Manfred Einsiedler
Thomas Ward

Ergodic Theory

with a view towards Number Theory

(first four Chapters and Appendices only,
for LMS–EPSRC Summer School July 2010)

– Monograph –

June 21, 2010

Springer

To the memory of Daniel Jay Rudolph
(1949 – 2010)

Preface

Many mathematicians are aware of some of the dramatic interactions between ergodic theory and other parts of the subject, notably Ramsey theory, infinite combinatorics, and Diophantine number theory. These notes are intended to provide a gentle route to a tiny sample of these results. The intended readership is expected to be mathematically sophisticated, with some background in measure theory and functional analysis, or to have the resilience to learn some of this material along the way from other sources.

In this volume we develop the beginnings of ergodic theory and dynamical systems. While the selection of topics has been made with the applications to number theory in mind, we also develop other material to aid motivation and to give a more rounded impression of ergodic theory. Different points of view on ergodic theory, with different kinds of examples, may be found in the monographs of Cornfeld, Fomin and Sinaĭ [60], Petersen [282], or Walters [373]. Ergodic theory is one facet of dynamical systems; for a broad perspective on dynamical systems see the books of Katok and Hasselblatt [182] or Brin and Stuck [44]. An overview of some of the more advanced topics we hope to pursue in a subsequent volume may be found in the lecture notes of Einsiedler and Lindenstrauss [80] in the Clay proceedings of the Pisa Summer school.

Fourier analysis of square-integrable functions on the circle is used extensively. The more general theory of Fourier analysis on compact groups is not essential, but is used in some examples and results. The ergodic theory of commuting automorphisms of compact groups is touched on using a few examples, but is not treated systematically. It is highly developed elsewhere: an extensive treatment may be found in the monograph by Schmidt [332]. Standard background material on measure theory, functional analysis and topological groups is collected in the appendices for convenience.

Among the many *lacunae*, some stand out: Entropy theory; the isomorphism theory of Ornstein, a convenient source being Rudolph [324]; the more advanced spectral theory of measure-preserving systems, a convenient source being Nadkarni [264]; finally Pesin theory and smooth ergodic theory, a source being Barreira and Pesin [19]. Of these omissions, entropy theory is perhaps the most fundamental for applications in number theory, and this was the reason for not including it here. There is simply too much to say about entropy to fit into this volume, so we will treat this important topic, both in general terms and in more detail in the algebraic context needed for number theory, in a subsequent volume. The notion is mentioned in one or two places in this volume, but is never used directly.

No Lie theory is assumed, and for that reason some arguments here may seem laborious in character and limited in scope. Our hope is that seeing the language of Lie theory emerge from explicit matrix manipulations allows a relatively painless route into the ergodic theory of homogeneous spaces. This will be carried further in a subsequent volume, where some of the deeper applications will be given.

Notation and Conventions

The symbols $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $\mathbb{Z}$ denote the natural numbers, non-negative integers and integers; $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the rational numbers, real numbers and complex numbers; $\mathbb{S}^1$, $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the multiplicative and additive circle respectively. The elements of $\mathbb{T}$ are thought of as the elements of $[0, 1)$ under addition modulo 1. The real and imaginary parts of a complex number are denoted $x = \Re(x+\mathrm{i}y)$ and $y = \Im(x+\mathrm{i}y)$. The order of growth of real- or complex-valued functions f, g defined on $\mathbb{N}$ or $\mathbb{R}$ with $g(x) \neq 0$ for large x is compared using Landau's notation:

$$f \sim g \text{ if } \left|\frac{f(x)}{g(x)}\right| \longrightarrow 1 \text{ as } x \to \infty;$$
$$f = \mathrm{o}(g) \text{ if } \left|\frac{f(x)}{g(x)}\right| \longrightarrow 0 \text{ as } x \to \infty.$$

For functions f, g defined on $\mathbb{N}$ or $\mathbb{R}$, and taking values in a normed space, we write $f = \mathrm{O}(g)$ if there is a constant $A > 0$ with $\|f(x)\| \leqslant A\|g(x)\|$ for all x. In particular, $f = \mathrm{O}(1)$ means that f is bounded. Where the dependence of the implied constant A on some set of parameters $\mathscr{A}$ is important, we write $f = \mathrm{O}_{\mathscr{A}}(g)$. The relation $f = \mathrm{O}(g)$ will also be written $f \ll g$, particularly when it is being used to express the fact that two functions are commensurate, $f \ll g \ll f$. A sequence $a_1, a_2, \dots$ will be denoted (a_n). Unadorned norms $\|x\|$ will only be used when x lives in a Hilbert space (usually L^2) and always refer to the Hilbert space norm. For a topological space X, $C(X)$, $C_{\mathbb{C}}(X)$, $C_c(X)$ denote the space of real-valued, complex-valued, compactly supported continuous functions on X respectively, with the supremum norm. For sets A, B, denote the set difference by

$$A \smallsetminus B = \{x \mid x \in A, x \notin B\}.$$

Additional specific notation is collected in an index of notation on page 471.

Statements and equations are numbered consecutively within chapters, and exercises are numbered in sections. Theorems without numbers in the main body of the text will not be proved; appendices contain background material in the form of numbered theorems that will not be proved here.

Several of the issues addressed in this book revolve around *measure rigidity*, in which there is a natural measure that other measures are compared with. These natural measures will usually be Haar measure on a compact or locally compact group, or measures constructed from Haar measures, and these will usually be denoted m.

We have not tried to be exhaustive in tracing the history of the ideas used here, but have tried to indicate some of the rich history of mathematical developments that have contributed to ergodic theory. Certain references to earlier and to related material is generally collected in endnotes at the end of each chapter; the presence of these references should not be viewed in any way as authoritative. Statements in these notes are informed throughout by a desire to remain rooted in the familiar territory of ergodic theory. The standing assumption is that, unless explicitly noted otherwise, metric spaces are complete and separable, compact groups are metrizable, discrete groups are countable, countable groups are discrete, and measure spaces are assumed to be Borel probability spaces (this assumption is only relevant starting with Section 5.3; see Definition 5.13 for the details). A convenient summary of the measure-theoretic background may be found in the work of Royden [320] or of Parthasarathy [280].

Acknowledgements

It is inevitable that we have borrowed ideas and used them inadvertently without citation, and certain that we have misunderstood, misrepresented or misattributed some historical developments; we apologize for any egregious instances of this. We are grateful to several people for their comments on drafts of sections, including Alex Abercrombie, Menny Aka, Sarah Bailey-Frick, Tania Barnett, Vitaly Bergelson, Michael Björklund, Florin Boca, Will Cavendish, Tushar Das, Jerry Day, Jingsong Chai, Alexander Fish, Anthony Flatters, Nikos Frantzikinakis, Jenny George, John Griesmer, Shirali Kadyrov, Cor Kraaikamp, Beverly Lytle, Fabrizio Polo, Christian Röttger, Nimish Shah, Ronggang Shi, Christoph Übersohn, Alex Ustian, Peter Varju and Barak Weiss; the second named author also thanks John and Sandy Phillips for sustaining him with coffee at Le Pas Opton in Summer 2006 and 2009.

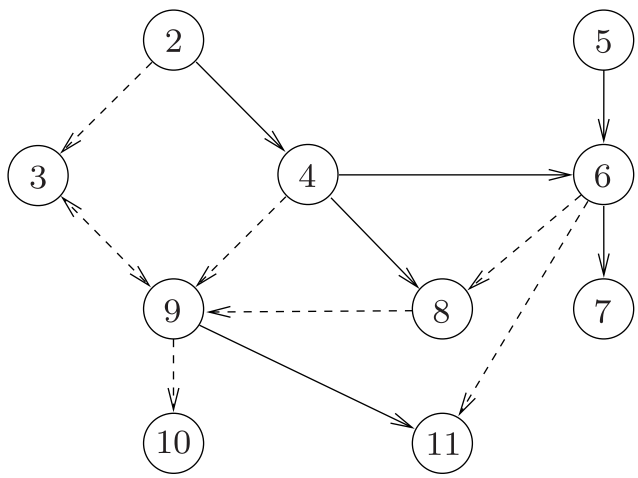
We both thank our previous and current home institutions Princeton University, the Clay Mathematics Institute, The Ohio State University, Eidgenössische Technische Hochschule Zürich, and the University of East Anglia, for support, including support for several visits, and for providing the rich mathematical environments that made this project possible. We also thank the National Science Foundation for support under NSF grant DMS-0554373.

June 21, 2010

Manfred Einsiedler, Zürich
Thomas Ward, Norwich

Leitfaden

The dependencies between the chapters is illustrated below, with solid lines indicating logical dependency and dotted lines indicating partial or motivational links.



Some possible shorter courses could be made up as follows.

- Chapters 2 & 4: A gentle introduction to ergodic theory and topological dynamics.
- Chapters 2 & 3: A gentle introduction to ergodic theory and the continued fraction map (the dotted line indicates that only parts of Chapter 2 are needed for Chapter 3).
- Chapters 2, 3, & 9: As above, with the connection between the Gauss map and hyperbolic surfaces, and ergodicity of the geodesic flow.
- Chapters 2, 4, & 8: An introduction to ergodic theory for group actions.

The highlights of this book are Chapters 7 and 11. Some more ambitious courses could be made up as follows.

- To Chapter 6: Ergodic theory up to conditional measures and the ergodic decomposition.
- To Chapter 7: Ergodic theory including the Furstenberg–Katznelson–Ornstein proof of Szemerédi's theorem.
- To Chapter 11: Ergodic theory and an introduction to dynamics on homogeneous spaces, including equidistribution of horocycle orbits. A minimal path to equidistribution of horocycle orbits on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ would include the discussions of ergodicity from Chapter 2, genericity from Chapter 4, Haar measure from Chapter 8, the hyperbolic plane from Chapter 9, and ergodicity and mixing from Chapter 11.

Contents

Chapter 1
Motivation

Our main motivation throughout the book will be to understand the applications of ergodic theory to certain problems outside of ergodic theory, in particular to problems in number theory. As we will see, this requires a good understanding of particular examples, which will often be of an algebraic nature. Therefore, we will start with a few concrete examples, and state a few theorems arising from ergodic theory, some of which we will prove within this volume. In Section 1.8 we will discuss ergodic theory as a subject in more general terms[(1)].

1.1 Examples of Ergodic Behavior

The *orbit* of a point $x \in X$ under a transformation $T : X \to X$ is the set $\{T^n(x) \mid n \in \mathbb{N}\}$. The structure of the orbit can say a great deal about the original point x. In particular, the behavior of the orbit will sometimes detect special properties of the point. A particularly simple instance of this appears in the next example.

Example 1.1. Write $\mathbb{T}$ for the quotient group $\mathbb{R}/\mathbb{Z} = \{x + \mathbb{Z} \mid x \in \mathbb{R}\}$, which can be identified with a circle (as a topological space, this can also be obtained as a quotient space of $[0, 1]$ by identifying 0 with 1); there is a natural bijection between $\mathbb{T}$ and the half-open interval $[0, 1)$ obtained by sending the coset $x+\mathbb{Z}$ to the fractional part of x. Let $T : \mathbb{T} \to \mathbb{T}$ be defined by $T(x) = 10x \pmod 1$. Then $x \in \mathbb{T}$ is rational if and only if the orbit of x under T is finite. To see this, assume first that $x = \frac{p}{q}$ is rational. In this case the orbit of x is some subset of $\{0, \frac{1}{q}, \dots, \frac{q-1}{q}\}$. Conversely, if the orbit is finite then there must be integers m, n with $1 \leqslant n < m$ for which $T^m(x) = T^n(x)$. It follows that $10^m x = 10^n x + k$ for some $k \in \mathbb{N}$, so x is rational.

Detecting the behavior of the orbit of a given point is usually not so straightforward. Ergodic theory generally has more to say about the orbit of "typical" points, as illustrated in the next example. Write χ_A for the indicator function of a set,

$$\chi_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A. \end{cases}$$

Example 1.2. This example recovers a result due to Borel [40]. We shall see later that the map $T : \mathbb{T} \to \mathbb{T}$ defined by $T(x) = 10x \pmod 1$ preserves Lebesgue measure m on $[0,1)$ (see Definition 2.1), and is *ergodic* with respect to m (see Definition 2.13). A consequence of the pointwise ergodic theorem (Theorem 2.30) is that for any interval

$$A(j,k) = [\tfrac{j}{10^k}, \tfrac{j+1}{10^k}),$$

we have

$$\frac{1}{N}\sum_{i=0}^{N-1} \chi_{A(j,k)}(T^i x) \longrightarrow \int_0^1 \chi_{A(j,k)}(x)\,\mathrm{d}m(x) = \frac{1}{10^k} \tag{1.1}$$

as $N \to \infty$, for almost every x (that is, for all x in the complement of a set of zero measure, which will be denoted a.e.). For any block $j_1 \dots j_k$ of k decimal digits, the convergence in equation (1.1) with $j = 10^{k-1}j_1 + 10^{k-2}j_2 + \dots + j_k$ shows that the block $j_1 \dots j_k$ appears with asymptotic frequency $\frac{1}{10^k}$ in the decimal expansion of almost every real number in $[0,1]$.

Even though the ergodic theorem only concerns the orbital behavior of typical points, there are situations where one is able to describe the orbits for *all* starting points.

Example 1.3. We show later that the circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ defined by $R_\alpha(t) = t + \alpha \pmod 1$ is *uniquely ergodic* if α is irrational (see Definition 4.9 and Example 4.11). A consequence of this is that for any interval $[a,b) \subseteq [0,1) = \mathbb{T}$,

$$\frac{1}{N}\sum_{n=0}^{N-1} \chi_{[a,b)}(R_\alpha^n(t)) \longrightarrow b - a \tag{1.2}$$

as $N \to \infty$ for *every* $t \in \mathbb{T}$ (see Theorem 4.10 and Lemma 4.17). As pointed out by Arnol′d and Avez [7] this equidistribution result may be used to find the density of appearance of the digits(2) in the sequence $1, 2, 4, 8, 1, 3, 6, 1, \dots$ of first digits of the powers of 2:

$$\mathbf{1}, \mathbf{2}, \mathbf{4}, \mathbf{8}, \mathbf{1}6, \mathbf{3}2, \mathbf{6}4, \mathbf{1}28, \mathbf{2}56, \mathbf{5}12, \mathbf{1}024, \dots.$$

A set $A \subseteq \mathbb{N}$ is said to have *density* $\mathbf{d}(A)$ if

$$\mathbf{d}(A) = \lim_{k\to\infty} \frac{1}{k}\left|A \cap [1,k]\right|$$

exists. Notice that 2^n has first digit k for some $k \in \{1, 2, \dots, 9\}$ if and only if

$$\log_{10} k \leqslant \{n \log_{10} 2\} < \log_{10}(k+1),$$

where we write $\{t\}$ for the fractional part of the real number t.

Since $\alpha = \log_{10} 2$ is irrational, we may apply equation (1.2) to deduce that

$$\begin{aligned} \frac{\left|\{n \,|\, 0 \leqslant n \leqslant N-1, \, 1^{\text{st}} \text{ digit of } 2^n \text{ is } k\}\right|}{N} &= \frac{1}{N}\sum_{n=0}^{N-1} \chi_{[\log_{10} k, \log_{10}(k+1))}(R_\alpha^n(0)) \\ &\to \log_{10}\left(\frac{k+1}{k}\right) \end{aligned}$$

as $N \to \infty$.

Thus the first digit $k \in \{1, \dots, 9\}$ appears with density $\log_{10}\left(\frac{k+1}{k}\right)$, and it follows in particular that the digit 1 is the most common leading digit in the sequence of powers of 2.

Exercises for Section 1.1

Exercise 1.1.1. A point $x \in X$ is said to be *periodic* for the map $T : X \to X$ if there is some $k \geqslant 1$ with $T^k(x) = x$, and *pre-periodic* if the orbit of x under T is finite. Describe the periodic points and the pre-periodic points for the map $x \mapsto 10x \pmod 1$ from Example 1.1.

Exercise 1.1.2. Prove that the orbit of any point $x \in \mathbb{T}$ under the map R_α on $\mathbb{T}$ for α irrational is dense (that is, for any $\varepsilon > 0$ and $t \in \mathbb{T}$ there is some $k \in \mathbb{N}$ for which $T^k x$ lies within ε of t). Deduce that for any finite block of decimal digits, there is some power of 2 that begins with that block of digits.

1.2 Equidistribution for Polynomials

A sequence $(a_n)_{n\in\mathbb{N}}$ of numbers in $[0, 1)$ is said to be equidistributed if

$$\mathbf{d}(\{n \in \mathbb{N} \mid a \leqslant a_n < b\}) = b - a$$

for all a, b with $0 \leqslant a < b \leqslant 1$. A classical result of Weyl [380] extends the equidistribution of the numbers $(n\alpha)_{n\in\mathbb{N}}$ modulo 1 for irrational α to the values of any polynomial with an irrational coefficient*.

* Numbered theorems like Theorem 1.4 in the main text are proved in this volume, but not necessarily in the chapter in which they first appear.

Theorem 1.4 (Weyl). *Let $p(n) = a_k n^k + \cdots + a_0$ be a real polynomial with at least one coefficient among $a_1, \dots, a_k$ irrational. Then the sequence $(p(n))$ is equidistributed modulo* 1*.*

Furstenberg extended unique ergodicity to a dynamically defined extension of the irrational circle rotation described in Example 1.3, giving an elegant ergodic-theoretic proof of Theorem 1.4. This approach will be discussed in Section 4.4.

Exercises for Section 1.2

Exercise 1.2.1. Describe what Theorem 1.4 can tell us about the leading digits of the powers of 2.

1.3 Szemerédi's Theorem

Szemerédi, in an intricate and difficult combinatorial argument, proved a long-standing conjecture of Erdős and Turán [85] in his paper [357]. A set S of integers is said to have *positive upper Banach density* if there are sequences (m_j) and (n_j) with $n_j - m_j \to \infty$ as $j \to \infty$ with the property that

$$\lim_{j\to\infty} \frac{|S \cap [m_j, n_j]|}{n_j - m_j} > 0.$$

Theorem 1.5 (Szemerédi). *Any subset of the integers with positive upper Banach density contains arbitrarily long arithmetic progressions.*

Furstenberg [102] (see also his book [103] and the article of Furstenberg, Katznelson and Ornstein [107]) showed that Szemerédi's theorem would follow from a generalization of Poincaré's recurrence theorem, and proved that generalization. The connection between recurrence and Szemerédi's theorem will be explained in Section 7.3, and Furstenberg's proof of the generalization of Poincaré recurrence needed will be presented in Chapter 7. There are a great many more theorems in this direction which we cannot cover, but it is worth noting that many of these further theorems to date only have proofs using ergodic theory.

More recently, Gowers [122] has given a different proof of Szemerédi's theorem, and in particular has found the following effective form of it*.

Theorem (Gowers). For every integer $s \geqslant 1$ and sufficiently large integer N, every subset of $\{1, 2, \dots, N\}$ with at least

* Theorems and other results that are not numbered will not be proved in this volume, but will also not be used in the main body of the text.

$$N(\log \log N)^{-2^{-2^{s+9}}}$$

elements contains an arithmetic progression of length s.

Typically proofs using ergodic theory are not effective: Theorem 1.5 easily implies a finitistic version of Szemerédi's theorem, which states that for every s and constant $c > 0$ and all sufficiently large $N = N(s, c)$, any subset of $\{1, \dots, N\}$ with at least cN elements contains an arithmetic progression of length s. However, the dependence of N on c is not known by this means, nor is it easily deduced from the proof of Theorem 1.5. Gowers' Theorem, proved by different methods, does give an explicit dependence.

We mention Gowers' Theorem to indicate some of the limitations of ergodic theory. While ergodic methods have many advantages, proving quite general theorems which often have no other proofs, they also have disadvantages, one of them being that they tend to be non-effective.

Subsequent development of the combinatorial and arithmetic ideas by Goldston, Pintz and Yıldırım [118][(3)] and Gowers, and of the ergodic method by Host and Kra [159] and Ziegler [392], has influenced some arguments of Green and Tao [127] in their proof of the following long-conjectured result. This is a good example of how asking for effective or quantitative versions of existing results can lead to new qualitative theorems.

Theorem (Green and Tao). The set of primes contains arbitrarily long arithmetic progressions.

1.4 Indefinite Quadratic Forms and Oppenheim's Conjecture

Our purpose here is to provide enough background in ergodic theory to quickly reach some understanding of a few deeper results in number theory and combinatorial number theory where ergodic theory has made a contribution. Along the way we will develop a good portion of ergodic theory as well as some other background material. In the rest of this introductory chapter, we mention some more highlights of the many connections between ergodic theory and number theory. The results in this section, and in Sections 1.5 and 1.6, will not be covered in this book, but we plan to discuss them in a subsequent volume.

The next theorem was conjectured in a weaker form by Oppenheim in 1929 and eventually proved by Margulis in the stronger form stated here in 1986 [247], [250]. In order to state the result, we recall some terminology for quadratic forms.

A *quadratic form* in n variables is a homogeneous polynomial $Q(x_1, \dots, x_n)$ of degree two. Equivalently, a quadratic form is a polynomial Q for which there is a symmetric $n \times n$ matrix A_Q with

$$Q(x_1, \ldots, x_n) = (x_1, \ldots, x_n) A_Q (x_1, \ldots, x_n)^{\mathrm{t}}.$$

Since A_Q is symmetric, there is an orthogonal matrix P for which $P^{\mathrm{t}} A_Q P$ is diagonal. This means there is a different coordinate system $y_1, \ldots, y_n$ for which

$$Q(x_1, \ldots, x_n) = c_1 y_1^2 + \cdots + c_n y_n^2.$$

The quadratic form is called *non-degenerate* if all the coefficients c_i are non-zero (equivalently, if $\det A_Q \neq 0$), and is called *indefinite* if the coefficients c_i do not all have the same sign. Finally, the quadratic form is said to be *rational* if its coefficients (equivalently, if the entries of the matrix A_Q) are rational*.

Theorem (Margulis). Let Q be an indefinite non-degenerate quadratic form in $n \geqslant 3$ variables that is not a multiple of a rational form. Then $Q(\mathbb{Z}^n)$ is a dense subset of $\mathbb{R}$.

It is easy to see that two of the stated conditions are necessary for the result: if the form Q is definite then the elements of $Q(\mathbb{Z}^n)$ all have the same sign, and if Q is a multiple of a rational form, then $Q(\mathbb{Z}^n)$ lies in a discrete subgroup of $\mathbb{R}$. The assumption that Q is non-degenerate and n is at least 3 are also necessary, though this is less obvious (requiring in particular the notion of badly approximable numbers from the theory of Diophantine approximation, which will be introduced in Section 3.3). This shows that the theorem as stated above is in the strongest possible form. Weaker forms of this result have been obtained by other methods, but the full strength of Margulis' Theorem at the moment requires dynamical arguments (for example, ergodic methods).

Proving the theorem involves understanding the behavior of *orbits* for the action of the subgroup $\mathrm{SO}(2,1) \leqslant \mathrm{SL}_3(\mathbb{R})$ on points $x \in \mathrm{SL}_3(\mathbb{Z}) \backslash \mathrm{SL}_3(\mathbb{R})$ (the space of right cosets of $\mathrm{SL}_3(\mathbb{Z})$ in $\mathrm{SL}_3(\mathbb{R})$); these may be thought of as sets of the form $x\,\mathrm{SO}(2,1)$. As it turns out (a consequence of Raghunathan's conjectures, discussed briefly in Section 1.7), such orbits are either closed subsets of $\mathrm{SL}_3(\mathbb{Z}) \backslash \mathrm{SL}_3(\mathbb{R})$ or are dense in $\mathrm{SL}_3(\mathbb{Z}) \backslash \mathrm{SL}_3(\mathbb{R})$. Moreover, the former case happens if and only if the point x corresponds in an appropriate sense to a rational quadratic form.

Margulis' Theorem may be viewed as an extension of Example 1.3 to higher degree in the following sense. The statement that every orbit under the map $R_\alpha(t) = t + \alpha \pmod 1$ is dense in $\mathbb{T}$ is equivalent to the statement that if L is a linear form in two variables that is not a multiple of a rational form, then $L(\mathbb{Z}^2)$ is dense in $\mathbb{R}$.

* Note that the rationality of Q cannot be detected using the coefficients $c_1, \ldots, c_n$ after the real coordinate change.

1.5 Littlewood's Conjecture

For a real number t, write $\langle t\rangle$ for the distance from t to the nearest integer,

$$\langle t\rangle = \min_{q\in\mathbb{Z}} |t-q|.$$

The theory of continued fractions (which will be described in Chapter 3) shows that for any real number u, there is a sequence (q_n) with $q_n \to \infty$ such that $q_n\langle q_n u\rangle < 1$ for all $n \geqslant 1$. Littlewood conjectured the following in the 1930s: for any real numbers u, v,

$$\liminf_{n\to\infty} n\langle nu\rangle\langle nv\rangle = 0.$$

Some progress was made on this for restricted classes of numbers u and v by Cassels and Swinnerton-Dyer [50], Pollington and Velani [290], and others, but the problem remains open. In 2003 Einsiedler, Katok and Lindenstrauss [79] used ergodic methods to prove that the set of exceptions to Littlewood's conjecture is extremely small.

Theorem (Einsiedler, Katok & Lindenstrauss). Let

$$\Theta = \left\{(u,v)\in\mathbb{R}^2 \mid \liminf_{n\to\infty} n\langle nu\rangle\langle nv\rangle > 0\right\}.$$

Then the Hausdorff dimension of Θ is zero.

In fact the result in [79] is a little stronger, showing that Θ satisfies a stronger property that implies it has Hausdorff dimension zero. The proof relies on a partial classification of certain invariant measures on $\mathrm{SL}_3(\mathbb{Z})\backslash\mathrm{SL}_3(\mathbb{R})$. This is part of the theory of *measure rigidity*, and the particular type of phenomenon seen has its origins in work of Furstenberg [100], who showed that the natural action $t \mapsto at \pmod 1$ of the semi-group generated by two multiplicatively independent natural numbers a_1 and a_2 on $\mathbb{T}$ has, apart from finite sets, no non-trivial closed invariant sets. He asked if this system could have any non-atomic ergodic invariant measures other than Lebesgue measure. Partial results on this and related generalizations led to the formulation of far-reaching conjectures by Margulis [251], by Furstenberg, and by Katok and Spatzier [183], [184]. A special case of these conjectures concerns actions of the group A of positive diagonal matrices in $\mathrm{SL}_k(\mathbb{R})$ for $k \geqslant 3$ on the space $\mathrm{SL}_k(\mathbb{Z})\backslash\mathrm{SL}_k(\mathbb{R})$: if μ is an A-invariant ergodic probability measure on this space, is there a closed connected group $L \geqslant A$ for which μ is the unique L-invariant measure on a single closed L-orbit (that is, is μ *homogeneous*)?

In the work of Einsiedler, Katok and Lindenstrauss the conjecture stated above is proved under the additional hypothesis that the measure μ gives positive entropy to some one-parameter subgroup of A, which leads to the theorem concerning Θ. A complete classification of these measures without entropy hypotheses would imply the full conjecture of Littlewood.

In this volume we will develop the minimal background needed for the ergodic approach to continued fractions (see Chapter 3) as well as the basic theorems concerning the action of the diagonal subgroup A on the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ (see Chapter 9). We will also describe the connection between these two topics, which will help us to prove results about the continued fraction expansion and about the action of A.

1.6 Integral Quadratic Forms

An important topic in number theory, both classical and modern, is that of integral quadratic forms. A quadratic form $Q(x_1,\dots,x_n)$ is said to be *integral* if its coefficients are integers.

A natural problem(4) is to describe the range $Q(\mathbb{Z}^n)$ of an integral quadratic form evaluated on the integers. A classical theorem of Lagrange(5) on the sum of four squares says that $Q_0(\mathbb{Z}^4)=\mathbb{N}_0$ if

$$Q_0(x_1,x_2,x_3,x_4)=x_1^2+x_2^2+x_3^2+x_4^2,$$

solving the problem for a particular form.

More generally, Kloosterman, in his dissertation of 1924, found an asymptotic formula for the number of expressions for an integer in terms of a positive definite quadratic form Q in five or more variables and deduced that any large integer lies in $Q(\mathbb{Z}^n)$ if it satisfies certain congruence conditions. The case of four variables is much deeper, and required him to make new deep developments in analytic number theory; special cases appeared in [201] and the full solution in [202], where he proved that an integral definite quadratic form Q in four variables represents all large enough integers a for which there is no *congruence obstruction*. Here we say that $a\in\mathbb{N}$ has a congruence obstruction for the quadratic form $Q(x_1,\dots,x_n)$ if a modulo d is not a value of $Q(x_1,\dots,x_n)$ modulo d for some $d\in\mathbb{N}$.

The methods that are usually applied to prove these theorems are purely number-theoretic. Ellenberg and Venkatesh [83] have introduced a method that combines number theory, algebraic group theory, and ergodic theory to prove results in this field, leading to a different proof of the following special case of Kloosterman's Theorem.

Theorem (Kloosterman). Let Q be a positive definite quadratic form with integer coefficients in at least 6 variables. Then all large enough integers that do not fail the congruence conditions can be represented by the form Q.

That is, if $a\in\mathbb{N}$ is larger than some constant that depends on Q and for every $d>0$ there exists some $x_d\in\mathbb{Z}^n$ with $Q(x_d)=a$ modulo d, then there exists some $x \in \mathbb{Z}^n$ with $Q(x) = a$. This theorem has purely number-theoretic proofs (see the survey by Schulze-Pillot [335]).

In fact Ellenberg and Venkatesh proved in [83] a different theorem that currently does not have a purely number-theoretic proof. They considered the problem of representing a quadratic form by another quadratic form: If Q is an integral positive definite(6) quadratic form in n variables and Q' is another such form in $m < n$ variables, then one can ask whether there is a subgroup $\Lambda \leqslant \mathbb{Z}^n$ generated by m elements such that when Q is restricted to Λ the resulting form is isomorphic to Q'. This question has, for instance, been studied by Gauss in the case of $m = 2$ and $n = 3$ in the *Disquisitiones Arithmeticae* [111]. As before, there can be congruence obstructions to this problem, which are best phrased in terms of p-adic numbers. Roughly speaking, Ellenberg and Venkatesh show that for a given integral definite quadratic form Q in n variables, every integral definite quadratic form Q' in $m \leqslant n-5$ variables(7) that does not have small image values can be represented by Q, unless there is a congruence obstruction. The assumption that the quadratic form Q' does not have small image means that $\min_{x\in\mathbb{Z}^m\smallsetminus\{0\}} Q'(x)$ should be bigger than some constant that depends on Q.

The ergodic theory used in [83] is related to Raghunathan's conjecture mentioned in Section 1.4 and discussed again in Section 1.7 below, and is the result of work by many people, including Margulis, Mozes, Ratner, Shah, and Tomanov.

1.7 Dynamics on Homogeneous Spaces

Let $G \leqslant \mathrm{SL}_n(\mathbb{R})$ be a closed linear group over the reals (or over a local field; see Section 9.3 for a precise definition), let $\Gamma < G$ be a discrete subgroup(8), and let $H < G$ be a closed subgroup. For example, the case $G = \mathrm{SL}_3(\mathbb{R})$ and $\Gamma = \mathrm{SL}_3(\mathbb{Z})$ arises in Section 1.4 with $H = \mathrm{SO}(2,1)$, and arises in Section 1.5 with $H = A$. Dynamical properties of the action of right multiplication by elements of H on the homogeneous space $X = \Gamma\backslash G$ is important for numerous problems(9). Indeed, all the results in Sections 1.4–1.6 may be proved by studying concrete instances of such systems. We do not want to go into the details here, but simply mention a few highlights of the theory.

There are many important and general results on the ergodicity and mixing behavior of natural measures on such quotients (see Chapter 2 for the definitions). These results (introduced in Chapters 9 and 11) are interesting in their own right, but have also found applications to the problem of counting integer (and, more recently, rational) points on groups (or certain other varieties). The first instance of this can be found in Margulis's thesis [252], where this approach is used to find the asymptotics for the number of closed geodesics on compact manifolds of negative curvature. Independently, Eskin and McMullen [86] found the same method and applied it to a counting problem in certain varieties, which re-proved certain cases of the theorems in the work of Duke, Rudnick and Sarnak [76] in a simpler manner. However, as discussed in Section 1.1, the most difficult – and sometimes most interesting – problem is to understand the orbit of a given point rather than the orbit of almost every point. Indeed, the solution of Oppenheim's conjecture in Section 1.4 by Margulis involved understanding the $\mathrm{SO}(2,1)$-orbit of a point in $\mathrm{SL}_3(\mathbb{Z})\backslash \mathrm{SL}_3(\mathbb{R})$ corresponding to the given quadratic form.

We need one more definition before we can state a general theorem in this direction. A subgroup $U < \mathrm{SL}_n(\mathbb{R})$ is called a *one-parameter unipotent subgroup* if U is the image of $\mathbb{R}w$ under the exponential map, for some matrix $w \in \mathrm{Mat}_{nn}$ satisfying $w^n = 0$ (that is, w is nilpotent and $\exp(tw)$ has only 1 as an eigenvalue, hence the name). For example, there is an index two subgroup $H \leqslant \mathrm{SO}(2,1)$ which is generated by one-parameter unipotent subgroups. However, notice that the diagonal subgroup A is not generated by one-parameter unipotent subgroups.

Raghunathan conjectured that if the subgroup H is generated by one-parameter unipotent subgroups, then the closures of orbits xH are always of the form xL for some closed connected subgroup L of G that contains H. This reduces the properties of orbit closures (a dynamical problem) to the algebraic problem of deciding for which closed connected subgroups L the orbit xL is closed.

Ratner [305] proved this important result using methods from ergodic theory. In fact, she deduced Raghunathan's conjecture from Dani's conjecture[(10)] regarding H-invariant measures, which she proved first in the series of papers [302], [303] and [304].

To date there have been numerous applications of the above theorem, and certain extensions of it. To name a few more seemingly unrelated applications, Elkies and McMullen [82] have applied these theorems to obtain the distribution of the gaps in the sequence of fractional parts of $\sqrt{n}$, and Vatsal [366] has studied values of certain L-functions using the p-adic version of the theorems. There are further applications of the theory too numerous to describe here, but the examples above show again the variety of fields that have connections to ergodic theory.

We will discuss a few special cases of the conjectures of Raghunathan and Dani. Example 1.3, Section 4.4, Chapter 10, Section 11.5, and Section 11.7 treat special cases, some of which were known before the conjectures were formulated.

1.8 An Overview of Ergodic Theory

Having seen some statements that qualify as being ergodic in nature, and some of the many important applications of ergodic theory to number theory, in this short section we give a brief overview of ergodic theory. If this is not already clear, notice that it is a rather diffuse subject with ill-defined boundaries[(11)].

Ergodic theory is the study of long-term behavior in dynamical systems from a statistical point of view. Its origins therefore are intimately connected with the time evolution of systems modeled by measure-preserving actions of the reals or the integers, with the action representing the passage of time. Related approaches, using probabilistic methods to study the evolution of systems, also arose in statistical physics, where other natural symmetries – typically reflected by the presence of a $\mathbb{Z}^d$-action – arise. The rich interaction between arithmetic and geometry present in measure-preserving actions of (lattices in) Lie groups quickly emerged, and it is now natural to view ergodic theory as the study of measure-preserving group actions, containing but not limited to several special branches:

(1) The classical study of single measure-preserving transformations.
(2) Measure-preserving actions of $\mathbb{Z}^d$; more generally of countable amenable groups.
(3) Measure-preserving actions of $\mathbb{R}^d$ and more general amenable groups, called flows.
(4) Measure-preserving and more general actions of groups, in particular of Lie groups and of lattices in Lie groups.

Some of the illuminating results in ergodic theory come from the existence of (counter-)examples. Nonetheless, there are many substantial theorems. In addition to fundamental results (the pointwise and mean ergodic theorems themselves, for example) and structural results (the isomorphism theorem of Ornstein, Krieger's theorem on the existence of generators, the isomorphism invariance of entropy), ergodic theory and its way of thinking have made dramatic contributions to many other fields.

Notes to Chapter 1

[(1)](Page 1) The origins of the word 'ergodic' are not entirely clear. Boltzmann coined the word *monode* (unique μὸνος, nature εἶδος) for a set of probability distributions on the phase space that are invariant under the time evolution of a Hamiltonian system, and *ergode* for a monode given by uniform distribution on a surface of constant energy. Ehrenfest and Ehrenfest (in an influential encyclopedia article of 1912, translated as [78]) called a system *ergodic* if each surface of constant energy comprised a single time orbit — a notion called *isodic* by Boltzmann (same ισος, path ὀδός) — and *quasi-ergodic* if each surface has dense orbits. The Ehrenfests themselves suggested that the etymology of the word *ergodic* lies in a different direction (work έργον, path ὀδός). This work stimulated interest in the mathematical foundations of statistical mechanics, leading eventually to Birkhoff's formulation of the *ergodic hypothesis* and the notion of systems for which almost every orbit in the sense of measure spends a proportion of time in a given set in the phase space in proportion to the measure of the set.

[(2)](Page 2) Questions of this sort were raised by Gel′fand; he considered the vector of first digits of the numbers $(2^n, 3^n, 4^n, 5^n, 6^n, 7^n, 8^n, 9^n)$ and asked if (for example) there is a value of $n > 1$ for which this vector is $(2, 3, 4, 5, 6, 7, 8, 9)$. This circle of problems is related to the classical Poncelet's porism, as explained in an article by King [194]. The influence of Poncelet's book [292] is discussed by Gray [126, Chap. 27].

(3) (Page 5) See also the account with some simplifications by Goldston, Motohashi, Pintz, and Yıldırım [117] and the survey by Goldston, Pintz and Yıldırım [119].

(4) (Page 8) In a more general form, this is the 11th of Hilbert's famous set of problems formulated for the 1900 International Congress of Mathematics.

(5) (Page 8) Bachet conjectured the result, and Diophantus stated it; there are suggestions that Fermat may have known it. The first published proof is that of Lagrange in 1770; a standard proof may be found in [87, Sect. 2.3.1] for example.

(6) (Page 9) For *indefinite* quadratic forms there is a very successful algebraic technique, namely strong approximation for algebraic groups (an account may be found in the monograph [286] of Platonov and Rapinchuk), so ergodic theory does not enter into the discussion.

(7) (Page 9) Under an additional congruence condition on Q' the method also works for $m \leqslant n - 3$.

(8) For some of the statements made here one actually has to assume that Γ is a *lattice*; see Section 9.4.3.

(9) (Page 9) Further readings from various perspectives on the ergodic theory of homogeneous spaces may be found in the books of Bekka and Mayer [21], Feres [90], Starkov [350], Witte Morris [384], [386] and Zimmer [393].

(10) (Page 10) For linear groups over local fields, and products of such groups, the conjectures of Dani (resp. Raghunathan) have been proved by Margulis and Tomanov [253] and independently by Ratner [306].

(11) (Page 11) Some of the many areas of ergodic theory that we do not treat in a substantial way, and other general sources on ergodic theory, may be found in the following books: the connection with information theory in the work of Billingsley [31] and Shields [342]; a wide-ranging overview of ergodic theory in that of Cornfeld, Fomin and Sinaĭ [60]; ergodic theory developed in the language of joinings in the work of Glasner [116]; more on the theory of entropy and generators in books by Parry [277], [279]; a thorough development of the fundamentals of the measurable theory, including the isomorphism and generator theory, in the book of Rudolph [324].

Chapter 2
Ergodicity, Recurrence and Mixing

In this chapter the basic objects studied in ergodic theory, measure-preserving transformations, are introduced. Some examples are given, and the relationship between various mixing properties is described. Background on measure theory appears in Appendix A.

