following multiple recurrence result for Cantor systems.

THEOREM 6.2. *Suppose that $(X, \mathscr{R}, \mu, T)$ is a Cantor system. Let $R \in \mathscr{R}$ be a set with $\mu(R) > 0$. Then there is some $d \geqslant 1$ such that $\mu(R \cap T^{-d}R \cap T^{-2d}R \cdots \cap T^{-(k-1)d}R) > 0$.*

Here is the deduction, which is known as Furstenberg's correspondence principle. If Szemerédi's theorem is false, we may find an infinite sequence $(A_j)_{j=1}^{\infty}$ of finite sets with $A_j \subset [N_j]$, $N_j \to \infty$, with $|A_j|/N_j \geqslant \delta$ for all j, but with no A_j containing k distinct elements in arithmetic progression.

Consider the Cantor space $X = \{0,1\}^{\mathbb{Z}}$. Note that X can be identified with the power set $\mathscr{P}(\mathbb{Z})$ of subsets of $\mathbb{Z}$, via the identification which maps $A \subset \mathbb{Z}$ to $\vec{x}_A$ where

$$(\vec{x}_A)_n = \begin{cases} 1 & n \in A \\ 0 & n \notin A. \end{cases}$$

Consider the *shift map* $T : X \to X$ defined by $(T\vec{x})_n = x_{n+1}$. To any $A \subset [N]$ we also associate[2] a probability measure μ_A on $X = \{0,1\}^{\mathbb{Z}}$ by setting

$$\mu_A(S) = \frac{1}{N}\#\{n \in [N] : T^n \vec{x}_A \in S\}.$$

That any such μ_A *is* a probability measure is easily verified: remember that it suffices to check finite additivity.

[2]Technically this depends on N as well as on A, but we suppress explicit mention of this dependence.

The measure μ_A captures various statistics of A. Consider, for example, the cylinder sets $S(a_1, \dots, a_k) \in \mathscr{R}$ given by

$$S(a_1, a_2, \dots, a_k) := \{\vec{x} \in X : x_{a_1} = \dots = x_{a_k} = 1\}.$$

In particular if $S = S(0) = \{\vec{x} : x_0 = 1\}$ then, since $(T^n \vec{x}_A)_0 = (\vec{x}_A)_n = 1$ if $n \in A$ and 0 otherwise, we see that

$$\mu_A(S(0)) = \frac{1}{N}\#\{n \in [N] : n \in A\}.$$

That is to say, $\mu_A(S(0))$ is precisely the density of A as a subset of $[N]$. Similarly

$$\mu_A(S(3,7,12)) = \frac{1}{N}\#\{n \in [N] : n+3, n+7, n+12 \in A\},$$

a certain "triple correlation density of A" and so on.

Define $\mu_j := \mu_{A_j}$. Now we come to a key idea: because the collection $\mathscr{R}$ of clopen sets is countable, we may apply a diagonal argument to pass to a subsequence of the μ_j which converges "weakly", that is to say such that $\mu_j(R)$ converges for all $R \in \mathscr{R}$. Indeed, if $\mathscr{R} = \{R_1, R_2, R_3, \dots\}$ then first pass to a subsequence $\{j_{1,1}, j_{1,2}, \dots\} \subset \mathbb{N}$ such that $\mu_{j_{1,i}}(R_1)$ converges as $i \to \infty$. Next pass to a further subsequence $\{j_{2,1}, j_{2,2}, \dots\} \subset \{j_{1,1}, j_{1,2}, \dots\}$ such that $\mu_{j_{2,i}}(R_2)$ converges, and so on. Finally consider the diagonal subsequence $\{j_{1,1}, j_{2,2}, j_{3,3}, \dots\}$: by construction the sequence $\mu_{j_{i,i}}(R_r)$ converges as $i \to \infty$, for every fixed r.

For simplicity of notation, let us assume that the sequence μ_j itself converges weakly, and define a function $\mu : \mathscr{R} \to [0,1]$ by $\mu(R) := \lim_{j\to\infty} \mu_j(R)$. It is easy to check that μ inherits the prooperty of finite additivity from the μ_j, and hence μ is a probability measure on $(X, \mathscr{R}, \mu, T)$.

LEMMA 6.1. *The probability measure μ is T-invariant, and so $(X, \mathscr{R}, \mu, T)$ is a Cantor system.*

PROOF. The idea here is that, while the approximating measures μ_j are not T-invariant, they are almost so and become truly invariant in the limit.

Indeed for any clopen set $S \in \mathscr{R}$ we have

$$\begin{aligned}
\mu_j(T^{-1}S) &= \frac{1}{N_j}\#\{n \in [N_j] : T^n \vec{x}_A \in T^{-1}S\} \\
&= \frac{1}{N_j}\#\{n \in [N_j] : T^{n+1} \vec{x}_A \in S\} \\
&= \frac{1}{N_j}\#\{n \in [N_j] - 1 : T^n \vec{x}_A \in S\}.
\end{aligned}$$

Since

$$\mu_j(S) = \frac{1}{N_j}\#\{n \in [N_j] : T^n \vec{x}_A \in S\},$$

it follows that

$$|\mu_j(S) - \mu_j(T^{-1}S)| \leqslant \frac{2}{N_j} \to 0$$

as $j \to \infty$, and therefore $\mu(S) = \mu(T^{-1}S)$, as required. □

LEMMA 6.2. *We have* $\mu(S(0)) > 0$.

PROOF. We have $\mu_j(S(0)) = N_j^{-1}|A_j| \geqslant \delta$. Since $S(0)$ is clopen, we have $\mu(S(0)) = \lim_{j\to\infty} \mu_j(S(0)) \geqslant \delta$ and the result follows. □

Now apply the multiple recurrence theorem for Cantor systems, Theorem 6.2, with $R = S(0)$. We obtain the existence of some $d > 0$ such that

$$\mu(S(0) \cap T^{-d}(S(0)) \cap T^{-2d}(S(0)) \cap \cdots \cap T^{-(k-1)d}(S(0))) > 0.$$

Note that

$$S(0) \cap T^{-d}(S(0)) \cap T^{-2d}(S(0)) \cap \cdots \cap T^{-(k-1)d}(S(0)) = S(0, d, \ldots, (k-1)d)$$

is another clopen set. It follows that

$$\lim_{j\to\infty} \mu_j(S(0, d, \ldots, (k-1)d) > 0,$$

and in particular

$$\mu_j(S(0, d, \ldots, (k-1)d)) > 0$$

for j sufficiently large. However, $\mu_j(S(0, d, \ldots, (k-1)d))$ is precisely

$$\frac{1}{N_j} \#\{n \in [N_j] : n, n+d, \ldots, n+(k-1)d \in A_j\}.$$

This instantly implies that A_j contains at least one k-term arithmetic progression with common difference d, for all sufficiently large j, contrary to assumption.

6.4. Cantor systems and ergodic theory

The name *Cantor system* is not standard terminology. The reason for this is that every Cantor system in fact has the structure of a measure-preserving system.

PROPOSITION 6.1. *Suppose that* $(X, \mathscr{R}, \mu, T)$ *is a Cantor system. Let* $\mathscr{B}$ *be the* σ*-algebra generated by* $\mathscr{R}$*. Then we may extend* μ *to a (countably additive) measure on* $\mathscr{B}$*, and thereby regard the Cantor system as the restriction of a measure-preserving system* $(X, \mathscr{B}, \mu, T)$*.*

The proof of this proposition is an instance of the Carathéodory Extension Theorem, discussed in Appendix A (it is not examinable in this course). The σ-algebra $\mathscr{B}$ is usually called the *Borel* σ-algebra, because when a natural metric topology is placed on X this does indeed consist of the Borel sets (see Exercises 1, Q10).

It turns out that the analogue of Theorem 6.2 holds in this measure-preserving system too.

THEOREM 6.3 (Furstenberg). *Consider the measure-preserving system* $(X, \mathscr{B}, \mu, T)$, *where* $X = \{0,1\}^{\mathbb{Z}}$ *and* $\mathscr{B}$ *is the Borel* σ*-algebra. Let* $E \in \mathscr{B}$ *be a set with* $\mu(E) > 0$. *Then there is some* $d \geqslant 1$ *such that* $\mu(E \cap T^{-d}E \cap \cdots \cap T^{-(k-1)d}E) > 0$.

In fact the same holds in any measure-preserving system, in a slightly stronger form.

THEOREM 6.4. *Suppose that* $(X, \mathscr{B}, \mu, T)$ *is an arbitrary measure-preserving system. Then for any* $f \in L^\infty(X)$ *with* $f \geqslant 0$ *and* f *not equal to 0 a.e. we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{0 \leqslant n < N} \int_X f(x) f(T^n x) \cdots f(T^{(k-1)n}x) d\mu(x) > 0.$$

Note that if $f = 1_E$ then

$$\int_X f(x) f(T^d x) \cdots f(T^{(k-1)d}x) d\mu = \mu(E \cap T^{-d}E \cap \cdots \cap T^{-(k-1)d}E).$$

Thus Theorem 6.4 implies Theorem 6.3, and in fact implies rather more: a positive proportion of all integers d satisfy the conclusion of that theorem.

It is convenient to give a name to the property being demanded in Theorem 6.4.

DEFINITION 6.2 (The SZ property). We say that a m.p.s. $(X, \mathscr{B}, \mu, T)$ has the SZ-property at level k if, for any function $f \in L^\infty(X)$ with $f \geqslant 0$ and f not equal to zero almost everywhere we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{0 \leqslant n < N} \int f(x) f(T^n x) \cdots f(T^{(k-1)n}x) d\mu(x) > 0.$$

Equivalently, writing $U = U_T$ for the Koopman operator $Uf(x) := f(Tx)$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{0 \leqslant n < N} \int f U^n f \ldots U^{(k-1)n} f d\mu > 0.$$

Thus Theorem 6.4 may be recast in the following form.

THEOREM 6.5. *Every measure-preserving system* $(X, \mathscr{B}, \mu, T)$ *has the SZ property at level* k.

We will discuss some aspects of the proof of this theorem in the rest of the course, taking the time to develop some interesting notions (such as weak-mixing) which come up, and proving the theorem for particular types of system. The broad structure of (one possible) proof of the general case is to isolate some chain of T-invariant sub-σ-algebras

$$\{0\} \subset \mathscr{B}_1 \subset \mathscr{B}_2 \subset \cdots \subset \mathscr{B}_M = \mathscr{B},$$

and then to prove, by induction on i, the statement that Theorem 6.4 holds for $(X, \mu, \mathscr{B}_i, T)$.

The crucial point to note is that these sub σ-algebras are *not* necessarily generated by subrings of the ring $\mathscr{R}$ of clopen sets, and to see these vital structures one must work in the setting of measure-preserving systems.

CHAPTER 7

Weak-mixing transformations

In this chapter we introduce a property that may be enjoyed by measure-preserving systems $(X, \mathscr{B}, \mu, T)$ called *weak-mixing*. This is an important class of measure-preserving systems in its own right, but our particular aim is to prove that all such systems have the SZ property, introduced in the last chapter.

In this chapter $L^2(X)$ always refers to the Hilbert space of *complex* square integrable functions. We adopt the convention that $n < N$ means $n \in \{0, 1, \dots, N-1\}$, and write $\mathbb{E}_{n<N}$ for $\frac{1}{N}\sum_{n<N}$. Occasionally we will write $o(1)$ for a quantity tending to 0 as $N \to \infty$.