2.1 Measure-Preserving Transformations

Definition 2.1. Let $(X, \mathscr{B}, \mu)$ and $(Y, \mathscr{C}, \nu)$ be probability spaces. A map* ϕ from X to Y is *measurable* if $\phi^{-1}(A) \in \mathscr{B}$ for any $A \in \mathscr{C}$, and is *measure-preserving* if it is measurable and $\mu(\phi^{-1}B) = \nu(B)$ for all $B \in \mathscr{C}$. If in addition ϕ^{-1} exists almost everywhere and is measurable, then ϕ is called an *invertible measure-preserving map.* If $T : (X, \mathscr{B}, \mu) \to (X, \mathscr{B}, \mu)$ is measure-preserving, then the measure μ is said to be T*-invariant*, $(X, \mathscr{B}, \mu, T)$ is called a *measure-preserving system* and T a *measure-preserving transformation.*

Notice that we work with pre-images of sets rather than images to define measure-preserving maps (just as pre-images of sets are used to define measurability of real-valued functions on a measure space). As pointed out in Example 2.4 and Exercise 2.1.3, it is essential to do this. In order to show that a measurable map is measure-preserving, it is sufficient to check this property on a family of sets whose disjoint unions approximate all measurable sets (see Appendix A for the details).

Most of the examples we will encounter are algebraic or are motivated by algebraic or number-theoretic questions. This is not representative of ergodic theory as a whole, where there are many more types of examples (two non-algebraic classes of examples are discussed on the website [81]).

* In this measurable setting, a map is allowed to be undefined on a set of zero measure. Definition 2.7 will give one way to view this: a measurable map undefined on a set of zero measure can be viewed as an everywhere-defined map on an isomorphic measure space.

We define the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ to be the set of cosets of $\mathbb{Z}$ in $\mathbb{R}$ with the quotient topology induced by the usual topology on $\mathbb{R}$. This topology is also given by the metric

$$\mathsf{d}(r + \mathbb{Z}, s + \mathbb{Z}) = \min_{m \in \mathbb{Z}} |r - s + m|,$$

and this makes $\mathbb{T}$ into a compact abelian group (see Appendix C). The interval $[0, 1) \subseteq \mathbb{R}$ is a fundamental domain for $\mathbb{Z}$: that is, every element of $\mathbb{T}$ may be written in the form $t + \mathbb{Z}$ for a unique $t \in [0, 1)$. We will frequently use $[0, 1)$ to define points (and subsets) in $\mathbb{T}$, by identifying $t \in [0, 1)$ with the unique coset $t + \mathbb{Z} \in \mathbb{T}$ defined by t.

Example 2.2. For any $\alpha \in \mathbb{R}$, define the *circle rotation by* α to be the map

$$R_\alpha : \mathbb{T} \to \mathbb{T},\ R_\alpha(t) = t + \alpha \pmod 1.$$

We claim that R_α preserves the Lebesgue measure $m_{\mathbb{T}}$ on the circle. By Theorem A.8, it is enough to prove it for intervals, where it is clear. Alternatively, we may note that Lebesgue measure is a Haar measure on the compact group $\mathbb{T}$, which is invariant under any translation by construction (see Sections 8.3 and C.2).

Example 2.3. A generalization of Example 2.2 is a rotation on a compact group. Let X be a compact group, and let g be an element of X. Then the map $T_g : X \to X$ defined by $T_g(x) = gx$ preserves the (left) Haar measure m_X on X. The Haar measure on a locally compact group is described in Appendix C, and may be thought of as the natural generalization of the Lebesgue measure to a general locally compact group.

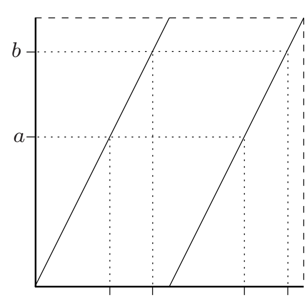


Fig. 2.1: The pre-image of $[a, b)$ under the circle-doubling map.

Example 2.4. The *circle-doubling map* is $T_2 : \mathbb{T} \to \mathbb{T}$, $T_2(t) = 2t \pmod 1$. We claim that T_2 preserves the Lebesgue measure $m_{\mathbb{T}}$ on the circle. By Theorem A.8, it is sufficient to check this on intervals, so let $B = [a, b) \subseteq [0, 1)$

be any interval. Then it is easy to check that

$$T_2^{-1}(B) = \left[\tfrac{a}{2}, \tfrac{b}{2}\right) \cup \left[\tfrac{a}{2} + \tfrac{1}{2}, \tfrac{b}{2} + \tfrac{1}{2}\right)$$

is a disjoint union (thinking of a and b as real numbers; see Figure 2.1), so

$$m_{\mathbb{T}}\left(T_2^{-1}(B)\right) = \tfrac{1}{2}(b-a) + \tfrac{1}{2}(b-a) = b-a = m_{\mathbb{T}}(B).$$

Notice that the measure-preserving property cannot be seen by studying forward iterates: if I is a small interval, then $T_2(I)$ is an interval* with total length $2(b-a)$.

Example 2.5. Generalizing Example 2.4, let X be a compact abelian group and let $T : X \to X$ be a surjective endomorphism. Then T preserves the Haar measure m_X on X by the following argument. Define a measure μ on X by $\mu(A) = m_X(T^{-1}A)$. Then, given any $x \in X$ pick y with $T(y) = x$ and notice that

$$\mu(A+x) = m_X(T^{-1}(A+x)) = m_X(T^{-1}A+y) = m_X(T^{-1}A) = \mu(A),$$

so μ is a translation-invariant Borel probability on X (this just means a probability measure defined on the Borel σ-algebra). Since the normalized Haar measure is the unique measure with this property, μ must be m_X, which means that T preserves the Haar measure m_X on X.

One of the ways in which a measure-preserving transformation may be studied is via its induced action on some natural space of functions. Given any function $f : X \to \mathbb{R}$ and map $T : X \to X$, write $f \circ T$ for the function defined by $(f \circ T)(x) = f(Tx)$. As usual we write L^1_μ for the space of (equivalence classes of) measurable functions $f : X \to \mathbb{R}$ with $\int |f| \,\mathrm{d}\mu < \infty$, $\mathscr{L}^\infty$ for the space of measurable bounded functions and $\mathscr{L}^1_\mu$ for the space of measurable integrable functions (in the usual sense of function, in particular defined everywhere; see Section A.3).

Lemma 2.6. *A measure μ on X is T-invariant if and only if*

$$\int f \,\mathrm{d}\mu = \int f \circ T \,\mathrm{d}\mu \tag{2.1}$$

for all $f \in \mathscr{L}^\infty$. Moreover, if μ is T-invariant, then equation (2.1) *holds for $f \in L^1_\mu$.*

Proof. If equation (2.1) holds, then for any measurable set B we may take $f = \chi_B$ to see that

---

* We say that a subset of $\mathbb{T}$ is an interval in $\mathbb{T}$ if it is the image of an interval in $\mathbb{R}$. An interval might therefore be represented in our chosen space of coset representatives $[0, 1)$ by the union of two intervals.

$$\mu(B) = \int \chi_B \,\mathrm{d}\mu = \int \chi_B \circ T \,\mathrm{d}\mu = \int \chi_{T^{-1}B} \,\mathrm{d}\mu = \mu(T^{-1}B),$$

so T preserves μ.

Conversely, if T preserves μ then equation (2.1) holds for any function of the form χ_B and hence for any simple function (see Section A.3). Let f be a non-negative real-valued function in $\mathscr{L}^1_\mu$. Choose a sequence of simple functions (f_n) increasing to f (see Section A.3). Then $(f_n \circ T)$ is a sequence of simple functions increasing to $f \circ T$, and so

$$\int f \circ T \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \circ T \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu,$$

showing that equation (2.1) holds for f. □

One part of ergodic theory is concerned with the structure and classification of measure-preserving transformations. The next definition gives the two basic relationships there may be between measure-preserving transformations[(12)].

Definition 2.7. Let $(X, \mathscr{B}_X, \mu, T)$ and $(Y, \mathscr{B}_Y, \nu, S)$ be measure-preserving systems on probability spaces.

(1) The system $(Y, \mathscr{B}_Y, \nu, S)$ is a *factor* of $(X, \mathscr{B}_X, \mu, T)$ if there are sets X' in $\mathscr{B}_X$ and Y' in $\mathscr{B}_Y$ with $\mu(X') = 1$, $\nu(Y') = 1$, $TX' \subseteq X'$, $SY' \subseteq Y'$ and a measure-preserving map $\phi : X' \to Y'$ with

$$\phi \circ T(x) = S \circ \phi(x)$$

for all $x \in X'$.

(2) The system $(Y, \mathscr{B}_Y, \nu, S)$ is *isomorphic to* $(X, \mathscr{B}_X, \mu, T)$ if there are sets X' in $\mathscr{B}_X$, Y' in $\mathscr{B}_Y$ with $\mu(X')=1$, $\nu(Y')=1$, $TX' \subseteq X'$, $SY' \subseteq Y'$, and an invertible measure-preserving map $\phi : X' \to Y'$ with

$$\phi \circ T(x) = S \circ \phi(x)$$

for all $x \in X'$.

In measure theory it is natural to simply ignore null sets, and we will sometimes loosely think of a factor as a measure-preserving map $\phi : X \to Y$ for which the diagram

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

is commutative, with the understanding that the map is not required to be defined everywhere.

A factor map

$$(X, \mathscr{B}_X, \mu, T) \longrightarrow (Y, \mathscr{B}_Y, \nu, S)$$

will also be described as an *extension* of $(Y, \mathscr{B}_Y, \nu, S)$. The factor $(Y, \mathscr{B}_Y, \nu, S)$ is called trivial if as a measure space Y comprises a single element; the extension is called trivial if ϕ is an isomorphism of measure spaces.

Example 2.8. Define the $(\frac{1}{2}, \frac{1}{2})$ measure $\mu_{(1/2,1/2)}$ on the finite set $\{0, 1\}$ by

$$\mu_{(1/2,1/2)}(\{0\}) = \mu_{(1/2,1/2)}(\{1\}) = \tfrac{1}{2}.$$

Let $X = \{0, 1\}^{\mathbb{N}}$ with the infinite product measure $\mu = \prod_{\mathbb{N}} \mu_{(1/2,1/2)}$ (see Section A.2 and Example 2.9 where we will generalize this example). This space is a natural model for the set of possible outcomes of the infinitely repeated toss of a fair coin. The *left shift map* $\sigma : X \to X$ defined by

$$\sigma(x_0, x_1, \dots) = (x_1, x_2, \dots)$$

preserves μ (since it preserves the measure of the cylinder sets described in Example 2.9). The map $\phi : X \to \mathbb{T}$ defined by

$$\phi(x_0, x_1, \dots) = \sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}}$$

is measure-preserving from (X, μ) to $(\mathbb{T}, m_{\mathbb{T}})$ and $\phi(\sigma(x)) = T_2(\phi(x))$. The map ϕ has a measurable inverse defined on all but the countable set of dyadic rationals $\mathbb{Z}[\frac{1}{2}]/\mathbb{Z}$, where

$$\mathbb{Z}[\tfrac{1}{2}] = \{\tfrac{m}{2^n} \mid m \in \mathbb{Z}, n \in \mathbb{N}\},$$

so this shows that (X, μ, σ) and $(\mathbb{T}, m_{\mathbb{T}}, T_2)$ are measurably isomorphic.

When the underlying space is a compact metric space, the σ-algebra is taken to be the Borel σ-algebra (the smallest σ-algebra containing all the open sets) unless explicitly stated otherwise. Notice that in both Example 2.8 and Example 2.9 the underlying space is indeed a compact metric space (see Section A.2).

Example 2.9. The shift map in Example 2.8 is an example of a one-sided Bernoulli shift. A more general[(13)] and natural two-sided definition is the following. Consider an infinitely repeated throw of a loaded n-sided die. The possible outcomes of each throw are $\{1, 2, \dots, n\}$, and these appear with probabilities given by the probability vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$ (probability vector means each $p_i \geqslant 0$ and $\sum_{i=1}^{n} p_i = 1$), so $\mathbf{p}$ defines a measure $\mu_{\mathbf{p}}$ on the finite sample space $\{1, 2, \dots, n\}$, which is given the discrete topology. The sample space for the die throw repeated infinitely often is

$$\begin{aligned} X &= \{1,2,\dots,n\}^{\mathbb{Z}} \\ &= \{x = (\dots, x_{-1}, x_0, x_1, \dots) \mid x_i \in \{1,2,\dots,n\} \text{ for all } i \in \mathbb{Z}\}. \end{aligned}$$

The measure on X is the infinite product measure $\mu = \prod_{\mathbb{Z}} \mu_{\mathbf{p}}$, and the σ-algebra $\mathscr{B}$ is the Borel σ-algebra for the compact metric space* X, or equivalently is the product σ–algebra defined below and in Section A.2.

A better description of the measure is given via *cylinder sets*. If I is a finite subset of $\mathbb{Z}$, and $\mathbf{a}$ is a map $I \to \{1,2,\dots,n\}$, then the cylinder set defined by I and $\mathbf{a}$ is

$$I(\mathbf{a}) = \{x \in X \mid x_j = \mathbf{a}(j) \text{ for all } j \in I\}.$$

It will be useful later to write $x|_I$ for the ordered block of coordinates

$$x_i x_{i+1} \cdots x_{i+s}$$

when $I = \{i, i+1, \dots, i+s\} = [i, i+s]$. The measure μ is uniquely determined by the property that

$$\mu\left(I(\mathbf{a})\right) = \prod_{i \in I} p_{a(i)},$$

and $\mathscr{B}$ is the smallest σ-algebra containing all cylinders (see Section A.2 for the details).

Now let σ be the (left) shift on X: $\sigma(x) = y$ where $y_j = x_{j+1}$ for all j in $\mathbb{Z}$. Then σ is μ-preserving and $\mathscr{B}$-measurable. So $(X, \mathscr{B}, \mu, \sigma)$ is a measure-preserving system, called the *Bernoulli scheme* or *Bernoulli shift* based on $\mathbf{p}$. A measure-preserving system measurably isomorphic to a Bernoulli shift is sometimes called a Bernoulli automorphism.

The next example, which we learned from Doug Lind, gives another example of a measurable isomorphism and reinforces the point that being a probability space is a finiteness property of the measure, rather than a metric boundedness property of the space. The measure μ on $\mathbb{R}$ described in Example 2.10 makes $(\mathbb{R}, \mu)$ into a probability space.

Example 2.10. Consider the 2-to-1 map $T : \mathbb{R} \to \mathbb{R}$ defined by

$$T(x) = \frac{1}{2}\left(x - \frac{1}{x}\right)$$

for $x \neq 0$, and $T(0) = 0$. For any L^1 function f, the substitution $y = T(x)$ shows that

---

* The topology on X is simply the product topology, which is also the metric topology given by the metric defined by $\mathsf{d}(x,y) = 2^{-k}$ where

$$k = \max\{j \mid x_i = y_i \text{ for } |j| \leqslant k\}$$

if $x \neq y$ and $\mathsf{d}(x,x) = 0$. In this metric, points are close together if they agree on a large block of indices around $0 \in \mathbb{Z}$.

$$\int_{-\infty}^{\infty} f\left(T(x)\right) \frac{\mathrm{d}x}{\pi(1+x^2)} = \int_{-\infty}^{\infty} f(y) \frac{\mathrm{d}y}{\pi(1+y^2)}$$

(in this calculation, note that T is only injective when restricted to $(0, \infty)$ or $(-\infty, 0)$). It follows by Lemma 2.6 that T preserves the probability measure μ defined by

$$\mu([a,b]) = \int_a^b \frac{\mathrm{d}x}{\pi(1+x^2)}.$$

The map $\phi(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$ from $\mathbb{R}$ to $\mathbb{T}$ is an invertible measure-preserving map from $(\mathbb{R}, \mu)$ to $(\mathbb{T}, m_{\mathbb{T}})$ where $m_{\mathbb{T}}$ denotes the Lebesgue measure on $\mathbb{T}$ (notice that the image of ϕ is the subset $(0,1) \subseteq \mathbb{T}$, but this is an invertible map in the measure-theoretic sense).

Define the map $T_2 : \mathbb{T} \to \mathbb{T}$ by $T_2(x) = 2x \pmod 1$ as in Example 2.4. The map ϕ is a measurable isomorphism from $(\mathbb{R}, \mu, T)$ to $(\mathbb{T}, m_{\mathbb{T}}, T_2)$. Example 2.8 shows in turn that $(\mathbb{R}, \mu, T)$ is isomorphic to the one-sided full 2-shift.

It is often more convenient to work with an invertible measure-preserving transformation as in Example 2.9 instead of a non-invertible transformation as in Examples 2.4 and 2.8. Exercise 2.1.7 gives a general construction of an invertible system from a non-invertible one.

Exercises for Section 2.1

Exercise 2.1.1. Show that the space $(\mathbb{T}, \mathscr{B}_{\mathbb{T}}, m_{\mathbb{T}})$ is isomorphic as a measure space to $(\mathbb{T}^2, \mathscr{B}_{\mathbb{T}^2}, m_{\mathbb{T}^2})$.

Exercise 2.1.2. Show that the measure-preserving system $(\mathbb{T}, \mathscr{B}_{\mathbb{T}}, m_{\mathbb{T}}, T_4)$, where $T_4(x) = 4x \pmod 1$, is measurably isomorphic to the product system $(\mathbb{T}^2, \mathscr{B}_{\mathbb{T}^2}, m_{\mathbb{T}^2}, T_2 \times T_2)$.

Exercise 2.1.3. For a map $T : X \to X$ and sets $A, B \subseteq X$, prove the following.

- $\chi_A(T(x)) = \chi_{T^{-1}(A)}(x)$;
- $T^{-1}(A \cap B) = T^{-1}(A) \cap T^{-1}(B)$;
- $T^{-1}(A \cup B) = T^{-1}(A) \cup T^{-1}(B)$;
- $T^{-1}(A \triangle B) = T^{-1}(A) \triangle T^{-1}(B)$.

Which of these properties also hold with the pre-image under T^{-1} replaced by the forward image under T?

Exercise 2.1.4. What happens to Example 2.5 if the map $T : X \to X$ is only required to be a continuous homomorphism?

Exercise 2.1.5. (a) Find a measure-preserving system $(X, \mathscr{B}, \mu, T)$ with a non-trivial factor map $\phi : X \to X$.
(b) Find an invertible measure-preserving system $(X, \mathscr{B}, \mu, T)$ with a non-trivial factor map $\phi : X \to X$.

Exercise 2.1.6. Prove that the circle rotation R_α from Example 2.2 is not measurably isomorphic to the circle-doubling map T_2 from Example 2.4.

Exercise 2.1.7. Let $\mathsf{X} = (X, \mathscr{B}, \mu, T)$ be any measure-preserving system. A sub-σ-algebra $\mathscr{A} \subseteq \mathscr{B}_X$ with $T^{-1}\mathscr{A} = \mathscr{A}$ modulo μ is called a *T-invariant sub-σ-algebra.* Show that the system $\widetilde{\mathsf{X}} = (\widetilde{X}, \widetilde{B}, \widetilde{\mu}, \widetilde{T})$ defined by

- $\widetilde{X} = \{x \in X^{\mathbb{Z}} \mid x_{k+1} = T(x_k) \text{ for all } k \in \mathbb{Z}\}$;
- $(\widetilde{T}(x))_k = x_{k+1}$ for all $k \in \mathbb{Z}$ and $x \in \widetilde{X}$;
- $\widetilde{\mu}\left(\{x \in \widetilde{X} \mid x_0 \in A\}\right) = \mu(A)$ for any $A \in \mathscr{B}$, and $\widetilde{\mu}$ is invariant under $\widetilde{T}$;
- $\widetilde{B}$ is the smallest $\widetilde{T}$-invariant σ-algebra for which the map $\pi : x \mapsto x_0$ from $\widetilde{X}$ to X is measurable;

is an invertible measure-preserving system, and that the map $\pi : x \mapsto x_0$ is a factor map. The system $\widetilde{\mathsf{X}}$ is called the *invertible extension* of X.

Exercise 2.1.8. Show that the invertible extension $\widetilde{\mathsf{X}}$ of a measure-preserving system X constructed in Exercise 2.1.7 has the following universal property. For any extension

$$\phi : (Y, \mathscr{B}_Y, \nu, S) \to (X, \mathscr{B}_X, \mu, T)$$

for which S is invertible, there exists a unique map

$$\widetilde{\phi} : (Y, \mathscr{B}_Y, \nu, S) \to (\widetilde{X}, \widetilde{B}, \widetilde{\mu}, \widetilde{T})$$

for which $\phi = \pi \circ \widetilde{\phi}$.

Exercise 2.1.9. (a) Show that the invertible extension of the circle-doubling map from Example 2.4,

$$X_2 = \{x \in \mathbb{T}^{\mathbb{Z}} \mid x_{k+1} = T_2 x_k \text{ for all } k \in \mathbb{Z}\},$$

is a compact abelian group with respect to the coordinate-wise addition defined by $(x + y)_k = x_k + y_k$ for all $k \in \mathbb{Z}$, and the topology inherited from the product topology on $\mathbb{T}^{\mathbb{Z}}$.
(b) Show that the diagonal embedding $\delta(r) = (r, r)$ embeds $\mathbb{Z}[\frac{1}{2}]$ as a discrete subgroup of $\mathbb{R} \times \mathbb{Q}_2$, and that $X_2 \cong \mathbb{R} \times \mathbb{Q}_2 / \delta(\mathbb{Z}[\frac{1}{2}]) \cong \mathbb{R} \times \mathbb{Z}_2 / \delta(\mathbb{Z})$ as compact abelian groups (see Appendix C for the definition of $\mathbb{Q}_p$ and $\mathbb{Z}_p$). In particular, the map $\widetilde{T_2}$ (which may be thought of as the left shift on X_2, or as the map that doubles in each coordinate) is conjugate to the map

$$(s, r) + \delta(\mathbb{Z}[\tfrac{1}{2}]) \mapsto (2s, 2r) + \delta(\mathbb{Z}[\tfrac{1}{2}])$$

on $\mathbb{R} \times \mathbb{Q}_2/\delta(\mathbb{Z}[\frac{1}{2}])$. The group X_2 constructed in this exercise is a simple example of a *solenoid*.

2.2 Recurrence

One of the central themes in ergodic theory is that of *recurrence*, which is a circle of results concerning how points in measurable dynamical systems return close to themselves under iteration. The first and most important of these is a result due to Poincaré [288] published in 1890; he proved this in the context of a natural invariant measure in the "three-body" problem of planetary orbits, before the creation of abstract measure theory(14). Poincaré recurrence is the pigeon-hole principle for ergodic theory; indeed on a finite measure space it is exactly the pigeon-hole principle.

Theorem 2.11 (Poincaré Recurrence). *Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathscr{B}, \mu)$, and let $E \subseteq X$ be a measurable set. Then almost every point $x \in E$ returns to E infinitely often. That is, there exists a measurable set $F \subseteq E$ with $\mu(F) = \mu(E)$ with the property that for every $x \in F$ there exist integers $0 < n_1 < n_2 < \cdots$ with $T^{n_i}x \in E$ for all $i \geqslant 1$.*

Proof. Let $B = \{x \in E \mid T^n x \notin E \text{ for any } n \geqslant 1\}$. Then

$$B = E \cap T^{-1}(X \smallsetminus E) \cap T^{-2}(X \smallsetminus E) \cap \cdots ,$$

so B is measurable. Now, for any $n \geqslant 1$,

$$T^{-n}B = T^{-n}E \cap T^{-n-1}(X \smallsetminus E) \cap \cdots ,$$

so the sets $B, T^{-1}B, T^{-2}B, \ldots$ are disjoint and all have measure $\mu(B)$ since T preserves μ. Thus $\mu(B) = 0$, so there is a set $F_1 \subseteq E$ with $\mu(F_1) = \mu(E)$ and for which every point of F_1 returns to E at least once under iterates of T. The same argument applied to the transformations T^2, T^3 and so on defines subsets $F_2, F_3, \ldots$ of E with $\mu(F_n) = \mu(E)$ and with every point of F_n returning to E under T^n for $n \geqslant 1$. The set

$$F = \bigcap_{n \geqslant 1} F_n \subseteq E$$

has $\mu(F) = \mu(E)$, and every point of F returns to E infinitely often. $\square$

Poincaré recurrence is entirely a consequence of the measure space being of finite measure, as shown in the next example.

Example 2.12. The map $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x + 1$ preserves the Lebesgue measure $m_{\mathbb{R}}$ on $\mathbb{R}$. Just as in Definition 2.1, this means that

$$m_{\mathbb{R}}(T^{-1}A) = m_{\mathbb{R}}(A)$$

for any measurable set $A \subseteq \mathbb{R}$. For any bounded set $E \subseteq \mathbb{R}$ and any $x \in E$, the set

$$\{n \geqslant 1 \mid T^n x \in E\}$$

is finite. Thus the map T exhibits no recurrence.

The absence of guaranteed recurrence in infinite measure spaces is one of the main reasons why we restrict attention to probability spaces. There is nonetheless a well-developed ergodic theory of transformations preserving an infinite measure, described in the monograph of Aaronson [1].

Theorem 2.11 may be applied when E is a set in some physical system preserving a finite measure that gives E positive measure. In this case it means that almost every orbit of such a dynamical system returns close to its starting point infinitely often (see Exercise 2.2.3(a)). A much deeper property that a dynamical system may have is that *almost every* orbit returns close to *almost every* point infinitely often, and this property is addressed in Section 2.3 (specifically, in Proposition 2.14).

Extending recurrence to multiple recurrence (where the images of a set of positive measure at many different future times is shown to have a non-trivial intersection) is the crucial idea behind the ergodic approach to Szemerédi's theorem (Theorem 1.5). This multiple recurrence generalization of Poincaré recurrence will be proved in Chapter 7.

Exercises for Section 2.2

Exercise 2.2.1. Prove the following version of Poincaré recurrence with a weaker hypothesis (finite additivity in place of countable additivity for the measure) and with a stronger conclusion (a bound on the return time). Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system with μ only assumed to be a finitely additive measure (see equation (A.1)), and let $A \in \mathscr{B}$ have $\mu(A) > 0$. Show that there is some positive $n \leqslant \frac{1}{\mu(A)}$ for which $\mu(A \cap T^{-n}A) > 0$.

Exercise 2.2.2. (a) Use Exercise 2.2.1 to show the following. If $A \subseteq \mathbb{N}$ has positive density, meaning that

$$\mathbf{d}(A) = \lim_{k\to\infty} \frac{1}{k} \left|A \cap [1, k]\right|$$

exists and is positive, prove that there is some $n \geqslant 1$ with $\overline{\mathbf{d}}\left(A \cap (A - n)\right) > 0$ (here $A - n = \{a - n \mid a \in A\}$), where

$$\overline{\mathbf{d}}(B) = \limsup_{k\to\infty} \frac{1}{k} \left|B \cap [1, k]\right|.$$

(b) Can you prove this starting with the weaker assumption that the upper density $\overline{\mathbf{d}}(A)$ is positive, and reaching the same conclusion?

Exercise 2.2.3. (a) Let (X, d) be a compact metric space and let $T : X \to X$ be a continuous map. Suppose that μ is a T-invariant probability measure defined on the Borel subsets of X. Prove that for μ-almost every $x \in X$ there is a sequence $n_k \to \infty$ with $T^{n_k}(x) \to x$ as $k \to \infty$.
(b) Prove that the same conclusion holds under the assumption that X is a metric space, $T : X \to X$ is Borel measurable, and μ is a T-invariant probability measure.

2.3 Ergodicity

Ergodicity is the natural notion of indecomposability in ergodic theory[(15)]. The definition of ergodicity for $(X, \mathscr{B}, \mu, T)$ means that it is impossible to split X into two subsets of positive measure each of which is invariant under T.

Definition 2.13. A measure-preserving transformation $T : X \to X$ of a probability space $(X, \mathscr{B}, \mu)$ is *ergodic* if for any* $B \in \mathscr{B}$,

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1. \tag{2.2}$$

When the emphasis is on the map $T : X \to X$, and we are studying different T-invariant measures, we will also say that μ is an ergodic measure for T. It is useful to have several different characterizations of ergodicity, and these are provided by the following proposition.

Proposition 2.14. *The following are equivalent properties for a measure-preserving transformation T of $(X, \mathscr{B}, \mu)$.*

(1) *T is ergodic.*
(2) *For any $B \in \mathscr{B}$, $\mu(T^{-1}B \triangle B) = 0$ implies that $\mu(B) = 0$ or $\mu(B) = 1$.*
(3) *For $A \in \mathscr{B}$, $\mu(A) > 0$ implies that $\mu\left(\bigcup_{n=1}^{\infty} T^{-n}A\right) = 1$.*
(4) *For $A, B \in \mathscr{B}$, $\mu(A)\mu(B) > 0$ implies that there exists $n \geqslant 1$ with*

$$\mu(T^{-n}A \cap B) > 0.$$

(5) *For $f : X \to \mathbb{C}$ measurable, $f \circ T = f$ almost everywhere implies that f is equal to a constant almost everywhere.*

In particular, for an ergodic transformation and countably many sets of positive measure, almost every point visits all of the sets infinitely often under iterations by the ergodic transformation.

* A set $B \in \mathscr{B}$ with $T^{-1}B = B$ is called *strictly invariant* under T.

Proof of Proposition 2.14. (1) $\implies$ (2): Assume that T is ergodic, so the implication (2.2) holds, and let B be an *almost invariant* measurable set – that is, a measurable set B with $\mu\left(T^{-1}B\triangle B\right)=0$. We wish to construct an invariant set from B, and this is achieved by means of the following limsup construction. Let

$$C=\bigcap_{N=0}^{\infty}\bigcup_{n=N}^{\infty}T^{-n}B.$$

For any $N\geqslant 0$,

$$B\triangle\bigcup_{n=N}^{\infty}T^{-n}B\subseteq\bigcup_{n=N}^{\infty}B\triangle T^{-n}B$$

and $\mu\left(B\triangle T^{-n}B\right)=0$ for all $n\geqslant 1$, since $B\triangle T^{-n}B$ is a subset of

$$\bigcup_{i=0}^{n-1}T^{-i}B\triangle T^{-(i+1)}B,$$

which has zero measure. Let $C_N=\bigcup_{n=N}^{\infty}T^{-n}B$; the sets C_N are nested,

$$C_0\supseteq C_1\supseteq\cdots,$$

and $\mu(C_N\triangle B)=0$ for each N. It follows that $\mu(C\triangle B)=0$, so

$$\mu(C)=\mu(B).$$

Moreover,

$$T^{-1}C=\bigcap_{N=0}^{\infty}\bigcup_{n=N}^{\infty}T^{-(n+1)}B=\bigcap_{N=0}^{\infty}\bigcup_{n=N+1}^{\infty}T^{-n}B=C.$$

Thus $T^{-1}C=C$, so by ergodicity $\mu(C)=0$ or 1, so $\mu(B)=0$ or 1.

(2) $\implies$ (3): Let A be a set with $\mu(A)>0$, and let $B=\bigcup_{n=1}^{\infty}T^{-n}A$. Then $T^{-1}B\subseteq B$; on the other hand $\mu\left(T^{-1}B\right)=\mu\left(B\right)$ so $\mu(T^{-1}B\triangle B)=0$. It follows that $\mu(B)=0$ or 1; since $T^{-1}A\subseteq B$ the former is impossible, so $\mu(B)=1$ as required.

(3) $\implies$ (4): Let A and B be sets of positive measure. By (3),

$$\mu\left(\bigcup_{n=1}^{\infty}T^{-n}A\right)=1,$$

so

$$0<\mu(B)=\mu\left(\bigcup_{n=1}^{\infty}B\cap T^{-n}A\right)\leqslant\sum_{n=1}^{\infty}\mu\left(B\cap T^{-n}A\right).$$

It follows that there must be some $n\geqslant 1$ with $\mu(B\cap T^{-n}A)>0$.

(4) $\implies$ (1): Let A be a set with $T^{-1}A = A$. Then

$$0 = \mu(A \cap X \smallsetminus A) = \mu(T^{-n}A \cap X \smallsetminus A)$$

for all $n \geqslant 1$ so, by (4), either $\mu(A) = 0$ or $\mu(X \smallsetminus A) = 0$.

(2) $\implies$ (5): We have seen that if (2) holds, then T is ergodic. Let f be a measurable complex-valued function on X, invariant under T in the stated sense. Since the real and the imaginary parts of f must also be invariant and measurable, we may assume without loss of generality that f is real-valued. Fix $k \in \mathbb{Z}$ and $n \geqslant 1$ and write

$$A_n^k = \{x \in X \mid f(x) \in [\tfrac{k}{n}, \tfrac{k+1}{n})\}.$$

Then $T^{-1}A_n^k \triangle A_n^k \subseteq \{x \in X \mid f \circ T(x) \neq f(x)\}$, a null set, so by (2)

$$\mu(A_n^k) \in \{0, 1\}.$$

For each n, X is the disjoint union $\bigsqcup_{k \in \mathbb{Z}} A_n^k$. It follows that there must be exactly one $k = k(n)$ with $\mu(A_n^{k(n)}) = 1$. Then f is constant on the set

$$Y = \bigcap_{n=1}^{\infty} A_n^{k(n)}$$

and $\mu(Y) = 1$, so f is constant almost everywhere.

(5) $\implies$ (2): If $\mu(T^{-1}B \triangle B) = 0$ then $f = \chi_B$ is a T-invariant measurable function, so by (5) χ_B is a constant almost everywhere. It follows that $\mu(B)$ is either 0 or 1. □

Proposition 2.15. *Bernoulli shifts are ergodic.*

Proof. Recall the measure-preserving transformation σ defined in Example 2.9 on the measure space $X = \{0, 1, \dots, n\}^{\mathbb{Z}}$ with the product measure μ. Let B denote a σ-invariant measurable set. Then given any $\varepsilon \in (0, 1)$ there is a finite union of cylinder sets A with $\mu(A \triangle B) < \varepsilon$, and hence with $|\mu(A) - \mu(B)| < \varepsilon$. This means A can be described as

$$A = \{x \in X \mid x|_{[-N,N]} \in F\}$$

for some N and some finite set $F \subseteq \{0, 1, \dots, n\}^{[-N,N]}$ (for brevity we write $[a, b]$ for the interval of integers $[a, b] \cap \mathbb{Z}$. It follows that for $M > 2N$,

$$\sigma^{-M}(A) = \{x \in X \mid x|_{[M-N,M+N]} \in F\},$$

where we think of $x|_{[M-N,M+N]}$ as a function on $[-N, N]$ in the natural way, is defined by conditions on a set of coordinates disjoint from $[-N, N]$, so

$$\mu(\sigma^{-M}A\smallsetminus A) = \mu(\sigma^{-M}A \cap X\smallsetminus A) = \mu(\sigma^{-M}A)\mu(X\smallsetminus A) = \mu(A)\mu(X\smallsetminus A). \tag{2.3}$$

Since B is σ-invariant, $\mu(B\triangle\sigma^{-1}B) = 0$. Now

$$\begin{aligned}\mu(\sigma^{-M}A\triangle B) &= \mu(\sigma^{-M}A\triangle\sigma^{-M}B)\\ &= \mu(A\triangle B) < \varepsilon,\end{aligned}$$

so $\mu(\sigma^{-M}A\triangle A) < 2\varepsilon$ and therefore

$$\mu(\sigma^{-M}A\triangle A) = \mu(A\smallsetminus\sigma^{-M}A) + \mu(\sigma^{-M}A\smallsetminus A) < 2\varepsilon. \tag{2.4}$$

Therefore, by equations (2.3) and (2.4),

$$\begin{aligned}\mu(B)\mu(X\smallsetminus B) &< (\mu(A)+\varepsilon)\,(\mu(X\smallsetminus A)+\varepsilon)\\ &= \mu(A)\mu(X\smallsetminus A) + \varepsilon\mu(A) + \varepsilon\mu(X\smallsetminus A) + \varepsilon^2\\ &< \mu(A)\mu(X\smallsetminus A) + 3\varepsilon < 5\varepsilon.\end{aligned}$$

Since ε was arbitrary, this implies that $\mu(B)\mu(X\smallsetminus B) = 0$, so $\mu(B) = 0$ or 1 as required. □

More general versions of this kind of approximation argument appear in Exercises 2.7.3 and 2.7.4.

Proposition 2.16. *The circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ is ergodic with respect to the Lebesgue measure $m_{\mathbb{T}}$ if and only if α is irrational.*

Proof. If $\alpha \in \mathbb{Q}$, then we may write $\alpha = \frac{p}{q}$ in lowest terms, so $R_\alpha^q = I_{\mathbb{T}}$ is the identity map. Pick any measurable set $A \subseteq \mathbb{T}$ with $0 < m_{\mathbb{T}}(A) < \frac{1}{q}$. Then

$$B = A \cup R_\alpha A \cup \cdots \cup R_\alpha^{q-1}A$$

is a measurable set invariant under R_α with $m_{\mathbb{T}}(B) \in (0,1)$, showing that R_α is not ergodic.

If $\alpha \notin \mathbb{Q}$ then for any $\varepsilon > 0$ there exist integers m, n, k with $m \neq n$ and $|m\alpha - n\alpha - k| < \varepsilon$. It follows that $\beta = (m-n)\alpha - k$ lies within ε of zero but is not zero, and so the set $\{0, \beta, 2\beta, \dots\}$ considered in $\mathbb{T}$ is ε-dense (that is, every point of $\mathbb{T}$ lies within ε of a point in this set). Thus $(\mathbb{Z}\alpha+\mathbb{Z})/\mathbb{Z} \subseteq \mathbb{T}$ is dense.

Now suppose that $B \subseteq \mathbb{T}$ is invariant under R_α. Then for any $\varepsilon > 0$ choose a function $f \in C(\mathbb{T})$ with $\|f - \chi_B\|_1 < \varepsilon$. By invariance of B we have

$$\|f \circ R_\alpha^n - f\|_1 < 2\varepsilon$$

for all n. Since f is continuous, it follows that

$$\|f \circ R_t - f\|_1 \leqslant 2\varepsilon$$

for all $t \in \mathbb{R}$. Thus, since $m_{\mathbb{T}}$ is rotation-invariant,

$$\left\| f - \int f(t)\,\mathrm{d}t \right\|_1 = \int \left| \int \left(f(x) - f(x+t)\right)\,\mathrm{d}t \right| \mathrm{d}x$$
$$\leqslant \iint \left| f(x) - f(x+t) \right| \mathrm{d}x\,\mathrm{d}t \leqslant 2\varepsilon$$

by Fubini's theorem (see Theorem A.13) and the triangle inequality for integrals. Therefore

$$\|\chi_B - \mu(B)\|_1 \leqslant \|\chi_B - f\|_1 + \left\| f - \int f(t)\,\mathrm{d}t \right\|_1 + \left\| \int f(t)\,\mathrm{d}t - \mu(B) \right\|_1 < 4\varepsilon.$$

Since this holds for every $\varepsilon > 0$ we deduce that χ_B is constant and therefore $\mu(B) \in \{0,1\}$. Thus for irrational α the transformation R_α is ergodic with respect to Lebesgue measure. □

Proposition 2.17. *The circle-doubling map $T_2 : \mathbb{T} \to \mathbb{T}$ from Example* 2.4 *is ergodic (with respect to Lebesgue measure).*

Proof. By Example 2.8, T_2 and the Bernoulli shift σ on $X = \{0,1\}^{\mathbb{N}}$ together with the fair coin-toss measure are measurably isomorphic. By Proposition 2.15 the latter is ergodic, and it is clear that measurably isomorphic systems are either both ergodic or both not ergodic. □

Ergodicity (indecomposability in the sense of measure theory) is a universal property of measure-preserving transformations in the sense that every measure-preserving transformation decomposes into ergodic components. This will be shown in Sections 4.2 and 6.1. In contrast the natural notion of indecomposability in topological dynamics – minimality – does not permit an analogous decomposition (see Exercise 4.2.3).

In Section 2.1 we pointed out that in order to check whether a map is measure-preserving it is enough to check this property on a family of sets that generates the σ-algebra. This is not the case when Definition 2.13 is used to establish ergodicity (see Exercise 2.3.2). Using a different characterization of ergodicity does allow this, as described in Exercise 2.7.3(3).

Exercises for Section 2.3

Exercise 2.3.1. Show that ergodicity is not preserved under direct products as follows. Find a pair of ergodic measure-preserving systems $(X, \mathscr{B}_X, \mu, T)$ and $(Y, \mathscr{B}_Y, \nu, S)$ for which $T \times S$ is not ergodic with respect to the product measure $\mu \times \nu$.

Exercise 2.3.2. Define a map $R : \mathbb{T}\times\mathbb{T} \to \mathbb{T}\times\mathbb{T}$ by $R(x, y) = (x+\alpha, y+\alpha)$ for an irrational α. Show that for any set of the form $A \times B$ with A, B measurable subsets of $\mathbb{T}$ (such a set is called a *measurable rectangle*) has the property of Definition 2.13, but the transformation R is not ergodic, even if α is irrational.

Exercise 2.3.3. (a) Find an arithmetic condition on α_1 and α_2 that is equivalent to the ergodicity of $R_{\alpha_1}\times R_{\alpha_2} : \mathbb{T}\times\mathbb{T} \to \mathbb{T}\times\mathbb{T}$ with respect to $m_{\mathbb{T}}\times m_{\mathbb{T}}$. (b) Generalize part (a) to characterize ergodicity of the rotation

$$R_{\alpha_1} \times \cdots \times R_{\alpha_n} : \mathbb{T}^n \to \mathbb{T}^n$$

with respect to $m_{\mathbb{T}^n}$.

Exercise 2.3.4. Prove that any factor of an ergodic measure-preserving system is ergodic.

Exercise 2.3.5. Extend Proposition 2.14 by showing that for each $p \in [1, \infty]$ a measure-preserving transformation T is ergodic if and only if for any L^p function f, $f \circ T = f$ almost everywhere implies that f is almost everywhere equal to a constant.

Exercise 2.3.6. Strengthen Proposition 2.14(5) by showing that a measure-preserving transformation T is ergodic if and only if any measurable function $f : X \to \mathbb{R}$ with $f(Tx) \geqslant f(x)$ almost everywhere is equal to a constant almost everywhere.

Exercise 2.3.7. Let X be a compact metric space and let $T : X \to X$ be continuous. Suppose that μ is a T-invariant ergodic probability measure defined on the Borel subsets of X. Prove that for μ-almost every $x \in X$ and every y in the support of μ there exists a sequence $n_k \nearrow \infty$ such that $T^{n_k}(x) \to y$ as $k \to \infty$. Here the support $\mathrm{Supp}(\mu)$ of μ is the smallest closed subset A of X with $\mu(A) = 1$; alternatively

$$\mathrm{Supp}(\mu) = X \smallsetminus \bigcup_{\substack{O\subseteq X \text{ open}, \\ \mu(O)=0}} O.$$

Notice that X has a countable base for its topology, so the union is still a μ-null set (see p. 406).

2.4 Associated Unitary Operators

A different kind of action[(16)] induced by a measure-preserving map T on a function space is the *associated operator* $U_T : L^2_\mu \to L^2_\mu$ defined by

$$U_T(f) = f \circ T.$$

Recall that L^2_μ is a Hilbert space, and for any functions $f_1, f_2 \in L^2_\mu$,

$$\begin{aligned}\langle U_T f_1, U_T f_2 \rangle &= \int f_1 \circ T \cdot \overline{f_2 \circ T} \,\mathrm{d}\mu \\ &= \int f_1 \overline{f_2} \,\mathrm{d}\mu \qquad \text{(since } \mu \text{ is } T\text{-invariant)} \\ &= \langle f_1, f_2 \rangle .\end{aligned}$$

Here it is natural to think of functions as being complex-valued; it will be clear from the context when members of L^2_μ are allowed to be complex-valued. Thus U_T is an isometry mapping L^2_μ into L^2_μ whenever $(X, \mathscr{B}_X, \mu, T)$ is a measure-preserving system.

If $U : \mathscr{H}_1 \to \mathscr{H}_2$ is a continuous linear operator from one Hilbert space to another then the relation

$$\langle Uf, g \rangle = \langle f, U^* g \rangle$$

defines an associated operator $U^* : \mathscr{H}_2 \to \mathscr{H}_1$ called the *adjoint* of U. The operator U is an *isometry* (that is, has $\|Uh\|_{\mathscr{H}_2} = \|h\|_{\mathscr{H}_1}$ for all $h \in \mathscr{H}_1$) if and only if

$$U^* U = I_{\mathscr{H}_1} \tag{2.5}$$

is the identity operator on $\mathscr{H}_1$ and

$$UU^* = P_{\operatorname{Im} U} \tag{2.6}$$

is the projection operator onto $\operatorname{Im} U$. Finally, an invertible linear operator U is called *unitary* if $U^{-1} = U^*$, or equivalently if U is invertible and

$$\langle Uh_1, Uh_2 \rangle = \langle h_1, h_2 \rangle \tag{2.7}$$

for all $h_1, h_2 \in \mathscr{H}_1$. If $U : \mathscr{H}_1 \to \mathscr{H}_2$ satisfies equation (2.7) then U is an isometry (even if it is not invertible). Thus for any measure-preserving transformation T, the associated operator U_T is an isometry, and if T is invertible then the associated operator U_T is a unitary operator, called the *associated unitary operator* of T or *Koopman operator* of T.

A property of a measure-preserving transformation is said to be a *spectral* or *unitary property* if it can be detected by studying the associated operator on L^2_μ.

Lemma 2.18. *A measure-preserving transformation T is ergodic if and only if* 1 *is a simple eigenvalue of the associated operator U_T. Hence ergodicity is a unitary property.*

Proof. This follows from the proof of the equivalence of (2) and (5) in Proposition 2.14 or via Exercise 2.3.5 applied with $p = 2$: an eigenfunction for the eigenvalue 1 is a T-invariant function, and ergodicity is characterized by the property that the only T-invariant functions are the constants. □

An isometry $U : \mathscr{H}_1 \to \mathscr{H}_2$ between Hilbert spaces[(17)] sends the expansion of an element

$$x = \sum_{n=1}^{\infty} c_n e_n$$

in terms of a complete orthonormal basis $\{e_n\}$ for $\mathscr{H}_1$ to a convergent expansion

$$U(x) = \sum_{n=1}^{\infty} c_n U(e_n)$$

in terms of the orthonormal set $\{U(e_n)\}$ in $\mathscr{H}_2$.

We will use this observation to study ergodicity of some of the examples using harmonic analysis rather than the geometrical arguments used earlier in this chapter.

Proof of Proposition 2.16 by Fourier analysis. Assume that α is irrational and let $f \in L^2(\mathbb{T})$ be a function invariant under R_α. Then f has a Fourier expansion $f(t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi i n t}$ (both equality and convergence are meant in $L^2(\mathbb{T})$). Now f is invariant, so $\|f \circ R_\alpha - f\|_2 = 0$. By uniqueness of Fourier coefficients, this requires that $c_n = c_n \mathrm{e}^{2\pi i n\alpha}$ for all $n \in \mathbb{Z}$. Since α is irrational, $\mathrm{e}^{2\pi i n\alpha}$ is only equal to 1 when $n = 0$, so this equation forces c_n to be 0 except when $n = 0$. Thus f is a constant almost everywhere, and hence R_α is ergodic.

If $\alpha \in \mathbb{Q}$ then write $\alpha = \frac{p}{q}$ in lowest terms. The function $g(t) = \mathrm{e}^{2\pi \mathrm{i} q t}$ is invariant under R_α but is not equal almost everywhere to a constant. □

Similar methods characterize ergodicity for endomorphisms.

Proof of Proposition 2.17 by Fourier analysis. Let $f \in L^2(\mathbb{T})$ be a function with $f \circ T_2 = f$ (equalities again are meant as elements of $L^2(\mathbb{T})$). Then f has a Fourier expansion $f(t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi i n t}$ with

$$\sum_{n\in\mathbb{Z}} |c_n|^2 = \|f\|_2^2 < \infty. \tag{2.8}$$

By invariance under T_2,

$$f(T_2 t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} 2 n t} = f(t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} n t},$$

so by uniqueness of Fourier coefficients we must have $c_{2n} = c_n$ for all $n \in \mathbb{Z}$. If there is some $n \neq 0$ with $c_n \neq 0$ then this contradicts equation (2.8), so we deduce that $c_n = 0$ for all $n \neq 0$. It follows that f is constant a.e., so T_2 is ergodic. □

The same argument gives the general abelian case, where Fourier analysis is replaced by character theory (see Section C.3 for the background). Notice that for a character $\chi : X \to \mathbb{S}^1$ on a compact abelian group and a continuous homomorphism $T : X \to X$, the map $\chi \circ T : X \to \mathbb{S}^1$ is also a character on X.

Theorem 2.19. *Let $T : X \to X$ be a continuous surjective homomorphism of a compact abelian group X. Then T is ergodic with respect to the Haar measure m_X if and only if the identity $\chi(T^n x) = \chi(x)$ for some $n > 0$ and character $\chi \in \widehat{X}$ implies that χ is the trivial character with $\chi(x) = 1$ for all $x \in X$.*

Proof. First assume that there is a non-trivial character χ with

$$\chi(T^n x) = \chi(x)$$

for some $n > 0$, chosen to be minimal with this property. Then the function

$$f(x) = \chi(x) + \chi(Tx) + \cdots + \chi(T^{n-1}x)$$

is invariant under T, and is non-constant since it is a sum of non-trivial distinct characters. It follows that T is not ergodic.

Conversely, assume that no non-trivial character is invariant under a non-zero power of T, and let $f \in L^2_{m_X}(X)$ be a function invariant under T . Then f has a Fourier expansion in $L^2_{m_X}$,

$$f = \sum_{\chi \in \widehat{X}} c_\chi \chi,$$

with $\sum_\chi |c_\chi|^2 = \|f\|_2^2 < \infty$. Since f is invariant, $c_\chi = c_{\chi \circ T} = c_{\chi \circ T^2} = \cdots$, so either $c_\chi = 0$ or there are only finitely many distinct characters among the $\chi \circ T^i$ (for otherwise $\sum_\chi |c_\chi|^2$ would be infinite). It follows that there are integers $p > q$ with $\chi \circ T^p = \chi \circ T^q$, which means that χ is invariant under T^{p-q} (the map $\chi \mapsto \chi \circ T$ from $\widehat{X}$ to $\widehat{X}$ is injective since T is surjective), so χ is trivial by hypothesis. It follows that the Fourier expansion of f is a constant, so T is ergodic. □

In particular, Theorem 2.19 may be applied to characterize ergodicity for endomorphisms of the torus.

Corollary 2.20. *Let $A \in \mathrm{Mat}_{dd}(\mathbb{Z})$ be an integer matrix with $\det(A) \neq 0$. Then A induces a surjective endomorphism T_A of $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ which preserves the Lebesgue measure $m_{\mathbb{T}^d}$. The transformation T_A is ergodic if and only if no eigenvalue of A is a root of unity.*

While harmonic analysis sometimes provides a short and readily understood proof of ergodic or mixing properties, these methods are in general less amenable to generalization than are the more geometric arguments.

Exercises for Section 2.4

Exercise 2.4.1. Give a different proof that the circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ is ergodic if α is irrational, using Lebesgue's density theorem (Theorem A.24) as follows. Suppose if possible that A and B are measurable invariant sets with $0 < m_{\mathbb{T}}(A), m_{\mathbb{T}}(B) < 1$ and $A \cap B = \varnothing$, and use the fact that the orbit of a point of density for A is dense to show that $A \cap B$ must be non-empty.

Exercise 2.4.2. Prove that an ergodic toral automorphism is not measurably isomorphic to an ergodic circle rotation.

Exercise 2.4.3. Extend Proposition 2.16 as follows. If X is a compact abelian group, prove that the group rotation $R_g(x) = gx$ is ergodic with respect to Haar measure if and only if the subgroup $\{g^n \mid n \in \mathbb{Z}\}$ generated by g is dense in X.

Exercise 2.4.4. In the notation of Corollary 2.20, prove that A is injective if and only if $|\det(A)| = 1$, and in general that $A : \mathbb{T}^d \to \mathbb{T}^d$ is $|\det(A)|$-to-one if $\det(A) \neq 0$. Prove Corollary 2.20 using Theorem 2.19 and the explicit description of characters on the torus from equation (C.3) on p. 436.

2.5 The Mean Ergodic Theorem

Ergodic theorems at their simplest express a relationship between averages taken along the orbit of a point under iteration of a measure-preserving map (in the physical origins of the subject, this represents an average over *time*) and averages taken over the measure space with respect to some invariant measure (an average over *space*). The averages taken are of *observables* in the physical sense, represented in our setting by measurable functions. Much of this way of viewing dynamical systems goes back to the seminal work of von Neumann [268].

We have already seen that ergodicity is a spectral property; the first and simplest ergodic theorem only uses properties of the operator U_T associated to a measure-preserving transformation T. Theorem 2.21 is due to von Neumann [267] and predates[(18)] the pointwise ergodic theorem (Theorem 2.30) of Birkhoff, despite the dates of the published versions.

Write $\underset{L^p_\mu}{\longrightarrow}$ for convergence in the L^p_μ norm.

Theorem 2.21 (Mean Ergodic Theorem). *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system, and let P_T denote the orthogonal projection onto the closed subspace*

$$I = \{g \in L^2_\mu \mid U_T g = g\} \subseteq L^2_\mu.$$

Then for any $f \in L^2_\mu$,

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f \underset{L^2_\mu}{\longrightarrow} P_T f.$$

PROOF. Let $B = \{U_T g - g \mid g \in L^2_\mu\}$. We claim that $B^\perp = I$. If

$$U_T f = f,$$

then

$$\langle f, U_T g - g\rangle = \langle U_T f, U_T g\rangle - \langle f, g\rangle = 0,$$

so $f \in B^\perp$. If

$$f \in B^\perp$$

then

$$\langle U_T g, f\rangle = \langle g, f\rangle$$

for all $g \in L^2_\mu$, so

$$U_T^* f = f. \tag{2.9}$$

Thus

$$\begin{aligned}
\|U_T f - f\|_2 &= \langle U_T f - f, U_T f - f\rangle \\
&= \|U_T f\|_2^2 - \langle f, U_T f\rangle - \langle U_T f, f\rangle + \|f\|_2^2 \\
&= 2\|f\|_2^2 - \langle U_T^* f, f\rangle - \langle f, U_T^* f\rangle \\
&= 0 \quad \text{by equation (2.9)},
\end{aligned}$$

so $f = U_T f$.

It follows that $L^2_\mu = I \oplus \overline{B}$, so any $f \in L^2_\mu$ decomposes as

$$f = P_T f + h, \tag{2.10}$$

with $h \in \overline{B}$. We claim that

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n h \underset{L^2_\mu}{\longrightarrow} 0.$$

This is clear for $h = U_T g - g \in B$, since

$$\begin{aligned}
\left\| \frac{1}{N}\sum_{n=0}^{N-1} U_T^n (U_T g - g) \right\|_2 &= \left\| \frac{1}{N}\big((U_T g - g) + (U_T^2 g - U_T g) + \cdots \right. \\
&\qquad \left. + (U_T^N g - U_T^{N-1} g)\big) \right\|_2 \\
&= \frac{1}{N}\left\|U_T^N g - g\right\|_2 \longrightarrow 0
\end{aligned} \tag{2.11}$$

as $N \to \infty$. All we know is that $h \in \overline{B}$, so let (g_i) be a sequence in L^2_μ with the property that $h_i = U_T g_i - g_i \to h$ as $i \to \infty$. Then for any $i \geqslant 1$,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 \leqslant \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (h - h_i) \right\|_2 + \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \right\|_2 . \tag{2.12}$$

Fix $\varepsilon > 0$ and choose, by the convergence (2.11), quantities i and N so large that

$$\|h - h_i\|_2 < \varepsilon$$

and

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \right\|_2 < \varepsilon.$$

Using these estimates in the inequality (2.12) gives

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 \leqslant 2\varepsilon$$

so

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \underset{L^2_\mu}{\longrightarrow} 0$$

as $N \to \infty$, for any $h \in \overline{B}$. The theorem follows by equation (2.10). $\square$

The quantity studied in Theorem 2.21 is an *ergodic average*, and it will be convenient to fix some notation for these. For a fixed measure-preserving system $(X, \mathscr{B}, \mu, T)$ and a function $f : X \to \mathbb{C}$ the Nth ergodic average of f is defined to be

$$\mathsf{A}_N = \mathsf{A}_N^f = \mathsf{A}_N(f) = \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n.$$

It is important to understand that this will be interpreted in several quite different ways.

- In Theorem 2.21 the function f is an element of the Hilbert space L^2_μ (that is, an equivalence class of measurable functions) and A_N^f is thought of as an element of L^2_μ.
- In Corollary 2.22 we will want to think of f as an element of L^1_μ, but evaluate the ergodic average A_N^f at points, sometimes writing $\mathsf{A}_N^f(x)$. Of course in this setting any statement can only be made almost everywhere with respect to μ, since f (and hence A_N^f) is only an equivalence class of functions, with two point functions identified if they agree almost everywhere.

- At times it will be useful to think of f as an element of $\mathscr{L}^p_\mu$ (that is, as a function rather than an equivalence class of functions) in which case A^f_N is defined everywhere. Also, if f is continuous, we will later ask whether the convergence of $\mathsf{A}^f_N(x)$ could be uniform across $x \in X$.

Corollary 2.22. (19) *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. Then for any function $f \in L^1_\mu$ the ergodic averages A^f_N converge in L^1_μ to a T-invariant function $f' \in L^1_\mu$.*

Proof. By the mean ergodic theorem (Theorem 2.21) we know that for any $g \in L^\infty_\mu \subseteq L^2_\mu$, the ergodic averages A^g_N converge in L^2_μ to some $g' \in L^2_\mu$. We claim that $g' \in L^\infty_\mu$. Indeed, $\|\mathsf{A}^g_N\|_\infty \leqslant \|g\|_\infty$ and so

$$|\langle \mathsf{A}^g_N, \chi_B \rangle| \leqslant \|g\|_\infty \mu(B)$$

for any $B \in \mathscr{B}$. Since $\mathsf{A}^g_N \to g'$ in L^2_μ, this implies that

$$|\langle g', \chi_B \rangle| \leqslant \|g\|_\infty \mu(B)$$

for $B \in \mathscr{B}$, so $\|g'\|_\infty \leqslant \|g\|_\infty$ as required.

Moreover, $\|\cdot\|_1 \leqslant \|\cdot\|_2$, so we deduce that

$$\mathsf{A}^g_N \underset{L^1_\mu}{\longrightarrow} g' \in L^\infty_\mu.$$

Thus the corollary holds for the dense set of functions $L^\infty_\mu \subseteq L^1_\mu$.

Let $f \in L^1_\mu$ and fix $\varepsilon > 0$; choose $g \in L^\infty_\mu$ with $\|f - g\|_1 < \varepsilon$. By averaging,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n \right\|_1 < \varepsilon,$$

and by the previous paragraph there exists g' and N_0 with

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n - g' \right\|_1 < \varepsilon$$

for $N \geqslant N_0$. Combining these gives

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N'} \sum_{n=0}^{N'-1} f \circ T^n \right\|_1 < 4\varepsilon$$

whenever $N, N' \geqslant N_0$. In other words, the ergodic averages form a Cauchy sequence in L^1_μ, and so they have a limit $f' \in L^1_\mu$ by the Riesz–Fischer theorem (Theorem A.23). Since

$$\left\| \left(\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \right) \circ T - \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \right\|_1 < \frac{2}{N} \|f\|_1$$

for all $N \geqslant 1$, the limit function f' must be T-invariant. □

Exercises for Section 2.5

Exercise 2.5.1. Show that a measure-preserving system $(X, \mathscr{B}, \mu, T)$ is ergodic if and only if, for any $f, g \in L^2_\mu$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle U_T^n f, g \rangle = \langle f, 1 \rangle \cdot \langle 1, g \rangle .$$

Exercise 2.5.2. Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. For any function f in L^p_μ, $1 \leqslant p < \infty$, prove that

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \underset{L^p_\mu}{\longrightarrow} f^*,$$

with $f^* \in L^p_\mu$ a T-invariant function.

Exercise 2.5.3. Show that a measure-preserving system $(X, \mathscr{B}, \mu, T)$ is ergodic if and only if $\mathsf{A}_N(f) \to \int f \, \mathrm{d}\mu$ as $N \to \infty$ for all f in a dense subset of L^1_μ.

Exercise 2.5.4. Extend Theorem 2.21 to a uniform mean ergodic theorem as follows. Under the assumptions and with the notation of Theorem 2.21, show that

$$\lim_{N - M \to \infty} \frac{1}{N - M} \sum_{n=M}^{N-1} U_T^n f \to P_T f.$$

Exercise 2.5.5. Apply Exercise 2.5.4 to strengthen Poincaré recurrence (Theorem 2.11) as follows. For any set B of positive measure in a measure-preserving system $(X, \mathscr{B}, \mu, T)$,

$$E = \{n \in \mathbb{N} \mid \mu(B \cap T^{-n} B) > 0\}$$

is syndetic: that is, there are finitely many integers $k_1, \dots, k_s$ with the property that $\mathbb{N} \subseteq \bigcup_{i=1}^{s} E - k_i$.

Exercise 2.5.6. Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. We say that T is *totally ergodic* if T^n is ergodic for all $n \geqslant 1$. Given $K \geqslant 1$ define a space $X^{(K)} = X \times \{1, \dots, K\}$ with measure $\mu^{(K)} = \mu \times \nu$ defined on the product σ-algebra $\mathscr{B}^{(K)}$, where $\nu(A) = \frac{1}{K}|A|$ is the normalized counting measure defined on any subset $A \subseteq \{1, \dots, K\}$, and a $\mu^{(K)}$-preserving transformation $T^{(K)}$ by

$$T^{(K)}(x,i) = \begin{cases} (x, i+1) & \text{if } 1 \leqslant i < K, \\ (Tx, 1) & \text{if } i = K \end{cases}$$

for all $x \in X$. Show that $T^{(K)}$ is ergodic with respect to $\mu^{(K)}$ if and only if T is ergodic with respect to μ, and that $T^{(K)}$ is not totally ergodic if $K > 1$.

2.6 Pointwise Ergodic Theorem

The conventional proof of the pointwise ergodic theorem involves two other important results, the maximal inequality and the maximal ergodic theorem. Roughly speaking, the maximal ergodic theorem may be used to show that the set of functions in L^1_μ for which the pointwise ergodic theorem holds is *closed* as a subset of L^1_μ; one then has to find a *dense* subset of L^1_μ for which the pointwise ergodic theorem holds. Examples 2.23 and 2.25 give another motivation for the maximal ergodic theorem.

Since the pointwise ergodic theorem involves evaluating a function along the orbit of individual points, it is most naturally phrased in terms of genuine functions (that is, elements of $\mathscr{L}^1_\mu$; see Section A.3 for the notation). We will normally apply it to a function in L^1_μ, where the meaning is that for any representative in $\mathscr{L}^1_\mu$ of the equivalence class in L^1_μ we have convergence almost everywhere.

2.6.1 The Maximal Ergodic Theorem

In order to see where the next result comes from, it is useful to ask how likely is it that the orbit of a point spends unexpectedly much time in a given small set (the ergodic theorem says that the orbit of a point spends a predictable amount of time in a given set).

Example 2.23. Let $(X, \mathscr{B}_X, \mu, T)$ be a measure-preserving system, and fix a small measurable set $B \in \mathscr{B}_X$ with $\mu(B) = \varepsilon > 0$. Consider the ergodic average

$$\mathsf{A}_N^{\chi_B} = \frac{1}{N} \sum_{n=0}^{N-1} \chi_B \circ T^n.$$

Since T preserves μ, $\int_X \chi_B \circ T^n \,\mathrm{d}\mu = \mu(B)$ for any $n \geqslant 0$, so

$$\int_X \mathsf{A}_N^{\chi_B}\,\mathrm{d}\mu = \int_X \chi_B\,\mathrm{d}\mu = \mu(B) = \varepsilon.$$

Now ask how likely is it that the orbit of a point x spends more than $\sqrt{\varepsilon} > \varepsilon$ of the time between 0 and $N-1$ in the set B. Notice that

$$\sqrt{\varepsilon}\mu\left(\{x \mid \mathsf{A}_N^{\chi_B}(x) > \sqrt{\varepsilon}\}\right) \leqslant \int_X \mathsf{A}_N^{\chi_B}\,\mathrm{d}\mu = \varepsilon,$$

since

$$\sqrt{\varepsilon}\chi_{\{y \mid \mathsf{A}_N^{\chi_B}(y) > \sqrt{\varepsilon}\}}(x) \leqslant \mathsf{A}_N^{\chi_B}(x)$$

for all $x \in X$. Thus on the fixed time scale $[0, N-1]$ the measure of the set B_ε^N of points that spend in proportion at least $\sqrt{\varepsilon}$ of the time between 0 and $N-1$ in the set B is no larger than $\sqrt{\varepsilon}$.

We would like to be able to say that one can find a set B_ε independent of N with similar properties for all N; as discussed below, this is a consequence of the maximal ergodic theorem[(20)].

Theorem 2.24 (Maximal Ergodic Theorem). *Consider the measure-preserving system $(X, \mathscr{B}, \mu, T)$ on a probability space and g a real-valued function in $\mathscr{L}_\mu^1$. Define*

$$E_\alpha = \left\{x \in X \,\Big|\, \sup_{n \geqslant 1} \frac{1}{n}\sum_{i=0}^{n-1} g(T^i x) > \alpha\right\}$$

for any $\alpha \in \mathbb{R}$. Then

$$\alpha\mu\left(E_\alpha\right) \leqslant \int_{E_\alpha} g\,\mathrm{d}\mu \leqslant \|g\|_1.$$

Moreover, $\alpha\mu\left(E_\alpha \cap A\right) \leqslant \int_{E_\alpha \cap A} g\,\mathrm{d}\mu$ whenever $T^{-1}A = A$.

Example 2.25. We continue the discussion from Example 2.23 by noting that if $B \subseteq X$ has $\mu(B) = \varepsilon > 0$ and $g = \chi_B$ is its characteristic function, then by applying the maximal ergodic theorem (Theorem 2.24) with $\alpha = \sqrt{\varepsilon}$ we get the following statement: There exists a set $B' \subseteq X$ with $\mu(B') \leqslant \sqrt{\varepsilon}$ such that for all $N \geqslant 1$ and all $x \in X \smallsetminus B'$ the orbit of the point x spends at most $\sqrt{\varepsilon}$ in proportion of the times between 0 and $N-1$ in the set B. Thus we have found a set as in Example 2.23, but independently of N.

2.6.2 Maximal Ergodic Theorem via Maximal Inequality

Notice that the operator U_T associated to a measure-preserving transformation T is a *positive* linear operator on each L_μ^p space (positive means that $f \geqslant 0$ implies $U_T f \geqslant 0$). A traditional proof of Theorem 2.24 starts with a maximal inequality for positive operators.

Proposition 2.26 (Maximal Inequality). *Let $U : L^1_\mu \to L^1_\mu$ be a positive linear operator with $\|U\| \leqslant 1$. For $f \in L^1_\mu$ a real-valued function, define inductively the functions*

$$\begin{aligned} f_0 &= 0 \\ f_1 &= f \\ f_2 &= f + Uf \\ &\vdots \\ f_n &= f + Uf + \cdots + U^{n-1} f \end{aligned}$$

for $n \geqslant 1$, and $F_N = \max\{f_n \mid 0 \leqslant n \leqslant N\}$ (all these functions are defined pointwise). Then

$$\int_{\{x \mid F_N(x) > 0\}} f \, \mathrm{d}\mu \geqslant 0$$

for all $N \geqslant 1$.

Proof. For each N, it is clear that $F_N \in L^1_\mu$. Since U is positive and linear, and since

$$F_N \geqslant f_n$$

for $0 \leqslant n \leqslant N$, we have

$$UF_N + f \geqslant Uf_n + f = f_{n+1}.$$

Hence

$$UF_N + f \geqslant \max_{1 \leqslant n \leqslant N} f_n.$$

For $x \in P = \{x \mid F_N(x) > 0\}$ we have

$$F_N(x) = \max_{0 \leqslant n \leqslant N} f_n(x) = \max_{1 \leqslant n \leqslant N} f_n(x)$$

since $f_0 = 0$. Therefore,

$$UF_N(x) + f(x) \geqslant F_N(x)$$

for $x \in P$, and so

$$f(x) \geqslant F_N(x) - UF_N(x) \tag{2.13}$$

for $x \in P$. Now $F_N(x) \geqslant 0$ for all x, so $UF_N(x) \geqslant 0$ for all x. Hence the inequality (2.13) implies that

$$\begin{aligned}\int_P f\,\mathrm{d}\mu &\geqslant \int_P F_N\,\mathrm{d}\mu - \int_P UF_N\,\mathrm{d}\mu\\ &= \int_X F_N\,\mathrm{d}\mu - \int_P UF_N\,\mathrm{d}\mu \qquad \text{(since } F_N(x)=0 \text{ for } x\notin P)\\ &\geqslant \int_X F_N\,\mathrm{d}\mu - \int_X UF_N\,\mathrm{d}\mu\\ &= \|F_N\|_1 - \|UF_N\|_1 \geqslant 0,\end{aligned}$$

since $\|U\|\leqslant 1$. □

First Proof of Theorem 2.24. Let $f=(g-\alpha)$ and $Uf=f\circ T$ for $f\in\mathscr{L}^1_\mu$ so that, in the notation of Proposition 2.26,

$$E_\alpha=\bigcup_{N=0}^{\infty}\{x\mid F_N(x)>0\}.$$

It follows that $\int_{E_\alpha} f\,\mathrm{d}\mu\geqslant 0$ and therefore $\int_{E_\alpha} g\,\mathrm{d}\mu\geqslant\alpha\mu(E_\alpha)$. For the last statement, apply the same argument to $f=(g-\alpha)$ on the measure-preserving system $(A,\mathscr{B}\big|_A,\frac{1}{\mu(A)}\mu\big|_A,T\big|_A)$. □

2.6.3 Maximal Ergodic Theorem via a Covering Lemma

In this subsection we use covering properties of intervals in $\mathbb{Z}$ to establish a version of the maximal ergodic theorem (Theorem 2.24). This demonstrates very clearly the strong link between the Lebesgue density theorem (Theorem A.24), whose proof involves the Hardy–Littlewood maximal inequality, and the pointwise ergodic theorem, whose proof involves the maximal ergodic theorem*. The material in this section illustrates some of the ideas used in the more extensive results of Bourgain [41]; a little of the history will be given in the note (83) on p. 275

We will obtain a formally weaker version of Theorem 2.24, by showing that

$$\alpha\mu(E_\alpha)\leqslant 3\|g\|_1 \tag{2.14}$$

in the notation of Theorem 2.24. This is sufficient for all our purposes. For future applications, we state the covering lemma(21) needed in a more general setting.

Lemma 2.27 (Finite Vitali covering lemma). *Let $B_{r_1}(a_1),\dots,B_{r_K}(a_K)$ be any collection of balls in a metric space. Then there exists a subcollec-*

* Additionally, this approach starts to reveal more about what properties of the acting group might be useful for obtaining more general ergodic theorems, and gives a method capable of generalization to ergodic averaging along other sets of integers.

tion $B_{r_{j(1)}}(a_{j(1)}),\dots,B_{r_{j(k)}}(a_{j(k)})$ *of those balls which are disjoint and satisfy*

$$B_{r_1}(a_1)\cup\dots\cup B_{r_K}(a_K)\subseteq B_{3r_{j(1)}}(a_{j(1)})\cup\dots\cup B_{3r_{j(k)}}(a_{j(k)}),$$

where in the right-hand side we have tripled the radii of the balls in the subcollection.

Proof. By reordering the balls if necessary, we may assume that

$$r_1\geqslant r_2\geqslant\dots\geqslant r_K.$$

Let $j(1)=1$. We choose the remaining disjoint balls by induction as follows. Assume that we have chosen $j(1),\dots,j(n)$ from the indices $\{1,\dots,\ell\}$, discarding those not chosen. If $B_{r_{\ell+1}}(a_{\ell+1})$ is disjoint from

$$B_{r_{j(1)}}(a_{j(1)})\cup\dots\cup B_{r_{j(n)}}(a_{j(n)})$$

we choose $j(n+1)=\ell+1$, and if not we discard $\ell+1$, and proceed with studying $\ell+2$, stopping if $\ell+1=K$. Suppose that $B_{r_{j(1)}}(a_{j(1)}),\dots,B_{r_{j(k)}}(a_{j(k)})$ are the balls chosen from all the balls considered, and let

$$V=B_{3r_{j(1)}}(a_{j(1)})\cup\dots\cup B_{3r_{j(k)}}(a_{j(k)}).$$

If $i\in\{j(1),\dots,j(k)\}$ then $B_{r_i}(a_i)\subseteq B_{3r_i}(a_i)\subseteq V$ by construction. If not, then by the construction there is some $n\in\{1,\dots,i-1\}\cap\{j(1),\dots,j(k)\}$ that was selected, such that

$$B_{r_i}(a_i)\cap B_{r_n}(a_n)\neq\varnothing,$$

and $r_n\geqslant r_i$ by the ordering of the indices. By the triangle inequality we therefore have

$$B_{r_i}(a_i)\subseteq B_{3r_n}(a_n)\subseteq V$$

as required. □

In the integers, the Vitali covering lemma may be formulated as follows (see Exercise 2.6.2).

Corollary 2.28. *For any collection of intervals*

$$I_1=[a_1,a_1+\ell(1)-1],\dots,I_K=[a_K,a_K+\ell(K)-1]$$

in $\mathbb{Z}$ *there is a disjoint subcollection* $I_{j(1)},\dots,I_{j(k)}$ *such that*

$$I_1\cup\dots\cup I_K\subseteq\bigcup_{m=1}^{k}[a_{j(m)}-\ell_{j(m)},a_{j(m)}+2\ell_{j(m)}-1].$$

Proof of the inequality (2.14). Let $(X,\mathscr{B},\mu,T)$ be a measure-preserving system, with $g\in\mathscr{L}^1_\mu$, and fix $\alpha>0$. Define

$$g^*(x) = \sup_{n \geqslant 1} \frac{1}{n} \sum_{i=0}^{n-1} g\left(T^i(x)\right)$$

and $E_\alpha = \{x \in X \mid g^*(x) > \alpha\}$ as before. We will deduce the inequality (2.14) from a similar estimate for the function

$$\phi(j) = \begin{cases} g(T^j x) & \text{for } j = 0, \dots, J; \\ 0 & \text{for } j < 0 \text{ or } j > J \end{cases} \tag{2.15}$$

for a fixed $x \in X$ and $J \geqslant 1$.

Lemma 2.29. *For any $\phi \in \ell^1(\mathbb{Z})$ and $\alpha > 0$, define*

$$\phi^*(a) = \sup_{n \geqslant 1} \frac{1}{n} \sum_{i=0}^{n-1} \phi(a+i),$$

and

$$E_\alpha^\phi = \{a \in \mathbb{Z} \mid \phi^*(a) > \alpha\}.$$

Then $\alpha |E_\alpha^\phi| \leqslant 3\|\phi\|_1$.

Proof of Lemma 2.29. Let $a_1, \dots, a_K$ be different elements of E_α^ϕ, and let $\ell(j)$ for $j = 1, \dots, K$ be chosen so that

$$\frac{1}{\ell(j)} \sum_{i=0}^{\ell(j)-1} \phi(a_j + i) > \alpha. \tag{2.16}$$

Define the intervals $I_j = [a_j, a_j + \ell(j) - 1]$ for $1 \leqslant j \leqslant K$ and use Corollary 2.28 to construct the subcollection $I_{j(1)}, \dots, I_{j(k)}$ as in the corollary. Since the intervals $I_{j(1)}, \dots, I_{j(k)}$ are disjoint, it follows that

$$\sum_{i=1}^{k} \sum_{m \in I_{j(i)}} \phi(m) \leqslant \|\phi\|_1, \tag{2.17}$$

where the left-hand side equals

$$\sum_{i=1}^{k} \ell(j(i)) \frac{1}{\ell(j(i))} \sum_{n=0}^{\ell(j(i))-1} \phi(a_j + n) > \sum_{i=1}^{k} \ell(j(i))\alpha \tag{2.18}$$

by the choice in equation (2.16) of the $\ell(j(i))$. However, since

$$\{a_1, \dots, a_K\} \subseteq \bigcup_{j=1}^{k} [a_{j(i)} - \ell_{j(i)}, a_{j(i)} + 2\ell_{j(i)} - 1]$$

by Corollary 2.28, we therefore have

$$K \leqslant 3 \sum_{i=1}^{k} \ell_{j(i)}. \tag{2.19}$$

Combining the inequalities (2.19), (2.18), and (2.17) in that order gives

$$\alpha K \leqslant 3 \sum_{i=1}^{k} \ell_{j(i)} \alpha < 3\|\phi\|_1,$$

which proves the lemma. □

Fix now some $M \geqslant 1$ (the parameter J will later be chosen much larger than M) and define

$$g_M^*(x) = \sup_{1 \leqslant n \leqslant M} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x),$$

and

$$E_{\alpha,M}^g = \{x \in X \mid g_M^*(x) > \alpha\}.$$

Using ϕ as in equation (2.15) and, suppressing the dependence on x as before, we also define

$$\phi_M^*(a) = \sup_{1 \leqslant n \leqslant M} \frac{1}{n} \sum_{i=0}^{n-1} \phi(a+i).$$

As $\phi(a+i) = g(T^{a+i}x)$ if $0 \leqslant a < J-M$ and $0 \leqslant i < M$, we have

$$\phi_M^*(a) = g_M^*(T^a x) \tag{2.20}$$

for $0 \leqslant a < J-M$. Also, for any $x \in X$ and $\alpha > 0$ we have

$$\alpha \left|\{a \in [0, J-1] \mid \phi_M^*(a) > \alpha\}\right| \leqslant 3\|\phi\|_1$$

by Lemma 2.29. Recalling the definition of ϕ and E_α and using equation (2.20), this may be written in a slightly weaker form as

$$\begin{aligned} \alpha \sum_{a=0}^{J-M-1} \chi_{E_{\alpha,M}^g}(T^a x) &= \alpha \left| \left\{ a \in [0, J-M-1] \mid g_M^*(T^a x) > \alpha \right\} \right| \\ &\leqslant 3 \sum_{i=0}^{J} |g(T^i x)|, \end{aligned}$$

which may be integrated over $x \in X$ to obtain

$$(J-M)\alpha\mu\left(E_{\alpha,M}^g\right) \leqslant 3(J+1)\|g\|_1,$$

where we have used the invariance of μ under T. Dividing by J and letting $J \to \infty$ gives $\alpha\mu\left(E^g_{\alpha,M}\right) \leqslant 3\|g\|_1$, and finally letting $M \to \infty$ gives inequality (2.14). □

2.6.4 The Pointwise Ergodic Theorem

We are now ready to give a proof of Birkhoff's pointwise ergodic theorem [33] using the maximal ergodic theorem(22). This precisely describes the relationship sought between the space average of a function and the time average along the orbit of a typical point.