7.1. Weak-mixing systems and their basic properties

Definition 7.1 (Weak-mixing systems). Suppose that $(X, \mathscr{B}, \mu, T)$ is a m.p.s. We say that this system is weak-mixing if, for all measurable sets $A, B \subset X$, we have

$$\lim_{N\to\infty} \mathbb{E}_{n<N}|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| = 0.$$

To give a little intuition, let us remark that weak-mixing implies ergodicity. Indeed if A is invariant, that is to say if $T^{-1}A = A$ up to measure zero, then the weak-mixing condition clearly implies that $\mu(A \cap B) = \mu(A)\mu(B)$ for all measurable B. Taking $B = A$, we see that $\mu(A) = 0$ or 1. Weak-mixing is a much stronger assumption than ergodicity, however. Our basic example of an ergodic system, the irrational circle rotation, fails to be weak-mixing as can be seen by taking $A = B = I$ for a suitable interval I (to establish this rigorously, for all irrational circle rotations, is on Sheet 4).

On the other hand the doubling map $Tx = 2x (\mathrm{mod}\, 1)$ *is* weakly-mixing. In fact it has a strong property called *mixing* (or sometimes *strongly mixing*:

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

for all measurable A, B. This essentially follows from the "L^1-style" proof that the doubling map is ergodic, given in Section 3.7. There, we showed that $\mu(T^{-n'}A \cap B) = \mu(A)\mu(B)$ whenever $n' \geqslant n$, if B is a union of dyadic intervals at level n. The claim now follows by approximating an arbitrary measurable B using unions of dyadic intervals, as done in Section 3.7 (and see also Lemma A.2).

It is traditional, and I think rather pleasant, to supply a list of properties that are equivalent to weak-mixing. Here, $U = U_T$ denotes the Koopman operator, namely $Uf(x) := f(Tx)$.

PROPOSITION 7.1 (Equivalent notions of weak-mixing). *Suppose that* $(X, \mu, \mathscr{B}, T)$ *is a measure-preserving system. Then the following are equivalent.*

(i) T *is weak-mixing;*
(ii) *Every* $f \in L^2(X)$ *with* $\int f d\mu = 0$ *satisfies* $\lim_{N\to\infty} \mathbb{E}_{n<N} |\langle f, U^n f\rangle| = 0$*;*
(iii) $\lim_{N\to\infty} \mathbb{E}_{n<N} |\langle f, U^n g\rangle - \int f d\mu \int \overline{g} d\mu| = 0$ *for all* $f, g \in L^2(X)$*;*
(iv) $T \times T$ *is weakly-mixing;*
(v) $T \times \tilde{T}$ *is ergodic on* $X \times \tilde{X}$ *for any ergodic m.p.s.* $(\tilde{X}, \tilde{\mu}, \tilde{\mathscr{B}}, \tilde{T})$*;*
(vi) $T \times T$ *is ergodic.*

PROOF. Evidently (iii) implies (ii). Conversely (ii) implies (iii) in the case $\int f d\mu = \int g d\mu = 0$ by using the "depolarization identity"

$$\langle f, U^n g\rangle = \frac{1}{4}\sum_{m=0}^{3} \langle g + i^m f, U^n(g + i^m f)\rangle i^{-m}.$$

The general case then follows straightforwardly from the observation that

$$\langle f - \int f d\mu, U^n(g - \int g d\mu)\rangle = \langle f, U^n g\rangle - \int f d\mu \int g d\mu.$$

Statement (iii) obviously implies (i), and conversely (i) implies (iii) by approximating f and g by simple-measurable functions in the usual way.

To see that (i) implies (iv), suppose that T is weak-mixing. We want to show that $T \times T$ is also weak-mixing, and so we must show that if $E, F \in \mathscr{B} \times \mathscr{B}$ then

$$\mathbb{E}_{n<N}(\mu \times \mu)((T \times T)^{-n} E \cap F) \to 0.$$

Suppose first that $E = A \times C$, $F = B \times D$ with $A, B, C, D \in \mathscr{B}$. Then we have

$$\begin{aligned}
&\mathbb{E}_{n<N}(\mu \times \mu)((T \times T)^{-n} E \cap F)\\
&\mathbb{E}_{n<N}|(\mu \times \mu)((T \times T)^{-n}(A \times C) \cap (B \times D) - \mu(A)\mu(B)\mu(C)\mu(D)|\\
&= \mathbb{E}_{n<N}|\mu(T^{-n}A \cap B)\mu(T^{-n}C \cap D) - \mu(A)\mu(B)\mu(C)\mu(D)|\\
&= \mathbb{E}_{n<N}\big|\big(\mu(T^{-n}A \cap B) - \mu(A)\mu(B)\big)\mu(T^{-n}C \cap D) + \mu(A)\mu(B)\big(\mu(T^{-n}C \cap D) - \mu(C)\mu(D)\big)\big|\\
&\leqslant \mathbb{E}_{n<N}\big|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)\big| + \mathbb{E}_{n<N}\big|\mu(T^{-n}C \cap D) - \mu(C)\mu(D)\big|\\
&\to 0.
\end{aligned}$$

The case of general E, F now follows by approximating each of these sets with a finite union of product sets $A \times C$, $B \times D$: see Appendix A.

That (iv) implies (i) is very straightforward and is left to the reader.

To see that (i) implies (v), suppose that T is weak-mixing and that $\tilde{T}$ is ergodic. We want to show that $T \times \tilde{T}$ is ergodic, to which end it suffices (and is necessary) to show that

$$\mathbb{E}_{n<N}(\mu \times \tilde{\mu})((T \times \tilde{T})^{-n} E \cap F) \to (\mu \times \tilde{\mu})(E)(\mu \times \tilde{\mu})(F)$$

for all measurable $E, F \in \mathscr{B} \times \tilde{\mathscr{B}}$. (Then take $E = F$ to be some $(T \times \tilde{T})$-invariant set and conclude that $(\mu \times \tilde{\mu})(E) = 0$ or 1.) One again we show this first in the special case $E = A \times C$ and $F = B \times D$, where $A, B \in \mathscr{B}$ and $C, D \in \tilde{\mathscr{B}}$. The general case again follows by an approximation argument.

The special case we need to prove is the following statement:

$$\mathbb{E}_{n<N}(\mu \times \tilde{\mu})((T \times \tilde{T})^{-n}(A \times C) \cap (B \times D)) \to \mu(A)\mu(B)\tilde{\mu}(C)\tilde{\mu}(D). \tag{7.1}$$

The left-hand side minus the right is simply

$$\mathbb{E}_{n<N}\big(\mu(T^{-n}A \cap B)\tilde{\mu}(\tilde{T}^{-n}C \cap D) - \mu(A)\mu(B)\tilde{\mu}(C)\tilde{\mu}(D)\big),$$

which equals

$$\begin{aligned}\mathbb{E}_{n<N}(\mu(T^{-n}A \cap B) &- \mu(A)\mu(B))\tilde{\mu}(\tilde{T}^{-n}C \cap D) \\ &+ \mu(A)\mu(B)(\tilde{\mu}(\tilde{T}^{-n}C \cap D) - \tilde{\mu}(C)\tilde{\mu}(D)).\end{aligned}$$

This is bounded by

$$\mathbb{E}_{n<N}|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| + \big|\mathbb{E}_{0\leqslant n<N}(\tilde{\mu}(\tilde{T}^{-n}C \cap D) - \tilde{\mu}(C)\tilde{\mu}(D))\big|$$

(note carefully the position of the modulus signs). Now since T is weak-mixing the first term tends to zero. The second term is

$$\langle \tilde{S}_N 1_C - \int 1_C d\tilde{\mu}, 1_D \rangle,$$

which by the Cauchy-Schwarz inequality is bounded by $\|\tilde{S}_N 1_C - \int 1_C d\tilde{\mu}\|_2$. By the mean ergodic theorem and the fact that $\tilde{T}$ is ergodic, this tends to 0 as $N \to \infty$.

To see that (v) implies (vi), first take $(\tilde{X}, \tilde{\mu}, \tilde{\mathscr{B}}, \tilde{T})$ to be the trivial one-point system to conclude that T itself is ergodic. Then, taking $\tilde{T} = T$, we see that $T \times T$ is ergodic.

Finally we must show that (vi) implies (i), to which end it suffices (by Cauchy-Schwarz) to show that

$$\mathbb{E}_{n<N}|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)|^2 \to 0.$$

Expanding out the square, this may (with a little effort) be rewritten as

$$\mathbb{E}_{n<N}(\mu\times\mu)((T\times T)^{-n}(A\times A)\cap(B\times B))- \\ 2\mu(A)\mu(B)\mathbb{E}_{n<N}(\mu\times\mu)((T\times T)^{-n}(A\times X)\cap(B\times X)) \\ +\mu(A)^2\mu(B)^2.$$

Writing $S_N^{\times}$ for the time averages with respect to $T\times T$, this may be written as

$$\langle S_N^{\times}1_{A\times A},1_{B\times B}\rangle-2\mu(A)\mu(B)\langle S_N^{\times}1_{A\times X},1_{B\times X}\rangle+\mu(A)^2\mu(B)^2,$$

where the inner products are on $X\times X$. By two applications of the mean ergodic theorem applied to $T\times T$ (which tells us that $S_N^{\times}1_{A\times A}\to\mu(A)^2$ and $S_N^{\times}1_{A\times X}\to\mu(A)$ in L^2) and the Cauchy-Schwarz inequality, this does indeed tend to zero. □

Remark. In the above argument we applied the mean ergodic theorem three times. A close inspection reveals that all we needed was the fact that $S_Nf\to\int f\,d\mu$ in *weak* L^2, which means that $\langle S_Nf-\int f\,d\mu,g\rangle=0$ for all $g\in L^2(X)$. This statement admits of a much "softer" proof than the mean ergodic theorem, at least if one knows a little functional analysis: see Exercise sheet 4.

An important fact is that weak-mixing is equivalent to a further property: that the Koopman operator U has no non-constant eigenfunctions. That is, if $f\in L^2(X)$ and if $f(Tx)=\lambda f(x)$ for a.e. x then $\lambda=1$ and f is constant a.e. We will not require this fact in these notes; the proof is not difficult, but does require a small amount of machinery in the form of the Bochner-Herglotz spectral theorem. For comparison note that T is ergodic if and only if the only eigenfunctions with $\lambda=1$ are the constants.

7.2. Multiple recurrence properties for weak-mixing systems

Our aim now is to establish that weakly-mixing systems have the SZ property.

THEOREM 7.1. *Let $(X,\mathscr{B},\mu,T)$ be an invertible weakly-mixing m.p.s. Then this system has the* SZ *property at level k for every k.*

We will in fact establish the following result.

PROPOSITION 7.2. *Let $r\geqslant 1$. Let $(X,\mathscr{B},\mu,T)$ be a weak-mixing m.p.s. Then*

$$\lim_{N\to\infty}\mathbb{E}_{n<N}U^nf_1\dots U^{rn}f_r\to\prod_{i=1}^{r}(\int f_i\,d\mu)$$

in $L^2(X)$, for every choice of $f_1,\dots,f_r\in L^\infty(X)$. The existence of the limit on the left is part of the claim.

The deduction of the SZ property at level k from the case $r = k-1$ of this proposition is a straightforward application of Cauchy-Schwarz. Indeed, taking $f_1 = f_2 = \cdots = f_{k-1} = g$ in the proposition we have

$$\begin{aligned}\mathbb{E}_{n<N}\int fU^n f\dots U^{(k-1)n}f d\mu - (\int f)^k &= \langle f, \mathbb{E}_{n<N}U^n f\dots U^{(k-1)n}f - (\int f)^{k-1}\rangle \\ &\leqslant \|f\|_2\|\mathbb{E}_{n<N}U^n f\dots U^{(k-1)n}f - (\int f)^{k-1}\|_2 \\ &\to 0\end{aligned}$$

as $N\to\infty$. In particular if $f \geqslant 0$ and f is not 0 a.e. then

$$\liminf_{N\to\infty}\mathbb{E}_{n<N}\int fU^n f\dots U^{(k-1)n}f d\mu = (\int f)^k > 0,$$

which is precisely the SZ property at level k.

The key ingredient in the proof of Proposition 7.2 is a result called the van der Corput lemma.

PROPOSITION 7.3. *Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence of vectors in some Hilbert space V, and that $\|x_n\| \leqslant 1$ for all n. Then*

$$\limsup_{N\to\infty}\|\mathbb{E}_{n<N}x_n\|^2 \leqslant \limsup_{H\to\infty}\mathbb{E}_{h<H}|\limsup_{N\to\infty}\mathbb{E}_{n<N}\langle x_n, x_{n+h}\rangle|.$$

PROOF. If H is some paranmeter, we write $o_{H;N\to\infty}(1)$ for a quantity tending to 0 (in the norm on V), the rate of convergence depending only on H. Different instances of the notation may indicate different quantities.

Fix H for the moment and suppose that $h < H$. Then we have the "shifting principle"

$$\mathbb{E}_{n<N}x_n = \mathbb{E}_{n<N}x_{n+h} + o_{H;N\to\infty}(1). \tag{7.2}$$

Indeed,

$$\|\mathbb{E}_{n<N}x_n - \mathbb{E}_{n<N}x_{n+h}\| = \|\frac{1}{N}(x_0 + \cdots + x_{h-1} - x_N - \cdots - x_{N+h-1})\| < \frac{2H}{N}.$$

Taking averages over $h < H$ yields

$$\mathbb{E}_{n<N}x_n = \mathbb{E}_{n<N}\mathbb{E}_{h<H}x_{n+h} + o_{H;N\to\infty}(1)$$

and so

$$\|\mathbb{E}_{n<N}x_n\|^2 = \|\mathbb{E}_{n<N}\mathbb{E}_{h<H}x_{n+h}\|^2 + o_{H;N\to\infty}(1). \tag{7.3}$$

(Indeed if $v = v' + w$ with $\|v\|, \|v'\| \leqslant 1$ then $\|v\|^2 - \|v'\|^2 = 2\langle v', w\rangle + \|w\|^2$, and $|\langle v', w\rangle| \leqslant \|w\|$ by the Cauchy-Schwarz inequality, so $\big|\|v\|^2 - \|v'\|^2\big| \leqslant 2\|w\| + \|w\|^2$.)

By the Cauchy-Schwarz inequality it follows from (7.3) that

$$\|\mathbb{E}_{n<N} x_n\|^2 \leqslant \mathbb{E}_{n<N} \|\mathbb{E}_{h<H} x_{n+h}\|^2 + o_{H;N\to\infty}(1).$$

Expanding out the square and swapping the order of summation gives

$$\|\mathbb{E}_{n<N} x_n\|^2 \leqslant \mathbb{E}_{h_1,h_2<H} |\mathbb{E}_{n<N} \langle x_{n+h_1}, x_{n+h_2}\rangle| + o_{H;N\to\infty}(1). \tag{7.4}$$

Note that by another application of the shifting principle

$$\mathbb{E}_{n<N} \langle x_{n+h_1}, x_{n+h_2}\rangle = \mathbb{E}_{n<N} \langle x_n, x_{n+h_2-h_1}\rangle + o_{H;N\to\infty}(1), \tag{7.5}$$

and also that

$$\begin{aligned} \mathbb{E}_{n<N} \langle x_{n+h_1}, x_{n+h_2}\rangle &= \mathbb{E}_{n<N} \langle x_{n+h_1-h_2}, x_n\rangle + o_{H;N\to\infty}(1) \\ &= \overline{\mathbb{E}_{n<N} \langle x_n, x_{n+h_1-h_2}\rangle} + o_{H;N\to\infty}(1). \end{aligned} \tag{7.6}$$

Substituting into (7.4) (using (7.5) when $h_2 \geqslant h_1$ and (7.6) when $h_2 < h_1$) we obtain

$$\|\mathbb{E}_{n<N} x_n\|^2 \leqslant \frac{1}{H^2} \sum_{h<H} w(h) |\mathbb{E}_{n<N} \langle x_n, x_{n+h}\rangle| + o_{H;N\to\infty}(1), \tag{7.7}$$

where $w(h)$ is the number of ways of writing $h = h_1 - h_2$ or $h = h_2 - h_1$ with $h_1, h_2 < H$, thus

$$w(h) = \begin{cases} H & h = 0 \\ 2(H-h) & 1 \leqslant h \leqslant H-1. \end{cases}$$

Taking limsups as $N \to \infty$ in (7.7), the $o(1)$ terms disappear and we obtain

$$\limsup_{N\to\infty} \|\mathbb{E}_{n<N} x_n\|^2 \leqslant \frac{1}{H^2} \sum_{h<H} w(h) \limsup_{N\to\infty} |\mathbb{E}_{n<N} \langle x_n, x_{n+h}\rangle|. \tag{7.8}$$

Finally, we must take limsups as $H \to \infty$. We claim that for any bounded sequence (y_h) we have

$$\limsup_{H\to\infty} \frac{1}{H^2} \sum_{h<H} w(h) y_h \leqslant \limsup_{H\to\infty} \mathbb{E}_{h<H} y_h. \tag{7.9}$$

Proposition 7.3 follows immediately from this claim and (7.8) upon taking $y_h = \limsup_{N\to\infty} |\mathbb{E}_{n<N} \langle x_n, x_{n+h}\rangle|$.

It remains to prove the claim. To do this, note the identity

$$\sum_{h<H} w(h) y_h = -H y_0 + 2 \sum_{H'=1}^{H} \sum_{h<H'} y_h. \tag{7.10}$$

Suppose that $\mathbb{E}_{h<H} y_h \leqslant C$ for all $H \geqslant H_0$. Then from (7.10) we have

$$\begin{aligned}\frac{1}{H^2}\sum_{h<H} w(h)y_h &= -\frac{y_0}{H} + 2\sum_{H'=1}^{H_0-1}\sum_{h<H'} y_h + 2\sum_{H'=H_0}^{H}\sum_{h<H'} y_h \\ &\leqslant \frac{1}{H} + 2\frac{H_0^2}{H^2} + \frac{2}{H^2}\sum_{H'=H_0}^{H} CH' \\ &\to C\end{aligned}$$

as $H \to \infty$.

The claim follows. □

Proof of Proposition 7.2. We proceed by induction on r. The case $r = 1$ is simply the mean ergodic theorem; note that T, being weak-mixing, is certainly ergodic. Suppose now that $r \geqslant 2$. The case in which f_r is constant reduces to the case $r-1$. Hence, replacing f_r by $f_r - \int f_r d\mu$, we may suppose in what follows that $\int f_r d\mu = 0$. We may also assume without loss of generality that $\|f_i\|_\infty \leqslant 1$ for all $i = 1, \ldots, r$. In what follows we will consider various "derivatives" $\Delta_h f_i := f_i \overline{U^h f_i}$. Note that $\|\Delta_h f_i\|_\infty \leqslant 1$ and hence $\|\Delta_h f_i\|_2 \leqslant 1$ for all h, i; note that this might not be the case if we assumed, in the statement of Proposition 7.2, only a bound on $\|f_i\|_2$.

Our task is to show that

$$\mathbb{E}_{n<N} U^n f_1 \ldots U^{rn} f_r \to 0$$

in L^2.

Taking $x_n := U^n f_1 \ldots U^{rn} f_r$ in the van der Corput lemma, and noting that

$$\langle x_n, x_{n+h}\rangle = \int (U^n \Delta_h f_1)\cdots(U^{rn}\Delta_{rh} f_r)d\mu,$$

it is enough to show that

$$\limsup_{H\to\infty} \mathbb{E}_{h<H} \limsup_{N\to\infty} |\mathbb{E}_{n<N} \int (U^n \Delta_h f_1)\cdots(U^{rn}\Delta_{rh} f_r)d\mu| = 0. \tag{7.11}$$

Now by the T-invariance of μ (and the fact that T is invertible) we have

$$\begin{aligned}&\int (U^n \Delta_h f_1)\cdots(U^{rn}\Delta_{rh} f_r)d\mu \\ &\qquad = \int (\Delta_h f_1)(U^n \Delta_{2h} f_2)\cdots(U^{(r-1)n}\Delta_{rh} f_r)d\mu,\end{aligned}$$

and hence by Cauchy-Schwarz

$$\begin{aligned}|\mathbb{E}_{n<N}\int(U^n\Delta_h f_1)\cdots(U^{rn}\Delta_{rh}f_r)d\mu| \\ &= |\int(\Delta_h f_1)\mathbb{E}_{n<N}(U^n\Delta_{2h}f_2)\cdots(U^{(r-1)n}\Delta_{rh}f_r)d\mu| \\ &\leqslant \|\mathbb{E}_{n<N}(U^n\Delta_{2h}f_2)\cdots(U^{(r-1)n}\Delta_{rh}f_r)\|_2.\end{aligned}$$

Taking limits as $N\to\infty$, and using the induction hypothesis, we have

$$\begin{aligned}\limsup_{N\to\infty}|\mathbb{E}_{n<N}\int(U^n\Delta_h f_1)\cdots(U^{rn}\Delta_{rh}f_r)d\mu| \\ &\leqslant |\prod_{j=2}^{r}(\int\Delta_{jh}f_j)| \leqslant |\int\Delta_{rh}f_r| = |\langle f_r, U^{rh}f_r\rangle|.\end{aligned}$$

To establish the desired statement (7.11), then, we need only show that

$$\limsup_{H\to\infty}\mathbb{E}_{h<H}|\langle f_r, U^{rh}f_r\rangle| = 0. \tag{7.12}$$

However, f_r has integral zero and hence satisfies item (ii) of Proposition 7.1: this means that

$$\limsup_{H'\to\infty}\mathbb{E}_{h'<H'}|\langle f_r, U^{h'}f_r\rangle| = 0,$$

from which (7.12) follows immediately upon taking $H' = rH$ and ignoring those h' which are not divisible by r. □

CHAPTER 8

Compact systems

8.1. Almost-periodic functions and compact systems

In the last set of notes we discussed weak-mixing systems. We begin this set of notes by introducing a very different class of system, the *compact* systems. We begin by defining the notion of an *almost periodic function.* Here, as in previous chapters, $(X, \mathscr{B}, \mu, T)$ will always be a measure-preserving system and $U = U_T : L^2(X) \to L^2(X)$ is the Koopman operator defined by $Uf(x) := f(Tx)$.

Definition 8.1 (Almost periodic functions). Let $(X, \mathscr{B}, \mu, T)$ be an measure-preserving system and suppose that $f \in L^2(X)$. Then we say that f is almost periodic if the forward orbit $(U^n f)_{n \geqslant 0}$ is precompact in $L^2(X)$, that is to say has compact closure.