Theorem 2.30 (Birkhoff). *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. If $f \in \mathscr{L}^1_\mu$, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = f^*(x)$$

converges almost everywhere and in L^1_μ to a T-invariant function $f^* \in \mathscr{L}^1_\mu$, and

$$\int f^* \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

If T is ergodic, then

$$f^*(x) = \int f \,\mathrm{d}\mu$$

almost everywhere.

Example 2.31. (23) In Example 1.2 we explained that almost every real number has the property that any block of length k of digits base 10 appears with asymptotic frequency $\frac{1}{10^k}$, thus almost every number is *normal* base 10. We now have all the material needed to justify this result: By Corollary 2.20, the map $x \mapsto Kx \pmod 1$ on the circle for $K \geqslant 2$ is ergodic, so the pointwise ergodic theorem (Theorem 2.30) may be applied to show that almost every number is normal to each base $K \geqslant 2$, and so (by taking the union of countably many null sets) almost every number is normal in *every* base $K \geqslant 2$.

As with the maximal ergodic theorem (Theorem 2.24), we will give two proofs(24) of the pointwise ergodic theorem. The first is a traditional one while the second is closer to the approach of Bourgain [41] for example, and is better adapted to generalization both of the acting group and of the sequence along which ergodic averages are formed.

Theorem 2.30 will be formulated differently in Theorem 6.1, and will be used in Theorem 6.2 to construct the ergodic decomposition.

2.6.5 Two Proofs of the Pointwise Ergodic Theorem

First Proof of Theorem 2.30. Recall that $(X, \mathscr{B}, \mu, T)$ is a measure-preserving system, $\mu(X) = 1$, and $f \in \mathscr{L}^1_\mu$. It is sufficient to prove the result for a real-valued function f. Define, for any $x \in X$,

$$f^*(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x),$$

$$f_*(x) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

Then

$$\frac{n+1}{n}\left(\frac{1}{n+1}\sum_{i=0}^{n} f(T^i x)\right) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(Tx)) + \frac{1}{n} f(x). \tag{2.21}$$

By taking the limit along a subsequence for which the left-hand side of equation (2.21) converges to the limsup, this shows that $f^* \leqslant f^* \circ T$. A limit along a subsequence for which the right-hand side of equation (2.21) converges to the limsup shows that $f^* \geqslant f^* \circ T$. A similar argument for f_* shows that

$$f^* \circ T = f^*, \quad f_* \circ T = f_*. \tag{2.22}$$

Now fix rationals $\alpha > \beta$, and write

$$E_\alpha^\beta = \{x \in X \mid f_*(x) < \beta \text{ and } f^*(x) > \alpha\}.$$

By equation (2.22), $T^{-1}E_\alpha^\beta = E_\alpha^\beta$ and $E_\alpha \supseteq E_\alpha^\beta$ where E_α is the set defined in Theorem 2.24 (with $g = f$). By Theorem 2.24,

$$\int_{E_\alpha^\beta} f \,\mathrm{d}\mu \geqslant \alpha\mu\left(E_\alpha^\beta\right). \tag{2.23}$$

After replacing f by $-f$, a similar argument shows that

$$\int_{E_\alpha^\beta} f \,\mathrm{d}\mu \leqslant \beta\mu\left(E_\alpha^\beta\right). \tag{2.24}$$

Now

$$\{x \mid f_*(x) < f^*(x)\} = \bigcup_{\substack{\alpha,\beta\in\mathbb{Q},\\ \alpha>\beta}} E_\alpha^\beta,$$

while the inequalities (2.23) and (2.24) show that $\mu(E_\alpha^\beta) = 0$ for $\alpha > \beta$. It follows that

$$\mu\Big(\bigcup_{\substack{\alpha,\beta\in\mathbb{Q},\\ \alpha>\beta}} E_\alpha^\beta\Big)=0,$$

so

$$f_*(x)=f^*(x) \text{ a.e.}$$

Thus

$$g_n(x)=\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\longrightarrow f^*(x) \text{ a.e.} \tag{2.25}$$

By Corollary 2.22 we also know that

$$g_n \underset{L^1_\mu}{\longrightarrow} f' \in \mathscr{L}^1_\mu. \tag{2.26}$$

By Corollary A.12, this implies that there is a subsequence $n_k\to\infty$ with

$$g_{n_k}(x)\longrightarrow f'(x) \text{ a.e.} \tag{2.27}$$

Putting equations (2.25), (2.26) and (2.27) together we see that $f^*=f'\in\mathscr{L}^1_\mu$ and that the convergence in equation (2.25) also happens in L^1_μ. Finally we also get

$$\int f\,\mathrm{d}\mu=\int g_n\,\mathrm{d}\mu=\int f^*\,\mathrm{d}\mu.$$

$\square$

A somewhat different approach is to use the maximal ergodic theorem (Theorem 2.24) to control the gap between mean convergence and pointwise convergence almost everywhere.

Second Proof of Theorem 2.30. Assume first that $f_0\in\mathscr{L}^\infty$. By the mean ergodic theorem in L^1 (Corollary 2.22) we know that the ergodic averages

$$\mathsf{A}_N(f_0)=\frac{1}{N}\sum_{n=0}^{N-1} f_0\circ T^n\to F_0$$

converge in L^1_μ as $N\to\infty$ to some T-invariant function $F_0\in\mathscr{L}^1_\mu$. Given $\varepsilon>0$ choose some M such that

$$\|F_0-\mathsf{A}_M(f_0)\|_1<\varepsilon^2.$$

By the maximal ergodic theorem (Theorem 2.24) applied to the function

$$g(x)=F_0(x)-\mathsf{A}_M(f_0)$$

we see that

$$\varepsilon\mu\Big(\{x\in X\mid \sup_{N\geqslant 1}|\mathsf{A}_N(F_0-\mathsf{A}_M(f_0))|>\varepsilon\}\Big)<\varepsilon^2.$$

Clearly $\mathsf{A}_N(F_0) = F_0$ since the limit function F_0 is T-invariant, while if M is fixed and $N \to \infty$ we have (see Exercise 2.6.4)

$$\begin{aligned} \mathsf{A}_N\left(\mathsf{A}_M(f_0)\right) &= \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f_0 \circ T^{n+m} \\ &= \mathsf{A}_N(f_0) + \mathrm{O}_M\left(\frac{\|f_0\|_\infty}{N}\right). \end{aligned} \tag{2.28}$$

Putting these together, we see that

$$\begin{aligned} \mu\big(\{x \mid \limsup_{N\to\infty} |F_0 - \mathsf{A}_N(f_0)| > \varepsilon\}\big) &= \mu\big(\{x \mid \limsup_{N\to\infty} |F_0 - \mathsf{A}_N(\mathsf{A}_M(f_0))| > \varepsilon\}\big) \\ &\leqslant \mu\big(\{x \mid \sup_{N\geqslant 1} |\mathsf{A}_N\left(F_0 - \mathsf{A}_M(f_0)\right)| > \varepsilon\}\big) \\ &< \varepsilon, \end{aligned}$$

which shows that $\mathsf{A}_N(f_0) \to F_0$ almost everywhere.

To prove convergence for any $f \in \mathscr{L}^1_\mu$, fix $\varepsilon > 0$ and choose some $f_0 \in \mathscr{L}^\infty$ with $\|f - f_0\|_1 < \varepsilon^2$. Write $F \in \mathscr{L}^1_\mu$ for the L^1-limit of $\mathsf{A}_N(f)$ and $F_0 \in \mathscr{L}^1_\mu$ for the L^1-limit of $\mathsf{A}_N(f_0)$. Since $\|\mathsf{A}_N(f) - \mathsf{A}_N(f_0)\|_1 \leqslant \|f - f_0\|_1$ we deduce that $\|F - F_0\|_1 < \varepsilon^2$. From this we get

$$\begin{aligned} &\mu\big(\{x \mid \limsup_{N\to\infty} |F - \mathsf{A}_N(f)| > 2\varepsilon\}\big) \\ &\quad\leqslant \mu\big(\{x \mid |F - F_0| + \limsup_{N\to\infty} |F_0 - \mathsf{A}_N(f_0)| + \sup_{N\geqslant 1} |\mathsf{A}_N(f_0 - f)| > 2\varepsilon\}\big) \\ &\quad\leqslant \mu\big(\{x \mid |F - F_0| > \varepsilon\big) + \mu\big(\{x \mid \sup_{N\geqslant 1} |\mathsf{A}_N(f_0 - f)| > \varepsilon\big) \\ &\quad\leqslant \varepsilon^{-1}\|F - F_0\|_1 + \varepsilon^{-1}\|f_0 - f\|_1 \leqslant 2\varepsilon \end{aligned} \tag{2.29}$$

by the maximal ergodic theorem (Theorem 2.24), which shows that $\mathsf{A}_N(f)$ converges almost everywhere as $N \to \infty$. □

Exercises for Section 2.6

Exercise 2.6.1. Prove the following version of the ergodic theorem for finite permutations (see the book of Nadkarni [263] where this is used to motivate a different approach to ergodic theorems). Let $X = \{x_1, \dots, x_r\}$ be a finite set, and let $\sigma : X \to X$ be a permutation of X. The orbit of x_j under σ is the set $\{\sigma^n(x_j)\}_{n\geqslant 0}$, and σ is called cyclic if there is an orbit of cardinality r.

(1) For a cyclic permutation σ and any function $f : X \to \mathbb{R}$, prove that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}f(\sigma^j x)=\frac{1}{r}\left(f(x_1)+\cdots+f(x_r)\right).$$

(2) More generally, prove that for any permutation σ and function $f:X\to\mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}f(\sigma^j x)=\frac{1}{p_x}\big(f(x)+f(\sigma(x))+\cdots+f(\sigma^{p_x-1}(x))\big)$$

where the orbit of x has cardinality p_x under σ.

Exercise 2.6.2. Mimic the proof of Lemma 2.27 (or give the details of a deduction) to prove Corollary 2.28.

Exercise 2.6.3. Let $(X,\mathscr{B},\mu,T)$ be an invertible measure-preserving system. Prove that, for any $f\in L^1_\mu$,

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

almost everywhere.

Exercise 2.6.4. Fill in the details to prove the estimate in (2.28).

Exercise 2.6.5. Formulate and prove a pointwise ergodic theorem for a measurable function $f\geqslant 0$ with $\int f\,\mathrm{d}\mu=\infty$, under the assumption of ergodicity.

2.7 Strong-mixing and Weak-mixing

In this section we step back from thinking of measure-preserving transformations through the functional-analytic prism of their action on L^p spaces to the more fundamental questions discussed in Sections 2.2 and 2.3. Namely, if A is a measurable set, what can be said about how the set $T^{-n}A$ is spread around the whole measure space for large n?

An easy consequence of the mean ergodic theorem is that a measure-preserving system $(X,\mathscr{B},\mu,T)$ is ergodic if and only if

$$\frac{1}{N}\sum_{n=0}^{N-1}f\circ T^n\underset{L^2_\mu}{\longrightarrow}\int f\,\mathrm{d}\mu$$

as $N\to\infty$ for every $f\in L^2_\mu$. It follows that $(X,\mathscr{B},\mu,T)$ is ergodic if and only if

$$\frac{1}{N}\sum_{n=0}^{N-1}\langle f\circ T^n,g\rangle\longrightarrow\int f\,\mathrm{d}\mu\int g\,\mathrm{d}\mu \tag{2.30}$$

as $N \to \infty$ for any $f, g \in L^2_\mu$. The characterization in (2.30) can be cast in terms of the behavior of sets to show that $(X, \mathscr{B}, \mu, T)$ is ergodic if and only if

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A \cap T^{-n} B\right) \longrightarrow \mu(A)\mu(B) \tag{2.31}$$

as $N \to \infty$ for all $A, B \in \mathscr{B}$. One direction is clear: if T is ergodic, then the convergence (2.30) may be applied with $g = \chi_A$ and $f = \chi_B$.

Conversely, if $T^{-1}B = B$ then the convergence (2.31) with $A = X \smallsetminus B$ implies that $\mu(X \smallsetminus B)\mu(B) = 0$, so T is ergodic.

There are several ways in which the convergence (2.31) might take place. Recall that measurable sets in $(X, \mathscr{B}, \mu)$ may be thought of as events in the sense of probability, and events $A, B \in \mathscr{B}$ are called *independent* if

$$\mu(A \cap B) = \mu(A)\mu(B).$$

Clearly if the action of T contrives to make $T^{-n}B$ and A become *independent* in the sense of probability for all large n, then the convergence (2.31) is assured. It turns out that this is too much to ask (see Exercise 2.7.1), but asking for $T^{-n}B$ and A to become *asymptotically independent* leads to the following non-trivial definition.

Definition 2.32. A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is *mixing* if

$$\mu\left(A \cap T^{-n} B\right) \longrightarrow \mu(A)\mu(B)$$

as $n \to \infty$, for all $A, B \in \mathscr{B}$.

Mixing is also sometimes called *strong-mixing*, in contrast to weak-mixing and mild-mixing.

Example 2.33. A circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ is not mixing. There is a sequence $n_j \to \infty$ for which $n_j \alpha \pmod 1 \to 0$ (if α is rational we may choose to have $n_j \alpha \pmod 1 = 0$). If $A = B = [0, \frac{1}{2}]$ then $m_{\mathbb{T}}(A \cap R_\alpha^{n_j} A) \to \frac{1}{2}$, so R_α is not mixing.

It is clear that some measure preserving systems make many sets become asymptotically independent as they move apart in time (that is, under iteration), leading to the following natural definition due to Rokhlin [316].

Definition 2.34. A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is *k-fold mixing*, *mixing of order k* or *mixing on $k+1$ sets* if

$$\mu\left(A_0 \cap T^{-n_1} A_1 \cap \cdots \cap T^{-n_k} A_k\right) \longrightarrow \mu(A_0) \cdots \mu(A_k)$$

as

$$n_1, n_2 - n_1, n_3 - n_2, \ldots, n_k - n_{k-1} \longrightarrow \infty$$

for any sets $A_0, \ldots, A_k \in \mathscr{B}$.

Thus mixing coincides with mixing of order 1. One of the outstanding open problems in classical ergodic theory is that it is not known(25) if mixing implies mixing of order k for every $k \geqslant 1$.

Despite the natural definition, mixing turns out to be a rather special property, less useful and less prevalent than a slightly weaker property called weak-mixing introduced by Koopman and von Neumann [209](26). Nonetheless, many natural examples are mixing of all orders (see the argument in Proposition 2.15 and Exercise 2.7.9 for example).

Definition 2.35. A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is *weak-mixing* if

$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\right| \longrightarrow 0$$

as $N \to \infty$, for all $A, B \in \mathscr{B}$.

Notice that for any sequence (a_n),

$$\lim_{n\to\infty} a_n = 0 \implies \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}|a_i| = 0,$$

but the converse does not hold because the second property permits $|a_n|$ to be large along an infinite but thin set of values of n. Thus at the level simply of sequences, weak-mixing seems to be strictly weaker than strong-mixing. It turns out that this is also true for measure-preserving transformations – there are weak-mixing transformations that are not mixing(27).

Weak-mixing and its generalizations will turn out to be central to Furstenberg's proof of Szemerédi's theorem presented in Chapter 7. The first intimation that weak-mixing is a natural property comes from the fact that it has many equivalent formulations, and we will start to define and explore some of these in Theorem 2.36 below.

For one of these equivalent properties, it will be useful to recall some terminology concerning the operator U_T on the Hilbert space L^2_μ associated to a measure-preserving transformation T of $(X, \mathscr{B}, \mu)$. An *eigenvalue* is a number $\lambda \in \mathbb{C}$ for which there is an *eigenfunction* $f \in L^2_\mu$ with $U_T f = \lambda f$ almost everywhere. Notice that 1 is always an eigenvalue, since a constant function f will satisfy $U_T f = f$. Any eigenvalue λ lies on $\mathbb{S}^1$, since U_T is an isometry of L^2_μ. A measure-preserving transformation T is said to have *continuous spectrum* if the only eigenvalue of T is 1 and the only eigenfunctions are the constant functions.

Recall that a set $J \subseteq \mathbb{N}$ is said to have *density*

$$\mathbf{d}(J) = \lim_{n\to\infty}\frac{1}{n}\left|\{j \in J \mid 1 \leqslant j \leqslant n\}\right|$$

if the limit exists.

Theorem 2.36. *The following properties of a system $(X, \mathscr{B}, \mu, T)$ are equivalent.*

(1) T *is weakly mixing.*
(2) $T \times T$ *is ergodic with respect to* $\mu \times \mu$.
(3) $T \times T$ *is weakly mixing with respect to* $\mu \times \mu$.
(4) *For any ergodic measure-preserving system* $(Y, \mathscr{B}_Y, \nu, S)$, *the system*

$$(X \times Y, \mathscr{B} \otimes \mathscr{B}_Y, \mu \times \nu, T \times S)$$

is ergodic.
(5) *The associated operator* U_T *has no non-constant measurable eigenfunctions (that is,* T *has continuous spectrum).*
(6) *For every* $A, B \in \mathscr{B}$, *there is a set* $J_{A,B} \subseteq \mathbb{N}$ *with density zero for which*

$$\mu\left(A \cap T^{-n}B\right) \longrightarrow \mu(A)\mu(B)$$

as $n \to \infty$ *with* $n \notin J_{A,B}$.
(7) *For every* $A, B \in \mathscr{B}$,

$$\frac{1}{N} \sum_{n=0}^{N-1} \left|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\right|^2 \longrightarrow 0$$

as $N \to \infty$.

The proof of Theorem 2.36 will be given in Section 2.8.

Corollary 2.37. *If* $(X, \mathscr{B}_X, \mu, T)$ *and* $(Y, \mathscr{B}_Y, \nu, S)$ *are both weak-mixing, then the product system* $(X \times Y, \mathscr{B} \otimes \mathscr{C}, \mu \times \nu, T \times S)$ *is weak-mixing.*

Corollary 2.38. *If* T *is weak-mixing, then for any* k *the* k*-fold Cartesian product* $T \times \cdots \times T$ *is weak-mixing with respect to* $\mu \times \cdots \times \mu$.

Corollary 2.39. *If* T *is weak-mixing, then for any* $n \geqslant 1$, *the* n*th iterate* T^n *is weak-mixing.*

Example 2.40. We know that the circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ defined by

$$R_\alpha(t) = t + \alpha \pmod 1$$

is not mixing, but is ergodic if $\alpha \notin \mathbb{Q}$ (cf. Proposition 2.16 and Example 2.33). It is also not weak-mixing; this may be seen using Theorem 2.36(2) since the function $(x, y) \mapsto \mathrm{e}^{2\pi\mathrm{i}(x-y)}$ from $\mathbb{T} \times \mathbb{T} \to \mathbb{S}^1$ is a non-constant function preserved by $R_\alpha \times R_\alpha$.

Exercises for Section 2.7

Exercise 2.7.1. Show that if a measure-preserving system $(X, \mathscr{B}, \mu, T)$ has the property that for any $A, B \in \mathscr{B}$ there exists N such that

$$\mu\left(A \cap T^{-n} B\right) = \mu(A)\mu(B)$$

for all $n \geqslant N$, then it is trivial in the sense that $\mu(A) = 0$ or 1 for every $A \in \mathscr{B}$.

Exercise 2.7.2. (28) Show that if a measure-preserving system $(X, \mathscr{B}, \mu, T)$ has the property that

$$\mu\left(A \cap T^{-n} B\right) \to \mu(A)\mu(B)$$

uniformly as $n \to \infty$ for every measurable $A \subseteq B \in \mathscr{B}$, then it is trivial in the sense that $\mu(A) = 0$ or 1 for every $A \in \mathscr{B}$.

Exercise 2.7.3. This exercise generalizes the argument used in the proof of Proposition 2.15 and relates to the material in Appendix A. A collection $\mathscr{A}$ of measurable sets in $(X, \mathscr{B}, \mu)$ is called a *semi-algebra* (cf. Appendix A) if

- $\mathscr{A}$ contains the empty set;
- for any $A \in \mathscr{A}$, $X \smallsetminus A$ is a finite union of pairwise disjoint members of $\mathscr{A}$;
- for any $A_1, \dots, A_r \in \mathscr{A}$, $A_1 \cap \cdots \cap A_r \in \mathscr{A}$.

The smallest σ-algebra containing $\mathscr{A}$ is called the σ-algebra generated by $\mathscr{A}$. Assume that $\mathscr{A}$ is a semi-algebra that generates $\mathscr{B}$, and prove the following characterizations of the basic mixing properties for a measure-preserving system $(X, \mathscr{B}, \mu, T)$:

(1) T is mixing if and only if

$$\mu\left(A \cap T^{-n} B\right) \longrightarrow \mu(A)\mu(B)$$

as $n \to \infty$ for all $A, B \in \mathscr{A}$.

(2) T is weak-mixing if and only if

$$\frac{1}{N} \sum_{n=0}^{N-1} \left|\mu(A \cap T^{-n} B) - \mu(A)\mu(B)\right| \longrightarrow 0$$

as $N \to \infty$ for all $A, B \in \mathscr{A}$.

(3) T is ergodic if and only if

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A \cap T^{-n} B\right) \longrightarrow \mu(A)\mu(B)$$

as $N \to \infty$ for all $A, B \in \mathscr{A}$.

Exercise 2.7.4. Let $\mathscr{A}$ be a generating semi-algebra in $\mathscr{B}$ (cf. Exercise 2.7.3), and assume that for $A \in \mathscr{A}$, $\mu\left(A \triangle T^{-1}A\right) = 0$ implies $\mu(A) = 0$ or 1. Does it follow that T is ergodic?

Exercise 2.7.5. Show that a measure-preserving system $(X, \mathscr{B}, \mu, T)$ is mixing if and only if

$$\lim_{n\to\infty} \langle U_T^n f, g\rangle = \langle f, 1\rangle \cdot \langle 1, g\rangle$$

for all f and g lying in a dense subset of L^2_μ.

Exercise 2.7.6. Use Exercise 2.7.5 and the technique from Theorem 2.19 to prove the following.

(1) An ergodic automorphism of a compact abelian group is mixing with respect to Haar measure.
(2) An ergodic automorphism of a compact abelian group is mixing of all orders with respect to Haar measure.

Exercise 2.7.7. Show that a measure-preserving system $(X,\mathscr{B},\mu,T)$ is weak-mixing if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, g\rangle - \langle f, 1\rangle \cdot \langle 1, g\rangle| = 0$$

for any $f, g \in L^2_\mu$

Exercise 2.7.8. Show that a measure-preserving system $(X,\mathscr{B},\mu,T)$ is weak-mixing if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle - \langle f, 1\rangle \cdot \langle 1, f\rangle| = 0$$

for any $f \in L^2_\mu$.

Exercise 2.7.9. Show that a Bernoulli shift (cf. Example 2.9) is mixing of order k for every $k \geqslant 1$.

Exercise 2.7.10. Prove the following result due to Rényi [308]: a measure-preserving transformation T is mixing if and only if

$$\mu(A \cap T^{-n}A) \to \mu(A)^2$$

for all $A \in \mathscr{B}$. Deduce that T is mixing if and only if $\langle U_T^n f, f\rangle \to 0$ as $n \to \infty$ for all f in a set of functions dense in the set of all L^2 functions of zero integral.

Exercise 2.7.11. Prove that a measure-preserving transformation T is weak-mixing if and only if for any measurable sets A, B, C with positive measure, there exists some $n \geqslant 1$ such that $T^{-n}A \cap B \neq \varnothing$ and $T^{-n}A \cap C \neq \varnothing$. (This is a result due to Furstenberg.)

Exercise 2.7.12. Write $T^{(k)}$ for the k-fold Cartesian product $T \times \cdots \times T$. Prove[29] that $T^{(k)}$ is ergodic for all $k \geqslant 2$ if and only if $T^{(2)}$ is ergodic.

Exercise 2.7.13. Let T be an ergodic endomorphism of $\mathbb{T}^d$. The following exponential error rate for the mixing property[30],

$$\left| \langle f_1, U_T^n f_2 \rangle - \int f_1 \int f_2 \right| \leqslant S(f_1) S(f_2) \theta^n$$

for some $\theta < 1$ depending on T and for a pair of constants $S(f_1), S(f_2)$ depending on $f_1, f_2 \in C^\infty(\mathbb{T}^d)$, is known to hold.
(a) Prove an exponential rate of mixing for the map $T_n : \mathbb{T} \to \mathbb{T}$ defined by $T_n(x) = nx \pmod 1$.
(b) Prove an exponential rate of mixing for the automorphism of $\mathbb{T}^2$ defined by $T : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y \\ x+y \end{pmatrix}$.
(c) Could an exponential rate of mixing hold for all continuous functions?

2.8 Proof of Weak-mixing Equivalences

Some of the implications in Theorem 2.36 require the development of additional material; after developing it we will end this section with a proof of Theorem 2.36. The first lemma needed is a general one from analysis, due to Koopman and von Neumann [209].

Lemma 2.41. *Let (a_n) be a bounded sequence of non-negative real numbers. Then the following are equivalent:*

(1) $\displaystyle\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} a_j = 0$;
(2) *there is a set $J = J((a_n)) \subseteq \mathbb{N}$ with density zero for which $a_n \xrightarrow[n \notin J]{} 0$;*
(3) $\displaystyle\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} a_j^2 = 0$.

PROOF. (1) $\Longrightarrow$ (2): Let $J_k = \{j \in \mathbb{N} \mid a_j > \frac{1}{k}\}$, so that

$$J_1 \subseteq J_2 \subseteq J_3 \subseteq \cdots . \tag{2.32}$$

For each $k \geqslant 1$,

$$\frac{1}{k} \left| J_k \cap [0,n) \right| < \sum_{\substack{i=0,\dots,n-1, \\ a_i > 1/k}} a_i \leqslant \sum_{i=0}^{n-1} a_i .$$

It follows that

$$\frac{1}{n}\left|J_k \cap [0,n)\right| \leqslant k\frac{1}{n}\sum_{i=0}^{n-1} a_i \longrightarrow 0$$

as $n \to \infty$ for each $k \geqslant 1$, so each J_k has zero density. We will construct the set J by taking a union of segments of each set J_k. Since each of the sets J_k has zero density, we may inductively choose numbers $0 < \ell_1 < \ell_2 < \cdots$ with the property that

$$\frac{1}{n}\left|J_k \cap [0,n)\right| \leqslant \frac{1}{k} \tag{2.33}$$

for $n \geqslant \ell_k$ and any $k \geqslant 1$. Define the set J by

$$J = \bigcup_{k=0}^{\infty} \left(J_k \cap [\ell_k, \ell_{k+1})\right).$$

We claim two properties for the set J, namely

- $a_n \underset{n \notin J}{\longrightarrow} 0$ as $n \to \infty$;
- J has density zero.

For the first claim, note that $J_k \cap [\ell_k, \infty) \subseteq J$ by equation (2.32), so if $J \not\ni n \geqslant \ell_k$ then $n \notin J_k$, and so $a_n \leqslant \frac{1}{k}$. This shows that $a_n \underset{n \notin J}{\longrightarrow} 0$ as claimed.

For the second claim, notice that if $n \in [\ell_k, \ell_{k+1})$ then again by equation (2.32) $J \cap [0,n) \subseteq J_k \cap [0,n)$ and so

$$\frac{1}{n}\left|J \cap [0,n)\right| \leqslant \frac{1}{k}$$

by equation (2.33), showing that J has density zero.

(2) $\implies$ (1): The sequence (a_n) is bounded, so there is some $R > 0$ with $a_n \leqslant R$ for all $n \geqslant 1$. For each $k \geqslant 1$ choose N_k so that

$$J \not\ni n \geqslant N_k \implies a_n < \frac{1}{k}$$

and so that

$$n \geqslant N_k \implies \frac{1}{n}\left|J \cap [0,n)\right| \leqslant \frac{1}{k}.$$

Then for $n \geqslant kN_k$,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} a_i &= \frac{1}{n}\left(\sum_{i=0}^{N_k-1} a_i + \sum_{\substack{i\in J,\\ N_k\leqslant i<n}} a_i + \sum_{\substack{i\notin J,\\ N_k\leqslant i<n}} a_i\right)\\
&< \frac{1}{n}\left(RN_k + R|J\cap[0,n)| + n\frac{1}{k}\right)\\
&\leqslant \frac{2R+1}{k},
\end{aligned}
$$

showing (1).

(3) $\Longleftrightarrow$ (1): This is clear from the characterization (2) of property (1). □

Proof of Theorem 2.36. Properties (1), (6) and (7) are equivalent by Lemma 2.41 applied with $a_n = |\mu(A\cap T^{-n}B) - \mu(A)\mu(B)|$.

(6) $\Longrightarrow$ (3): Given sets $A_1, B_1, A_2, B_2 \in \mathscr{B}$, property (6) gives sets J_1 and J_2 of density zero with

$$
\mu\left(A_1\cap T^{-n}B_1\right) \underset{n\notin J_1}{\longrightarrow} \mu(A_1)\mu(B_1)
$$

and

$$
\mu\left(A_2\cap T^{-n}B_2\right) \underset{n\notin J_2}{\longrightarrow} \mu(A_2)\mu(B_2).
$$

Let $J = J_1\cup J_2$; this still has density zero and

$$
\begin{aligned}
\lim_{J\not\ni n\to\infty} \big|(\mu\times\mu)\left((A_1\times A_2)\cap(T\times T)^{-n}(B_1\times B_2)\right)&\\
-(\mu\times\mu)(A_1\times A_2)\cdot(\mu\times\mu)(B_1\times B_2)\big|&\\
= \lim_{J\not\ni n\to\infty} \big|\mu(A_1\cap T^{-n}B_1)\cdot\mu(A_2\cap T^{-n}B_2)&\\
-\mu(A_1)\mu(A_2)\mu(B_1)\mu(B_2)\big|&\\
= 0,&
\end{aligned}
$$

so $T\times T$ is weak-mixing since the measurable rectangles generate $\mathscr{B}\times\mathscr{B}$.

(3) $\Longrightarrow$ (1): If $T\times T$ is weak-mixing, then property (1) holds in particular for subsets of $X\times X$ of the form $A\times X$ and $B\times X$, which shows that (1) holds for T, so T is weak-mixing.

(1) $\Longrightarrow$ (4): Let $(Y, \mathscr{B}_Y, \nu, S)$ be an ergodic system and assume that T is weak-mixing. For measurable sets $A_1, B_1 \in \mathscr{B}$ and $A_2, B_2 \in \mathscr{B}_Y$,

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1}&(\mu\times\nu)\left(A_1\times A_2\cap(T\times S)^{-n}(B_1\times B_2)\right)\\
&=\frac{1}{N}\sum_{n=0}^{N-1}\mu(A_1\cap T^{-n}B_1)\nu(A_2\cap S^{-n}B_2)\\
&=\frac{1}{N}\sum_{n=0}^{N-1}\mu(A_1)\mu(B_1)\nu(A_2\cap S^{-n}B_2)\\
&+\frac{1}{N}\sum_{n=0}^{N-1}\left[\mu(A_1\cap T^{-n}B_1)-\mu(A_1)\mu(B_1)\right]\nu(A_2\cap S^{-n}B_2). \quad (2.34)
\end{aligned}$$

By the characterization in equation (2.31) and ergodicity of S, the expression on the right in equation (2.34) converges to

$$\mu(A_1)\mu(B_1)\nu(A_2)\nu(B_2).$$

The second term in equation (2.34) is dominated by

$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(A_1\cap T^{-n}B_1)-\mu(A_1)\mu(B_1)\right|$$

which converges to 0 since T is weak-mixing. It follows that

$$\frac{1}{N}\sum_{n=0}^{N-1}(\mu\times\nu)\left(A_1\times A_2\cap(T\times S)^{-n}(B_1\times B_2)\right)\longrightarrow\mu(A_1)\mu(B_1)\nu(A_2)\nu(B_2)$$

so $T\times S$ is ergodic by the characterization in equation (2.31).

(4) $\implies$ (2): Let $(Y,\mathscr{B}_Y,\nu,S)$ be the ergodic system defined by the identity map on the singleton $Y=\{y\}$. Then $T\times S$ is isomorphic to T, so (4) shows that T is ergodic. Invoking (4) again now shows that $T\times T$ is ergodic, proving (2).

(2) $\implies$ (7): We must show that

$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\right|^2\longrightarrow 0$$

as $N\to\infty$, for every $A,B\in\mathscr{B}$. Let μ^2 denote the product measure $\mu\times\mu$ on $(X\times X,\mathscr{B}\otimes\mathscr{B})$. By the ergodicity of $T\times T$,

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A\cap T^{-n}B\right) &= \frac{1}{N}\sum_{n=0}^{N-1}\mu^2\left((A\times X)\cap(T\times T)^{-n}(B\times X)\right)\\
&\longrightarrow\mu^2\left(A\times X\right)\cdot\mu^2\left(B\times X\right)=\mu(A)\mu(B)
\end{aligned}$$

and

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\left(\mu\left(A\cap T^{-n}B\right)\right)^2 &= \frac{1}{N}\sum_{n=0}^{N-1}\mu^2\left((A\times A)\cap(T\times T)^{-n}(B\times B)\right)\\ &\longrightarrow \mu^2(A\times A)\cdot\mu^2(B\times B)=\mu(A)^2\mu(B)^2.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\left[\mu\left(A\cap T^{-n}B\right)-\mu(A)\mu(B)\right]^2 &= \frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A\cap T^{-n}B\right)^2\\ &\quad+\mu(A)^2\mu(B)^2\\ &\quad-2\mu(A)\mu(B)\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(A\cap T^{-n}B\right)\\ &\to 2\mu(A)^2\mu(B)^2-2\mu(A)^2\mu(B)^2=0,\end{aligned}$$

so (7) holds.

(2) $\implies$ (5): Suppose that f is a measurable eigenfunction for T, so

$$U_T f=\lambda f$$

for some $\lambda\in\mathbb{S}^1$. Define a measurable function on $X\times X$ by

$$g(x_1,x_2)=f(x_1)\overline{f(x_2)};$$

then

$$U_{T\times T}g(x,y)=g(Tx,Ty)=\lambda\overline{\lambda}g(x,y)=g(x,y)$$

so by ergodicity of $T\times T$, g (and hence f) must be constant almost everywhere.

All that remains is to prove that (5) $\implies$ (2), and this is considerably more difficult. There are several different proofs, each of which uses a non-trivial result from functional analysis$^{(31)}$. Assume that $T\times T$ is not ergodic, so there is a non-constant function $f\in L^2_{\mu^2}(X\times X)$ that is almost everywhere invariant under $T\times T$. We would like to have the additional symmetry property $f(x,y)=\overline{f(y,x)}$ for all $(x,y)\in X\times X$. To obtain this additional property, consider the functions

$$(x,y)\mapsto f(x,y)+\overline{f(y,x)}$$

and

$$(x,y)\mapsto \mathrm{i}(f(x,y)-\overline{f(y,x)}).$$

Notice that if both of these functions are constant, then f must be constant. It follows that one of them must be non-constant. So without loss of generality we may assume that f satisfies $f(x,y)=\overline{f(y,x)}$. We may further suppose (by subtracting $\int f \,\mathrm{d}\mu^2$) that $\int f \,\mathrm{d}\mu^2 = 0$. It follows that the operator F on L^2_μ defined by

$$(F(g))(x) = \int_X f(x,y)g(y)\,\mathrm{d}\mu(y)$$

is a non-trivial self-adjoint compact[(32)] operator, and so by Theorem B.3 has at least one non-zero eigenvalue λ whose corresponding eigenspace V_λ is finite-dimensional. We claim that the finite-dimensional space $V_\lambda \subseteq L^2_\mu$ is invariant under T. To see this, assume that $F(g) = \lambda g$. Then

$$\begin{aligned}\lambda g(Tx) &= \int_X f(Tx,y)g(y)\,\mathrm{d}\mu(y)\\ &= \int_X f(Tx,Ty)g(Ty)\,\mathrm{d}\mu(y) \quad \text{(since } \mu \text{ is } T\text{-invariant)}\\ &= \int_X f(x,y)g(Ty)\,\mathrm{d}\mu(y),\end{aligned}$$

since f is $T\times T$-invariant, so $F(g\circ T) = \lambda(g\circ T)$ and thus $g\circ T \in V_\lambda$. It follows that U_T restricted to V_λ is a non-trivial linear map of a finite-dimensional linear space, and therefore has a non-trivial eigenvector. Since $\int f\,\mathrm{d}\mu^2 = 0$, any such eigenvector is non-constant. □

2.8.1 Continuous Spectrum and Weak-Mixing

A more conventional proof of the difficult step in Theorem 2.36, which may be taken to be (5) $\implies$ (1), proceeds via the Spectral theorem (Theorem B.4) in the following form.

Alternative proof of (5) $\implies$ (1) in Theorem 2.36. Definition 2.35 is clearly equivalent to the property that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\langle U_T^n f, g\rangle - \langle f,1\rangle \cdot \langle 1,g\rangle| = 0$$

for any $f, g \in L^2_\mu$, and by polarization this is in turn equivalent to

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle - \langle f,1\rangle \cdot \langle 1,f\rangle| = 0$$

for any $f \in L^2_\mu$ (see Exercise 2.7.8 and page 441). By subtracting $\int_X f\,\mathrm{d}\mu$ from f, it is therefore enough to show that if $f \in L^2_\mu$ has $\int_X f\,\mathrm{d}\mu = 0$, then

$$\frac{1}{N}\sum_{n=0}^{N-1}|\langle U_T^n f, f\rangle|^2 \longrightarrow 0$$

as $N\to\infty$. By equation (B.1), it is enough to show that for the non-atomic measure μ_f on $\mathbb{S}^1$,

$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\int_{\mathbb{S}^1} z^n \,\mathrm{d}\mu_f(z)\right|^2 \longrightarrow 0 \tag{2.35}$$

as $N\to\infty$. Since $\overline{z^n}=z^{-n}$ for $z\in\mathbb{S}^1$ the product in equation (2.35) may be expanded to give

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1}\left|\int_{\mathbb{S}^1} z^n \,\mathrm{d}\mu_f(z)\right|^2 &= \frac{1}{N}\sum_{n=0}^{N-1}\left(\int_{\mathbb{S}^1} z^n \,\mathrm{d}\mu_f(z)\cdot\int_{\mathbb{S}^1} w^{-n}\,\mathrm{d}\mu_f(w)\right)\\
&= \frac{1}{N}\sum_{n=0}^{N-1}\int_{\mathbb{S}^1\times\mathbb{S}^1}(z/w)^n \,\mathrm{d}\mu_f^2(z,w) \quad \text{(by Fubini)}\\
&= \int_{\mathbb{S}^1\times\mathbb{S}^1}\left(\frac{1}{N}\sum_{n=0}^{N-1}(z/w)^n\right)\mathrm{d}\mu_f^2(z,w).
\end{aligned}$$

The measure μ_f is non-atomic so the diagonal set $\{(z,z)\mid z\in\mathbb{S}^1\}\subseteq\mathbb{S}^1\times\mathbb{S}^1$ has zero μ_f^2-measure. For $z\neq w$,

$$\frac{1}{N}\sum_{n=0}^{N-1}(z/w)^n = \frac{1}{N}\left(\frac{1-(z/w)^N}{1-(z/w)}\right)\longrightarrow 0$$

as $N\to\infty$, so the convergence (2.35) holds by the dominated convergence theorem (Theorem A.18). □

Exercises for Section 2.8

Exercise 2.8.1. Is the hypothesis that the sequence (a_n) be bounded necessary in Lemma 2.41?

Exercise 2.8.2. Give an alternative proof of (1) $\implies$ (5) in Theorem 2.36 by proving the following statements:

(1) Any factor of a weak-mixing transformation is weak-mixing.
(2) A complex-valued eigenfunction f of U_T has constant modulus.
(3) If f is an eigenfunction of U_T, then $x\mapsto\arg\left(f(x)/|f(x)|\right)$ is a factor map from $(X,\mathscr{B},\mu,T)$ to $(\mathbb{T},\mathscr{B}_{\mathbb{T}},m_{\mathbb{T}},R_\alpha)$ for some α.

Exercise 2.8.3. Show the following converse to Exercise 2.5.6: if a measure-preserving system $(Y,\mathscr{B}_Y,\nu,S)$ is not totally ergodic then there exists a measure-preserving system $(X, \mathscr{B}, \mu, T)$ and a $K > 1$ with the property that $(Y, \mathscr{B}_Y, \nu, S)$ is measurably isomorphic to the system

$$(X^{(K)}, \mathscr{B}^{(K)}, \mu^{(K)}, T^{(K)})$$

constructed in Exercise 2.5.6.

Exercise 2.8.4. Give a different proof(33) of the mean ergodic theorem (Theorem 2.21) as follows. For a measure-preserving system $(X, \mathscr{B}, \mu, T)$ and function $f \in L^2_\mu$, show that the function $n \mapsto \langle U_T^n f, f\rangle$ is positive-definite (see Section C.3). Apply the Herglotz–Bochner theorem (Theorem C.9) to translate the problem into one concerned with functions on $\mathbb{S}^1$, and there use the fact that $\frac{1}{N}\sum_{n=1}^{N} \rho^n$ converges for $\rho \in \mathbb{S}^1$ (to zero, unless $\rho = 1$).

2.9 Induced Transformations

Poincaré recurrence gives rise to an important inducing construction introduced by Kakutani [172]. Throughout this section, $(X, \mathscr{B}, \mu, T)$ denotes an invertible measure-preserving system(34).

Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system, and let A be a measurable set with $\mu(A) > 0$. By Poincaré recurrence, the first return time to A, defined by

$$r_A(x) = \inf_{n \geqslant 1}\{n \mid T^n(x) \in A\} \tag{2.36}$$

exists (that is, is finite) almost everywhere.

Definition 2.42. The map $T_A : A \to A$ defined (almost everywhere) by

$$T_A(x) = T^{r_A(x)}(x)$$

is called the transformation *induced* by T on the set A.

Notice that both $r_A : X \to \mathbb{N}$ and $T_A : A \to A$ are measurable by the following argument. For $n \geqslant 1$, write $A_n = \{x \in A \mid r_a(x) = n\}$. Then the sets

$$\begin{aligned}
A_1 &= A \cap T^{-1}A,\\
A_2 &= A \cap T^{-2}A \smallsetminus A_1,\\
&\ \ \vdots\\
A_n &= A \cap T^{-n}A \smallsetminus \bigcup_{i<n} A_i
\end{aligned}$$

are all measurable, as is

$$T^n A_n = A \cap T^n A \smallsetminus \left(TA \cup T^2 A \cup \cdots \cup T^{n-1}A\right),$$

since T is invertible by assumption.

Lemma 2.43. *The induced transformation T_A is a measure-preserving transformation on the space $(A, \mathscr{B}\big|_A, \mu_A = \frac{1}{\mu(A)}\mu\big|_A, T_A)$. If T is ergodic with respect to μ then T_A is ergodic with respect to μ_A.*

The notation means that the σ-algebra consists of $\mathscr{B}\big|_A = \{B \cap A \mid B \in \mathscr{B}\}$ and the measure is defined for $B \in \mathscr{B}\big|_A$ by $\mu_A(B) = \frac{1}{\mu(A)}\mu(B)$. The effect of T_A is seen in the *Kakutani skyscraper* Figure 2.2. The original transformation T sends any point with a floor above it to the point immediately above on the next floor, and any point on a top floor is moved somewhere to the base floor A. The induced transformation T_A is the map defined almost everywhere on the bottom floor by sending each point to the point obtained by going through all the floors above it and returning to A.

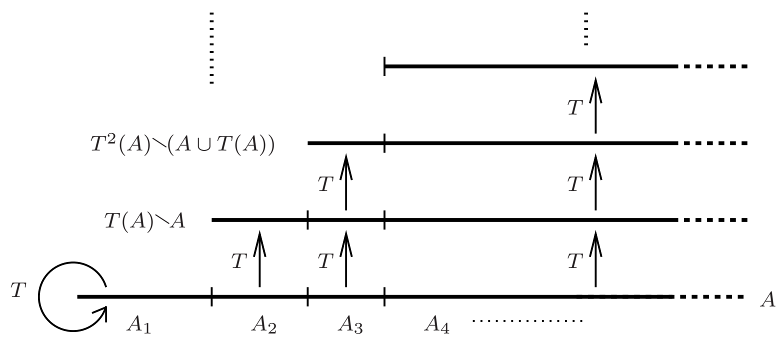


Fig. 2.2: The induced transformation T_A.

Proof of Lemma 2.43. If $B \subseteq A$ is measurable, then $B = \bigsqcup_{n \geqslant 1} B \cap A_n$ is a disjoint union so

$$\mu_A(B) = \frac{1}{\mu(A)} \sum_{n \geqslant 1} \mu(B \cap A_n). \tag{2.37}$$

Now

$$T_A(B) = \bigsqcup_{n \geqslant 1} T_A(B \cap A_n) = \bigsqcup_{n \geqslant 1} T^n(B \cap A_n),$$

so

$$
\begin{aligned}
\mu_A(T_A(B)) &= \frac{1}{\mu(A)} \sum_{n\geqslant 1} \mu(T^n(B\cap A_n)) \\
&= \frac{1}{\mu(A)} \sum_{n\geqslant 1} \mu(B\cap A_n) \qquad \text{(since } T \text{ preserves } \mu\text{)} \\
&= \mu(B)
\end{aligned}
$$

by equation (2.37).

If T_A is not ergodic, then there is a T_A-invariant measurable set $B \subseteq A$ with $0 < \mu(B) < \mu(A)$; it follows that $\bigcup_{n\geqslant 1}\bigcup_{j=0}^{n-1} T^j(B\cap A_n)$ is a non-trivial T-invariant set, showing that T is not ergodic. □

Poincaré recurrence (Theorem 2.11) says that for any measure-preserving system $(X, \mathscr{B}, \mu, T)$ and set A of positive measure, almost every point on the ground floor of the associated Kakutani skyscraper returns to the ground floor at some point. Ergodicity strengthens this statement to say that almost every point of the entire space X lies on some floor of the skyscraper. This enables a quantitative version of Poincaré recurrence to be found, a result due to Kac [168].

Theorem 2.44 (Kac). *Let $(X, \mathscr{B}, \mu, T)$ be an ergodic measure-preserving system and let $A \in \mathscr{B}$ have $\mu(A) > 0$. Then the expected return time to A is $\frac{1}{\mu(A)}$; equivalently*

$$\int_A r_A \,\mathrm{d}\mu = 1.$$

Proof[(35)]. Referring to Figure 2.2, each column

$$A_n \sqcup T(A_n) \sqcup \cdots \sqcup T^{n-1}(A_n)$$

comprises n disjoint sets each of measure $\mu(A_n)$, and the entire skyscraper contains almost all of X by ergodicity and Proposition 2.14(3) applied to the transformation T^{-1}. It follows that

$$1 = \mu(X) = \sum_{n\geqslant 1} n\mu(A_n) = \int_A r_A \,\mathrm{d}\mu$$

by the monotone convergence theorem (Theorem A.16), since r_A is the increasing limit of the functions $\sum_{k=1}^{n} k\chi_{A_k}$ as $n \to \infty$. □

Kakutani skyscrapers are a powerful tool in ergodic theory. A simple application is to prove the Kakutani–Rokhlin lemma (Lemma 2.45) proved by Kakutani [172] and Rokhlin [315].

Lemma 2.45 (Kakutani–Rokhlin). *Let $(X, \mathscr{B}, \mu, T)$ be an invertible ergodic measure-preserving system and assume that μ is non-atomic (that is, $\mu(\{x\}) = 0$ for all $x \in X$). Then for any $n \geqslant 1$ and $\varepsilon > 0$ there is a set $B \in \mathscr{B}$ with the property that*

$$B, T(B), \dots, T^{n-1}(B)$$

are disjoint sets, and

$$\mu\left(B \sqcup T(B) \sqcup \cdots \sqcup T^{n-1}(B)\right) > 1 - \varepsilon.$$

As the proof will show, the lemma uses only division (constructing a quotient and remainder) and the Kakutani skyscraper.

PROOF OF LEMMA 2.45. Let A be a measurable set with $0 < \mu(A) < \varepsilon/n$ (such a set exists by the assumption that μ is non-atomic) and form the Kakutani skyscraper over A. Then X decomposes into a union of disjoint columns of the form

$$A_k \sqcup T(A_k) \sqcup \cdots \sqcup T^{k-1}(A_k)$$

for $k \geqslant 1$, as in Figure 2.2. Now let

$$B = \bigsqcup_{k \geqslant n} \bigsqcup_{j=0}^{\lfloor k/n \rfloor - 1} T^{jn}(A_k),$$

the set obtained by grouping together that part of the ground floor made up of the sets A_k with $k \geqslant n$ together with every nth floor above that part of the ground floor (stopping before the top of the skyscraper). By construction the sets $B, T(B), \dots, T^{n-1}(B)$ are disjoint, and together they cover all of X apart from a set comprising no more than n of the floors in each of the towers, which therefore has measure no more than $n \sum_{k=1}^{\infty} \mu(A_k) \leqslant n\mu(A) < \varepsilon$. $\square$

One often refers to the structure given by Lemma 2.45 as a *Rokhlin tower* of *height* n with *base* B and *residual set of size* ε.

Exercises for Section 2.9

Exercise 2.9.1. Show that the inducing construction can be reversed in the following sense. Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system, and let $r : X \to \mathbb{N}_0$ be a map in L^1_μ. The *suspension* defined by r is the system $(X^{(r)}, \mathscr{B}^{(r)}, \mu^{(r)}, T^{(r)})$, where:

- $X^{(r)} = \{(x, n) \mid 0 \leqslant n < r(x)\}$;
- $\mathscr{B}^{(r)}$ is the product σ-algebra of $\mathscr{B}$ and the Borel σ-algebra on $\mathbb{N}$ (which comprises all subsets);
- $\mu^{(r)}$ is defined by $\mu^{(r)}(A \times N) = \frac{1}{\int r \, \mathrm{d}\mu} \mu(A) \times |N|$ for $A \in \mathscr{B}$ and $N \subseteq \mathbb{N}$; and
- $T^{(r)}(x, n) = \begin{cases} (x, n+1) & \text{if } n+1 < r(x); \\ (T(x), 0) & \text{if } n+1 = r(x). \end{cases}$

(a) Verify that this defines a finite measure-preserving system.
(b) Show that the induced map on the set $A = \{(x, 0) \mid x \in X\}$ is isomorphic to the original system $(X, \mathscr{B}, \mu, T)$.

Exercise 2.9.2. [(36)] The hypothesis of ergodicity in Lemma 2.45 can be weakened as follows. An invertible measure-preserving system $(X, \mathscr{B}, \mu, T)$ is called *aperiodic* if $\mu\left(\{x \in X \mid T^k(x) = x\}\right) = 0$ for all $k \in \mathbb{Z}\smallsetminus\{0\}$.
(a) Show that an ergodic transformation on a non-atomic space is aperiodic.
(b) Find an example of an aperiodic transformation on a non-atomic space that is not ergodic.
(c) Prove Lemma 2.45 for an invertible aperiodic transformation on a non-atomic space.

Exercise 2.9.3. [(37)]Show that the Kakutani–Rokhlin lemma (Lemma 2.45) does not hold for arbitrary sequences of iterates of the map T. Specifically, show that for an ergodic measure-preserving system $(X, \mathscr{B}, \mu, T)$, sequence $a_1, \dots, a_n$ of distinct integers, and $\varepsilon > 0$ it is not always possible to find a measurable set A with the properties that $T^{a_1}(A), \dots, T^{a_n}(A)$ are disjoint and $\mu\left(\bigcup_{i=1}^n T^{a_i}(A)\right) > \varepsilon$.

Exercise 2.9.4. Use Exercise 2.9.2 above to prove the following result of Steele [351]. Let $(X, \mathscr{B}, \mu, T)$ be an invertible aperiodic measure-preserving system on a non-atomic space. Then, for any $\varepsilon > 0$, there is a set $A \in \mathscr{B}$ with $\mu(A) < \varepsilon$ with the property that for any finite set $F \subseteq X$, there is some $j = j(F)$ with $F \subseteq T^{-j}(A)$.

Notes to Chapter 2

[(12)](Page 16) A measurable isomorphism is also sometimes called a *conjugacy*; conjugacy is also used to describe an isomorphism between the measure algebras that implies isomorphism on sufficiently well-behaved probability spaces. This is discussed in Walters [373, Sect. 2.2] and Royden [320].
[(13)](Page 17) The shift maps constructed here are measure-preserving transformations, but they are also homeomorphisms of a compact metric space in a natural way. The study of the dynamics of closed shift-invariant subsets of these systems comprises *symbolic dynamics* and is a rich theory in itself. A gentle introduction may be found in the book of Lind and Marcus [230] or Kitchens [197]; further reading in the collection edited by Berthé, Ferenczi, Mauduit and Siegel [93].
[(14)](Page 21) Poincaré's formulation in [288, Th. I, p. 69] is as follows:

> "Supposons que le point P reste à distance finie, et que le volume
>
> $$\int \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3$$
>
> soit un invariant intégral; si l'on considère une région r_0 quelconque, quelque petite que soite cette région, il y aura des trajectoires qui la traverseront une infinité de fois. [...] En effet le point P restant à distance finie, ne sortira jamais d'une région limitée R."

The modern abstract measure-theoretic statement in Theorem 2.11 appears in a paper of Carathéodory [49].

(15)(Page 23) The notion of ergodicity predates the ergodic theorems of the 1930s, in various guises. These include the seminal work of Borel [40], described by Doob as being

> "characterized by convenient neglect of error terms in asymptotics, incorrect reasoning, and correct results,"

as well as that of Knopp [205]; a striking remark of Novikoff and Barone [273] is that a result implicit in the work of van Vleck [369] on non-measurable subsets of $[0,1]$ is that any measurable subset of $[0,1]$ invariant under the map $x \mapsto 2x \pmod 1$ has measure zero or one, a prototypical ergodic statement. The general formulation was given by Birkhoff and Smith [35].

(16)(Page 28) These operators are usually called Koopman operators; Koopman [208] used the then-recent development of functional analysis and Hilbert space by von Neumann [266] and Stone [354] to use these operators in the setting of flows arising in classical Hamiltonian mechanics.

(17)(Page 30) Even though this is not necessary here, we assume for simplicity that Hilbert spaces are separable, and as a result that they have countable orthonormal bases. As discussed in Section A.6, we only need the separable case.

(18)(Page 32) For a recent account of the history of the relationship between the two results and the account of how they came to be published as and when they did, see Zund [394]. The issue has also been discussed by Ulam [364] and others. The note [25] by Bergelson discusses both the history and how the two results relate to more recent developments.

(19)(Page 35) This result is simply one of many extensions and generalizations of the mean ergodic theorem (Theorem 2.21) to other complete function spaces. It is a special instance of the mean ergodic theorem for Banach spaces, due to Kakutani and Yosida [171], [390], [391].

(20)(Page 38) The maximal ergodic theorem is due to Wiener [381] and was also proved by Yosida and Kakutani [391].

(21)(Page 40) Covering lemmas of this sort were introduced by Vitali [368], and later became important tools in the proof of the Hardy–Littlewood maximal inequality, and thence of the Lebesgue density and differentiation theorems (Theorems A.24 and A.25).

(22)(Page 44) Birkhoff based his proof on a weaker maximal inequality concerning the set of points on which $\limsup_{n\to\infty} \mathsf{A}_n^f \geqslant \alpha$, and initially formulated his result for indicator functions in the setting of a closed analytic manifold with a finite invariant measure. Khinchin [189] showed that Birkhoff's result applies to integrable functions on abstract finite measure spaces, but made clear that the idea of the proof is precisely that used by Birkhoff. A natural question concerning Theorem 2.30, or indeed any convergence result, is whether anything can be said about the rate of convergence. An important special case is the law of the iterated logarithm due to Hartman and Wintner [141]: if $\|f\|_2 = 1$, $\int f\,\mathrm{d}\mu = 0$ and the functions $f, U_T f, U_T^2 f, \dots$ are all independent, then

$$\limsup_{n\to\infty} \mathsf{A}_n^f / \sqrt{(2\log\log n)/n} = 1$$

almost everywhere (and $\liminf = -1$ by symmetry). It follows that

$$\mathsf{A}_n^f = \mathrm{O}\left((\tfrac{1}{n}\log\log n)^{1/2} \right)$$

almost everywhere. However, the hypothesis of independence is essential: Krengel [210] showed that for any ergodic Lebesgue measure-preserving transformation T of $[0,1]$ and sequence (a_n) with $a_n \to 0$ as $n \to \infty$, there is a continuous function $f : [0,1] \to \mathbb{R}$ for which

$$\limsup_{n\to\infty} \frac{1}{a_n} \left| \mathsf{A}_n^f - \int f\,\mathrm{d}m \right| = \infty$$

almost everywhere, and

$$\limsup_{n\to\infty} \frac{1}{a_n} \left\| \mathsf{A}_n^f - \int f \, \mathrm{d}m \right\|_p = \infty$$

for $1 \leqslant p \leqslant \infty$. An extensive treatment of ergodic theorems may be found in the monograph of Krengel [211].

Despite the absence of any general rate bounds in the ergodic theorem, the constructive approach to mathematics has produced rate results in a different sense, which may lead to effective versions of results like the multiple recurrence theorem. Bishop's work [36] included a form of ergodic theorem, and Spitters [348] found constructive characterizations of the ergodic theorem. As an application of 'proof mining', Avigad, Gerhardy and Towsner [12] gave bounds on the rate of convergence that can be explicitly computed in terms of the initial data (T and f) under a weak hypotheses, while earlier work of Simic and Avigad [13], [346] showed that, in general, it is impossible to compute such a bound. An overview of this area and its potential may be found in the survey [11] by Avigad.

(23) (Page 44) Despite the impressive result in Example 2.31, the numbers known to be normal to every base have been constructed to meet the definition of normality (with the remarkable exception of Chaitin's constant [53]). Champernowne [54] showed that the specific number $0.123456789101112131415\ldots$ is normal in base 10, and Sierpiński [345] constructed a number normal to every base. Sierpiński's construction was reformulated to be recursive by Becher and Figueira [20], giving a computable number normal to every base. The irrational numbers arising naturally in other fields, like $\pi, \mathrm{e}, \zeta(3), \sqrt{2}$, and so on, are not known to be normal to any base.

(24) (Page 44) There are many proofs of the pointwise ergodic theorem; in addition to that of Birkhoff [33] there is a more elementary (though intricate) argument due to Katznelson and Weiss [186], motivated by a paper of Kamae [177]. A different proof is given by Jones [167].

(25) (Page 50) This conjectured result — the "Rokhlin problem" — has been shown in important special cases by Host [158], Kalikow [176], King [193], Ryzhikov [328], del Junco and Yassawi [68], [389] and others, but the general case is open.

(26) (Page 50) The definition used by Koopman and von Neumann is the spectral one that will be given in Theorem 2.36(5), and was called by them the absence of "angle variables"; they also considered flows (measure-preserving actions of $\mathbb{R}$ rather than actions of $\mathbb{Z}$ or $\mathbb{N}$). In physical terms, they characterized lack of ergodicity as barriers that are never passed, and the presence of an angle variable as a clock that never changes, under the dynamics.

(27) (Page 50) Examples of such systems were constructed using Gaussian processes by Maruyama [255]; Kakutani [174] gave a direct combinatorial construction of an example (this example is described in detail in the book of Petersen [282, Sect. 4.5]). Other examples were found by Chacon [51], [52] and Katok and Stepin [185]. Indeed, there is a reasonable way of viewing the collection of all measure-preserving transformations of a fixed space in which a typical transformation is weak-mixing but not mixing (see papers of Rokhlin [315] and Halmos [135] or Halmos' book [138, pp. 77–80]).

(28) (Page 52) This more subtle version of Exercise 2.7.1 appears in a paper of Halmos [136], and is attributed to Ambrose, Halmos and Kakutani in Petersen's book [282].

(29) (Page 54) This is shown in the notes of Halmos [138]. Ergodicity also makes sense for transformations preserving an infinite measure; in that setting Kakutani and Parry [175] used random walk examples of Gillis [115] to show that for any $k \geqslant 1$ there is an infinite measure-preserving transformation T with $T^{(k)}$ ergodic and $T^{(k+1)}$ not ergodic.

(30) (Page 54) This is also known as exponential or effective rate of mixing or decay of correlations; see Baladi [15] for an overview of dynamical settings where it is known.

(31) (Page 58) A more constructive proof of the difficult step in Theorem 2.36 (which may be taken to be (5) $\Longrightarrow$ (1)) exploiting properties of almost-periodic functions on compact groups, and giving more insight into the structure of ergodic measure-preserving transformations that are not weak-mixing, may be found in Petersen [282, Sect. 4.1].

(32) (Page 59) This is an example of a Hilbert–Schmidt operator [331]; a convenient source for this material is the book of Rudin [321] or Appendix B.

(33) (Page 61) This way of viewing ergodic theorems lies at the start of a sophisticated investigation of ergodic theorems along arithmetic sets of integers by Bourgain [41]. This exercise already points at a relationship between ergodic theorems and equidistribution on the circle.
(34) (Page 61) Notice that the assumption that $(X, \mathscr{B}, \mu, T)$ is invertible also implies that T is *forward measurable*, that is $T(A) \in \mathscr{B}$ for any $A \in \mathscr{B}$. Heinemann and Schmitt [146] prove the Rokhlin lemma for an aperiodic measure-preserving transformation on a Borel probability space using Exercise 5.3.2 and Poincaré recurrence instead of a Kakutani tower (aperiodic is defined in Exercise 2.9.2; for Borel probability space see Definition 5.13). A non-invertible Rokhlin lemma is also developed by Rosenthal [317] in his work on topological models for measure-preserving systems and by Hoffman and Rudolph [155] in their extension of the Bernoulli theory to non-invertible systems.
(35) (Page 63) This short proof comes from a paper of Wright [388], in which Kac's theorem is extended to measurable transformations.
(36) (Page 65) The extension in Exercise 2.9.2 appears in the notes of Halmos [138, p. 71].
(37) (Page 65) Exercise 2.9.3 is taken from a paper of Keane and Michel [188]; they also show that the supremum of $\mu\left(\bigcup_{i=1}^{n} T^{a_i}(A)\right)$ over sets A for which

$$T^{a_1}(A), \dots, T^{a_n}(A)$$

are disjoint is a rational number, and show how this can be computed from the integers $a_1, \dots, a_n$.

Chapter 3
Continued Fractions

The continued fraction decomposition of real numbers grows naturally from the Euclidean algorithm, and continued fractions have been used in some form for thousands of years. One goal of this volume is to show how they relate to a natural action on a homogeneous space. To start there would be to willfully reverse their historical development: We start instead with their basic properties[38] from an elementary point of view in Section 3.1, then show how continued fractions are related to an explicit measure-preserving transformation in Section 3.2. In Chapter 9 we will see how the continued fraction map fits into the more general framework of actions on homogeneous spaces.

Let us mention one result proved in this chapter. We will show that for every irrational $x \in \mathbb{R}$ there is a sequence of 'best rational approximations' $\frac{p_n(x)}{q_n(x)} \in \mathbb{Q}$, defined by the continued fraction expansion of x. Moreover, for almost every x we have

$$\lim_{n\to\infty} \frac{1}{n} \log \left| x - \frac{p_n(x)}{q_n(x)} \right| \longrightarrow -\frac{\pi^2}{6\log 2},$$

which gives a precise description of the expected speed of approximation along this sequence.

3.1 Elementary Properties

A (simple) *continued fraction* is a formal expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}}} \tag{3.1}$$

which we will also denote by

$$[a_0; a_1, a_2, a_3, \dots]$$

with $a_n \in \mathbb{N}$ for $n \geqslant 1$ and $a_0 \in \mathbb{N}_0$. Also write

$$[a_0; a_1, a_2, \dots, a_n]$$

for the finite fraction

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}.$$

Thus, for example

$$[a_0; a_1, a_2, \dots, a_n] = a_0 + \frac{1}{[a_1; a_2, \dots, a_n]}.$$

We will see later that the expression in equation (3.1) – when suitably interpreted – converges, and therefore defines a real number. The numbers a_n are the *partial quotients* of the continued fraction. The following simple lemma is crucial for many of the basic properties of the continued fraction expansion.

Lemma 3.1. *Fix a sequence $(a_n)_{n\geqslant 0}$ with $a_0 \in \mathbb{N}_0$ and $a_n \in \mathbb{N}$ for $n \geqslant 1$. Then the rational numbers*

$$\frac{p_n}{q_n} = [a_0; a_1, a_2, \dots, a_n] \tag{3.2}$$

for $n \geqslant 0$ with coprime numerator $p_n \geqslant 1$ and denominator $q_n \geqslant 1$ can be found recursively from the relation

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \text{ for } n \geqslant 0. \tag{3.3}$$

In particular, we set $p_{-1} = 1, q_{-1} = 0, p_0 = a_0,$ and $q_0 = 1$.

Proof. Notice first that the sequence $(a_n)_{n\geqslant 0}$ defines the sequences $(p_n)_{n\geqslant -1}$ and $(q_n)_{n\geqslant -1}$. The claim of the lemma is proved by induction on n. Assume that equation (3.3) holds for $0 \leqslant n \leqslant k-1$ and p_n, q_n as defined by equation (3.2) for any sequence $(a_0, a_1, \dots)$. This is clear for $n = 0$. Thus, on replacing the first k terms of the sequence $(a_n)_{n\geqslant 0}$ with the first k terms of the sequence $(a_n)_{n\geqslant 1}$, we have

$$\frac{x}{y} = [a_1; a_2, \dots, a_k]$$

as a fraction in lowest terms where x and y are defined by

$$\begin{pmatrix} x & x' \\ y & y' \end{pmatrix} = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x & x' \\ y & y' \end{pmatrix} = \begin{pmatrix} p_k & p_{k-1} \\ q_k & q_{k-1} \end{pmatrix} = \begin{pmatrix} a_0 x + y & a_0 x' + y' \\ x & x' \end{pmatrix},$$

so

$$\frac{p_k}{q_k} = \frac{a_0 x + y}{x} = a_0 + \frac{y}{x} = a_0 + \frac{1}{[a_1; a_2, \dots, a_k]} = [a_0; a_1, \dots, a_k],$$

which shows that equation (3.2) holds for $n = k$ also. □

An immediate consequence of Lemma 3.1 is a pair of recursive formulas

$$p_{n+1} = a_{n+1} p_n + p_{n-1}$$

and

$$q_{n+1} = a_{n+1} q_n + q_{n-1} \tag{3.4}$$

for any $n \geqslant 1$, since

$$\begin{pmatrix} p_{n+1} & p_n \\ q_{n+1} & q_n \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} a_{n+1} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a_{n+1} p_n + p_{n-1} & p_n \\ a_{n+1} q_n + q_{n-1} & q_n \end{pmatrix}.$$

It follows that

$$1 = q_0 \leqslant q_1 < q_2 < \cdots \tag{3.5}$$

since $a_n \geqslant 1$ for all $n \geqslant 1$; by induction

$$q_n \geqslant 2^{(n-2)/2} \tag{3.6}$$

and similarly

$$p_n \geqslant 2^{(n-2)/2} \tag{3.7}$$

for all $n \geqslant 1$. Taking determinants in equation (3.3) shows that

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n+1} \tag{3.8}$$

and hence $\frac{p_1}{q_1} = a_0 + \frac{1}{q_0 q_1}$, $\frac{p_2}{q_2} = \frac{p_1}{q_1} - \frac{1}{q_1 q_2} = a_0 + \frac{1}{q_0 q_1} - \frac{1}{q_1 q_2}$ and

$$\begin{aligned} \frac{p_n}{q_n} &= \frac{p_{n-1}}{q_{n-1}} + (-1)^{n+1} \frac{1}{q_{n-1} q_n} \\ &= a_0 + \frac{1}{q_0 q_1} - \frac{1}{q_1 q_2} + \cdots + (-1)^{n+1} \frac{1}{q_{n-1} q_n} \end{aligned} \tag{3.9}$$

for all $n \geqslant 1$ by induction.

This shows that an infinite continued fraction is not just a formal object, it in fact converges to a real number. Namely,

$$\begin{aligned} u = [a_0; a_1, a_2, \dots] &= \lim_{n\to\infty} [a_0; a_1, \dots, a_n] \\ &= \lim_{n\to\infty} \frac{p_n}{q_n} = a_0 + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{q_{n-1}q_n}, \end{aligned} \tag{3.10}$$

is always convergent (indeed, is absolutely convergent) by the inequality (3.6). Moreover, an immediate consequence of equation (3.10) and equation (3.5) is a sequence of inequalities describing how the continued fraction converges: if $a_n \in \mathbb{N}$ for $n \geqslant 1$ then

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < u < \cdots < \frac{p_{2m+1}}{q_{2m+1}} < \cdots < \frac{p_3}{q_3} < \frac{p_1}{q_1}. \tag{3.11}$$

We say that $[a_0; a_1, \dots]$ is the *continued fraction expansion* for u. The name suggests that the expansion is (almost) unique and that it always exists. We will see that in fact any irrational number u has a continued fraction expansion, and that it is unique (Lemmas 3.6 and 3.4).

The rational numbers $\frac{p_n}{q_n}$ are called the *convergents* of the continued fraction for u and they provide very rapid rational approximations to u. Indeed,

$$u - \frac{p_n}{q_n} = (-1)^n \left[\frac{1}{q_n q_{n+1}} - \frac{1}{q_{n+1}q_{n+2}} + \cdots \right] \tag{3.12}$$

so by equation (3.5) we have[(39)]

$$\left| u - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}. \tag{3.13}$$

By equation (3.4) we deduce that

$$\left| u - \frac{p_n}{q_n} \right| < \frac{1}{a_{n+1} q_n^2} \leqslant \frac{1}{q_n^2}. \tag{3.14}$$

Recall from Section 1.5 that we write

$$\langle t \rangle = \min_{q\in\mathbb{Z}} |t - q|$$

for the distance from t to the nearest integer. The inequality (3.14) gives one explanation* for the comment made on p. 7: using the fact that any irrational has a continued fraction expansion, it follows that for any real number u, there is a sequence (q_n) with $q_n \to \infty$ such that $q_n \langle q_n u \rangle < 1$.

---

* This can also be seen more directly as a consequence of the Dirichlet principle (see Exercise 3.1.3).

Lemma 3.2. *Let $a_n \in \mathbb{N}$ for all $n \geqslant 0$. Then the limit in equation* (3.10) *is irrational.*

PROOF. Suppose that $u = \frac{a}{b} \in \mathbb{Q}$. Then, by equation (3.14),

$$|q_n a - b p_n| < \frac{b}{a_{n+1} q_n} \leqslant \frac{b}{q_n}.$$

Since $q_n \to \infty$ by the inequality (3.6) and $q_n a - b p_n \in \mathbb{Z}$ we see that

$$q_n a - b p_n = 0$$

and hence $u = \frac{a}{b} = \frac{p_n}{q_n}$ for large enough n. However, by Lemma 3.1 p_n and q_n are coprime, so this contradicts the fact that $q_n \to \infty$ as $n \to \infty$. Thus u is irrational. □

The continued fraction convergents to a given irrational not only provide good rational approximants. In fact, they provide *optimal* rational approximants in the following sense (see Exercise 3.1.4).

Proposition 3.3. *Let $u = [a_0; a_1, \dots] \in \mathbb{R} \smallsetminus \mathbb{Q}$ as in equation* (3.10)*. For any $n > 1$ and p, q with $0 < q \leqslant q_n$, if $\frac{p}{q} \neq \frac{p_n}{q_n}$, then*

$$|p_n - q_n u| < |p - qu|.$$

In particular,

$$\left| \frac{p_n}{q_n} - u \right| < \left| \frac{p}{q} - u \right|.$$

PROOF. Note that $|p_n - q_n u| < |p - qu|$ and $0 < q \leqslant q_n$ together imply that

$$\frac{1}{q} \left| \frac{p_n}{q_n} - u \right| < \frac{1}{q_n} \left| \frac{p}{q} - u \right| \leqslant \frac{1}{q} \left| \frac{p}{q} - u \right|,$$

giving the second statement of the proposition. It is enough therefore to prove the first inequality. Recall from equation (3.13) that

$$\left| u - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}$$

and

$$\left| u - \frac{p_{n+1}}{q_{n+1}} \right| < \frac{1}{q_{n+1} q_{n+2}}.$$

By the alternating behavior of the convergents in equation (3.11), each of the three bracketed expressions in the identity

$$\left(u - \frac{p_n}{q_n} \right) = \left(\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right) - \left(\frac{p_{n+1}}{q_{n+1}} - u \right)$$

is positive (if n is even) or negative (if n is odd). It follows that

$$\left|u - \frac{p_n}{q_n}\right| = \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| - \left|\frac{p_{n+1}}{q_{n+1}} - u\right|,$$

so

$$\left|u - \frac{p_n}{q_n}\right| > \frac{1}{q_n q_{n+1}} - \frac{1}{q_{n+1} q_{n+2}} = \frac{q_{n+2} - q_n}{q_n q_{n+1} q_{n+2}} = \frac{a_{n+2}}{q_n q_{n+2}}$$

by equations (3.4) and (3.14). It follows that

$$\frac{1}{q_{n+2}} < |p_n - q_n u| < \frac{1}{q_{n+1}} \tag{3.15}$$

for $n \geqslant 1$.

By the inequalities (3.15),

$$|q_n u - p_n| < \frac{1}{q_{n+1}} < |q_{n-1} u - p_{n-1}|$$

so we may assume that $q_{n-1} < q \leqslant q_n$ (if not, use downwards induction on n).

If $q = q_n$, then $|\frac{p_n}{q_n} - \frac{p}{q}| \geqslant \frac{1}{q_n}$, while

$$\left|\frac{p_n}{q_n} - u\right| < \frac{1}{q_n q_{n+1}} \leqslant \frac{1}{2q_n},$$

since $q_{n+1} \geqslant 2$ for all $n \geqslant 1$. Therefore,

$$\left|\frac{p}{q} - u\right| \geqslant \frac{1}{2q_n} = \frac{1}{2q}$$

and so $|q_n u - p_n| < |qu - p|$.

Assume now that $q_{n-1} < q < q_n$ and write

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix},$$

so that $a, b \in \mathbb{Z}$ by equation (3.8). Clearly $ab \neq 0$ since otherwise $q = q_{n-1}$ or $q = q_n$. Now $q = aq_n + bq_{n-1} < q_n$, so $ab < 0$; by equation (3.11) we also know that $p_n - q_n u$ and $p_{n-1} - q_{n-1} u$ are of opposite signs. It follows that $a(p_n - q_n u)$ and $b(p_{n-1} - q_{n-1} u)$ are of the same sign, so the fact that

$$p - qu = a(p_n - q_n u) + b(p_{n-1} - q_{n-1} u)$$

implies that

$$|p - qu| > |p_{n-1} - q_{n-1} u| > |p_n - q_n u|$$

as required. $\square$

We end this section with the uniqueness of the continued fraction expansion.

Lemma 3.4. *The map that sends the sequence*

$$(a_0, a_1, \dots) \in \mathbb{N}_0 \times \mathbb{N}^{\mathbb{N}}$$

to the limit in equation (3.10) *is injective.*

Proof. Let $u = (a_0, a_1, \dots) \in \mathbb{N}_0 \times \mathbb{N}^{\mathbb{N}}$ be given. Then it is clear that

$$u = [a_0; a_1, \dots]$$

is positive. Applying this to $(a_1, a_2, \dots)$ and the inductive relation

$$u = a_0 + \frac{1}{[a_1; a_2, \dots]}$$

we see that

$$u \in (a_0, a_0 + \tfrac{1}{a_1}) \subseteq (a_0, a_0 + 1).$$

It follows that u uniquely determines a_0. Using the inductive relation again, we have

$$[a_1; a_2, \dots] = \frac{1}{u - a_0},$$

which by the argument above shows that u uniquely determines a_1. Iterating the procedure shows that all the terms in the continued fraction can be reconstructed from u. □

The argument used in the proof of Lemma 3.4 also suggests a way to find the continued fraction expansion of a given irrational number $u \in \mathbb{R} \smallsetminus \mathbb{Q}$. This will be pursued further in the next section.

Exercises for Section 3.1

Exercise 3.1.1. Show that any positive rational number has exactly two continued fraction expansions, both of which are finite.

Exercise 3.1.2. Show that a continued fraction in which some of the digits are allowed to be zero (but that is not allowed to end with infinitely many zeros) can always be rewritten with digits in $\mathbb{N}$.

Exercise 3.1.3. [Dirichlet principle] For a given $u \in \mathbb{R}$ and $n \geqslant 1$ consider the points $0, u, 2u, \dots, nu$ (mod 1) as elements of the circle $\mathbb{T}$. Show that for some k, $0 < k < n$ we have $\langle ku \rangle \leqslant \frac{1}{n}$, and deduce that there exists a sequence $q_n \to \infty$ with $q_n \langle q_n u \rangle < 1$.

Exercise 3.1.4. Extend Proposition 3.3 in the following way. Given u as in equation (3.10), and the nth convergent $\frac{p_n}{q_n}$, the $(n+1)$th convergent $\frac{p_{n+1}}{q_{n+1}}$ is characterized by being the ratio of the unique pair of positive integers (p_{n+1}, q_{n+1}) for which $|p_{n+1} - q_{n+1}u| < |p_n - q_n u|$ with $q_{n+1} > q_n$ minimal. Notice that the same cannot be said when using the expression $\left|u - \frac{p_n}{q_n}\right|$, as becomes clear in the case where $u > \frac{1}{3}$ is very close to $\frac{1}{3}$, in which case the first approximation is not $\frac{1}{2}$.

Exercise 3.1.5. Let $u = [a_0; a_1, \dots]$ with convergents $\frac{p_n}{q_n}$. Show that

$$\frac{1}{2q_{n+1}} \leqslant |p_n - q_n u| < \frac{1}{q_{n+1}}.$$

3.2 The Continued Fraction Map and the Gauss Measure

Let $Y = [0,1] \smallsetminus \mathbb{Q}$, and define a map $T : Y \to Y$ by

$$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor,$$

where $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t. Thus $T(x)$ is the fractional part $\left\{\frac{1}{x}\right\}$ of $\frac{1}{x}$. The graph of this so-called *continued fraction* or *Gauss map* is shown in Figure 3.1.

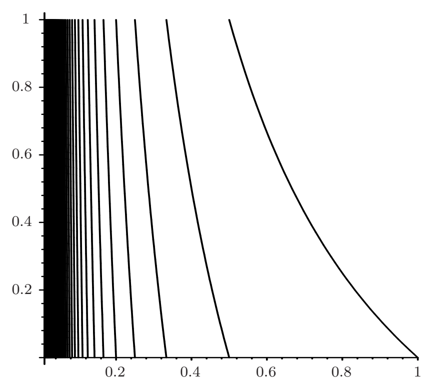


Fig. 3.1: The Gauss map.

Gauss observed in 1845 that T preserves[(40)] the probability measure given by

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x}\,\mathrm{d}x,$$

by showing that the Lebesgue measure of $T^{-n}I$ converges to $\mu(I)$ for each interval I.

This map will be studied via a geometric model (for its invertible extension) in Chapter 9; in this section we assemble some basic facts from an elementary point of view, showing that the Gauss measure is T-invariant and ergodic. Since the measure defined in Lemma 3.5 is non-atomic, we may extend the map to include the points 0 and 1 in any way without affecting the measurable structure of the system.

Lemma 3.5. *The continued fraction map* $T(x) = \left\{\frac{1}{x}\right\}$ *on* $(0,1)$ *preserves the Gauss measure* μ *given by*

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x}\,\mathrm{d}x$$

for any Borel measurable set $A \subseteq [0,1]$.