To begin to get a handle on this definition, note first that $L^2(X)$ itself will usually *not* be compact. For example, the functions $f_n(x) = e^{2\pi i n x}$ all lie in $L^2(\mathbb{R}/\mathbb{Z})$, but since $\|f_n - f_m\|_2 = \sqrt{2}$ whenever $n \neq m$, this sequence of functions has no convergent subsequence. Since $L^2(X)$ is a complete metric space, it follows from basic material in analysis that a set $S \subset L^2(X)$ being precompact is equivalent to that set being *totally bounded*: for every $\varepsilon > 0$ there exists a finite collection $g_1, \dots, g_J \in S$ such that for every $f \in S$ there is a j such that $\|f - g_j\|_2 \leqslant \varepsilon$.

Definition 8.2 (Compact systems). Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. Then we say that this system is compact if every function in $L^2(X)$ is almost periodic.

The archetypal example of a compact system is a circle rotation, as we shall see in Lemma 8.2 below. Before establishing this, however, let us prove some basic properties of almost periodic functions.

8.2. Basic properties of almost-periodic functions

In this section we collect together some basic properties of almost periodic functions, and then use this information to conclude that circle rotations are compact.

Lemma 8.1. *Write* AP(X) *for the collection of almost periodic functions in* $L^2(X)$*. Then*

(i) $\mathrm{AP}(X)$ *is a* T*-invariant;*
(ii) $\mathrm{AP}(X)$ *is closed in the* L^2*-norm;*
(iii) $\mathrm{AP}(X)$ *is a vector subspace of* $L^2(X)$*;*
(iv) *If* $f \in \mathrm{AP}(X)$ *then* $f_+ := \max(f, 0) \in \mathrm{AP}(X)$.

Proof. (i) It is clear that $\mathrm{AP}(X)$ is T-invariant since the forward orbit of Uf is contained in that of f.

(ii) Suppose that f_m, $m = 1, 2, \dots$ are in $\mathrm{AP}(X)$ and that $f_m \to f$ in L^2. Let $\varepsilon > 0$. Then there is some m such that $\|f_m - f\|_2 \leqslant \varepsilon/3$. Since the forward orbit $(U^n f_m)_{n \geqslant 0}$ is precompact there is some $J = J(\varepsilon)$ such that, for any $n \geqslant 0$, there is an $n' \leqslant J(\varepsilon)$ for which $\|U^n f_m - U^{n'} f_m\|_2 \leqslant \varepsilon/3$. Since U is an isometry, we have $\|U^n f - U^n f_m\|_2 \leqslant \varepsilon/3$ and $\|U^{n'} f - U^{n'} f_m\|_2 \leqslant \varepsilon/3$. By the triangle inequality it follows that $\|U^n f - U^{n'} f\|_2 \leqslant \varepsilon$. Since ε was arbitrary, this means that $f \in \mathrm{AP}(X)$, and so $\mathrm{AP}(X)$ is indeed topologically closed.

(iii) Closure under scalar multiplication is obvious, so we just prove closure under addition. Suppose, then, that $f, g \in \mathrm{AP}(X)$. Take functions $f_i, g_j \in L^2(X)$, $1 \leqslant i, j \leqslant k$, with the property that for any n there are i, j such that $\|U^n f - f_i\|_2, \|U^n g - g_j\|_2 \leqslant \varepsilon/4$. Thus for any n there are i, j such that

$$\|U^n(f + g) - (f_i + g_j)\|_2 \leqslant \varepsilon/2. \tag{8.1}$$

For each i, j take n_{ij} to be the smallest n corresponding to a given i, j, if such an n exists. Then if (8.1) holds we have

$$\|U^n(f + g) - U^{n_{ij}}(f + g)\|_2 \leqslant \varepsilon,$$

and so indeed $f + g \in \mathrm{AP}(X)$.

(iv) This is straightforward. Suppose that $f \in \mathrm{AP}(X)$. Then there is $J(\varepsilon)$ such that for any n there is $n' \leqslant J(\varepsilon)$ for which $\|U^n f - U^{n'} f\|_2 \leqslant \varepsilon$. But then, noting that $(U^i f)_+ = U^i(f_+)$, we see that $\|U^n f_+ - U^{n'} f_+\|_2 \leqslant \varepsilon$ as well. □

Lemma 8.2. *Let* $(X, \mathscr{B}, \mu, T)$ *be a circle rotation. That is, suppose that* $X = \mathbb{R}/\mathbb{Z}$*, that* $\mathscr{B}$ *is the Borel* σ*-algebra, that* μ *is Lebesgue measure and that* $T : X \to X$ *for the map* $x \mapsto x + \alpha (\mathrm{mod}\, 1)$ *for some* $\alpha \in R$*. Then the system* $(X, \mathscr{B}, \mu, T)$ *is compact.*

Proof. By the previous lemma we need only exhibit a collection of almost periodic functions which are dense in $L^2(X)$. A suitable set of such functions is the set of trigonometric polynomials. Since $\mathrm{AP}(X)$ is a vector subspace, it is enough to check that each exponential function $f(x) = e^{2\pi i \alpha m x}$ is almost-periodic. This is clear when $\alpha \in \mathbb{Q}$, so suppose that α is irrational. Then we have

$$\|U^n f - U^{n'} f\|_2 = |e^{2\pi i \alpha m (n - n')} - 1| \leqslant 10 \|\alpha m (n - n')\|_{\mathbb{R}/\mathbb{Z}},$$

where $\|t\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from t to the nearest integer. Now the points $(\alpha m n')_{n' \in \mathbb{N}}$ are dense in $\mathbb{R}/\mathbb{Z}$, and so for fixed ε there is some $J = J(\varepsilon)$ such that the points $(\alpha m n')_{n'=1,\ldots,J}$ come within $\varepsilon/10$ of every point in $\mathbb{R}/\mathbb{Z}$. Therefore for any n there is some $n' \leqslant J$ such that $\|U^n f - U^{n'} f\|_2 \leqslant \varepsilon$. This proves the lemma. □

8.3. Multiple recurrence for compact systems

PROPOSITION 8.1. *Suppose that $(X, \mathscr{B}, \mu, T)$ is a measure-preserving system. Let $f \in L^\infty(X)$ be almost-periodic, and suppose that $f \geqslant 0$ and that f is not 0 a.e. Then*

$$\liminf_{N \to \infty} \mathbb{E}_{n<N} \int f \cdot U^n f \cdot \cdots \cdot U^{(k-1)n} f d\mu > 0.$$

In particular if $(X, \mathscr{B}, \mu, T)$ is a compact system then it satisfies the SZ *property at level k for every $k \geqslant 2$.*

PROOF. Suppose that $f \in L^\infty(X)$ is a function with $f \geqslant 0$ and f not zero a.e. Without loss of generality we may suppose that $\|f\|_\infty = 1$. Set $\varepsilon := \frac{1}{2} \int f^k d\mu$; then $\varepsilon > 0$. Set $\varepsilon' := \varepsilon / k2^k$.

Now since f is almost periodic the closure of the forward orbit $(U^n f)_{n \in \mathbb{N}}$ is compact, and so there is some $J = J(\varepsilon)$ such that, for every $n \in \mathbb{N}$, there is some $n' \leqslant J$ such that $\|U^n f - U^{n'} f\|_2 \leqslant \varepsilon$. Since U is an isometry, this means that $\|U^{n-n'} f - f\|_2 \leqslant \varepsilon'$. In other words, the maximum gap between consecutive elements of the set $R := \{n \in \mathbb{N} : \|U^n f - f\|_2 \leqslant \varepsilon'\}$ is bounded by J, a property known as R being *syndetic*.

Let $n \in R$. Since U is an isometry we have $\|U^{jn} f - U^{(j+1)n} f\|_2 \leqslant \varepsilon'$ for all positive integers j, and hence by the triangle inequality that $\|f - U^{jn} f\|_2 \leqslant k\varepsilon'$ for $j = 1, \ldots, k-1$. Write $g_j := f - U^{jn} f$; then we have

$$\int f \cdot U^n f \cdots U^{(k-1)n} f \, d\mu = \int f (f - g_1) \ldots (f - g_{k-1}) \, d\mu.$$

This may be split as $\int f^k \, d\mu$ plus a sum of at most 2^k other terms, each of the form $\langle g_j, F \rangle$ for some function F with $\|F\|_\infty \leqslant 1$. Each such term is bounded by ε. Therefore

$$\int f \cdot U^n f \cdots U^{(k-1)n} f \, d\mu \geqslant \int f^k \, d\mu - \varepsilon = \frac{1}{2} \int f^k d\mu.$$

Since this is true for all $n \in R$, a syndetic set, the result follows. □

CHAPTER 9

The case $k = 3$ of Szemerédi's theorem

9.1. Weak mixing vs compactness

It is not the case that every system is either weak-mixing or compact (the 2-dimensional skew torus is an example: see Sheet 4). However, it turns out that a system which fails to be weakly mixing does at least have a small piece of compactness.

THEOREM 9.1. *Suppose that $(X, \mathscr{B}, \mu, T)$ is not weakly mixing. Then there is a non-constant almost periodic function in $L^2(X)$.*

One way of proving this is to show that if $(X, \mathscr{B}, \mu, T)$ is not weakly-mixing then the Koopman operator $U = U_T$ has a non-trivial eigenfunction: that is, there is some $f \in L^2(X)$ and a λ such that $Uf = \lambda f$, i.e. $f(Tx) = \lambda f(x)$ for a.e. x. An eigenfunction for the Koopman operator is necessarily almost-periodic: every iterate $U^n f$ comes within ε in the L^2-norm of one of the functions $e^{2\pi i j/M} f(x)$, $j = 1, \ldots, M$, provided $M = M(\varepsilon)$ is sufficiently large. Unfortunately the proof that there is an eigenfunction in the non weak-mixing case does require a basic result in spectral theory, the Bochner-Hergoltz spectral theorem. While this is, in a sense, just a kind of limiting version of the Fourier analysis results of Appendix B, it falls outside the scope of this course. The following more elementary proposition does not give an eigenfunction, but it does yield an almost-periodic function.

PROPOSITION 9.1. *Suppose that $(X, \mathscr{B}, \mu, T)$ is a measure-preserving system. Suppose that $f \in L^2(X)$ is not weak-mixing. Then there is an almost periodic function $\phi \in L^2(X)$ such that $\langle f, \phi \rangle \neq 0$.*

PROOF. Note that this proposition has little content when $\int f \neq 0$, as in that case we may take $\phi = 1$. Suppose, then, that $\int f = 0$. Without loss of generality we may normalise so that $\|f\|_2 \leqslant 1$.

Define a function $K \in L^2(X \times X)$ by

$$K(x, y) = \lim_{N \to \infty} \mathbb{E}_{n<N} U^n f(x) U^n f(y).$$

The limit exists by the L^2-ergodic theorem in $(X \times X, \mathscr{B} \times \mathscr{B}, \mu \times \mu, T \times T)$: in fact K is the orthogonal projection of $f \otimes f \in L^2(X \times X)$ to the space of $(T \times T)$-invariant functions. In particular, K is $(T \times T)$-invariant.

Now define

$$\phi(x) := \int K(x,y)f(y)d\mu(y).$$

We have

$$\begin{aligned}\langle f, \phi\rangle &= \int f(x)\big(\int \lim_{N\to\infty} \mathbb{E}_{n<N} U^n f(x) U^n f(y)\big) f(y) d\mu(y)\big) d\mu(x) \\ &= \lim_{N\to\infty} \mathbb{E}_{n<N}\big(\int f(x)U^n f(x) d\mu(x)\big)\big(\int f(y) U^n f(y) d\mu(y)\big) \\ &= \lim_{N\to\infty} \mathbb{E}_{n<N} |\langle f, U^n f\rangle|^2.\end{aligned}$$

Since f is not a weak-mixing function, the right hand side is not zero.

To conclude the proof, we must show that ϕ is almost periodic. To do this, first note that

$$\begin{aligned}U^n\phi(x) &= \int K(T^n x, y) f(y) d\mu(y) \\ &= \int K(T^n x, T^n y) U^n f(y) d\mu(y) \\ &= \int K(x,y) U^n f(y) d\mu(y),\end{aligned}$$

where the last step follows from the fact that K is $(T\times T)$-invariant. Therefore the orbit $(U^n\phi)_{n\in\mathbb{N}}$ is contained in the set

$$\{\int K(x,y) g(y) d\mu(y) : \|g\|_2 \leqslant 1.\} \tag{9.1}$$

We are required to show that the set on the left has compact closure in $L^2(X)$. This is a well-known type of statement from Hilbert space theory: the operator $\Psi_K : L^2(X) \to L^2(X)$ defined by $\Psi_K g(y) := \int K(x,y) g(y) d\mu(y)$ is Hilbert-Schmidt and hence compact. However, for our purposes here we need essentially no general theory since the definition of Ψ_K makes it clear that it is a limit of finite rank operators. Let us turn to the details.

We note that if $L \in L^2(X\times X)$ then

$$\|\Psi_L g\|_2 \leqslant \|L\|_2 \|g\|_2.$$

Indeed,

$$\begin{aligned}\|\Psi_L g\|_2^2 &= \int |\int L(x,y) g(y) dy|^2 dx \\ &\leqslant \int \big(\int |L(x,y)|^2 dy\big)\big(\int |g(y)|^2 dy\big) dx \\ &= \|L\|_2^2 \|g\|_2^2\end{aligned}$$

by the Cauchy-Schwarz inequality. Now $K_N \to K$ in $L^2(X\times X)$, where

$$K_N(x,y) := \mathbb{E}_{n<N} U^n f(x) U^n f(y).$$

Choosing N so large that $\|K - K_N\|_2 \leqslant \varepsilon$, it follows that if $\|g\|_2 \leqslant 1$ then

$$\|\Psi_{K_N} g - \Psi_K g\|_2 \leqslant \varepsilon$$

whenever $\|g\|_2 \leqslant 1$. Thus every point of the set (9.1) lies within ε of a point of the set

$$\{\int K_N(x,y)g(y)d\mu(y) : \|g\|_\infty \leqslant 1\}.$$

However,

$$\int K_N(x,y)g(y)d\mu(y) = \mathbb{E}_{n<N} c_n U^n f(x),$$

where $c_n = \langle U^n f, g\rangle$, so this function lies in the convex unit ball spanned by the finite collection of functions $(U^n f)_{n<N}$, a compact subset of $L^2(X)$. (In other words, Ψ_{K_N} is a finite rank operator.)

It follows that (9.1) is covered by finitely many balls of radius ε, thereby concluding the proof. □

9.2. A decomposition theorem

A slight modification of the van der Corput lemma. In what follows, we need the van der Corput lemma, Lemma **??**. Although the lemma was only stated for functions b_n satisfying $|b_n(x)| \leqslant 1$ pointwise, it turns out that almost the same argument gives the same conclusion under the weaker assumption $\|b_n\|_2 \leqslant 1$ (see Exercise sheet 4. Admittedly I should have done this earlier in the course; this modification is not examinable.) Here is the restated lemma.

LEMMA 9.1. *Suppose that $(X, \mathscr{B}, \mu)$ is a probability space and that $(b_n)_{n\geqslant 0}$ is a sequence of functions in the unit ball of $L^2(X)$. Suppose that*

$$\limsup_{H\to\infty} \mathbb{E}_{h<H} \limsup_{n<N} |\mathbb{E}_{n<N}\langle b_n, b_{n+h}\rangle| = 0.$$

Then $\mathbb{E}_{n<N} b_n \to 0$ in L^2.

In Proposition 7.1 (ii) we looked at functions $f \in L^2(X)$ with the property that $\int f d\mu = 0$ and $\lim_{N\to\infty} \mathbb{E}_{n<N}|\langle f, U^n f\rangle| = 0$. In that proposition, $(X, \mathscr{B}, \mu, T)$ was a weak-mixing system, and the conclusion was that all functions f with integral zero have this property. Following Tao, we shall call these "weak-mixing functions", and write $\mathrm{WM}(X)$ for the set of all weak-mixing functions.

LEMMA 9.2. *Suppose that $f \in \mathrm{WM}(X)$, and let $g \in L^2(X)$ be arbitrary. Then $\mathbb{E}_{n<N}|\langle g, U^n f\rangle| \to 0$ as $N \to \infty$.*

PROOF. By the Cauchy-Schwarz inequality, it is enough to prove that

$$\mathbb{E}_{n<N}|\langle g, U^n f\rangle|^2 \to 0 \tag{9.2}$$

(which is a new fact for us even when $g = f$). Since

$$\mathbb{E}_{n<N}|\langle g, U^n f\rangle|^2 = \langle g, \mathbb{E}_{n<N} U^n f \langle g, U^n f\rangle\rangle,$$

it is enough to show that

$$\mathbb{E}_{n<N} U^n f \langle g, U^n f\rangle \to 0$$

in L^2. For this we can use van der Corput's lemma with $b_n = U^n f\langle g, U^n f\rangle$. The lemma tells us that it is enough to prove that

$$\limsup_{H\to\infty} \mathbb{E}_{h<H} \limsup_{N\to\infty} |\mathbb{E}_{n<N}\langle b_n, b_{n+h}\rangle| = 0. \tag{9.3}$$

However,

$$\begin{aligned}\langle b_n, b_{n+h}\rangle &= \langle U^n f, U^{n+h} f\rangle\langle g, U^n f\rangle\langle g, U^{n+h} f\rangle \\ &= \langle f, U^h f\rangle\langle g, U^n f\rangle\langle g, U^{n+h} f\rangle,\end{aligned}$$

and so

$$|\langle b_n, b_{n+h}\rangle| \leqslant |\langle f, U^h f\rangle|$$

uniformly in n. Therefore establishing (9.3) reduces to establishing that

$$\limsup_{h<H} \mathbb{E}_{h<H}|\langle f, U^h f\rangle| = 0,$$

which is precisely the statement that $f \in \mathrm{WM}(X)$. □

COROLLARY 9.1. *We have* $\mathrm{WM}(X) = \mathrm{AP}(X)^\perp$, *and so* $L^2(X) = \mathrm{WM}(X) + \mathrm{AP}(X)$. *Furthermore, if* $\pi : L^2(X) \to \mathrm{AP}(X)$ *denotes projection onto the space of almost periodic functions then* π *preserves non-negativity: if* $f(x) \geqslant 0$ *a.e. then* $\pi(f)(x) \geqslant 0$ *a.e. More generally if* $a \leqslant f(x) \leqslant b$ *for a.e.* x *then* $a \leqslant \pi(f)(x) \leqslant b$ *for a.e.* x, *and* $\int f d\mu = \int \pi(f) d\mu$.

PROOF. Suppose $f \in \mathrm{AP}(X)^\perp$. Then, by Proposition **??**, f is weak-mixing, $f \in \mathrm{WM}(X)$. Conversely we must show that if $f \in \mathrm{WM}(X)$ then $f \in \mathrm{AP}(X)^\perp$. Suppose that $g \in \mathrm{AP}(X)$. Let $\varepsilon > 0$, and let $K = K(\varepsilon)$ be such that for all n there is some $k(n) \leqslant K$ with $\|U^n g - U^{k(n)} g\|_2 \leqslant \varepsilon$. Then we have

$$\begin{aligned}|\langle f, g\rangle| &= |\mathbb{E}_{n<N}\langle U^n f, U^n g\rangle| \\ &\leqslant \varepsilon + |\mathbb{E}_{n<N}\langle U^n f, U^{k(n)} g\rangle| \\ &\leqslant \varepsilon + \sup_{k\leqslant K} |\mathbb{E}_{n<N}\langle U^n f, U^k g\rangle|.\end{aligned}$$

By Lemma 9.2 the expression on the right tends to 0 as $N \to \infty$, so $|\langle f, g\rangle| \leqslant \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we in fact have $\langle f, g\rangle = 0$.

Now we show that projection onto $\mathrm{AP}(X)$ preserves non-negativity. This follows from the (Hilbert space) fact that $\pi(f)$ is the unique closest almost-periodic function to f, in the L^2-norm. If $f \geqslant 0$ a.e. then $\|f - \pi(f)\|_2 \leqslant \|f - \pi(f)_+\|_2$. Since $\pi(f)_+ \in \mathrm{AP}(X)$, it follows that $\pi(f) = \pi(f)_+$ a.e., or in other words $\pi(f)$ is non-negative (a.e.). That π preserves the property $a \leqslant f \leqslant b$ follows quickly from this.

Finally, $f - \pi(f)$ is orthogonal to all almost-periodic functions, and that includes the constant function 1. This implies that $\int f d\mu = \int \pi(f) d\mu$. □

A consequence of this corollary is that $\mathrm{WM}(X)$ is closed in the L^2 topology and also closed under addition. However, the first of these facts is easy to show directly, and the second is a striaghtforward consequence of Lemma 9.2.

9.3. The SZ property at level 3

In this final section of the course we *almost* prove that all measure-preserving systems $(X, \mathscr{B}, \mu, T)$ have the SZ property at level 3. We must in fact make two assumptions: that T is invertible (and T^{-1} is measure-preserving), and that T is ergodic. The first assumption is not very serious at all, and in fact the systems generated by the correspondence principle in Chapter **??** will automatically be invertible. The assumption of ergodicity is a little more serious, but it too may be circumvented using the machinery of *ergodic decomposition*. Whilst this is unfortunately outside the scope of this course, we make some remarks on it in the last section.

Theorem 9.2. *Suppose that $(X, \mathscr{B}, \mu, T)$ is ergodic and invertible. Then it has the SZ property at level* 3.

Proof. We use the decomposition $L^2(X) = \mathrm{WM}(X) \oplus \mathrm{AP}(X)$ just established. Let $\pi : L^2(X) \to \mathrm{AP}(X)$ be the orthogonal projection. Suppose that $f \in L^\infty(X)$ has $f \geqslant 0$ a.e. We need to show that

$$\mathrm{Sz}_3(f, f, f) > 0,$$

where here for any three functions $g_0, g_1, g_2 \in L^\infty(X)$ we set

$$\mathrm{Sz}_3(g_0, g_1, g_2) := \liminf_{N \to \infty} \mathbb{E}_{n < N} \int g_0 U^n g_1 U^{2n} g_2.$$

It follows from Proposition 8.1 (together with the basic facts in Corollary 9.1) that

$$\mathrm{Sz}_3(\pi(f), \pi(f), \pi(f)) > 0.$$

Using the fact that Sz_3 is trilinear, we have

$$\mathrm{Sz}_3(f, f, f) = \mathrm{Sz}_3(\pi(f), \pi(f), \pi(f)) + \mathrm{Sz}_3(f - \pi(f), \pi(f), \pi(f)) \\ + \mathrm{Sz}_3(f, f - \pi(f), \pi(f)) + \mathrm{Sz}_3(f, f, f - \pi(f)).$$

We claim that the second, third and fourth terms are all zero, which is enough to prove the theorem. The second term is

$$\langle f - \pi(f), \mathbb{E}_{n<N} U^n \pi(f) U^{2n} \pi(f) \rangle.$$

Since $\mathrm{AP}(X)$ is closed under U and is a subspace, this is the inner product of $f - \pi(f)$ with an almost-periodic function, and hence is identically zero.