A geometric and less formal proof of this will be given on page 93 using basic properties of the invertible extension of the continued fraction map in Proposition 3.15.

Proof of Lemma 3.5. It is sufficient to show that $\mu\left(T^{-1}[0,s]\right) = \mu\left([0,s]\right)$ for every $s > 0$. Clearly

$$T^{-1}[0,s] = \{x \mid 0 \leqslant T(x) \leqslant s\} = \bigsqcup_{n=1}^{\infty} \left[\frac{1}{s+n}, \frac{1}{n}\right]$$

is a disjoint union. It follows that

$$\begin{aligned}
\mu\left(T^{-1}[0,s]\right) &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{1/(s+n)}^{1/n} \frac{1}{1+x}\,\mathrm{d}x \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log(1+\tfrac{1}{n}) - \log(1+\tfrac{1}{s+n})\right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log(1+\tfrac{s}{n}) - \log(1+\tfrac{s}{n+1})\right) \qquad (3.16) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{s/(n+1)}^{s/n} \frac{1}{1+x}\,\mathrm{d}x \\
&= \mu\left([0,s]\right),
\end{aligned}$$

completing the proof. The identity used in equation (3.16) amounts to

$$\frac{1+\frac{s}{n}}{1+\frac{s}{n+1}} = \frac{1+\frac{1}{n}}{1+\frac{1}{s+n}},$$

which may be seen by multiplying numerator and denominator of the left-hand side by $\frac{n+1}{n+s}$, and the interchange of integral and sum is justified by absolute convergence. $\square$

Thus Lemma 3.5 shows that $\big([0,1], \mathscr{B}_{[0,1]}, \mu, T\big)$ is a measure-preserving system.

Define for $x \in Y = [0,1]\smallsetminus\mathbb{Q}$ and $n \geqslant 1$ the sequence of natural numbers $(a_n) = (a_n(x))$ by

$$\frac{1}{1+a_n} < T^{n-1}(x) < \frac{1}{a_n}, \tag{3.17}$$

or equivalently by

$$a_n(x) = \left\lfloor \frac{1}{T^{n-1}x} \right\rfloor \in \mathbb{N}. \tag{3.18}$$

For any sequence $(a_n)_{n\geqslant 1}$ of natural numbers we define the continued fraction $[a_1, a_2, \dots]$ just as in equation (3.1) with $a_0 = 0$.

Lemma 3.6. *For any irrational $x \in [0,1]\smallsetminus\mathbb{Q}$ the sequence $(a_n(x))$ defined in equation* (3.18) *gives the digits of the continued fraction expansion to x. That is,*

$$x = [a_1(x), a_2(x), \dots].$$

Proof. Define $a_n = a_n(x)$ and let $u = [a_1, a_2, \dots]$ be the limit as in equation (3.10) with $a_0 = 0$. By equation (3.11) we have

$$\frac{p_{2n}}{q_{2n}} < u < \frac{p_{2n+1}}{q_{2n+1}}$$

and by equation (3.8) and the inequality (3.6) we have

$$\frac{p_{2n+1}}{q_{2n+1}} - \frac{p_{2n}}{q_{2n}} = \frac{1}{q_{2n}q_{2n-1}} \leqslant \frac{1}{2^{2n-2}}.$$

We now show by induction that

$$[a_1, \dots, a_{2n}] = \frac{p_{2n}}{q_{2n}} < x < \frac{p_{2n+1}}{q_{2n+1}} = [a_1, \dots, a_{2n+1}], \tag{3.19}$$

which together with the above shows that $u = x$.

Recall that $\frac{p_0}{q_0} = 0$ and $\frac{p_1}{q_1} = \frac{1}{a_1}$, so equation (3.19) holds for $n = 0$ because of the definition of a_1 in equation (3.18). Now assume that the inequality (3.19) holds for a given n and all $x \in [0,1]$. In particular, we may apply it to $T(x)$ to get

$$[a_2, \dots, a_{2n+1}] < T(x) < [a_2, \dots, a_{2n+2}].$$

Since $T(x) = \frac{1}{x} - a_1$ we get

$$a_1 + [a_2, \dots, a_{2n+1}] < \frac{1}{x} < a_1 + [a_2, \dots, a_{2n+2}]$$

and therefore

$$[a_1, \dots, a_{2n+2}] = \frac{1}{a_1 + [a_2, \dots, a_{2n+2}]} < x,$$

$$x < \frac{1}{a_1 + [a_2, \dots, a_{2n+1}]} = [a_1, \dots, a_{2n+1}]$$

as required. $\square$

This gives a description of the continued fraction map as a shift map: the list of digits in the continued fraction expansion of $x \in [0,1]\smallsetminus\mathbb{Q}$ defines a unique element of $\mathbb{N}^{\mathbb{N}}$, and the diagram

$$\begin{array}{ccc} \mathbb{N}^{\mathbb{N}} & \xrightarrow{\sigma} & \mathbb{N}^{\mathbb{N}} \\ \downarrow & & \downarrow \\ (0,1) & \xrightarrow[T]{} & (0,1) \end{array}$$

commutes, where σ is the left shift and the vertical map sends a sequence of digits $(a_n)_{n\geqslant 1}$ to the real irrational number defined by the continued fraction expansion.

In Corollary 3.8 we will draw some easy consequences[(41)] of ergodicity for the Gauss measure μ in terms of properties of the continued fraction expansion for almost every real number. Given a continued fraction expansion, recall that the *convergents* are the terms of the sequence of rationals $\frac{p_n(x)}{q_n(x)}$ in lowest terms defined by

$$\frac{p_n(x)}{q_n(x)} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_n}}}}.$$

Theorem 3.7. *The continued fraction map $T(x) = \{\frac{1}{x}\}$ on $(0,1)$ is ergodic with respect to the Gauss measure μ.*

Before proving this[(42)] we develop some more of the basic identities for continued fractions. Given a continued fraction expansion $u = [a_0; a_1, \dots]$ of an irrational number u, we write $u_n = [a_n; a_{n+1}, \dots]$ for the nth tail of the expansion. By Lemma 3.1 applied twice, we have

$$\begin{pmatrix} p_{n+k} \\ q_{n+k} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n+k} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} a_{n+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n+k} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Writing $p_k(u_{n+1})$ and $q_k(u_{n+1})$ for the numerator and denominator of the kth convergents to u_{n+1}, we can apply Lemma 3.1 again to deduce that

$$\begin{pmatrix} p_{n+k} \\ q_{n+k} \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} p_{k-1}(u_{n+1}) & p_{k-2}(u_{n+1}) \\ q_{k-1}(u_{n+1}) & q_{k-2}(u_{n+1}) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

so

$$\frac{p_{n+k}}{q_{n+k}} = \frac{p_n \frac{p_{k-1}(u_{n+1})}{q_{k-1}(u_{n+1})} + p_{n-1}}{q_n \frac{p_{k-1}(u_{n+1})}{q_{k-1}(u_{n+1})} + q_{n-1}},$$

which gives

$$u = \frac{p_n u_{n+1} + p_{n-1}}{q_n u_{n+1} + q_{n-1}} \tag{3.20}$$

in the limit as $k \to \infty$. Notice that the above formulas are derived for a general positive irrational number u. If $u = [a_1, \dots] \in (0,1)$, then $u_{n+1} = (T^n(u))^{-1}$ so that

$$u = \frac{p_n + p_{n-1}T^n(u)}{q_n + q_{n-1}T^n(u)}. \tag{3.21}$$

Proof of Theorem 3.7. The description of the continued fraction map as a shift on the space $\mathbb{N}^{\mathbb{N}}$ described above suggests the method of proof: the measure μ corresponds to a rather complicated measure on the shift space, but if we can control the measure of cylinder sets (and their intersections) well enough then we may prove ergodicity along the lines of the proof of ergodicity for Bernoulli shifts in Proposition 2.15. For two expressions f, g we write $f \asymp g$ to mean that there are absolute constants $C_1, C_2 > 0$ such that

$$C_1 f \leqslant g \leqslant C_2 f.$$

Given a vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ of length $|\mathbf{a}| = n$, define a set

$$I(\mathbf{a}) = \{[x_1, x_2, \dots] \mid x_i = a_i \text{ for } 1 \leqslant i \leqslant n\}$$

(which may be thought of as an interval in $(0,1)$, or as a cylinder set in $\mathbb{N}^{\mathbb{N}}$).

The main step towards the proof of the theorem is to show that

$$\mu\left(T^{-n}A \cap I(\mathbf{a})\right) \asymp \mu(A)\mu(I(\mathbf{a})) \tag{3.22}$$

for any measurable set A. Notice that for the proof of equation (3.22) it is sufficient to show it for any interval $A = [d, e]$; the case of a general Borel set then follows by a standard approximation argument (the set of Borel sets satisfying equation (3.22) with a fixed choice of constants is easily seen to be a monotone class, so Theorem A.4 may be applied.)

Now define $\frac{p_n}{q_n} = [a_1, \dots, a_n]$ and $\frac{p_{n-1}}{q_{n-1}} = [a_1, \dots, a_{n-1}]$. Then $u \in I(\mathbf{a})$ if and only if $u = [a_1, \dots, a_n, a_{n+1}(u), \dots]$, and so $u \in I(\mathbf{a}) \cap T^{-n}A$ if and only if u can be written as in equation (3.21), with $T^n(u) \in A = [d, e]$. As T^n restricted to $I(\mathbf{a})$ is continuous and monotone (increasing if n is even, and decreasing if n is odd), it follows that $I(\mathbf{a}) \cap T^{-n}A$ is an interval with endpoints given by

$$\frac{p_n + p_{n-1}d}{q_n + q_{n-1}d}$$

and

$$\frac{p_n + p_{n-1}e}{q_n + q_{n-1}e}.$$

Thus the Lebesgue measure of $I(\mathbf{a}) \cap T^{-n}A$,

$$\left| \frac{p_n + p_{n-1}d}{q_n + q_{n-1}d} - \frac{p_n + p_{n-1}e}{q_n + q_{n-1}e} \right|,$$

expands to

$$\begin{aligned} &\left| \frac{(p_n + p_{n-1}d)(q_n + q_{n-1}e) - (p_n + p_{n-1}e)(q_n + q_{n-1}d)}{(q_n + q_{n-1}d)(q_n + q_{n-1}e)} \right| \\ &= \left| \frac{p_n q_{n-1} e + p_{n-1} q_n d - p_n q_{n-1} d - p_{n-1} q_n e}{(q_n + q_{n-1}d)(q_n + q_{n-1}e)} \right| \\ &= (e-d) \frac{|p_n q_{n-1} - p_{n-1} q_n|}{(q_n + q_{n-1}e)(q_n + q_{n-1}f)} = (e-d) \frac{1}{(q_n + q_{n-1}e)(q_n + q_{n-1}f)} \end{aligned}$$

by equation (3.8). On the other hand, the Lebesgue measure of $I(\mathbf{a})$ is

$$\left| \frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}} \right| = \frac{|p_n q_{n-1} - p_{n-1} q_n|}{q_n(q_n + q_{n-1})} = \frac{1}{q_n(q_n + q_{n-1})} \tag{3.23}$$

again by equation (3.8), which implies that

$$\begin{aligned} m(I(\mathbf{a}) \cap T^{-n}A) &= m(A) m(I(\mathbf{a})) \frac{q_n(q_n + q_{n-1})}{(q_n + q_{n-1}e)(q_n + q_{n-1}f)} \\ &\asymp m(A) m(I(\mathbf{a})), \end{aligned} \tag{3.24}$$

where m denotes Lebesgue measure on $(0, 1)$. Next notice that

$$\frac{m(B)}{2 \log 2} \leqslant \mu(B) \leqslant \frac{m(B)}{\log 2}$$

for any Borel set $B \subseteq (0, 1)$, which together with equation (3.24) gives equation (3.22).

Now assume that $A \subseteq (0,1)$ is a Borel set with $T^{-1}A = A$. For such a set, the estimate in equation (3.22) reads as

$$\mu(A \cap I(\mathbf{a})) \asymp \mu(A)\mu(I(\mathbf{a}))$$

for any interval $I(\mathbf{a})$ defined by $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ and any n. However, for a fixed n the intervals $I(\mathbf{a})$ partition $(0,1)$ (as $\mathbf{a}$ varies in $\mathbb{N}^n$), and by equation (3.23)

$$\begin{aligned}\operatorname{diam}(I(\mathbf{a})) &= \frac{1}{q_n(q_n + q_{n-1})} \\ &\leqslant \frac{1}{2^{n-2}} \qquad \text{(by (3.6))},\end{aligned}$$

so the lengths of the sets in this partition shrink to zero uniformly as $n \to \infty$. Therefore, the intervals $I(\mathbf{a})$ generate the Borel σ-algebra, and so

$$\mu(A \cap B) \asymp \mu(A)\mu(B)$$

for any Borel subset $B \subseteq (0,1)$ (again by Theorem A.4). We apply this to the set $B = (0,1) \smallsetminus A$ and obtain $0 \asymp \mu(A)\mu(B)$, which shows that either $\mu(A) = 0$ or $\mu((0,1) \smallsetminus A) = 0$, as needed. □

We will use the ergodicity of the Gauss map in Corollary 3.8 to deduce statements about the digits of the continued fraction expansion of a typical real number. Just as Borel's normal number theorem (Example 1.2) gives precise statistical information about the decimal expansion of almost every real number, ergodicity of the Gauss map gives precise statistical information about the continued fraction digits of almost every real number. Of course the form of the conclusion is necessarily different. For example, since there are infinitely many different digits in the continued fraction expansion, they cannot all occur with equal frequency, and equation (3.25) makes precise the way in which small digits occur more frequently than large ones. We also obtain information on the geometric and arithmetic mean of the digits a_n in equations (3.26) and (3.27), the growth rate of the denominators q_n in equation (3.28), and the rate at which the convergents $\frac{p_n}{q_n}$ approximate a typical real number in equation (3.29).

In particular, equations (3.28) and (3.29) together say that the digit a_{n+1} appearing in the estimate (3.14) does not affect the logarithmic rate of approximation of an irrational by the continued fraction partial quotients significantly.

Corollary 3.8. *For almost every real number $x = [a_1, a_2, \dots] \in (0,1)$, the digit j appears in the continued fraction with density*

$$\frac{2\log(1+j) - \log j - \log(2+j)}{\log 2}, \tag{3.25}$$

$$\lim_{n\to\infty} (a_1 a_2 \dots a_n)^{1/n} = \prod_{a=1}^{\infty} \left(\frac{(a+1)^2}{a(a+2)} \right)^{\log a / \log 2}, \tag{3.26}$$

$$\lim_{n\to\infty} \frac{1}{n} (a_1 + a_2 + \dots + a_n) = \infty, \tag{3.27}$$

$$\lim_{n\to\infty} \frac{1}{n} \log q_n(x) = \frac{\pi^2}{12 \log 2}, \tag{3.28}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \left| x - \frac{p_n(x)}{q_n(x)} \right| \longrightarrow -\frac{\pi^2}{6 \log 2}. \tag{3.29}$$

Proof. The digit j appears in the first N digits with frequency

$$\begin{aligned} \frac{1}{N} |\{i \mid i \leqslant N, a_i = j\}| &= \frac{1}{N} |\{i \mid i \leqslant N, T^i x \in (\tfrac{1}{j+1}, \tfrac{1}{j})\}| \\ &\to \frac{1}{\log 2} \int_{1/(j+1)}^{1/j} \frac{1}{1+y} \,\mathrm{d}y \\ &= \frac{2\log(1+j) - \log j - \log(2+j)}{\log 2}, \end{aligned}$$

which proves equation (3.25).

Define a function f on $(0,1)$ by $f(x) = \log a$ for $x \in \left(\frac{1}{a+1}, \frac{1}{a}\right)$. Then

$$\begin{aligned} \int_0^1 f(x) \,\mathrm{d}x &= \sum_{a=1}^{\infty} \left(\frac{1}{a} - \frac{1}{a+1} \right) \log a \\ &\leqslant \sum_{a=1}^{\infty} \frac{1}{a^2} \log a < \infty, \end{aligned}$$

so $\int_0^1 f \,\mathrm{d}\mu < \infty$ also, since the density $\frac{\mathrm{d}\mu}{\mathrm{d}x} = \frac{1}{(1+x)\log 2}$ is bounded on $[0,1]$. By the pointwise ergodic theorem (Theorem 2.30) we therefore have, for almost every x,

$$\frac{1}{n} \sum_{j=0}^{n-1} \log a_j = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \longrightarrow \int f(x) \,\mathrm{d}\mu.$$

This shows equation (3.26) since

$$\int_0^1 f \,\mathrm{d}\mu = \sum_{a=1}^{\infty} \frac{\log a}{\log 2} \int_{1/(1+a)}^{1/a} \frac{1}{1+x} \,\mathrm{d}x.$$

Now consider the function $g(x) = \mathrm{e}^{f(x)}$ (so $g(x) = a_1$ is the first digit in the continued fraction expansion of x). We have

$$\frac{1}{n}\left(a_1+\cdots+a_n\right)=\frac{1}{n}\sum_{j=0}^{n-1}g(T^jx),$$

but the pointwise ergodic theorem cannot be applied to g since $\int_0^1 g\,\mathrm{d}\mu=\infty$ (the result needed is Exercise 2.6.5(2); the argument here shows how to do this exercise). However, for any fixed N the truncated function

$$g_N(x)=\begin{cases} g(x) & \text{if } g(x)\leqslant N;\\ 0 & \text{if not}\end{cases}$$

is in L^1_μ since

$$\int g_N\,\mathrm{d}\mu=\frac{1}{\log 2}\sum_{a=1}^{N}\int_{1/(a+1)}^{1/a}a\,\mathrm{d}x=\frac{1}{\log 2}\sum_{a=1}^{N}\frac{1}{a+1}.$$

Notice that $\int_0^1 g_N\,\mathrm{d}\mu\to\infty$ as $N\to\infty$. By the ergodic theorem,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}g(T^jx)&\geqslant\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}g_N(T^jx)\\ &=\int_0^1 g_N\,\mathrm{d}\mu\to\infty\end{aligned}$$

as $N\to\infty$, showing equation (3.27).

The proofs of (3.25) and (3.26) were straightforward applications of the ergodic theorem, and (3.27) only required a simple extension to measurable functions. Proving (3.28) and (3.29) takes a little more effort.

First notice that

$$\begin{aligned}\frac{p_n(x)}{q_n(x)}&=\frac{1}{a_1+[a_2,\dots,a_n]}\\ &=\frac{1}{a_1+\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}}\\ &=\frac{q_{n-1}(Tx)}{p_{n-1}(Tx)+q_{n-1}(Tx)a_1},\end{aligned}$$

so $p_n(x)=q_{n-1}(Tx)$ since the convergents are in lowest terms. Recall that we always have $p_1=q_0=1$. It follows that

$$\frac{1}{q_n(x)}=\frac{p_n(x)}{q_n(x)}\cdot\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}\cdots\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)},$$

so

$$-\frac{1}{n}\log q_n(x) = \frac{1}{n}\sum_{j=0}^{n-1}\log\left[\frac{p_{n-j}(T^jx)}{q_{n-j}(T^jx)}\right].$$

Let $h(x) = \log x$ (so $h \in L^1_\mu$). Then

$$-\frac{1}{n}\log q_n(x) = \underbrace{\frac{1}{n}\sum_{j=0}^{n-1}h(T^jx)}_{S_n} - \underbrace{\frac{1}{n}\sum_{j=0}^{n-1}\left[\log(T^jx) - \log\left(\frac{p_{n-j}(T^jx)}{q_{n-j}(T^jx)}\right)\right]}_{R_n}$$

gives a splitting of $-\frac{1}{n}\log q_n(x)$ into an ergodic average $S_n = \mathsf{A}_h^n$ and a remainder term R_n. By the ergodic theorem,

$$\lim_{n\to\infty}\frac{1}{n}S_n = \frac{1}{\log 2}\int_0^1 \frac{\log x}{1+x}\,\mathrm{d}x = -\frac{\pi^2}{12\log 2}.$$

To complete the proof of equation (3.28), we need to show that $\frac{1}{n}R_n \to 0$ as $n \to \infty$. This will follow from the observation that $\frac{p_{n-j}(T^jx)}{q_{n-j}(T^jx)}$ is a good approximation to T^jx if $(n-j)$ is large enough. Recall from equations (3.7) and (3.6) that

$$p_k \geqslant 2^{(k-2)/2},\ q_k \geqslant 2^{(k-1)/2},$$

so, by using the inequality (3.13),

$$\left|\frac{x}{p_k/q_k} - 1\right| = \frac{q_k}{p_k}\left|x - \frac{p_k}{q_k}\right| \leqslant \frac{1}{p_kq_{k+1}} \leqslant \frac{1}{2^{k-1}}.$$

By using this together with the fact that $|\log u| \leqslant 2|u-1|$ whenever $u \in [\frac{1}{2}, \frac{3}{2}]$ (which applies in the sum below with $j \leqslant n-2$), we get

$$\begin{aligned}|R_n| &\leqslant \sum_{j=0}^{n-1}\left|\log\frac{T^jx}{p_{n-j}(T^jx)/q_{n-j}(T^jx)}\right|\\ &\leqslant \underbrace{2\sum_{j=0}^{n-2}\left|\frac{T^jx}{p_{n-j}(T^jx)/q_{n-j}(T^jx)} - 1\right|}_{T_n} + \underbrace{\left|\log\frac{T^{n-1}x}{p_1(T^{n-1}x)/q_1(T^{n-1}x)}\right|}_{U_n}.\end{aligned}$$

Now

$$T_n \leqslant \sum_{j=0}^{n-2}\frac{2}{2^{n-j-1}} \leqslant 2$$

for all n. For the second term, notice that

$$U_n = \left|\log\left[\left(T^{n-1}x\right)a_1\left(T^{n-1}x\right)\right]\right|,$$

and by the inequality (3.17) we have

$$1 \geqslant \left(T^{n-1}x\right) a_1 \left(T^{n-1}x\right) \geqslant \frac{a_1 \left(T^{n-1}x\right)}{1 + a_1 \left(T^{n-1}x\right)} \geqslant \frac{1}{2}$$

since $a_1(T^{n-1}x) \geqslant 1$. Therefore,

$$\left| \log \left[\left(T^{n-1}x\right) a_1 \left(T^{n-1}x\right) \right] \right| \leqslant \log 2,$$

which completes the proof that

$$\frac{1}{n} R_n \to 0$$

as $n \to \infty$, and hence shows equation (3.28).

Equation (3.29) follows from equation (3.28), since from the inequalities (3.13) and (3.15) we have

$$\log q_n + \log q_{n+1} \leqslant - \log \left| x - \frac{p_n}{q_n} \right| \leqslant \log q_n + \log q_{n+2}.$$

□

Exercises for Section 3.2

Exercise 3.2.1. Use the idea in the proof of equation (3.27) to extend the pointwise ergodic theorem (Theorem 2.30) to the case of a measurable function $f \geqslant 0$ with $\int_X f \, \mathrm{d}\mu = \infty$ without the assumption of ergodicity.

Exercise 3.2.2. Show that the map from $\mathbb{N}^{\mathbb{N}}$ to $[0,1] \smallsetminus \mathbb{Q}$ sending $(a_1, a_2, \dots)$ to $[a_1, a_2, \dots]$ is a homeomorphism with respect to the discrete topology on $\mathbb{N}$ and the product topology on $\mathbb{N}^{\mathbb{N}}$.

Exercise 3.2.3. Let $\mathbf{p} = (p_1, p_2, \dots)$ be an infinite probability vector (this means that $p_i \geqslant 0$ for all i, and $\sum_i p_i = 1$). Show that $\mathbf{p}$ gives rise to a σ-invariant and ergodic probability measure $\mathbf{p}^{\mathbb{N}}$ on $\mathbb{N}^{\mathbb{N}}$.

Exercise 3.2.4. Let $\phi : \mathbb{N}^{\mathbb{N}} \to (0,1) \smallsetminus \mathbb{Q}$ be the map discussed on page 79, and let μ be the Gauss measure on $[0,1]$. Show that $\phi_*^{-1}\mu$ is not of the form $\mathbf{p}^{\mathbb{N}}$ for any infinite probability vector $\mathbf{p}$.

3.3 Badly Approximable Numbers

While Corollary 3.8 gives precise information about the behavior of typical real numbers, it does not say anything about the behavior of all real numbers. In this section we discuss a special class of real numbers that behave very differently to typical real numbers.

Definition 3.9. A real number $u = [a_1, a_2, \dots] \in (0,1)$ is called *badly approximable* if there is some bound M with the property that $a_n \leqslant M$ for all $n \geqslant 1$.

Clearly a badly approximable number cannot satisfy equation (3.27). It follows that the set of all badly approximable numbers in $(0,1)$ is a null set with respect to the Gauss measure, and hence is a null set with respect to Lebesgue measure[(43)]. The next result explains the terminology: badly approximable numbers cannot be approximated very well by rationals.

Proposition 3.10. *A number $u \in (0,1)$ is badly approximable if and only if there exists some $\varepsilon > 0$ with the property that*

$$\left| u - \frac{p}{q} \right| \geqslant \frac{\varepsilon}{q^2}$$

for all rational numbers $\frac{p}{q}$.

Proof. If u is badly approximable, then equation (3.4) shows that

$$q_{n+1} \leqslant (M+1) q_n$$

for all $n \geqslant 0$. For any q there is some n with $q \in (q_{n-1}, q_n]$, and by Proposition 3.3 and equation (3.15) we therefore have

$$\left| \frac{p}{q} - u \right| > \left| \frac{p_n}{q_n} - u \right| > \frac{1}{q_n q_{n+2}} > \frac{1}{(M+1)^4 q^2}$$

as required.

Conversely, if

$$\left| u - \frac{p}{q} \right| \geqslant \frac{\varepsilon}{q^2}$$

for all rational numbers $\frac{p}{q}$ then, in particular,

$$\frac{\varepsilon}{q_n^2} \leqslant \left| u - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}$$

by equation (3.13). This implies that

$$a_{n+1} q_n < a_{n+1} q_n + q_{n-1} = q_{n+1} < \frac{1}{\varepsilon} q_n,$$

so $a_{n+1} \leqslant \frac{1}{\varepsilon}$ for all $n \geqslant 1$. □

Example 3.11. Notice that $\frac{2}{\sqrt{5}-1} = \frac{\sqrt{5}+1}{2} \in (1,2)$ and $\frac{\sqrt{5}+1}{2} - 1 = \frac{\sqrt{5}-1}{2}$. It follows that if

$$\frac{\sqrt{5}-1}{2} = [a_1, a_2, \dots]$$

then $a_1 + [a_2, a_3, \dots] \in (1,2)$, so $a_1 = 1$, and hence

$$[a_2, a_3, \dots] = \frac{\sqrt{5}+1}{2} - 1 = \frac{\sqrt{5}-1}{2} = [a_1, a_2, \dots].$$

We deduce by the uniqueness of the continued fraction digits that

$$\frac{\sqrt{5}-1}{2} = [1, 1, 1, \dots],$$

so $\frac{\sqrt{5}-1}{2}$ is badly approximable.

Indeed, the specific number in Example 3.11 is, in a precise sense, the most badly approximable real number in $(0,1)$. In the next section we generalize this example to show that all quadratic irrationals are badly approximable.

3.3.1 Lagrange's Theorem

The periodicity of the continued fraction expansion seen in Example 3.11 is a general property of quadratics. A real number u is called a *quadratic irrational* if $u \notin \mathbb{Q}$ and there are integers a, b, c with $au^2 + bu + c = 0$. Notice that u is a quadratic irrational if and only if $\mathbb{Q}(u)$ is a subfield of $\mathbb{R}$ of degree 2 over $\mathbb{Q}$.

Definition 3.12. A continued fraction $[a_0; a_1, \dots]$ is *eventually periodic* if there are numbers $N \geqslant 0$ and $k \geqslant 1$ with $a_{n+k} = a_n$ for all $n \geqslant N$. Such a continued fraction will be written

$$[a_0; a_1, \dots, a_{N-1}, \overline{a_N, \dots, a_{N+k}}].$$

The main result describing the special properties of quadratic irrationals is Lagrange's Theorem [218, Sect. 34].

Theorem 3.13 (Lagrange). *Let u be an irrational positive real number. Then the continued fraction expansion of u is eventually periodic if and only if u is a quadratic irrational.*

Proof. Assume first that $u = [\overline{a_0; a_1, \dots, a_k}]$ has a strictly periodic continued fraction expansion, so that $u_{k+1} = u_0 = u$. Thus

$$u = \frac{up_k + p_{k-1}}{uq_k + q_{k-1}}$$

by equation (3.20), so

$$u^2 q_k + u(q_{k-1} - p_k) - p_{k-1} = 0$$

and u is a quadratic irrational (u cannot be rational, since it has an infinite continued fraction; alternatively notice that the quadratic equation satisfied by u has discriminant $(q_{k-1} - p_k)^2 + 4q_k p_{k-1} = (q_{k-1} + p_k)^2 - 4(-1)^k$ by equation (3.8), so cannot be a square).

Now assume that

$$u = [a_0; \dots, a_{N-1}, \overline{a_N, \dots, a_{N+k}}].$$

Then, by equation (3.20),

$$u = \frac{[\overline{a_N; a_{N+1}, \dots, a_{N+k}}]p_{N-1} + p_{N-2}}{[\overline{a_N; a_{N+1}, \dots, a_{N+k}}]q_{N-1} + q_{N-2}},$$

so $\mathbb{Q}(u) = \mathbb{Q}([\overline{a_N; a_{N+1}, \dots, a_{N+k}}])$, and therefore u is a quadratic irrational.

The converse is more involved[(44)]. Assume now that u is a quadratic irrational, with

$$f_0(u) = \alpha_0 u^2 + \beta_0 u + \gamma_0 = 0$$

for some $\alpha_0, \beta_0, \gamma_0 \in \mathbb{Z}$ and $\delta = \beta_0^2 - 4\alpha_0\gamma_0$ not a square. We claim that for each $n \geqslant 0$ there is a polynomial

$$f_n(x) = \alpha_n x^2 + \beta_n + \gamma_n$$

with

$$\beta_n^2 - 4\alpha_n\gamma_n = \delta$$

and with the property that $f_n(u_n) = 0$. This claim again follows from the fact that $\mathbb{Q}(u) = \mathbb{Q}(u_n)$, but we will need specific properties of the numbers $\alpha_n, \beta_n, \gamma_n$, so we proceed by induction.

Assume such a polynomial exists for some $n \geqslant 0$. Since $u_n = a_n + \frac{1}{u_{n+1}}$, we therefore have

$$u_{n+1}^2 f_n\left(a_n + \frac{1}{u_{n+1}}\right) = 0.$$

The resulting relation for u_{n+1} may be written in the form

$$f_{n+1}(x) = \alpha_{n+1}x^2 + \beta_{n+1}x + \gamma_{n+1}$$

where

$$\alpha_{n+1} = a_n^2\alpha_n + a_n\beta_n + \gamma_n,$$
$$\beta_{n+1} = 2a_n\alpha_n + \beta_n, \tag{3.30}$$
$$\gamma_{n+1} = \alpha_n. \tag{3.31}$$

It is clear that $\alpha_{n+1}, \beta_{n+1}, \gamma_{n+1} \in \mathbb{Z}$, and a simple calculation shows that

$$\beta_{n+1}^2 - 4\alpha_{n+1}\gamma_{n+1} = \beta_n^2 - 4\alpha_n\gamma_n,$$

proving the claim.

Notice that all the polynomials f_n have the same discriminant δ, which is not a square, so $\alpha_n \neq 0$ for $n \geqslant 0$. If there is some N with $\alpha_n > 0$ for all $n \geqslant N$, then equation (3.30) shows that the sequence $\beta_N, \beta_{N+1}, \dots$ is increasing since $a_n > 0$ for $n \geqslant 1$. Thus for large enough n, by equation (3.31), all three of α_n, β_n and γ_n are positive. This is impossible, since $f_n(u_n) = 0$ and $u_n > 0$. A similar argument shows that there is no N with $\alpha_n < 0$ for all $n \geqslant N$. We deduce that α_n must change in sign infinitely often, so in particular there is an infinite set $A \subseteq \mathbb{N}$ with the property that $\alpha_n\alpha_{n-1} < 0$ for all $n \in A$. By equation (3.31), it follows that $\alpha_n\gamma_n < 0$ for all $n \in A$. Now $\beta_n^2 - 4\alpha_n\gamma_n = \delta$, so for $n \in A$ we must have

$$|\alpha_n| \leqslant \tfrac{1}{4}\delta,$$
$$|\beta_n| < \sqrt{\delta},$$

and

$$|\gamma_n| \leqslant \tfrac{1}{4}\delta.$$

It follows that as n runs through the infinite set A there are only finitely many possibilities for the polynomials f_n, so there must be some $n_0 < n_1 < n_2$ with $f_{n_0} = f_{n_1} = f_{n_2}$. Since a quadratic polynomial has only two zeros, and $u_{n_0}, u_{n_1}, u_{n_2}$ are all zeros of the same polynomial, we see that two of them coincide so the continued fraction expansion of u is eventually periodic. □

Corollary 3.14. *Any quadratic irrational is badly approximable.*

Proof. This is an immediate consequence of Theorem 3.13 and Definition 3.9. □

It is not known if any other algebraic numbers are badly approximable.

Exercises for Section 3.3

Exercise 3.3.1. [45] Show that $\mathbb{Q}(\sqrt{5})$ contains infinitely many elements with a uniform bound on their partial quotients, by checking that the numbers $[\overline{1^{k+1},4,2,1^k,3}]$ for $k \geqslant 0$ all lie in $\mathbb{Q}(\sqrt{5})$ (here 1^k denotes the string $1,1,\dots,1$ of length k). Can you find a similar pattern in any real quadratic field $\mathbb{Q}(\sqrt{d})$?

Exercise 3.3.2. A number $u \in (0,1)$ is called *very well approximable* if there is some $\delta > 0$ with the property that there are infinitely many rational numbers $\frac{p}{q}$ with $\gcd(p,q)=1$ for which

$$\left|u - \frac{p}{q}\right| \leqslant \frac{1}{q^{2+\delta}}.$$

(a) Show that u is very well approximable if and only if there is some $\varepsilon > 0$ with the property that $a_{n+1} \geqslant q_n^{\varepsilon}$ for infinitely many values of n.
(b) Show that for any very well approximable number the convergence in equation (3.28) fails.

Exercise 3.3.3. Prove Liouville's Theorem[46]: if u is a real algebraic number of degree $d \geqslant 2$, then there is some constant $c(u) > 0$ with the property that

$$\frac{c(u)}{q^d} < \left|u - \frac{p}{q}\right|$$

for any rational number $\frac{p}{q}$.

Exercise 3.3.4. Use Liouville's Theorem from Exercise 3.3.3 to show that the number

$$u = \sum_{n=1}^{\infty} 10^{-n!}$$

is transcendental (that is, u is not a zero of any integral polynomial)[47].

Exercise 3.3.5. Prove that the theorem of Margulis from p. 6 does not hold for quadratic forms in 2 variables.

3.4 Invertible Extension of the Continued Fraction Map

We are interested in finding a geometrically convenient invertible extension of the non-invertible map T, and in Section 9.6 will re-prove the ergodicity of the Gauss measure in that context.

Define a set

$$\overline{Y} = \{(y,z) \in [0,1)^2 \mid 0 \leqslant z \leqslant \frac{1}{1+y}\}$$

(this set is illustrated in Figure 3.2) and a map $\overline{T} : \overline{Y} \to \overline{Y}$ by

$$\overline{T}(y,z) = (Ty, y(1-yz)).$$

The map $\overline{T}$ will also be called the Gauss map.

Proposition 3.15. *The map $\overline{T}:\overline{Y}\to\overline{Y}$ is an area-preserving bijection off a null set. More precisely, there is a countable union N of lines and curves in $\overline{Y}$ with the property that $T|_{\overline{Y}\smallsetminus N}:\overline{Y}\smallsetminus N\to\overline{Y}\smallsetminus N$ is a bijection preserving the Lebesgue measure.*

Proof. The derivative of the map $\overline{T}$ is

$$\begin{pmatrix} -\frac{1}{y^2} & 0 \\ 1-2yz & -y^2 \end{pmatrix},$$

with determinant 1. It follows that $\overline{T}$ preserves area locally. To see that the map is a bijection, define regions A_n and B_n in $\overline{Y}$ by

$$A_n=\{(y,z)\in\overline{Y}\mid \frac{1}{n+1}<y<\frac{1}{n}\}$$

and

$$B_n=\{(y,z)\in\overline{Y}\mid \frac{1}{n+1+y}<z<\frac{1}{n+y}\text{ and } y>0\}.$$

These sets are shown in Figure 3.2. Both

$$\{A_n\mid n=1,2,\dots\}$$

and

$$\{B_n\mid n=1,2,\dots\}$$

define partitions of $\overline{Y}$ after removing countably many vertical lines (or curves in the case of $\{B_n\}$). Since this is a Lebesgue null set, it is enough to show that $\overline{T}|_{A_n}:A_n\to B_n$ is a bijection for each $n\geqslant 1$, for then

$$\overline{T}|_{\cup_{n\geqslant 1}A_n}:\bigcup_{n\geqslant 1}A_n\longrightarrow\bigcup_{n\geqslant 1}B_n$$

is also a bijection, and we can take for the null set N the set of all images and pre-images of

$$\Big(\overline{Y}\smallsetminus\bigcup_{n\geqslant 1}A_n\Big)\cup\Big(\overline{Y}\smallsetminus\bigcup_{n\geqslant 1}B_n\Big).$$

Notice that $y>0$ and $0<z<\frac{1}{1+y}$ implies that

$$0<yz<\frac{y}{1+y},$$

$$\frac{1}{1+y}<(1-yz)<1,$$

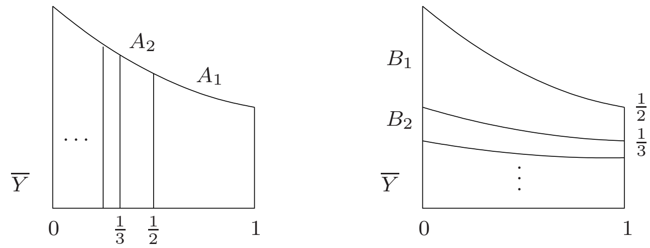


Fig. 3.2: The Gauss map is a bijection between $\overline{Y}$ and $\overline{Y}$, sending the subset $A_n \subseteq \overline{Y}$ to the subset $B_n \subseteq \overline{Y}$ for each $n \geqslant 1$.

and

$$\frac{y}{1+y} < y(1-yz) < y. \tag{3.32}$$

If now $(y,z) \in A_n$ for some $n \geqslant 1$ then $y = \frac{1}{n+y_1}$ for $\overline{T}(y,z) = (y_1, z_1)$ and the inequality (3.32) becomes

$$\frac{1}{n+1+y_1} = \frac{y}{1+y} < z_1 = y(1-yz) < y = \frac{1}{n+y_1},$$

so that $(y_1, z_1) \in B_n$ and therefore $\overline{T}(A_n) \subseteq B_n$. To see that the restriction to A_n is a bijection, fix $(y_1, z_1) \in B_n$. Then $y = \frac{1}{n+y_1}$ is uniquely determined, and the equation $z_1 = y(1-yz)$ then determines z uniquely. Clearly

$$y \in \left(\frac{1}{n+1}, \frac{1}{n}\right)$$

since $y_1 \in (0,1)$, and by reversing the argument above (or by a straightforward calculation) we see that

$$\frac{y}{1+y} = \frac{1}{n+1+y_1} < z_1 < \frac{1}{n+y_1} = y$$

implies $0 < z < \frac{1}{1+y}$ so that $(y,z) \in A_n$. □

Lemma 3.5 gives no indication of where the Gauss measure might have came from. The invertible extension, which preserves Lebesgue measure, gives an alternative proof that the Gauss measure is invariant, and gives one explanation of where it might come from.