We prove that the fourth term $\mathrm{Sz}_3(f, f, f - \pi(f))$ is zero; the argument for the third term is very similar and is left as an exercise for the reader (See Sheet 4 – this is where the invertibility of T is used). For ease of notation write $g := f - \pi(f)$. We proceed exactly as in the proof of Proposition 7.2. The application of van der Corput shows that it is enough to prove that

$$\limsup_{H\to\infty} \mathbb{E}_{h<H} \limsup_{N\to\infty} \|\mathbb{E}_{n<N} U^n \Delta_{2h} g\|_2 = 0.$$

However, since T is ergodic we have, by the L^2 ergodic theorem,

$$\lim_{N\to\infty} \|\mathbb{E}_{n<N} U^n \Delta_{2h} g\|_2 = |\int \Delta_{2h} g| = |\langle g, U^{2h} g\rangle|,$$

and so all we must do is show that

$$\limsup_{H\to\infty} \mathbb{E}_{h<H} |\langle g, U^{2h} g\rangle| = 0.$$

This, however, follows from the fact that $g \in \mathrm{WM}(X)$. □

9.4. Further reading

I mentioned at the end of the last lecture that the assumption that T is ergodic is not very serious in this result, because one can reduce the general case to the ergodic case by the technique of *ergodic decomposition*, in which an arbitrary T-invariant measure μ is written as a "convex combination" of ergodic T-invariant measures. Further discussion is outside the scope of the course, but students may wish to read up on this in (for example) Einsiedler and Ward.

APPENDIX A

Measure theory

In this chapter we collect some facts about measure theory used in the main part of the course. Let's begin by recalling the definition of σ-algebra.

The collection $\mathscr{B}$ is required to be a σ-algebra, which means that it contains the empty set $\emptyset$ and X, and it is closed under complements, countable intersections and countable unions. To spell it out:

- We have $\emptyset, X \in \mathscr{B}$;
- If $A \in \mathscr{B}$ then $X \setminus A \in \mathscr{B}$;
- If $A_1, A_2, \dots \in \mathscr{B}$ then $\bigcap_{n=1}^{\infty} A_n \in \mathscr{B}$;
- If $A_1, A_2, \dots \in \mathscr{B}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathscr{B}$.

Next, we recall the definition of a probability measure μ. This satisfies the following properties whenever $A, A', A_1, A_2, A_3, \dots$:

- We have $\mu(\emptyset) = 0$ and $\mu(X) = 1$;
- (Additivity) If A, A' are disjoint then $\mu(A \cup A') = \mu(A) + \mu(A')$;
- (Limits) If $A_1 \subset A_2 \subset A_3 \subset \dots$ then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} A_n$

Note that the these conditions imply the descending version of the limit property, namely

- If $A_1 \supset A_2 \supset A_3 \supset \dots$ then $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

A.1. Construction of measures

We used the fact that Lebesgue measure exists without proof in several places, and we also used the fact that there is a natural *Cantor measure* on the Bernouilli space $\{0,1\}^{\mathbb{N}}$. Both of these facts are consequences of the *Carathéodory extension theorem.* The Carathéodory extension theorem concerns probability measures defined on *rings* of sets. A ring $\mathscr{R}$ of sets is a weaker concept than a σ-algebra, in that we only require closure under *finite* unions. Thus

- We have $\emptyset, X \in \mathscr{R}$;
- If $A \in \mathscr{R}$ then $X \setminus A \in \mathscr{R}$;
- If $A_1, A_2, \dots, A_k \in \mathscr{R}$ then $\bigcap_{n=1}^{k} A_n \in \mathscr{R}$;
- If $A_1, A_2, \dots, A_k \in \mathscr{R}$ then $\bigcup_{n=1}^{k} A_n \in \mathscr{R}$.

To define what is meant by a probability measure on a ring $\mathscr{R}$, we need to tweak the definition for σ-algebras very slightly. The definition is the same, except the limit condition becomes

- (Limits) If $A_1 \subset A_2 \subset A_3 \subset \dots$ are in $\mathscr{R}$ *and* $\bigcup_{n=1}^{\infty} A_n \in \mathscr{R}$ then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

Here is the Carathéodory extension theorem.

Theorem A.1 (Carathéodory). *Suppose that $\mathscr{R}$ is a ring which generates $\mathscr{B}$ as a σ-algebra. Then any probability measure $\mu : \mathscr{R} \to [0,1]$ has a unique extension to a probability measure on $\mathscr{B}$.*

The most pleasant setting in which to apply this result is probably that of the Cantor space $X = \{0,1\}^{\mathbb{N}}$. Here we take $\mathscr{R}$ to be the ring generated by cylinder sets, that is to say sets of the form $\{\vec{x} \in X : x_{i_1} = \varepsilon_{i_1}, \dots, x_{i_k} = \varepsilon_{i_k}\}$, with the measure μ of such a set being defined to be 2^{-k} and extended to $\mathscr{R}$ using additivity. It must be checked that μ is a probability measure, and to do that we must show that if $A_1 \subset A_2 \subset \dots \in \mathscr{R}$ and $\bigcup_{n=1}^{\infty} A_n \in \mathscr{R}$ then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$. A key observation here is that, with the product topology on X, every set in $\mathscr{R}$ is both open and closed and X is compact. Therefore any union $\bigcup_{n=1}^{\infty} A_n$ is a covering of a compact set (itself) by open sets, and hence is in fact equal to a finite subcover $\bigcup_{n=1}^{k} A_n$, and so in this case the limit property is a consequence of additivity.

The existence of Cantor measure, as used in the course, now follows from the Carathéodory extension theorem.

A.2. Littlewood's first principle. Approximation arguments

Littlewood's three basic principles are a very useful guide to dealing with measure theory in practice. It is interesting that they were formulated (by J. E. Littlewood) at a time when heuristic reasoning and "conceptual" explanations were perhaps not so much to the fore as they are in contemporary mathematical exposition. Littlewood's three principles are:

(i) A measurable set is nearly an open set;
(ii) A measurable function is nearly a continuous function;
(iii) A convergent sequence of functions is nearly uniformly convergent.

Rigorous versions of these principles apply to any Borel probability measure μ on a compact metric space X.

In this section we'll discuss the first Littlewood principle. Invocations of rigorous versions of this are referred to in the main text as "approximation arguments". The most basic type of approximation argument is the following.

LEMMA A.1. *Suppose that $\mathscr{R}$ is a ring of subsets of an underlying set X, and that $\mu : \mathscr{R} \to [0,1]$ is a probability measure. Let $\mathscr{B}$ be the σ-algebra generated by $\mathscr{R}$, and let $\mu : \mathscr{B} \to [0,1]$ be the probability measure whose existence is guaranteed by the Carathéodory extension theorem. Then elements of $\mathscr{B}$ are well-approximated by elements of $\mathscr{R}$ in the following sense: if $B \in \mathscr{B}$ and $\varepsilon > 0$ then there is some $R \in \mathscr{R}$ with $\mu(B \triangle R) < \varepsilon$.*

PROOF. (Sketch) Define $\mathscr{B}_0$ to be the set of all $B \in \mathscr{B}$ with the stated property. Obviously $\mathscr{R} \subset \mathscr{B}_0$. We claim that $\mathscr{B}_0$ is a σ-algebra, which is enough to conclude the stated claim. To see that $\mathscr{B}_0$ is closed under countable unions, it is clearly enough to consider countable *nested* unions $\bigcup_{n=1}^{\infty} B_n$ with $B_1 \subset B_2 \subset \dots$ lying in $\mathscr{B}_0$. For each n there is some $R_n \in \mathscr{R}$ such that $\mu(B_n \triangle R_n) < \varepsilon/2$. Furthermore, by the limit property of measures, we have $\lim_{n\to\infty} \mu(B_n) = \mu(B)$. Choosing n so large that $|\mu(B) - \mu(B_n)| < \varepsilon/2$, we obviously have $\mu(B \triangle R_n) < \varepsilon$. This concludes the proof. $\square$

We turn now to approximation arguments specific to the setting in which X is a compact metric space such as $[0,1]$ or $\mathbb{R}/\mathbb{Z}$. These satisfy a strong type of approximation principle known as *regularity*, which means that if $E \subset X$ is measurable and if $\varepsilon > 0$ then there is an open set $U \supset E$ with $\mu(U \setminus E) < \varepsilon$. In the main text we only ever need the weaker property that $\mu(E \triangle U) < \varepsilon$, which of course follows from this.

In the case of Lebesgue measure on $X = \mathbb{R}/\mathbb{Z}$ or $[0,1]$ this property is in fact built in to the standard *construction* given in Chapter 2, that is to say

$$\mu(E) = \inf\{\sum_{j=1}^{\infty} \mu(I_j) : E \subset \bigcup_{j=1}^{\infty} I_j, I_j \text{ open intervals}\}.$$

Note that by applying this principle to E^c (and changing ε to $\varepsilon/2$) we see that if $E \subset X$ is measurable and if $\varepsilon > 0$ then there are sets K, U with K compact, U open, $K \subset E \subset U$ and $\mu(K \setminus U) < \varepsilon$. One will usually see the existence of such K and U given as the definition of what it means for a measure μ to be regular, but in the case that X is a compact metric space it is equivalent to the one-sided condition involving only U.

To prove that *all* Borel probability measures on a compact metric space X are regular, the key idea is consider the set of all E for which K, U as above exist for all $\varepsilon > 0$. One checks that open sets E have this property, and then that having the property is a condition closed under complements, countable unions and intersections. Hence all Borel sets have the property.

Here is a fairly general result on the interval $[0,1]$ (or on $\mathbb{R}/\mathbb{Z}$) which covers most of the approximation arguments we needed in the main part of the course.

LEMMA A.2. *Suppose that we have collections of open intervals* $(U_{n,i})_{i=1,2,\dots}$, $n = 1, 2, \dots$ *such that*

(i) $\sup_i \mu(U_{n,i}) \to 0$ *as* $n \to \infty$*;*
(ii) *For fixed* n*, the* $U_{n,i}$ *are disjointl*
(iii) $[0,1] \setminus \bigcup_i U_{n,i}$ *has measure 0.*

The any Borel set E *can be approximated arbitrarily well by finite unions of these intervals, all with the same* n*. In fact for any* $\varepsilon > 0$*, for all* n *sufficiently large there is a set* $\Sigma_n \subset \mathbb{N}$ *such that, if we write* $U := \bigcup_{i \in \Sigma_n} U_{n,i}$*, then* $\mu(E \triangle U) < \varepsilon$*.*

PROOF. Since μ is regular, there is some open set $U \subset [0,1]$ such that $E \subset U$ and $\mu(U \setminus E) < \varepsilon/3$. The set U, like every open subset of $[0,1]$, is a countable union $\bigcup_{j=1}^{\infty} I_j$ of open intervals (proof: define an equivalence relation on U by setting $x \sim y$ iff $(\min(x,y), \max(x,y)) \subset U$: the equivalence classes are open intervals, and each contains a rational point so there are only countably many), and so by the limit property of μ there is some finite J such that $\mu(U \setminus \bigcup_{j=1}^{J} I_j) < \varepsilon$. Hence, $\mu(E \triangle \bigcup_{j=1}^{J} I_j) < 2\varepsilon/3$. Now choose n so large that the longest length of any $U_{n,i}$ is at most $\varepsilon/6J$. Then, taking $\Sigma_n = \{i : U_{n,i} \cap I_j \neq 0 \text{ for some } j = 1, \dots, J\}$, we see that $\bigcup_{i \in \Sigma_n} U_{n,i} \supset \bigcup_{j \in J} I_j$ and $\mu(\bigcup_{i \in \Sigma_n} U_{n,i} \setminus \bigcup_{j \in J} I_j) < \varepsilon/3$. Putting everything together, the result follows. □

A straightforward consequence of this lemma is the Lebesgue density theorem. We used this in one of our proofs that the irrational circle rotation is ergodic.

LEMMA A.3 (Lebesgue density theorem). *Let* μ *be Lebesgue measure on* $[0,1]$ *and suppose that* $E \subset [0,1]$ *is measurable and has* $\mu(E) > 0$*. Then for every* $\varepsilon > 0$ *there is an open interval* $I \subset [0,1]$ *such that the density of* E *on* I*,* $\mu(E \cap I)/\mu(I)$*, is at least* $1 - \varepsilon$*.*

PROOF. Take $(U_{n,i})_{i=0,\dots,n-1}$ to be the standard intervals of length $1/n$. Let $\varepsilon' > 0$ be a quantity to be selected later. By the last lemma, for all n sufficiently large there is a (finite) union $\bigcup_{i \in \Sigma_n} U_{n,i}$ whose symmetric difference with E has measure at most ε'. Suppose that the density of E on each $U_{n,i}$ is less than $1 - \varepsilon$. Then we have

$$\begin{aligned} \mu(E) &\leqslant (1-\varepsilon) \sum_{i \in \Sigma_n} \mu(U_{n,i}) + \varepsilon' \\ &\leqslant (1-\varepsilon)(\mu(E) + \varepsilon') + \varepsilon'. \end{aligned}$$

If $\varepsilon' > 0$ is sufficiently small, this is a contradiction. □

A.3. On integration

We did not make very heavy use of integration in the course, but we did talk about $L^1(X)$ and $L^2(X)$, in particular using the fact that when $X = [0,1]$ the continuous functions are dense in both these spaces, and also using the fact that $L^2(X)$ is complete (and hence a Hilbert space).

Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $f : X \to \mathbb{R}$ be a function. We say that f is measurable if $f^{-1}(a,b)$ is measurable for every half-open interval (a,b). We only try to make sense of the integral of measurable functions. In the first instance, suppose that $f(x) \geqslant 0$ for a.e. x. Then for each n we may define

$$f_n(x) := \sum_{i=1}^{n2^n} (i-1)2^{-n} 1_{(i-1)2^{-n} \leqslant f(x) < i2^{-n}}(x).$$

Each such function f_n is an example of a *simple measurable function*, that is to say a linear combination of characteristic functions 1_E with E measurable, and it is clear that we "should" define $\int_X f_n d\mu$ by

$$\int_X f_n d\mu := \sum_{i=1}^{n2^n} (i-1)2^{-n} \mu(\{x : (i-1)2^{-n} \leqslant f(x) < i2^{-n}\}).$$

Note that we have the nesting $f_1(x) \leqslant f_2(x) \leqslant \dots \geqslant f(x)$, and so

$$\int_X f_1 d\mu \leqslant \int_X f_2 d\mu \leqslant \dots.$$

Also the f_n approximate f increasingly well in a certain sense, namely we have $|f_n(x) - f(x)| \leqslant 2^{-n}$ if $|f(x)| \leqslant n$. All this means that it makes sense to define

$$\int_X f d\mu = \lim_{n\to\infty} \int_X f_n d\mu,$$

defining this to be ∞ if the integrals $\int_X f_n d\mu$ grow without bound.

Suppose now that $f : X \to \mathbb{R}$ is an arbitrary measurable function, not necessarily a non-negative one. Then we write $f = f_+ - f_-$ with $f_+, f_- \geqslant 0$ everywhere. If both of the integrals $\int f_+ d\mu$ and $\int f_- d\mu$ are finite then we say that $f \in L^1(X)$, and define $\int f d\mu := \int f_+ d\mu - \int f_- d\mu$. Finally we handle complex-valued functions $f : X \to \mathbb{C}$ by splitting into real and imaginary parts.

In the main text we used the fact that if $f : [0,1] \to \mathbb{R}$ is continuous then the Lebesgue integral $\int_X f d\mu$ coincides with the usual Riemann integral of f from 0 to 1. This is more-or-less obvious from the definition.

A.4. Littlewood's second principle

(Almost) built into the definition of the integral we just gave is the following statement, used at several points in the main text.

LEMMA A.4. *The simple measurable functions are dense in $L^1(X)$. That is, if $f \in L^1(X)$ and if $\varepsilon > 0$ then there is a simple measurable function $\tilde{f}$ such that $\int |f - \tilde{f}| d\mu < \varepsilon$.*

Focussing on the specific case $X = [0,1]$ and μ equal to Lebesgue measure, it follows straightforwardly from Lemma A.4 and Lemma A.2 (with any permissible choice of the $U_{n,i}$, for example the standard intervals $(\frac{i-1}{n}, \frac{i}{n})$) that the functions of the form $f = \sum_{j=1}^{K} c_j 1_{I_j}$, where $I_j \subset [0,1]$ is an open interval, are dense in $L^1(X)$. Finally one may use the construction discussed in Chapter 1 to approximate 1_I arbitrarily well in $L^1(X)$ by continuous functions, and thereby conclude the following result.

PROPOSITION A.1. *Let $X = [0,1]$ with Lebesgue measure μ. Then the continuous functions $C(X)$ are dense in $L^1(X)$.*

In fact the same holds for $(X, \mathscr{B}, \mu)$ with X any compact metric space, $\mathscr{B}$ the Borel σ-algebra and $\mu : \mathscr{B} \to [0,1]$ a Borel probability measure. To construct a continuous function f approximating the characteristic function 1_E of a Borel set, first use the regularity of μ to locate an open set U and a compact set K with $K \subset E \subset U$ and $\mu(U \setminus K) < \varepsilon$. Then consider the continuous function $f(x) = \frac{\mathrm{dist}(x,U^c)}{\mathrm{dist}(x,U^c)+\mathrm{dist}(x,K)}$, which satisfies $0 \leqslant f(x) \leqslant 1$, $f(x) = 1$ for $x \in K$ and $f(x) = 0$ for $x \notin U$, and thus satisfies $\|1_E - f\|_1 \leqslant \varepsilon$.

A.5. On product measures

In various places in the main text we talked about product measures. For example, we talked about the skew torus system, in which the underlying space is $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$. Then, in Chapter 7, we considered more general products of probability spaces $(X, \mathscr{B}, \mu)$ and $(\tilde{X}, \tilde{\mathscr{B}}, \tilde{\mu})$. Here we give a few remarks concerning the construction of the product space $(X \times \tilde{X}, \mathscr{B} \times \mathscr{B}', \mu \times \tilde{\mu})$. Giving all the details of this construction is rather tedious, and we do not.

The first key point to make is that $\mathscr{B} \times \tilde{\mathscr{B}}$ does *not* mean the set $\{E \times \tilde{E} : E \in \mathscr{B}, \tilde{E} \in \mathscr{B}\}$, but rather the σ-algebra generated by this set. Thus, for example, it contains all countable unions $\bigcup_{n=1}^{\infty}(E_n \times \tilde{E}_n)$ with $E_n, \tilde{E}_n \in \mathscr{B}$. Inside this product σ-algebra we distinguish the ring $\mathscr{R}$ of sets which are finite unions of this type. It is easy to see that any such set S may be written as a *disjoint* finite union $\bigcup_{n=1}^{k}(E_n \times \tilde{E}_n)$, in which case we define $(\mu \times \tilde{\mu})(S) := \sum_{n=1}^{k} \mu(E_n)\tilde{\mu}(\tilde{E}_n)$. One may then verify that this map $\mu \times \tilde{\mu} : \mathscr{R} \to [0,1]$ is a probability measure on $\mathscr{R}$, and it then follows from the Carathéodory extension theorem that $\mu \times \tilde{\mu}$ may be extended to a probability measure on $\mathscr{B} \times \tilde{\mathscr{B}}$.

A fact we used in several places in Chapter 7 was the following assertion.

LEMMA A.5. *Suppose that $E \in \mathscr{B} \times \tilde{\mathscr{B}}$. Then E may be approximated arbitrarily closely by disjoint finite unions of products $A \times B$, with $A \in \mathscr{B}$ and $B \in \tilde{\mathscr{B}}$.*

This lemma is an immediate consequence of Lemma A.1.

A.6. Littlewood's third principle

The third of Littlewood's principles refers to Egorov's theorem: If (f_n) is a sequence of measurable functions which converge pointwise on X, and if $\varepsilon > 0$, then there is a measurable set $X' \subset X$, $\mu(X \setminus X') \leqslant \varepsilon$, such that the f_n converge uniformly on X'. We did not make any use of Egorov's theorem in this course.

A.7. Further reading.

I rather like the introduction to Lebesgue measure on $\mathbb{R}$ in the book of Stein and Shakarchi, *Real analysis: measure theory, integration and Hilbert spaces.* For a comprehensive introduction to the more general setting that we need here one might consult Rudin's "red" book, *Real and complex analysis.* He works with locally compact Hausdorff spaces X rather than simply compact metric spaces as we discuss here.

APPENDIX B

Basic facts about Fourier analysis on the circle

In this appendix we gather together some statements about Fourier analysis of functions $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$. We give just what was needed in the main text, together with a tiny bit of extra context here and there. The reader interested in a more leisurely discussion should consult [].

Throughout this section we write

$$e_r(x) := e^{2\pi i r x}$$

for $r \in \mathbb{Z}$.

Suppose that $f \in L^1(\mathbb{R}/\mathbb{Z})$. Then we may define the Fourier coefficients

$$\hat{f}(r) := \langle f, e_r \rangle = \int_0^1 f(x) e^{-2\pi i r x} dx$$

for $r \in \mathbb{Z}$. The Fourier series of f is then the (purely formal at this stage) sum

$$\sum_r \hat{f}(r) e_r(x).$$

Much of the elementary theory of Fourier series on the circle is concerned with the question of whether this formal sum makes any sense and, if so, in what sense it can be thought to "represent" f. In studying this question it is natural to introduce the partial sums

$$\sigma_M f(x) := \sum_{|r| \leqslant M} \hat{f}(r) e_r(x).$$

It also turns out to be expedient to introduce the *Cesaro sums*

$$\tilde{\sigma}_M f(x) := \sum_{|r| \leqslant M} (1 - \frac{|r|}{M}) \hat{f}(r) e_r(x) = \sum_r (1 - \frac{|r|}{M})_+ \hat{f}(r) e_r(x).$$

The Cesaro sums should be thought of as a "smoothed" method of Fourier summation.