Second proof of Lemma 3.5. Let $\pi : \overline{Y} \to Y$ be the projection

$$\pi(y,z) = y \tag{3.33}$$

onto Y. The Gauss measure μ on Y is the measure defined* by

* This construction of μ from m is called the *push-forward* of m by π.

$$\mu(B) = m(\pi^{-1}B)$$

where m is the normalized Lebesgue measure on $\overline{Y}$. Since $\overline{T} : \overline{Y} \to \overline{Y}$ preserves m by Proposition 3.15 and $\pi \circ \overline{T} = T \circ \pi$, the measure μ is T-invariant. □

The projection map $\pi : \overline{Y} \to Y$ defined in equation (3.33) shows that $\overline{T}$ on $\overline{Y}$ is an invertible extension of the non-invertible map T on Y.

Notes to Chapter 3

(38)(Page 69) The material in Section 3.1 may be found in many places; a convenient source for the path followed here using matrices is a note of van der Poorten [294].

(39)(Page 72) In particular, we have Dirichlet's theorem: for any $u \in \mathbb{R}$ and $Q \in \mathbb{N}$, there exists a rational number $\frac{p}{q}$ with $0 < q \leqslant Q$ and $|u - \frac{p}{q}| \leqslant \frac{1}{q(Q+1)}$, which can also be seen via the pigeon-hole principle.

(40)(Page 77) A broad overview of continued fractions from an ergodic perspective may be found in the monograph of Iosifescu and Kraaikamp [161]. Kraaikamp and others have suggested ways in which Gauss could have arrived at this measure; see also Keane [187]. Other approaches to the Gauss measure are described in the book of Khinchin [191]. The ergodic approach to continued fractions has a long history. Knopp [205] showed that the Gauss measure is ergodic (in different language); Kuz′min [217] found results on the rate of mixing of the Gauss measure; Doeblin [71] showed ergodicity; Ryll-Nardzewski [326] also showed this (that the Gauss measure is "indecomposable") and used the ergodic theorem to deduce results like equation (3.26). This had also been shown earlier by Khinchin [190]. Lévy [227] showed equation (3.25), an implicitly ergodic result, in 1936 (using the language of probability rather than ergodic theory).

(41)(Page 79) These results are indeed easily seen given both the ergodic theorem and the ergodicity of the Gauss map; their original proofs by other methods are not easy. For other results on the continued fraction expansion from the ergodic perspective, see Cornfeld, Fomin and Sinaĭ [60, Chap. 7] and from a number-theoretic perspective, see Khinchin [191]. The limit in equation (3.26), approximately 2.685, is known as Khinchin's constant; the problem of estimating it numerically is considered by Bailey, Borwein and Crandall [14]. Little is known about its arithmetical properties. The (exponential of the) constant appearing in equation (3.28) is usually called the Khinchin–Lévy constant. Just as in Example 2.31, it is a quite different problem to exhibit any specific number that satisfies these almost everywhere results: Adler, Keane and Smorodinsky exhibit a normal number for the continued fraction map in [2].

(42)(Page 79) This is proved here directly, using estimates for conditional measures on cylinder sets; see Billingsley [31] for example. We will re-prove it in Proposition 9.25 on p. 323 using a geometrical argument.

(43)(Page 87) Most of this section is devoted to quadratic irrationals, but it is clear there are uncountably many badly approximable numbers; the survey of Shallit [340] describes some of the many settings in which these numbers appear, gives other families of such numbers, and has an extensive bibliography on these numbers (which are also called numbers of *constant type*). For example, Kmošek [203] and Shallit [339] showed that if

$$\sum_{n=0}^{\infty} k^{-2^n} = [a_1^{(k)}, a_2^{(k)}, \dots],$$

then $\sup_{n\geqslant 1}\{a_n^{(2)}\} = 6$ and $\sup_{n\geqslant 1}\{a_n^{(k)}\} = n+2$ for $k \geqslant 3$.

(44)(Page 89) There are many ways to prove this; we follow the argument of Steinig [352] here.

(45)(Page 90) This remarkable uniformity in Definition 3.9 was shown by Woods [387] for $\mathbb{Q}(\sqrt{5})$ and by Wilson [383] in general, who showed that any real quadratic field $\mathbb{Q}(\sqrt{d})$ contains infinitely many numbers of the form $[\overline{a_1, a_2, \dots, a_k}]$ with $1 \leqslant a_n \leqslant M_d$ for all $n \geqslant 1$. McMullen [259] has explained these phenomena in terms of closed geodesics; the connection between continued fractions and closed geodesics will be developed in Chapter 9. Exercise 3.3.1 shows that we may take $M_5 = 4$, and the question is raised in [259] of whether there is a tighter bound allowing M_d to be taken equal to 2 for all d.

(46)(Page 91) Liouville's Theorem [234], [236] (on Diophantine approximation; there are several important results bearing his name) marked the start of an important series of advances in Diophantine approximation, attempting to sharpen the lower bound. These results may be summarized as follows. The statement that for any algebraic number u of degree d there is a constant $c(u)$ so that for all rationals p/q we have $|u - p/q| > c(u)/q^{\lambda(u)}$ holds: for $\lambda(u) = d$ (Liouville 1844); for any $\lambda(u) > \frac{1}{2}d + 1$ (Thue [360], 1909); for any $\lambda(u) > 2\sqrt{d}$ (Siegel [343], 1921); for any $\lambda(u) > \sqrt{2d}$ (Dyson [77], 1947); finally, and definitively, for any $\lambda(u) > 2$ (Roth [319], 1955).

(47)(Page 91) This observation of Liouville [235] dates from 1844 and seems to be the earliest construction of a transcendental number; in 1874 Cantor [47] used set theory to show that the set of algebraic numbers is countable, deducing that there are uncountably many transcendental real numbers (as pointed out by Herstein and Kaplansky [150, p. 238], and despite what is often taught, Cantor's proof can be used to exhibit many explicit transcendental numbers). In a different direction, many important constants were shown to be transcendental. Examples include: e (Hermite [149], 1873); π (Lindemann [232], 1882); α^β for α algebraic and not equal to 0 or 1 and β algebraic and irrational (Gelfond [113] and Schneider [334], 1934).

Chapter 4
Invariant Measures for Continuous Maps

One of the natural ways in which measure-preserving transformations arise is from continuous maps on compact metric spaces. Let (X, d) be a compact metric space, and let $T : X \to X$ be a continuous map. Recall that the dual space $C(X)^*$ of continuous real functionals on the space $C(X)$ of continuous functions $X \to \mathbb{R}$ can be naturally identified with the space of finite signed measures on X equipped with the weak*-topology. Our main interest is in the space $\mathscr{M}(X)$ of Borel probability measures on X. The main properties of $\mathscr{M}(X)$ needed are described in Section B.5.

Any continuous map $T : X \to X$ induces a continuous map

$$T_* : \mathscr{M}(X) \to \mathscr{M}(X)$$

defined by $T_*(\mu)(A) = \mu(T^{-1}A)$ for any Borel set $A \subseteq X$. Each point $x \in X$ defines a measure δ_x by

$$\delta_x(A) = \begin{cases} 1 \text{ if } x \in A; \\ 0 \text{ if } x \notin A. \end{cases}$$

We claim that $T_*(\delta_x) = \delta_{T(x)}$ for any $x \in X$. To see this, let $A \subseteq X$ be any measurable set, and notice that

$$(T_*\delta_x)(A) = \delta_x(T^{-1}A) = \delta_{T(x)}(A).$$

This suggests that we should think of the space of measures $\mathscr{M}(X)$ as generalized points, and the transformation $T_* : \mathscr{M}(X) \to \mathscr{M}(X)$ as a natural extension of the map T from the copy $\{\delta_x \mid x \in X\}$ of X to the larger set $\mathscr{M}(X)$. For $f \in C(X)$ and $\mu \in \mathscr{M}(X)$,

$$\int_X f \,\mathrm{d}(T_*\mu) = \int_X f \circ T \,\mathrm{d}\mu,$$

and this property characterizes T_* by equation (B.2) and Lemma B.12.

The map T_* is continuous and affine, so the set $\mathscr{M}^T(X)$ of T-invariant measures is a closed convex subset of $\mathscr{M}(X)$; in the next section[(48)] we will see that it is always non-empty.

4.1 Existence of Invariant Measures

The connection between ergodic theory and the dynamics of continuous maps on compact metric spaces begins with the next result, which shows that invariant measures can always be found.

Theorem 4.1. *Let $T : X \to X$ be a continuous map of a compact metric space, and let (ν_n) be any sequence in $\mathscr{M}(X)$. Then any weak*-limit point of the sequence (μ_n) defined by $\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} T_*^j \nu_n$ is a member of $\mathscr{M}^T(X)$.*

An immediate consequence is the following important general statement, which shows that measure-preserving transformations are ubiquitous. It is known as the Kryloff–Bogoliouboff Theorem [214].

Corollary 4.2 (Kryloff–Bogoliouboff). *Under the hypotheses of Theorem* 4.1*, $\mathscr{M}^T(X)$ is non-empty.*

PROOF. Since $\mathscr{M}(X)$ is weak*-compact, the sequence (μ_n) must have a limit point. □

Write $\|f\|_\infty = \sup\{|f(x)| \mid x \in X\}$ as usual.

PROOF OF THEOREM 4.1. Let $\mu_{n(j)} \to \mu$ be a convergent subsequence of (μ_n) and let $f \in C(X)$. Then, by applying the definition of $T_*\mu_n$, we get

$$\begin{aligned}\left|\int f \circ T \,\mathrm{d}\mu_{n(j)} - \int f \,\mathrm{d}\mu_{n(j)}\right| &= \frac{1}{n(j)}\left|\int \sum_{i=0}^{n(j)-1} \left(f \circ T^{i+1} - f \circ T^i\right) \mathrm{d}\nu_{n(j)}\right| \\ &= \frac{1}{n(j)}\left|\int \left(f \circ T^{n(j)+1} - f\right) \mathrm{d}\nu_{n(j)}\right| \\ &\leqslant \frac{2}{n(j)}\|f\|_\infty \longrightarrow 0\end{aligned}$$

as $j \to \infty$, for all $f \in C(X)$. It follows that $\int f \circ T \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$, so μ is a member of $\mathscr{M}^T(X)$ by Lemma B.12. □

Thus $\mathscr{M}^T(X)$ is a non-empty compact convex set, since convex combinations of elements of $\mathscr{M}^T(X)$ belong to $\mathscr{M}^T(X)$. It follows that $\mathscr{M}^T(X)$ is an infinite set unless it comprises a single element. For many maps it is difficult to describe the space of invariant measures. The next example has very few ergodic invariant measures, and we shall see later many maps that have only one invariant measure.

Example 4.3 (North–South map). Define the stereographic projection π from the circle $X = \{z \in \mathbb{C} \mid |z - \mathrm{i}| = 1\}$ to the real axis by continuing the line from $2\mathrm{i}$ through a unique point on $X \smallsetminus \{2\mathrm{i}\}$ until it meets the line $\Im(z) = 0$ (see Figure 4.1).

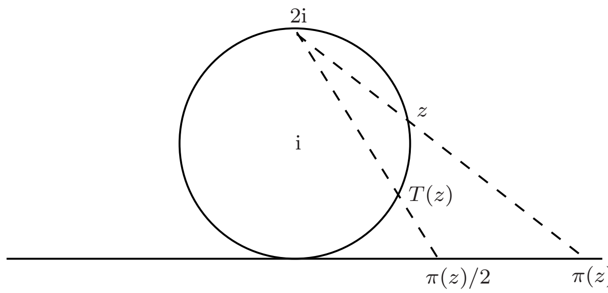


Fig. 4.1: The North-South map on the circle; for $z \neq 2\mathrm{i}$, $T^n z \to 0$ as $n \to \infty$.

The "North–South" map $T : X \to X$ is defined by

$$T(z) = \begin{cases} 2\mathrm{i} & \text{if } z = 2\mathrm{i}; \\ \pi^{-1}(\pi(z)/2) & \text{if } z \neq 2\mathrm{i} \end{cases}$$

as shown in Figure 4.1. Using Poincaré recurrence (Theorem 2.11) it is easy to show that $\mathscr{M}^T(X)$ comprises the measures $p\delta_{2\mathrm{i}} + (1-p)\delta_0$, $p \in [0, 1]$ that are supported on the two points $2\mathrm{i}$ and 0. Only the measures corresponding to $p = 0$ and $p = 1$ are ergodic.

It is in general difficult to identify measures with specific properties, but the ergodic measures are readily characterized in terms of the geometry of the space of invariant measures.

Theorem 4.4. *Let X be a compact metric space and let $T : X \to X$ be a measurable map. The ergodic elements of $\mathscr{M}^T(X)$ are exactly the extreme points of $\mathscr{M}^T(X)$.*

That is, T is ergodic with respect to an invariant probability measure if and only if that measure cannot be expressed as a strict convex combination of two different T-invariant probability measures. For any measurable set A, define $\mu\big|_A$ by $\mu\big|_A(C) = \mu(A \cap C)$. If T is not assumed to be continuous, then we do not know that $\mathscr{M}^T(X) \neq \varnothing$, so without the assumption of continuity Theorem 4.4 may be true but vacuous (see Exercise 4.1.1).

Proof of Theorem 4.4. Let $\mu \in \mathscr{M}^T(X)$ be a non-ergodic measure. Then there is a measurable set B with $\mu(B) \in (0, 1)$ and with $T^{-1}B = B$. It follows that

$$\frac{1}{\mu(B)}\mu\big|_B, \frac{1}{\mu(X\smallsetminus B)}\mu\big|_{X\smallsetminus B} \in \mathscr{M}^T(X),$$

so

$$\mu = \mu(B)\left(\frac{1}{\mu(B)}\mu\big|_B\right) + \mu(X\smallsetminus B)\left(\frac{1}{\mu(X\smallsetminus B)}\mu\big|_{X\smallsetminus B}\right)$$

expresses μ as a strict convex combination of the invariant probability measures

$$\frac{1}{\mu(B)}\mu\big|_B$$

and

$$\frac{1}{\mu(X\smallsetminus B)}\mu\big|_{X\smallsetminus B},$$

which are different since they give different measures to the set B.

Conversely, let μ be an ergodic measure and assume that

$$\mu = s\nu_1 + (1-s)\nu_2$$

expresses μ as a strict convex combination of the invariant measures ν_1 and ν_2. Since $s > 0$, $\nu_1 \ll \mu$, so there is a positive function $f \in L^1_\mu$ (f is the Radon–Nikodym derivative $\frac{d\nu_1}{d\mu}$; see Theorem A.15) with the property that

$$\nu_1(A) = \int_A f \,\mathrm{d}\mu \tag{4.1}$$

for any measurable set A. The set $B = \{x \in X \mid f(x) < 1\}$ is measurable since f is measurable, and

$$\begin{aligned}\int_{B\cap T^{-1}B} f \,\mathrm{d}\mu + \int_{B\smallsetminus T^{-1}B} f \,\mathrm{d}\mu &= \nu_1(B)\\ &= \nu_1(T^{-1}B)\\ &= \int_{B\cap T^{-1}B} f \,\mathrm{d}\mu + \int_{(T^{-1}B)\smallsetminus B} f \,\mathrm{d}\mu,\end{aligned}$$

so

$$\int_{B\smallsetminus T^{-1}B} f \,\mathrm{d}\mu = \int_{(T^{-1}B)\smallsetminus B} f \,\mathrm{d}\mu. \tag{4.2}$$

By definition, $f(x) < 1$ for $x \in B\smallsetminus(T^{-1}B)$ while $f(x) \geqslant 1$ for $x \in T^{-1}B\smallsetminus B$. On the other hand,

$$\begin{aligned}\mu((T^{-1}B)\smallsetminus B) &= \mu(T^{-1}B) - \mu((T^{-1}B)\cap B)\\ &= \mu(B) - \mu((T^{-1}B)\cap B)\\ &= \mu(B\smallsetminus T^{-1}B)\end{aligned}$$

so equation (4.2) implies that $\mu(B\smallsetminus T^{-1}B) = 0$ and $\mu((T^{-1}B)\smallsetminus B) = 0$. Therefore $\mu((T^{-1}B)\triangle B) = 0$, so by ergodicity of μ we must have $\mu(B) = 0$ or 1. If $\mu(B) = 1$ then

$$\nu_1(X) = \int_X f \,\mathrm{d}\mu < \mu(B) = 1,$$

which is impossible. So $\mu(B) = 0$.

A similar argument shows that $\mu(\{x \in X \mid f(x) > 1\}) = 0$, so $f(x) = 1$ almost everywhere with respect to μ. By equation (4.1), this shows that

$$\nu_1 = \mu,$$

so μ is an extreme point in $\mathscr{M}^T(X)$. $\square$

Write $\mathscr{E}^T(X)$ for the set of extreme points in $\mathscr{M}^T(X)$ – by Theorem 4.4, this is the set of ergodic measures for T.

Example 4.5. Let $X = \{1,\dots,r\}^{\mathbb{Z}}$ and let $T : X \to X$ be the left shift map. In Example 2.9 we defined for any probability vector $\mathbf{p} = (p_1,\dots,p_r)$ a T-invariant probability measure $\mu = \mu_{\mathbf{p}}$ on X, and by Proposition 2.15 all these measures are ergodic. Thus for this example the space $\mathscr{E}^T(X)$ of ergodic invariant measures is uncountable. This collection of measures is an inconceivably tiny subset of the set of all ergodic measures – there is no hope of describing all of them.

Measures μ_1 and μ_2 are called *mutually singular* if there exist disjoint measurable sets A and B with $A \cup B = X$ for which $\mu_1(B) = \mu_2(A) = 0$ (see Section A.4).

Lemma 4.6. *If $\mu_1, \mu_2 \in \mathscr{E}^T(X)$ and $\mu_1 \neq \mu_2$ then μ_1 and μ_2 are mutually singular.*

Proof. Let $f \in C(X)$ be chosen with $\int f \,\mathrm{d}\mu_1 \neq \int f \,\mathrm{d}\mu_2$ (such a function exists by Theorem B.11). Then by the ergodic theorem (Theorem 2.30)

$$\mathsf{A}_n^f(x) \to \int f \,\mathrm{d}\mu_1 \tag{4.3}$$

for μ_1-almost every $x \in X$, and

$$\mathsf{A}_n^f(x) \to \int f \,\mathrm{d}\mu_2$$

for μ_2-almost every $x \in X$. It follows that the set $A = \{x \in X \mid (4.3) \text{ holds}\}$ is measurable and has $\mu_1(A) = 1$ but $\mu_2(A) = 0$. $\square$

Some of the problems for this section make use of the topological analog of Definition 2.7, which will be used later.

Definition 4.7. Let $T : X \to X$ and $S : Y \to Y$ be continuous maps of compact metric spaces (that is, topological dynamical systems). Then a homeomorphism $\theta : X \to Y$ with $\theta \circ T = S \circ \theta$ is called a *topological conjugacy*, and if there such a conjugacy then T and S are *topologically conjugate*. A continuous surjective map $\phi : X \to Y$ with $\phi \circ T = S \circ \phi$ is called a *topological factor* map, and in this case S is said to be a *factor* of T.

Exercises for Section 4.1

Exercise 4.1.1. Let $X = \{0, \frac{1}{n} \mid n \geqslant 1\}$ with the compact topology inherited from the reals. Since X is countable, there is a bijection $\theta : X \to \mathbb{Z}$. Show that the map $T : X \to X$ defined by $T(x) = \theta^{-1}(\theta(x) + 1)$ is measurable with respect to the Borel σ-algebra on X but has no invariant probability measures.

Exercise 4.1.2. Show that a weak*-limit of ergodic measures need not be an ergodic measure by the following steps. Start with a point x in the full 2-shift $\sigma : X \to X$ with the property that any finite block of symbols of length ℓ appears in x with asymptotic frequency $\frac{1}{2^\ell}$ (such points certainly exist; indeed the ergodic theorem says that almost every point with respect to the $(1/2, 1/2)$ Bernoulli measure will do). Write $(x_1 \dots x_n 0 \dots 0)^\infty$ for the point $y \in \{0, 1\}^{\mathbb{Z}}$ determined by the two conditions

$$y|_{[0,2n-1]} = x_1 \dots x_n 0 \dots 0$$

and $\sigma^{2n}(y) = y$. Now for each n construct an ergodic σ-invariant measure μ_n supported on the orbit of the periodic point $(x_1 \dots x_n 0 \dots 0)^\infty$ in which there are n 0 symbols in every cycle of the periodic point under the shift. Show that μ_n converges to some limit ν and use Theorem 4.4 to deduce that ν is not ergodic.

Exercise 4.1.3. For a continuous map $T : X \to X$ of a compact metric space (X, d), define the invertible extension $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ as follows. Let

- $\widetilde{X} = \{x \in X^{\mathbb{Z}} \mid x_{k+1} = T x_k \text{ for all } k \in \mathbb{Z}\}$;
- $(\widetilde{T}x)_k = x_{k+1}$ for all $k \in \mathbb{Z}$ and $x \in \widetilde{X}$;

with metric $\widetilde{\mathsf{d}}(x, y) = \sum_{k \in \mathbb{Z}} 2^{-|k|} \mathsf{d}(x_k, y_k)$. Write $\pi : \widetilde{X} \to X$ for the map sending x to x_0. Prove the following.

(1) $\widetilde{T}$ is a homeomorphism of a compact metric space, and $\pi : \widetilde{X} \to X$ is a topological factor map.
(2) If (Y, S) is any homeomorphism of a compact metric space with the property that there is a topological factor map $(Y, S) \to (X, T)$, then $(\widetilde{X}, \widetilde{T})$ is a topological factor of (Y, S).

(3) $\pi_* \mathscr{M}^{\widetilde{T}}(\widetilde{X}) = \mathscr{M}^T(X)$.
(4) $\pi_* \mathscr{E}^{\widetilde{T}}(\widetilde{X}) = \mathscr{E}^T(X)$.

Exercise 4.1.4. Show that the ergodic Bernoulli measures discussed in Example 4.5 do not exhaust all ergodic measures for the full shift as follows.

(1) Show that any periodic orbit supports an ergodic measure which is not a Bernoulli measure.
(2) Show that there are ergodic measures on the full shift that are neither Bernoulli nor supported on a periodic orbit.

Exercise 4.1.5. Give a different proof of Lemma 4.6 using the Radon–Nikodym derivative (Theorem A.15) and the Lebesgue decomposition theorem (Theorem A.14), instead of the pointwise ergodic theorem.

Exercise 4.1.6. Prove that the ergodic measures for the circle-doubling map $T_2 : x \mapsto 2x \pmod 1$ are dense in the space of all invariant measures.

4.2 Ergodic Decomposition

An important consequence of the fact that $\mathscr{M}^T(X)$ is a compact convex set is that the Choquet representation theorem may be applied[(49)] to it. This generalizes the simple geometrical fact that in a finite-dimensional convex simplex, every point is a unique convex combination of the extreme points, to an infinite-dimensional result. In our setting, this gives a way to decompose any invariant measure into ergodic components.

Theorem 4.8 (Ergodic decomposition). *Let X be a compact metric space and $T : X \to X$ a continuous map. Then for any $\mu \in \mathscr{M}^T(X)$ there is a unique probability measure λ defined on the Borel subsets of the compact metric space $\mathscr{M}^T(X)$ with the properties that*

(1) $\lambda(\mathscr{E}^T(X)) = 1$, *and*
(2) $\displaystyle\int_X f \,\mathrm{d}\mu = \int_{\mathscr{E}^T(X)} \left(\int_X f \,\mathrm{d}\nu \right) \mathrm{d}\lambda(\nu)$ *for any* $f \in C(X)$.

Proof. This follows from Choquet's theorem [55] (see also the notes of Phelps [283]). A different proof will be given later (cf. p. 154), and a nontrivial example may be seen in Example 4.13. □

In fact Choquet's theorem is more general than we need: in our setting, X is a compact metric space so $C(X)$ is separable, and hence $\mathscr{M}^T(X)$ is metrizable (see equation (B.3) for an explicit metric on $\mathscr{M}(X)$ built from a dense set of continuous functions). The picture of the space of invariant measures given by this result is similar to the familiar picture of a finite-dimensional simplex, but in fact few continuous maps[(50)] have a finite-dimensional space of invariant measures. Indeed, as we have seen in Exercise 4.1.2, the set of ergodic measures is in general not a closed subset of the set of invariant measures.

We will see some non-trivial examples of ergodic decompositions in Section 4.3. The existence of the ergodic decomposition is one of the reasons that ergodicity is such a powerful tool: any property that is preserved by the integration in Theorem 4.8(2) which holds for ergodic systems holds for any measure-preserving transformation. A particularly striking case of this general principle will come up in connection with the ergodic proof of Szemerédi's theorem (see Section 7.2.3). There is no real topological analog of this decomposition (see Exercises 4.2.3 and 4.2.4).

Exercises for Section 4.2

Exercise 4.2.1. A homeomorphism $T : X \to X$ of a compact metric space (a *topological dynamical system* or *cascade*) is called *minimal* if the only non-empty closed T-invariant subset of X is X itself.
(a) Show that (X,T) is minimal if and only if the orbit of each point in X is dense.
(b) Show that (X,T) is minimal if and only if $\bigcup_{n\in\mathbb{Z}} T^n O = X$ for every non-empty open set $O \subseteq X$.
(c) Show that any topological dynamical system (X,T) has a *minimal set*: that is, a closed T-invariant set A with the property that $T : A \to A$ is minimal.

Exercise 4.2.2. Use Exercise 4.2.1(c) to prove Birkhoff's recurrence theorem(51): every topological dynamical system (X,T) contains a point x for which there is a sequence $n_k \to \infty$ with $T^{n_k}x \to x$ as $k \to \infty$. Such a point is called *recurrent* under T.

Exercise 4.2.3. Show that in general a topological dynamical system is not a disjoint union of closed minimal subsystems.

Exercise 4.2.4. A homeomorphism $T : X \to X$ of a compact metric space is called *topologically ergodic* if every closed proper T-invariant subset of X has empty interior. Show that the following properties are equivalent:

- (X,T) is topologically ergodic;
- there is a point in X with a dense orbit;
- for any non-empty open sets O_1 and O_2 in X, there is some $n \geqslant 0$ for which $O_1 \cap T^n O_2 \neq \varnothing$.

Show that in general a topological dynamical system is not a disjoint union of closed topologically ergodic subsystems.

Exercise 4.2.5. Let $T : X \to X$ be a continuous map on a compact metric space. Show that the measures in $\mathscr{E}^T(X)$ constrain all the ergodic averages in the following sense. For $f \in C(X)$, define

$$m(f) = \inf_{\mu \in \mathscr{E}^T(X)} \left\{ \int f \,\mathrm{d}\mu \right\}$$

and

$$M(f) = \sup_{\mu \in \mathscr{E}^T(X)} \left\{ \int f \,\mathrm{d}\mu \right\}.$$

Prove that

$$m(f) \leqslant \liminf_{N\to\infty} \mathsf{A}_N^f(x) \leqslant \limsup_{N\to\infty} \mathsf{A}_N^f(x) \leqslant M(f)$$

for any $x \in X$.

4.3 Unique Ergodicity

A natural distinguished class of transformations are those for which there is only one invariant Borel measure. This measure is automatically ergodic, and the uniqueness of this measure has several powerful consequences.

Definition 4.9. Let X be a compact metric space and let $T : X \to X$ be a continuous map. Then T is said to be *uniquely ergodic* if $\mathscr{M}^T(X)$ comprises a single measure.

Theorem 4.10. *For a continuous map $T : X \to X$ on a compact metric space, the following properties are equivalent.*

(1) *T is uniquely ergodic.*
(2) $|\mathscr{E}^T(X)| = 1$.
(3) *For every $f \in C(X)$,*

$$\mathsf{A}_N^f = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \longrightarrow C_f, \tag{4.4}$$

where C_f is a constant independent of x.

(4) *For every $f \in C(X)$, the convergence* (4.4) *is uniform across X.*
(5) *The convergence* (4.4) *holds for every f in a dense subset of $C(X)$.*

Under any of these assumptions, the constant C_f in (4.4) *is $\int_X f \,\mathrm{d}\mu$, where μ is the unique invariant measure.*

We will make use of Theorem 4.8 for the equivalence of (1) and (2); the equivalence between (1) and (3)–(5) is independent of it.

Proof of Theorem 4.10. (1) $\iff$ (2): If T is uniquely ergodic and μ is the only T-invariant probability measure on X, then μ must be ergodic by Theorem 4.4. If there is only one ergodic invariant probability measure on X, then by Theorem 4.8, it is the only invariant probability measure on X.

(1) $\implies$ (3): Let μ be the unique invariant measure for T, and apply Theorem 4.1 to the constant sequence (δ_x). Since there is only one possible limit point and $\mathscr{M}(X)$ is compact, we must have

$$\frac{1}{N}\sum_{n=0}^{N-1}\delta_{T^n x}\longrightarrow\mu$$

in the weak*-topology, so for any $f\in C(X)$

$$\frac{1}{N}\sum_{n=0}^{N-1}f(T^n x)\longrightarrow\int_X f\,\mathrm{d}\mu.$$

(3) $\implies$ (1): Let $\mu\in\mathscr{M}^T(X)$. Then by the dominated convergence theorem, (4.4) implies that

$$\int_X f\,\mathrm{d}\mu=\int_X\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}f(T^n x)\,\mathrm{d}\mu=C_f$$

for all $f\in C(X)$. It follows that C_f is the integral of f with respect to any measure in $\mathscr{M}^T(X)$, so $\mathscr{M}^T(X)$ can only contain a single measure.

Notice that this also shows $C_f=\int_X f\,\mathrm{d}\mu$ for the unique measure μ.

(1) $\implies$ (4): Let $\mu\in\mathscr{M}^T(X)$, and notice that we must have $C_f=\int f\,\mathrm{d}\mu$ as above. If the convergence is not uniform, then there is a function g in $C(X)$ and an $\varepsilon>0$ such that for every N_0 there is an $N>N_0$ and a point $x_j\in X$ for which

$$\left|\frac{1}{N}\sum_{n=0}^{N-1}g(T^n x_j)-C_g\right|\geqslant\varepsilon.$$

Let $\mu_N=\frac{1}{N}\sum_{n=0}^{N-1}\delta_{T^n x_j}$, so that

$$\left|\int_X g\,\mathrm{d}\mu_N-C_g\right|\geqslant\varepsilon. \tag{4.5}$$

By weak*-compactness the sequence (μ_N) has a subsequence $\left(\mu_{N(k)}\right)$ with

$$\mu_{N(k)}\to\nu$$

as $k\to\infty$. Then $\nu\in\mathscr{M}^T(X)$ by Theorem 4.1, and

$$\left|\int_X g\,\mathrm{d}\nu-C_g\right|\geqslant\varepsilon$$

by equation (4.5). However, this shows that $\mu \neq \nu$, which contradicts (1).

(4) $\implies$ (5): This is clear.

(5) $\implies$ (1): If $\mu, \nu \in \mathscr{E}^T(X)$ then, just as in the proof that (3) $\implies$ (1),

$$\int_X f \, \mathrm{d}\nu = C_f = \int_X f \, \mathrm{d}\mu$$

for any function f in a dense subset of $C(X)$, so $\nu = \mu$. □

The equivalence of (1) and (3) in Theorem 4.10 appeared first in the paper of Kryloff and Bogoliouboff [214] in the context of uniquely ergodic flows.

Example 4.11. The circle rotation $R_\alpha : \mathbb{T} \to \mathbb{T}$ is uniquely ergodic if and only if α is irrational. The unique invariant measure in this case is the Lebesgue measure $m_{\mathbb{T}}$. This may be proved using property (5) of Theorem 4.10 (or using property (1); see Theorem 4.14). Assume first that α is irrational, so $\mathrm{e}^{2\pi \mathrm{i} k\alpha} = 1$ only if $k = 0$. If $f(t) = \mathrm{e}^{2\pi \mathrm{i} kt}$ for some $k \in \mathbb{Z}$, then

$$\frac{1}{N}\sum_{n=0}^{N-1} f(R_\alpha^n t) = \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{e}^{2\pi \mathrm{i} k(t+n\alpha)} = \begin{cases} 1 & \text{if } k = 0; \\ \dfrac{1}{N}\mathrm{e}^{2\pi \mathrm{i} kt}\dfrac{\mathrm{e}^{2\pi \mathrm{i} Nk\alpha} - 1}{\mathrm{e}^{2\pi \mathrm{i} k\alpha} - 1} & \text{if } k \neq 0. \end{cases} \tag{4.6}$$

Equation (4.6) shows that

$$\frac{1}{N}\sum_{n=0}^{N-1} f(R_\alpha^n t) \longrightarrow \int f \, \mathrm{d}m_{\mathbb{T}} = \begin{cases} 1 \text{ if } k = 0; \\ 0 \text{ if } k \neq 0. \end{cases}$$

By linearity, the same convergence will hold for any trigonometric polynomial, and therefore property (5) of Theorem 4.10 holds. For a curious application of this result, see Example 1.3.

If α is rational, then Lebesgue measure is invariant but not ergodic, so there must be other invariant measures.

Example 4.11 may be used to illustrate the ergodic decomposition of a particularly simple dynamical system.

Example 4.12. Let $X = \{z \in \mathbb{C} \mid |z| = 1 \text{ or } 2\}$, let α be an irrational number, and define a continuous map $T : X \to X$ by $T(z) = \mathrm{e}^{2\pi \mathrm{i}\alpha} z$. By unique ergodicity on each circle, any invariant measure μ takes the form

$$\mu = s m_1 + (1 - s) m_2,$$

where m_1 and m_2 denote Lebesgue measures on the two circles comprising X. Thus $\mathscr{M}^T(X) = \{s m_1 + (1-s) m_2 \mid s \in [0, 1]\}$, with the two ergodic measures given by the extreme points $s = 0$ and $s = 1$. The decomposition of μ is described by the measure $\nu = s\delta_{m_1} + (1 - s)\delta_{m_2}$. A convenient notation for this is $\mu = \int_{\mathscr{M}^T(X)} m \, \mathrm{d}\nu(m)$.

Example 4.13. A more sophisticated version of Example 4.12 is a rotation on the disk. Let $\mathbb{D} = \{z \in \mathbb{C} \mid |z| \leqslant 1\}$, let α be an irrational number, and define a continuous map $T : \mathbb{D} \to \mathbb{D}$ by $T(z) = \mathrm{e}^{2\pi \mathrm{i}\alpha} z$. For each $r \in (0, 1]$, let m_r denote the normalized Lebesgue measure on the circle $\{z \in \mathbb{C} \mid |z| = r\}$ and let $m_0 = \delta_0$ (these are the ergodic measures). Then the decomposition of $\mu \in \mathscr{M}^T(X)$ is a measure ν on $\{m_r \mid r \in [0, 1]\}$, and

$$\mu(A) = \int_{\mathscr{M}^T(X)} m_r(A) \,\mathrm{d}\nu(m_r).$$

Both Proposition 2.16 and Example 4.11 are special cases of the following more general result about unique ergodicity for rotations on compact groups.

Theorem 4.14. *Let X be a compact metrizable group and $R_g(x) = gx$ the rotation by a fixed element $g \in X$. Then the following are equivalent.*

(1) *R_g is uniquely ergodic (with the unique invariant measure being m_X, the Haar measure on X).*
(2) *R_g is ergodic with respect to m_X.*
(3) *The subgroup $\{g^n\}_{n\in\mathbb{Z}}$ generated by g is dense in X.*
(4) *X is abelian, and $\chi(g) \neq 1$ for any non-trivial character $\chi \in \widehat{X}$.*

Proof. (1) $\implies$ (2): This is clear.

(2) $\implies$ (3): Let Y denote the closure of the subgroup generated by g. If $Y \neq X$ then there is a continuous non-constant function on X that is constant on each coset of Y: in fact if d is a bi-invariant metric on X giving the topology, then

$$\mathsf{d}_Y(x) = \min\{\mathsf{d}(x, y) \mid y \in Y\}$$

defines such a function (an invariant metric exists by Lemma C.2). Such a function is invariant under R_g, showing that R_g is not ergodic.

(3) $\implies$ (1): If $Y = X$ then X is abelian (since it contains a dense abelian subgroup), and any probability measure μ invariant under R_g is invariant under translation by a dense subgroup. This implies that μ is invariant under translation by any $y \in X$ by the following argument. Let $f \in C(X)$ be any continuous function, and fix $\varepsilon > 0$. Then for every $\delta > 0$ there is some n with $\mathsf{d}(y, g^n) < \delta$, so by an appropriate choice of δ we have

$$|f(g^n x) - f(yx)| < \varepsilon$$

for all $x \in X$. Since

$$\int f(x) \,\mathrm{d}\mu(x) = \int f(g^n x) \,\mathrm{d}\mu(x),$$

it follows that

$$\left| \int f(yx) \,\mathrm{d}\mu(x) - \int f(x) \,\mathrm{d}\mu(x) \right| = \left| \int \left(f(yx) - f(g^n x)\right) \,\mathrm{d}\mu(x) \right| < \varepsilon$$

for all $\varepsilon > 0$, so R_y preserves μ. Since this holds for all $y \in X$, μ must be the Haar measure. It follows that R_g is uniquely ergodic.

(4) $\implies$ (2): Assume now that X is abelian and $\chi(g) \neq 1$ for every non-trivial character $\chi \in \widehat{X}$. If $f \in L^2(X)$ is invariant under R_g, then the Fourier series

$$f = \sum_{\chi \in \widehat{X}} c_\chi \chi$$

satisfies

$$f = U_{R_g} f = \sum_{\chi \in \widehat{X}} c_\chi \chi(g) \chi,$$

and so f is constant as required.

(2) $\implies$ (4): By (3) it follows that X is abelian. If now $\chi \in \widehat{X}$ is a character with $\chi(g) = 1$, then

$$\chi(R_g x) = \chi(g)\chi(x) = \chi(x)$$

is invariant, which by (2) implies that χ is itself a constant almost everywhere and so is trivial. $\square$

Corollary 4.15. *Let $X = \mathbb{T}^\ell$, and let $g = (\alpha_1, \alpha_2, \dots, \alpha_\ell) \in \mathbb{R}^\ell$. Then the toral rotation $R_g : \mathbb{T}^\ell \to \mathbb{T}^\ell$ given by $R_g(x) = x + g$ is uniquely ergodic if and only if $1, \alpha_1, \dots, \alpha_\ell$ are linearly independent over $\mathbb{Q}$.*

Theorems 2.19 and 4.14 have been generalized to give characterizations of ergodicity for affine maps on compact abelian groups by Hahn and Parry [131] and Parry [278], and on non-abelian groups by Chu [57].

Exercises for Section 4.3

Exercise 4.3.1. Prove that (3) $\implies$ (1) in Theorem 4.14 using Pontryagin duality.

Exercise 4.3.2. Show that a surjective homomorphism $T : X \to X$ of a compact group X is uniquely ergodic if and only if $|X| = 1$.

Exercise 4.3.3. Extend Theorem 4.14 by using the quotient space $Y\backslash X$ of a compact group X to classify the probability measures on X invariant under the rotation R_g when $Y \neq X$.

Exercise 4.3.4. Show that for any Riemann-integrable function $f : \mathbb{T} \to \mathbb{R}$ and $\varepsilon > 0$ there are trigonometric polynomials p^- and p^+ such that

$$p^-(t) < f(t) < p^+(t)$$

for all $t \in \mathbb{T}$, and $\int_0^1 (p^+(t) - p^-(t))\,\mathrm{d}t < \varepsilon$. Use this to show that if α is irrational then for any Riemann-integrable function $f : \mathbb{T} \to \mathbb{R}$,

$$\frac{1}{N}\sum_{n=0}^{N-1} f(R_\alpha^n t) \to \int f \,\mathrm{d}m_{\mathbb{T}}$$

for all $t \in \mathbb{T}$.