Here are the key facts from basic Fourier analysis relevant to the main text.

THEOREM B.1. *We have the following statements.*

(i) (Bessel) *We have* $\sum_{r \in \mathbb{Z}} |\hat{f}(r)|^2 \leqslant \|f\|_2^2$.
(ii) (Riemann–Lebesgue) *If* $f \in L^2(\mathbb{R}/\mathbb{Z})$ *then* $\hat{f}(r) \to 0$ *as* $r \to \infty$.

(iii) *If f is continuous then we have $\tilde{\sigma}_M f(x) \to f(x)$ uniformly in $x \in \mathbb{R}/\mathbb{Z}$. In particular, the trigonometric polynomials are dense in $C(X)$ with the $L^\infty(X)$ norm.*

(iv) *We do have $\sigma_M f \to f$ in L^2 for every $f \in L^2(\mathbb{R}/\mathbb{Z})$, that is to say $\|\sigma_M f - f\|_2 \to 0$.*

(v) (Uniqueness) *If $f \in L^2(\mathbb{R}/\mathbb{Z})$ and if $\hat{f}(r) = 0$ for all r then $f = 0$ a.e.*

(vi) (Plancherel) *We in fact have equality in Bessel's inequality.*

(vii) *There exist continuous f with $\sigma_M f(0) \nrightarrow f(0)$.*

Part (vii) was not needed in the main text but are included as context and general culture (in particular, (iii) would be false if $\tilde{\sigma}_M$ were replaced by σ_M).

PROOF. (i) Fix $f \in L^2(\mathbb{R}/\mathbb{Z})$. Now $L^2(\mathbb{R}/\mathbb{Z})$ is a Hilbert space, and inside it we consider the closed linear subspace V_M of all trigonometric polynomials of degree at most M. We claim that $\sigma_M : L^2(\mathbb{R}/\mathbb{Z}) \to V_M$ is the orthogonal projection onto this subspace. To see this, it is enough to check that $\sigma_M f$ is the orthogonal projection of f to V_M, that is to say $f - \sigma_M f \perp V_M$ and $\sigma_M f = f$ if $f \in V_M$. However one may check that

$$\langle \sigma_M f, e_r \rangle = \hat{f}(r) = \langle f, e_r \rangle$$

for all $|r| \leqslant M$, and so the assertion follows. A particular consequence of this (and Pythagoras theorem) is that

$$\|\sigma_M f\|_2^2 = \|f\|_2^2 - \|f - \sigma_M f\|_2^2 \leqslant \|f\|_2^2.$$

However, by direct calculation we have

$$\begin{aligned}\|\sigma_M f\|_2^2 &= \int_0^1 |\sum_{|r| \leqslant M} \hat{f}(r) e_r(x)|^2 dx \\ &= \sum_{|r_1|, |r_2| \leqslant M} \hat{f}(r_1)\overline{\hat{f}(r_2)} \int_0^1 e_{r_1}(x)\overline{e_{r_2}(x)} dx \\ &= \sum_{|r| \leqslant M} |\hat{f}(r)|^2.\end{aligned}$$

Bessel's inequality follows.

(ii) This follows immediately from Bessel's inequality.

Parts (iv), (v) and (vi) of the theorem require the "completeness" of the trigonometric system $\{e_r\}_{r \in \mathbb{Z}}$ in $L^2(\mathbb{R}/\mathbb{Z})$, that is to say the fact that trigonometric polynomials (finite weighted sums of the e_r) are dense in L^2. One way to establish this is via (iii), which is the deepest part of the theorem. We prove this now. The crucial point here is that there is an expression

$$\tilde{\sigma}_M f(x) = \int_0^1 f(x-y) K_M(y) dy = f * K_M(x) \tag{B.1}$$

for a certain "kernel" K_M, the Fejér kernel. One can guess that this should be the case by taking the Fourier transform of both sides: on the one hand we have

$$\widehat{\tilde{\sigma}_M}(r') = \sum_{|r}(1 - \frac{|r|}{M})_+ \hat{f}(r) \int_0^1 e_r(x)\overline{e_{r'}(x)}dx = (1 - \frac{|r'|}{M})_+ \hat{f}(r'),$$

whilst on the other hand

$$\widehat{f * K_M}(r') = \int_0^1 \int_0^1 f(x-y)\overline{e_{r'}(x-y)} K_M(y) \overline{e_{r'}(y)} dy dx = \widehat{K_M}(r')\hat{f}(r').$$

Thus we think to choose K_M so that

$$\widehat{K_M}(r') = (1 - \frac{|r'|}{M})_+$$

for all r'. It is easy to find a function K_M with this property, namely

$$K_M(x) := \sum_r (1 - \frac{|r|}{M})_+ e_r(x). \tag{B.2}$$

(We leave the easy proof as an exercise.) These last few lines, leading to the idea that we should define K_M as in (B.2), have just been for motivation. Now that we have a definition of K_M, we may check directly that (B.1) holds:

$$\begin{aligned} f * K_M(x) &= \sum_r (1 - \frac{|r|}{M})_+ \int_0^1 f(x-y) e_r(y) dx \\ &= \sum_r (1 - \frac{|r|}{M})_+ \int_0^1 f(x-y)\overline{e_r(x-y)} e_r(x) \\ &= \sum_r (1 - \frac{|r|}{M})_+ \hat{f}(r) e_r(x) = \tilde{\sigma}_M f(x). \end{aligned}$$

To make use of these observations, we note a different formula for the Fejér kernel K_M, namely

$$K_M(x) = \frac{1}{M}|\sum_{r=0}^{M-1} e_r(x)|^2 = \frac{1}{M}\Big|\frac{1 - e^{2\pi i r x M}}{1 - e^{2\pi i r x}}\Big|^2 = \frac{1}{M}\Big|\frac{\sin \pi r x M}{\sin \pi r x}\Big|^2. \tag{B.3}$$

The first equality may be verified directly using (B.2), the second follows by summing the geometric progression, and the third uses the relation $|1 - e^{i\theta}| = 2|\sin \frac{1}{2}\theta|$. Using this, we note the following facts:

(1) K_M is symmetric: $K_M(t) = K_M(-t)$;
(2) $K_M(t) \geqslant 0$ and $\int_0^1 |K_M(t)| dt = \int_0^1 K_M(t) dt = 1$.
(3) For any fixed $\delta > 0$, $\lim_{M\to\infty} \int_{|t|>\delta} K_M(t) dt = 0$;

(1) and the first part of (2) are obvious. The second part of (2) is immediate from (B.2). To prove (3), we use the inequalities $1 \geqslant |\sin t| \geqslant c|t|$ to conclude that $|K_M(x)| \ll \frac{1}{M|x|^2}$, from which the result follows easily.

We are now ready to show that $\tilde{\sigma}_M f(0) \to f(0)$. The proof may be easily modified to establish that $\tilde{\sigma}_M f(x) \to f(x)$ uniformly in x, and hence to prove (iv). Using (1) and (2), it is easy to see that we have

$$\tilde{\sigma}_M(f)(0) - f(0) = \int_0^1 (f(t) - f(0))K_M(t)dt.$$

Now let $\delta > 0$ be arbitrary, and split the range of integration into the two parts $|t| \leqslant \delta$ and $|t| > \delta$. On the first part we have

$$|\int_{|t|\leqslant\delta} (f(t) - f(0))K_M(t)dt| \leqslant \sup_{|t|\leqslant\delta} |f(t) - f(0)|,$$

by (2). Since f is continuous, this tends to zero as $\delta \to 0$. On the second part we have

$$|\int_{|t|>\delta} (f(t) - f(0))K_M(t)dt| \leqslant 2\|f\|_\infty \int_{|t|>\delta} K_M(t)dt.$$

For fixed δ, this tends to zero as $M \to \infty$ by (3). The result follows.

Now we turn to (iv). In the proof of this we will use a key consequence of (iii): the trigonometric polynomials are dense in $L^2(\mathbb{R}/\mathbb{Z})$. Indeed we showed in (iii) that the trigonometric polynomials are dense in $C(\mathbb{R}/\mathbb{Z})$ with the $L^\infty(\mathbb{R}/\mathbb{Z})$ norm, and hence also with the $L^2(\mathbb{R}/\mathbb{Z})$ norm. But $C(\mathbb{R}/\mathbb{Z})$ is dense in $L^2(\mathbb{R}/\mathbb{Z})$. It follows that, for any fixed $f \in L^2(\mathbb{R}/\mathbb{Z})$, the L^2-distance of f to the space V_M spanned by trigonometric functions $e_r(x)$, $|r| \leqslant M$, tends to zero. However we have already seen that $\sigma_M f$ is the orthogonal projection of f on to this space, and so $\|f - \sigma_M f\|_2 \to 0$, as required.

Part (v) is an immediate consequence of (iv): if $\hat{f}(r) = 0$ for all $r \in \mathbb{Z}$ then $\sigma_M f(x) = 0$ for all x and for all M. But $\sigma_M f \to f$ in L^2, and thus $f = 0$ a.e.

Part (vi) is immediate from (iv) and the fact that $\|\sigma_M f\|_2^2 = \sum_{|r|\leqslant M} |\hat{f}(r)|^2$, established above.

As we stated, part (vii) was not important in the main part of the book, so we simply offer some brief comments. As a point of departure, we can try to see where the proof in (iii) that $\tilde{\sigma}_M f(0) \to f(0)$ would go wrong if we tried to adapt it to prove that $\sigma_M f(0) \to f(0)$. It turns out that there are formulae to parallel (B.1), (B.2) and (B.3): namely,

$$\sigma_M f(x) = f * D_M(x),$$

where

$$D_M(x) = \sum_{|r| \leqslant M} e_r(x) = e^{-2\pi i M x} \Big| \frac{1 - e^{2\pi i(2m+1)rx}}{1 - e^{2\pi i r x}} \Big|.$$

Unfortunately, this kernel D_M (a variant of the Dirichlet kernel) does not enjoy all of the nice properties (1), (2), (3). Specifically, we do not have $D_M(t) \geqslant 0$, and it is not too hard to show that (2) fails somewhat dramatically:

$$\int_0^1 |D_M(t)| dt \gg \log M.$$

This means that the norm of the functional $\phi_M : C(\mathbb{R}/\mathbb{Z}) \to \mathbb{C}$ defined by $\phi_M f := \sigma_M f(0) = \int_0^1 f(t) D_M(t) dt$ is $\gg \log M$ (take $f(t)$ to be a continuous approximation to $\mathrm{sgn}(D_M(t))$). By the uniform boundedness principle of functional analysis [] there is some $f \in C(\mathbb{R}/\mathbb{Z})$ for which $\sigma_M f(0) \to \infty$. $\square$

To conclude this section we remark that the completeness of the trigonometric system $\{e_r\}_{r \in \mathbb{Z}}$ may be established in other ways, for example via the Stone-Weierstrass theorem, or by showing pointwise convergence $\sigma_M f \to f$ when f is sufficiently smooth, then appealing to the fact that smooth functions are dense in $L^2(\mathbb{R}/\mathbb{Z})$. Items (iv), (v) and (vi) may then be proven in the same manner as we did above.

Bibliography

[1] https://terrytao.wordpress.com/2009/01/03/254a-notes-0a-an-alternate-approach-to-the-caratheodory-extension-theorem/#more-1368

[2] https://www.maths.ox.ac.uk/system/files/coursematerial/2013/983/70/PartA-Integration-2014-Mar-10.pdf