Exercise 4.3.5. Prove Corollary 4.15
(a) using Theorem 4.14;
(b) using Theorem 4.10(5).

Exercise 4.3.6. [52] Let X be a compact metric space, and let $T : X \to X$ be a continuous map. Assume that $\mu \in \mathscr{E}^T(X)$, and that for every $x \in X$ there exists a constant $C = C(x)$ such that for every $f \in C(X)$, $f \geqslant 0$,

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \leqslant C \int f \,\mathrm{d}\mu.$$

Show that T is uniquely ergodic.

4.4 Measure Rigidity and Equidistribution

A natural question in number theory concerns how a sequence of real numbers is distributed when reduced modulo 1. When the terms of the sequence are generated by some dynamical process, then the expressions resemble ergodic averages, and it is natural to expect that ergodic theory will have something to offer.

4.4.1 Equidistribution on the Interval

Ergodic theorems give conditions under which all or most orbits in a dynamical system spend a proportion of time in a given set proportional to the measure of the set. In this section we consider a more abstract notion of equidistribution[53] in the specific setting of Lebesgue measure on the unit interval.

Definition 4.16. A sequence (x_n) with $x_n \in [0, 1]$ for all n is said to be *equidistributed* or *uniformly distributed* if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) = \int_0^1 f(x)\,\mathrm{d}x \tag{4.7}$$

for any $f \in C([0,1])$.

A more intuitive formulation (developed in Lemma 4.17) of equidistribution requires that the terms of the sequence fall in an interval with the correct frequency, just as the pointwise ergodic theorem (Theorem 2.30) says that almost every orbit under an ergodic transformation falls in a measurable set with the correct frequency.

Lemma 4.17. (54) *For a sequence (x_n) of elements of $[0,1]$, the following properties are equivalent.*

(1) *The sequence (x_n) is equidistributed.*
(2) *For any $k \neq 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathrm{e}^{2\pi \mathrm{i} k x_j} = 0.$$

(3) *For any numbers a, b with $0 \leqslant a < b \leqslant 1$,*

$$\frac{1}{n} \left| \{ j \mid 1 \leqslant j \leqslant n, x_j \in [a,b] \} \right| \longrightarrow (b-a)$$

as $n \to \infty$.

Proof. (1) $\iff$ (3): Assume (1) and fix a, b with $0 \leqslant a < b \leqslant 1$. Given a sufficiently small $\varepsilon > 0$, define continuous functions that approximate the indicator function $\chi_{[a,b]}$ by

$$f^+(x) = \begin{cases} 1 & \text{if } a \leqslant x \leqslant b; \\ \big(x - (a-\varepsilon)\big)/\varepsilon & \text{if } \max\{0, a-\varepsilon\} \leqslant x < a; \\ \big((b+\varepsilon) - x\big)/\varepsilon & \text{if } b < x \leqslant \min\{b+\varepsilon, 1\}; \\ 0 & \text{for other } x, \end{cases}$$

and

$$f^-(x) = \begin{cases} 1 & \text{if } a+\varepsilon \leqslant x \leqslant b-\varepsilon; \\ (x-a)/\varepsilon & \text{if } a \leqslant x < a+\varepsilon; \\ (b-x)/\varepsilon & \text{if } b-\varepsilon < x \leqslant b; \\ 0 & \text{for other } x. \end{cases}$$

Notice that $f^-(x) \leqslant \chi_{[a,b]}(x) \leqslant f^+(x)$ for all $x \in [0,1]$, and

$$\int_0^1 \big(f^+(x) - f^-(x)\big)\,\mathrm{d}x \leqslant 2\varepsilon.$$

For small ε and $0 < a < b < 1$, these functions are illustrated in Figure 4.2.

It follows that

$$\frac{1}{n}\sum_{j=1}^{n} f^-(x_j) \leqslant \frac{1}{n}\sum_{j=1}^{n} \chi_{[a,b]}(x_j) \leqslant \frac{1}{n}\sum_{j=1}^{n} f^+(x_j).$$

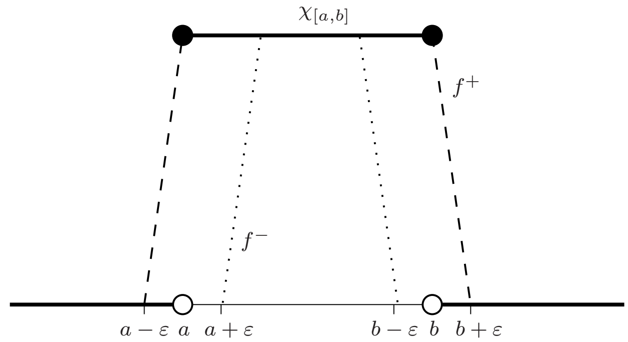


Fig. 4.2: The function $\chi_{[a,b]}$ and the approximations f^- (dots) and f^+ (dashes).

By equidistribution, this implies that

$$\begin{aligned} b-a-2\varepsilon \leqslant \int_0^1 f^- \,\mathrm{d}x \leqslant \liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \chi_{[a,b]}(x_j) \\ \leqslant \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \chi_{[a,b]}(x_j) \leqslant \int_0^1 f^+ \,\mathrm{d}x \leqslant b-a+2\varepsilon. \end{aligned}$$

Thus

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \chi_{[a,b]}(x_j) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \chi_{[a,b]}(x_j) = b-a$$

as required.

Conversely, if (3) holds then (1) holds since any continuous function may be approximated uniformly by a finite linear combination of indicators of intervals*.

(1) $\Longleftrightarrow$ (2): In one direction this is clear; to see that (2) implies (1) it is enough to notice that finite trigonometric polynomials are dense in $C([0,1])$ in the uniform metric. □

Notice that equidistribution of (x_n) does not imply that equation (4.7) holds for measurable functions (but see Exercise 4.4.7).

Example 4.18. [(55)] A consequence of Theorem 4.10 and Example 4.11 is that for any irrational number α, and any initial point $x \in \mathbb{T}$, the orbit $x, R_\alpha x, R_\alpha^2 x, \dots$ under the circle rotation is an equidistributed sequence. Note that this is proved in Example 4.11 by using property (2) of Lemma 4.17.

* We note that the two implications (3) $\implies$ (1) and (2) $\implies$ (1) rely on the same argument, which will be explained in detail in the proof of Corollary 4.20.

4.4.2 Equidistribution and Generic Points

Definition 4.19. If X is a compact metric space, and μ is a Borel probability measure on X, then a sequence (x_n) of elements of X is equidistributed with respect to μ if for any $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f(x_j) = \int_X f(x)\,\mathrm{d}\mu(x).$$

Equivalently, (x_n) is equidistributed if

$$\frac{1}{n}\sum_{j=1}^{n} \delta_{x_j} \longrightarrow \mu$$

in the weak*-topology.

For a continuous transformation $T : X \to X$ and an invariant measure μ we say that $x \in X$ is *generic* (with respect to μ and T) if the sequence of points along the orbit $(T^n x)$ is equidistributed with respect to μ. Notice that if x is generic with respect to one invariant probability measure for T, then x cannot be generic with respect to any other invariant probability measure for T. The following is an easy consequence of the ergodic theorem (Theorem 2.30).

Corollary 4.20. *Let X be a compact metric space, let $T : X \to X$ be a continuous map, and let μ be a T-invariant ergodic probability measure. Then μ-almost every point in X is generic with respect to T and μ.*

Proof. Recall that $C(X)$ is a separable metric space with respect to the uniform norm

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in X\}$$

by Lemma B.8. Let $(f_n)_{n\geqslant 1}$ be a dense sequence in $C(X)$. By applying Theorem 2.30 to each of these functions we obtain one set $X' \subseteq X$ of full measure with the property that

$$\frac{1}{N}\sum_{n=0}^{N-1} f_i(T^n x) \longrightarrow \int_X f_i\,\mathrm{d}\mu$$

for all $i \geqslant 1$ and $x \in X'$. Now let $f \in C(X)$ be any function and fix $\varepsilon > 0$. By the uniform density of the sequence, we may find an $i \in \mathbb{N}$ for which

$$|f(x) - f_i(x)| < \varepsilon$$

for all $x \in X$. Then

$$\int f \,\mathrm{d}\mu - 2\varepsilon \leqslant \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \leqslant \limsup_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \leqslant \int f \,\mathrm{d}\mu + 2\varepsilon,$$

showing convergence of the ergodic averages for f at any $x \in X'$. The limit must be $\int f \,\mathrm{d}\mu$ since $|\int f \,\mathrm{d}\mu - \int f_i \,\mathrm{d}\mu| \leqslant \varepsilon$, so x is a generic point. □

4.4.3 Equidistribution for Irrational Polynomials

Example 4.18 may be thought of as a statement in number theory: for an irrational α, the values of the polynomial $p(n) = x + \alpha n$, when reduced modulo 1, form an equidistributed sequence for any value of x. Weyl [380] generalized this to more general polynomials, and Furstenberg [98] found that this result could also be understood using ergodic theory. We recall the statement of Weyl's polynomial equidistribution Theorem (Theorem 1.4 on p. 4): Let $p(n) = a_k n^k + \cdots + a_0$ be a real polynomial with at least one coefficient among $a_1, \dots, a_k$ irrational. Then the sequence $(p(n))$ is equidistributed modulo 1.

As indicated in Example 4.18, the unique ergodicity of irrational circle rotations proves Theorem 1.4 for $k = 1$. More generally, Theorem 4.10 shows that the orbits of any transformation of the circle for which the Lebesgue measure is the unique invariant measure are equidistributed. In order to apply this to the case of polynomials, we turn to a structural result of Furstenberg [99] that allows more complicated transformations to be built up from simpler ones while preserving a dynamical property (in Chapter 7 a similar approach will be used for another application of ergodic theory).

Notice that by Theorem 4.10, orbits of a uniquely ergodic transformation are equidistributed with respect to the unique invariant measure.

Theorem 4.21 (Furstenberg). *Let $T : X \to X$ be a uniquely ergodic homeomorphism of a compact metric space with unique invariant measure μ. Let G be a compact group* with Haar measure m_G, and let $c : X \to G$ be a continuous map. Define the skew-product map S on $Y = X \times G$ by*

$$S(x, g) = (T(x), c(x)g).$$

If S is ergodic with respect to $\mu \times m_G$, then it is uniquely ergodic.

Proof. To see that S preserves $\mu \times m_G$, let $f \in C(Y)$. Then, by Fubini's theorem,

* The reader may replace G by a torus $\mathbb{T}^k$ with group operation written additively, together with Lebesgue measure $m_{\mathbb{T}^k}$. Notice that in any case the Haar measure is invariant under multiplication on the right or the left since G is compact (see Section C.2).

$$\begin{aligned}\int_Y f\circ S \,\mathrm{d}(\mu\times m_G) &= \int_X\int_G f(Tx, c(x)g)\,\mathrm{d}m_G(g)\,\mathrm{d}\mu(x)\\ &= \int_X\int_G f(Tx,g)\,\mathrm{d}m_G(g)\,\mathrm{d}\mu(x)\\ &= \int_X\int_G f(x,g)\,\mathrm{d}m_G(g)\,\mathrm{d}\mu(x) = \int_Y f\,\mathrm{d}(\mu\times m_G).\end{aligned}$$

Assume that S is ergodic. Let

$$E = \{(x,g) \mid (x,g) \text{ is generic w.r.t. } \mu\times m_G\}.$$

By Corollary 4.20, $\mu\times m_G(E)=1$. We claim that E is invariant under the map $(x,g)\mapsto(x,gh)$. To see this, notice that $(x,g)\in E$ means that

$$\frac{1}{N}\sum_{n=0}^{N-1} f\left(S^n(x,g)\right) \longrightarrow \int f\,\mathrm{d}(\mu\times m_G)$$

for all $f\in C(X\times G)$. Writing $f_h(\cdot,g)=f(\cdot,gh)$, it follows that

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1} f\left(S^n(x,gh)\right) &= \frac{1}{N}\sum_{n=0}^{N-1} f_h\left(S^n(x,g)\right)\\ &\longrightarrow \int f_h\,\mathrm{d}(\mu\times m_G) = \int f\,\mathrm{d}(\mu\times m_G)\end{aligned}$$

since m_G is invariant under multiplication on the right, so $(x,gh)\in E$ also. It follows that $E=E_1\times G$ for some set $E_1\subseteq X, \mu(E_1)=1$. Now assume that ν is an S-invariant ergodic measure on Y. Write $\pi : Y\to X$ for the projection $\pi(x,g)=x$. Then $\pi_*\nu$ is a T-invariant measure, so by unique ergodicity $\pi_*\nu=\mu$. In particular, $\nu(E)=\nu(E_1\times G)=\mu(E_1)=1$. By Corollary 4.20, ν-almost every point is generic with respect to ν. Thus there must be a point $(x,g)\in E$ generic with respect to ν. By definition of E, it follows that $\nu=\mu\times m_G$. □

Corollary 4.22. *Let α be an irrational number. Then the map $S:\mathbb{T}^k\to\mathbb{T}^k$ defined by*

$$S: \begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_k\end{pmatrix} \longmapsto \begin{pmatrix} x_1+\alpha\\ x_2+x_1\\ \vdots\\ x_k+x_{k-1}\end{pmatrix}$$

is uniquely ergodic.

Proof. Notice that the transformation S is built up from the irrational circle map by taking $(k-1)$ skew-product extensions as in Theorem 4.21. By Theorem 4.21, it is sufficient to prove that S is ergodic with respect to

Lebesgue measure on $\mathbb{T}^k$. Let $f \in L^2(\mathbb{T}^k)$ be an S-invariant function, and write

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^k} c_{\mathbf{n}} \mathrm{e}^{2\pi\mathrm{i}\mathbf{n}\cdot\mathbf{x}}$$

for the Fourier expansion of f. Then, since $f(\mathbf{x}) = f(S\mathbf{x})$, we have

$$\sum_{\mathbf{n}\in\mathbb{Z}^k} c_{\mathbf{n}} \mathrm{e}^{2\pi\mathrm{i}\mathbf{n}\cdot S\mathbf{x}} = \sum_{\mathbf{n}\in\mathbb{Z}^k} c_{\mathbf{n}} \mathrm{e}^{2\pi\mathrm{i}n_1\alpha} \mathrm{e}^{2\pi\mathrm{i}S'\mathbf{n}\cdot\mathbf{x}}$$

where

$$S' : \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{k-1} \\ n_k \end{pmatrix} \longmapsto \begin{pmatrix} n_1 + n_2 \\ n_2 + n_3 \\ \vdots \\ n_{k-1} + n_k \\ n_k \end{pmatrix}$$

is an automorphism of $\mathbb{Z}^k$. By the uniqueness of Fourier coefficients,

$$c_{S'\mathbf{n}} = \mathrm{e}^{2\pi\mathrm{i}\alpha n_1} c_{\mathbf{n}}, \tag{4.8}$$

and in particular $|c_{S'\mathbf{n}}| = |c_{\mathbf{n}}|$ for all $\mathbf{n}$. Thus for each $\mathbf{n} \in \mathbb{Z}^k$ we either have $\mathbf{n}, S'\mathbf{n}, (S')^2\mathbf{n}, \dots$ all distinct (in which case $c_{\mathbf{n}} = 0$ since $\sum_{\mathbf{n}} |c_{\mathbf{n}}|^2 < \infty$) or $(S')^p\mathbf{n} = (S')^q\mathbf{n}$ for some $p > q$, so $n_2 = n_3 = \dots = n_k = 0$ (by downward induction on k, for example). Now for $\mathbf{n} = (n_1, 0, \dots, 0)$, equation (4.8) simplifies to $c_{\mathbf{n}} = \mathrm{e}^{2\pi\mathrm{i}n_1\alpha} c_{\mathbf{n}}$, so $n_1 = 0$ or $c_{\mathbf{n}} = 0$. We deduce that f is constant, so S is ergodic. □

Proof of Theorem 1.4. Assume that Theorem 1.4 holds for all polynomials of degree strictly less than k. If a_k is rational, then $qa_k \in \mathbb{Z}$ for some integer q. Then the quantities $p(qn+j)$ modulo 1 for varying n and fixed $j = 0, \dots, q-1$, coincide with the values of polynomials of degree strictly less than k satisfying the hypothesis of the theorem. It follows that the values of each of those polynomials are equidistributed, so the values of the original polynomial are equidistributed modulo 1 by induction. Therefore, we may assume without loss of generality that the leading coefficient a_k is irrational.

A convenient description of the transformation S in Corollary 4.22 comes from viewing $\mathbb{T}^k$ as $\{\alpha\} \times \mathbb{T}^k$ with a map defined by

$$\begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ & 1 & 1 & & \\ & & & \ddots & \\ & & & & 1\ 1 \end{pmatrix} \begin{pmatrix} \alpha \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} \alpha \\ x_1 + \alpha \\ x_2 + x_1 \\ \vdots \\ x_k + x_{k-1} \end{pmatrix}.$$

Iterating this map gives

$$\begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ & 1 & 1 & & \\ & & & \ddots & \\ & & & 1 & 1 \end{pmatrix}^n \begin{pmatrix} \alpha \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ n & 1 & & & \\ \binom{n}{2} & n & 1 & & \\ \vdots & & \ddots & \ddots & \\ \binom{n}{k} & \dots & & n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$$
$$= \begin{pmatrix} \alpha \\ n\alpha + x_1 \\ \binom{n}{2}\alpha + nx_1 + x_2 \\ \vdots \\ \binom{n}{k}\alpha + \binom{n}{k-1}x_1 + \cdots + nx_{k-1} + x_k \end{pmatrix}.$$

Now define $\alpha = k!a_k$, and choose points $x_1, \dots, x_k$ so that

$$p(n) = \binom{n}{k}\alpha + \binom{n}{k-1}x_1 + \cdots + nx_{k-1} + x_k.$$

Then by Corollary 4.22, the orbits of this map are equidistributed on $\mathbb{T}^k$, so the same holds for its last component, which coincides with the sequence of values of $p(n)$ reduced modulo 1 in $\mathbb{T}$. $\square$

An alternative approach in the quadratic case will be described in Exercise 7.4.2.

Exercises for Section 4.4

Exercise 4.4.1. Consider the circle-doubling map $T_2 : x \mapsto 2x \pmod 1$ on $\mathbb{T}$ with Lebesgue measure $m_{\mathbb{T}}$.
(a) Construct a point that is generic for $m_{\mathbb{T}}$.
(b) Construct a point that is generic for a T_2-invariant ergodic measure other than $m_{\mathbb{T}}$.
(c) Construct a point that is generic for a non-ergodic T_2-invariant measure.
(d) Construct a point that is not generic for any T_2-invariant measure.

Exercise 4.4.2. Extend Lemma 4.17 to show that equation (4.7) holds for Riemann-integrable functions (cf. Exercise 4.3.4). Could it hold for Lebesgue-integrable functions?

Exercise 4.4.3. Use Exercise 4.3.4 to show that the fractional parts of the sequence $(n\alpha)$ are uniformly distributed in $[0, 1]$. That is,

$$\frac{|\{n \mid 0 \leqslant n < N, n\alpha - \lfloor n\alpha \rfloor \in [a, b)\}|}{N} \to (b - a)$$

as $N \to \infty$, for any $0 \leqslant a < b \leqslant 1$.

Exercise 4.4.4. Carry out the procedure used in the proof of Theorem 1.4 to prove that the sequence (x_n) defined by $x_n = \begin{pmatrix} \alpha_1 n \\ \alpha_2 n^2 \end{pmatrix}$ is equidistributed in $\mathbb{T}^2$ if and only if $\alpha_1, \alpha_2 \notin \mathbb{Q}$.

Exercise 4.4.5. A number α is called a *Liouville number* if there is an infinite sequence $(\frac{p_n}{q_n})_{n\geqslant 1}$ of rationals with the property that

$$\left|\frac{p_n}{q_n} - \alpha\right| < \frac{1}{q_n^n}$$

for all $n \geqslant 1$. Notice that Exercise 3.3.3 shows that algebraic numbers are not Liouville numbers.
(a) Assuming that α is not a Liouville number, prove the following error rate in the equidistribution of the sequence $(x + n\alpha)_{n\geqslant 1}$ modulo 1:

$$\left|\frac{1}{N}\sum_{n=0}^{N-1} f(x+n\alpha) - \int_0^1 f(x)dx\right| \leqslant S(\alpha, f)\frac{1}{N},$$

for $f \in C^\infty(\mathbb{T})$ and some constant $S(\alpha, f)$ depending on α and f.
(b) Formulate and prove a generalization to rotations of $\mathbb{T}^d$.

Exercise 4.4.6. Use the ideas from Exercise 2.8.4 to prove a mean ergodic theorem along the squares: for a measure-preserving system $(X, \mathscr{B}, \mu, T)$ and $f \in L^2_\mu$, show that

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^{n^2} f$$

converges in L^2_μ. Under the assumption that T is totally ergodic (see Exercise 2.5.6), show that the limit is $\int f \,\mathrm{d}\mu$.

Exercise 4.4.7. Let X be a compact metric space, and assume that $\nu_n \to \mu$ in the weak*-topology on $\mathscr{M}(X)$. Show that for a Borel set B with $\mu(\partial B) = 0$,

$$\lim_{n\to\infty} \nu_n(B) = \mu(B).$$

Notes to Chapter 4

(48) (Page 98) The fact that $\mathscr{M}^T(X)$ is non-empty may also be seen as a result of various fixed-point theorems that generalize the Brouwer fixed point theorem to an infinite-dimensional setting; the argument used in Section 4.1 is attractive because it is elementary and is connected directly to the dynamics.
(49) (Page 103) A convenient source for the Choquet representation theorem is the updated lecture notes by Phelps [283]; the original papers are those of Choquet [55], [56].

(50)(Page 103) Notice that the space of invariant measures for a given continuous map is a topological attribute rather than a measurable one: measurably isomorphic systems may have entirely unrelated spaces of invariant measures. In particular, the Jewett–Krieger theorem shows that any ergodic measure-preserving system $(X, \mathscr{B}, \mu, T)$ on a Lebesgue space is measurably isomorphic to a minimal, uniquely ergodic homeomorphism on a compact metric space (a continuous map on a compact metric space is called *minimal* if every point has a dense orbit; see Exercise 4.2.1). This deep result was found by Jewett [166] for weakly-mixing transformations, and was extended to ergodic systems by Krieger [213] using his proof of the existence of generators [212]. Thus having a model (up to measurable isomorphism) as a uniquely ergodic map on a compact metric space carries no information about a given measurable dynamical system. Among the many extensions and modifications of this important result, Bellow and Furstenberg [22], Hansel and Raoult [140] and Denker [69] gave different proofs; Jakobs [164] and Denker and Eberlein [70] extended the result to flows; Lind and Thouvenot [231] showed that any finite entropy ergodic transformation is isomorphic to a homeomorphism of the torus $\mathbb{T}^2$ preserving Lebesgue measure; Lehrer [222] showed that the homeomorphism can always be chosen to be topologically mixing (a homeomorphism $S : Y \to Y$ of a compact metric space is topologically mixing if for any open sets $U, V \subseteq Y$, there is an $N = N(U, V)$ with $U \cap S^n V \neq \varnothing$ for $n \geqslant N$); Weiss [378] extended to certain group actions and to diagrams of measure-preserving systems; Rosenthal [317] removed the assumption of invertibility. In a different direction, Downarowicz [74] has shown that every possible Choquet simplex arises as the space of invariant measures of a map even in a highly restricted class of continuous maps.

(51)(Page 104) Birkhoff's recurrence theorem may be thought of as a topological analog of Poincaré recurrence (Theorem 2.11), with the essential hypothesis of finite measure replaced by compactness. Furstenberg and Weiss [109] showed that there is also a topological analog of the ergodic multiple recurrence theorem (Theorem 7.4): if (X, T) is minimal and $U \subseteq X$ is open and non-empty, then for any $k > 1$ there is some $n \geqslant 1$ with

$$U \cap T^n U \cap \cdots \cap T^{(k-1)n} U \neq \varnothing.$$

(52)(Page 110) This characterization is due to Pjateckiĭ-Šapiro [285], who showed it as a property characterizing normality for orbits under the map $x \mapsto ax \pmod 1$.

(53)(Page 110) The theory of equidistribution from the viewpoint of number theory is a large and sophisticated one. Extensive overviews of this theory in three different decades may be found in the monographs of Kuipers and Niederreiter [215], Hlawka [154], and Drmota and Tichy [75].

(54)(Page 111) The formulation in (2) is the Weyl criterion for equidistribution; it appears in his paper [380]. Weyl really established the principle that equidistribution can be shown using a sufficiently rich set of test functions; in particular on a compact group it is sufficient to use an appropriate orthonormal basis of L^2. Thus a more general formulation of the Weyl criterion is as follows. Let G be a compact metrizable group and let $G^\sharp$ denote the set of conjugacy classes in G. Then a sequence (g_n) of elements of $G^\sharp$ is equidistributed with respect to Haar measure if and only if

$$\sum_{j=1}^{n} \operatorname{tr}\left(\pi(g_j)\right) = \mathrm{o}(n)$$

as $n \to \infty$, for any non-trivial irreducible unitary representation $\pi : G \to \mathrm{GL}_k(\mathbb{C})$. For more about equidistribution in the number-theoretic context, see the monograph of Iwaniec and Kowalski [162, Ch. 21].

(55)(Page 112) This equidistribution result was proved independently by several people, including Weyl [379], Bohl [39] and Sierpiński [344].

Appendix A: Measure Theory

Complete treatments of the results stated in this appendix may be found in any measure theory book; see for example Parthasarathy [280], Royden [320] or Kingman and Taylor [195]. A similar summary of measure theory without proofs may be found in Walters [373, Chap. 0]. Some of this appendix will use terminology from Appendix B.

A.1 Measure Spaces

Let X be a set, which will usually be infinite, and denote by $\mathbb{P}(X)$ the collection of all subsets of X.

Definition A.1. A set $\mathscr{S} \subseteq \mathbb{P}(X)$ is called a *semi-algebra* if

(1) $\varnothing \in \mathscr{S}$,
(2) $A, B \in \mathscr{S}$ implies that $A \cap B \in \mathscr{S}$, and
(3) if $A \in \mathscr{S}$ then the complement $X \smallsetminus A$ is a finite union of pairwise disjoint elements in $\mathscr{S}$;

if in addition

(4) $A \in \mathscr{S}$ implies that $X \smallsetminus A \in \mathscr{S}$,

then it is called an *algebra*. If $\mathscr{S}$ satisfies the additional property

(5) $A_1, A_2, \dots \in \mathscr{S}$ implies that $\bigcup_{n=1}^{\infty} A_n \in \mathscr{S}$,

then $\mathscr{S}$ is called a σ-algebra. For any collection of sets $\mathscr{A}$, write $\sigma(\mathscr{A})$ for the smallest σ-algebra containing $\mathscr{A}$ (this is possible since the intersection of σ-algebras is a σ-algebra).

Example A.2. The collection of intervals in $[0, 1]$ forms a semi-algebra.

Definition A.3. A collection $\mathscr{M} \subseteq \mathbb{P}(X)$ is called a *monotone class* if

$$A_1 \subseteq A_2 \subseteq \cdots \text{ and } A_n \in \mathscr{M} \text{ for all } n \geqslant 1 \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{M}$$

and

$$B_1 \supseteq B_2 \supseteq \cdots \text{ and } B_n \in \mathscr{M} \text{ for all } n \geqslant 1 \implies \bigcap_{n=1}^{\infty} B_n \in \mathscr{M}.$$

The intersection of two monotone classes is a monotone class, so there is a well-defined smallest monotone class $\mathscr{M}(\mathscr{A})$ containing any given collection of sets $\mathscr{A}$. This gives an alternative characterization of the σ-algebra generated by an algebra.

Theorem A.4. *Let $\mathscr{A}$ be an algebra. Then the smallest monotone class containing $\mathscr{A}$ is $\sigma(\mathscr{A})$.*

A function $\mu : \mathscr{S} \to \mathbb{R}_{\geqslant 0} \cup \{\infty\}$ is *finitely additive* if $\mu(\varnothing) = 0$ and*

$$\mu(A \cup B) = \mu(A) + \mu(B) \tag{A.1}$$

for any disjoint elements A and B of $\mathscr{S}$ with $A \sqcup B \in \mathscr{S}$, and is *countably additive* if

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

if $\{A_n\}$ is a collection of disjoint elements of $\mathscr{S}$ with $\bigsqcup_{n=1}^{\infty} A_n \in \mathscr{S}$.

The main structure of interest in ergodic theory is that of a *probability space* or *finite measure space*.

Definition A.5. A triple $(X, \mathscr{B}, \mu)$ is called a finite measure space if $\mathscr{B}$ is a σ-algebra and μ is a countably additive measure defined on $\mathscr{B}$ with $\mu(X) < \infty$. A triple $(X, \mathscr{B}, \mu)$ is called a σ-finite measure space if X is a countable union of elements of $\mathscr{B}$ of finite measure. If $\mu(X) = 1$ then a finite measure space is called a probability space.

A probability measure μ is said to be *concentrated* on a measurable set A if $\mu(A) = 1$.

Theorem A.6. *If $\mu : \mathscr{S} \to \mathbb{R}_{\geqslant 0}$ is a countably additive measure defined on a semi-algebra, then there is a unique countably additive measure defined on $\sigma(\mathscr{S})$ which extends μ.*

* The conventions concerning the symbol ∞ in this setting are that $\infty + c = \infty$ for any c in $\mathbb{R}_{\geqslant 0} \cup \{\infty\}$, $c \cdot \infty = \infty$ for any $c > 0$, and $0 \cdot \infty = 0$.

Theorem A.7. *Let $\mathscr{A} \subseteq \mathscr{B}$ be an algebra in a probability space $(X, \mathscr{B}, \mu)$. Then the collection of sets B with the property that for any $\varepsilon > 0$ there is an $A \in \mathscr{A}$ with $\mu(A \triangle B) < \varepsilon$ is a σ-algebra.*

As discussed in Section 2.1, the basic objects of ergodic theory are measure-preserving maps (see Definition 2.1). The next result gives a convenient way to check whether a transformation is measure-preserving.

Theorem A.8. *Let $(X, \mathscr{B}_X, \mu)$ and $(Y, \mathscr{B}_Y, \nu)$ be probability spaces, and let $\mathscr{S}$ be a semi-algebra which generates $\mathscr{B}_Y$. A measurable map $\phi : X \to Y$ is* measure-preserving *if and only if*

$$\mu(\phi^{-1}B) = \nu(B)$$

for all $B \in \mathscr{S}$.

Proof. Let

$$\mathscr{S}' = \{B \in \mathscr{B}_Y \mid \phi^{-1}(B) \in \mathscr{B}_X, \mu(\phi^{-1}B) = \nu(B)\}.$$

Then $\mathscr{S} \subseteq \mathscr{S}'$, and (since each member of the algebra generated by $\mathscr{S}$ is a finite disjoint union of elements of $\mathscr{S}$) the algebra generated by $\mathscr{S}$ lies in $\mathscr{S}'$. It is clear that $\mathscr{S}'$ is a monotone class, so Theorem A.4 shows that $\mathscr{S}' = \mathscr{B}_Y$ as required. □

The next result is an important lemma from probability; what it means is that if the sum of the probabilities of a sequence of events is finite, then the probability that infinitely many of them occur is zero.

Theorem A.9 (Borel–Cantelli[102]). *Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $(A_n)_{n \geqslant 1}$ be a sequence of measurable sets with $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. Then*

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right)\right) = 0.$$

If the sequence of sets are pairwise independent, that is if

$$\mu(A_i \cap A_j) = \mu(A_i)\mu(A_j)$$

for all $i \neq j$, then $\sum_{n=1}^{\infty} \mu(A_n) = \infty$ implies that

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right)\right) = 1.$$

The elements of a σ-algebra are typically very complex, and it is often enough to approximate sets by a convenient smaller collection of sets.

Theorem A.10. *If $(X, \mathscr{B}, \mu)$ is a probability space and $\mathscr{A}$ is an algebra which generates $\mathscr{B}$ (that is, with $\sigma(\mathscr{A}) = \mathscr{B}$), then for any $B \in \mathscr{B}$ and $\varepsilon > 0$ there is an $A \in \mathscr{A}$ with $\mu(A \triangle B) < \varepsilon$.*

A measure space is called *complete* if any subset of a null set is measurable.

If X is a topological space, then there is a distinguished collection of sets to start with, namely the open sets. The σ-algebra generated by the open sets is called the *Borel σ-algebra*. If the space is second countable, then the *support* of a measure is the largest closed set with the property that every open neighborhood of every point in the set has positive measure; equivalently the support of a measure is the complement of the largest open set of zero measure.

If X is a metric space, then any Borel probability measure μ on X (that is, any probability measure defined on the Borel σ-algebra $\mathscr{B}$ of X) is *regular*[(103)]: for any Borel set $B \subseteq X$ and $\varepsilon > 0$ there is an open set O and a closed set C with $C \subseteq B \subseteq O$ and $\mu(O \smallsetminus C) < \varepsilon$.

A.2 Product Spaces

Let $I \subseteq \mathbb{Z}$ and assume that for each $i \in I$ a probability space $\mathsf{X}_i = (X_i, \mathscr{B}_i, \mu_i)$ is given. Then the product space $X = \prod_{i \in I} X_i$ may be given the structure of a probability space $(X, \mathscr{B}, \mu)$ as follows. Any set of the form

$$\prod_{i \in I, i < \min(F)} X_i \times \prod_{i \in F} A_i \times \prod_{i \in I, i > \max(F)} X_i,$$

or equivalently of the form

$$\{x = (x_i)_{i \in I} \in X \mid x_i \in A_i \text{ for } i \in F\},$$

for some finite set $F \subseteq I$, is called a *measurable rectangle*. The collection of all measurable rectangles forms a semi-algebra $\mathscr{S}$, and the product σ-algebra is $\mathscr{B} = \sigma(\mathscr{S})$. The product measure μ is obtained by defining the measure of the measurable rectangle above to be $\prod_{i \in F} \mu_i(A_i)$ and then extending to $\mathscr{B}$.

The main extension result in this setting is the Kolmogorov consistency theorem, which allows measures on infinite product spaces to be built up from measures on finite product spaces.

Theorem A.11. *Let $X = \prod_{i \in I} X_i$ with $I \subseteq \mathbb{Z}$ and each X_i a probability space. Suppose that for every finite subset $F \subseteq I$ there is a probability measure μ_F defined on $X_F = \prod_{i \in F} X_i$, and that these measures are consistent in the sense that if $E \subseteq F$ then the projection map*

$$\left(\prod_{i \in F} X_i, \mu_F\right) \longrightarrow \left(\prod_{i \in E} X_i, \mu_E\right)$$

is measure-preserving. Then there is a unique probability measure μ on the probability space $\prod_{i \in I} X_i$ with the property that for any $F \subseteq I$ the projection *map*

$$\left(\prod_{i\in I} X_i, \mu\right) \longrightarrow \left(\prod_{i\in F} X_i, \mu_F\right)$$

is measure-preserving.

In the construction of an infinite product $\prod_{i\in I}\mu_i$ of probability measures above, the finite products $\mu_F = \prod_{i\in F}\mu_i$ satisfy the compatibility conditions needed in Theorem A.11.

In many situations each $X_i = (X_i, \mathsf{d}_i)$ is a fixed compact metric space with $0 < \operatorname{diam}(X_i) < \infty$. In this case the product space $X = \prod_{n\in\mathbb{Z}} X_n$ is also a compact metric space with respect to the metric

$$\mathsf{d}(x,y) = \sum_{n\in\mathbb{Z}} \frac{d_n(x_n,y_n)}{2^n \operatorname{diam}(X_n)},$$

and the Borel σ-algebra of X coincides with the product σ-algebra defined above.

A.3 Measurable Functions

Let $(X,\mathscr{B},\mu)$ be a probability space. Natural classes of measurable functions on X are built up from simpler functions, just as the σ-algebra $\mathscr{B}$ may be built up from simpler collections of sets.

A function $f : X \to \mathbb{R}$ is called *simple* if

$$f(x) = \sum_{j=1}^{m} c_j \chi_{A_j}(x)$$

for constants $c_j \in \mathbb{R}$ and disjoint sets $A_j \in \mathscr{B}$. The integral of f is then defined to be

$$\int f \,\mathrm{d}\mu = \sum_{j=1}^{m} c_j \mu(A_j).$$

A function $g : X \to \mathbb{R}$ is called *measurable* if $g^{-1}(A) \in \mathscr{B}$ for any (Borel) measurable set $A \subseteq \mathbb{R}$. The basic approximation result states that for any measurable function $g : X \to \mathbb{R}_{\geqslant 0}$ there is a pointwise increasing sequence of simple functions $(f_n)_{n\geqslant 1}$ with $f_n(x) \nearrow g(x)$ for each $x \in X$. This allows us to define

$$\int g \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu,$$

which is guaranteed to exist since

$$f_n(x) \leqslant f_{n+1}(x)$$

for all $n \geqslant 1$ and $x \in X$ (in contrast to the usual terminology from calculus, we include the possibility that the integral and the limit are infinite). It may be shown that this is well-defined (independent of the choice of the sequence of simple functions).

A measurable function $g : X \to \mathbb{R}_{\geqslant 0}$ is *integrable* if $\int g \,\mathrm{d}\mu < \infty$. In general, a measurable function $g : X \to \mathbb{R}$ has a unique decomposition into $g = g^+ - g^-$ with $g^+(x) = \max\{g(x), 0\}$; both g^+ and g^- are measurable. The function g is said to be integrable if both g^+ and g^- are integrable, and the integral is defined by $\int g \,\mathrm{d}\mu = \int g^+ \,\mathrm{d}\mu - \int g^- \,\mathrm{d}\mu$. If f is integrable and g is measurable with $|g| \leqslant f$, then g is integrable. The integral of an integrable function f over a measurable set A is defined by

$$\int_A f \,\mathrm{d}\mu = \int f \chi_A \,\mathrm{d}\mu.$$

For $1 \leqslant p < \infty$, the space $\mathscr{L}^p_\mu$ (or $\mathscr{L}^p(X)$, $\mathscr{L}^p(X, \mu)$ and so on) comprises the measurable functions $f : X \to \mathbb{R}$ with $\int |f|^p \,\mathrm{d}\mu < \infty$. Define an equivalence relation on $\mathscr{L}^p_\mu$ by $f \sim g$ if $\int |f - g|^p \,\mathrm{d}\mu = 0$ and write $L^p_\mu = \mathscr{L}^p_\mu / \sim$ for the space of equivalence classes. Elements of L^p_μ will be described as functions rather than equivalence classes, but it is important to remember that this is an abuse of notation (for example, in the construction of conditional measures on page 138). In particular the value of an element of L^p_μ at a specific point does not make sense, unless that point itself has positive μ-measure. The function $\|\cdot\|_p$ defined by

$$\|f\|_p = \left(\int |f|^p \,\mathrm{d}\mu \right)^{1/p}$$

is a norm (see Appendix B), and under this norm L^p is a Banach space.

The case $p = \infty$ is distinguished: the *essential supremum* is the generalization to measurable functions of the supremum of a continuous function, and is defined by

$$\|f\|_\infty = \inf \left\{ \alpha \mid \mu\left(\{x \in X \mid f(x) > \alpha\}\right) = 0 \right\}.$$

The space $\mathscr{L}^\infty_\mu$ is then defined to be the space of measurable functions f with $\|f\|_\infty < \infty$, and once again L^∞_μ is defined to be $\mathscr{L}^\infty_\mu / \sim$. The norm $\|\cdot\|_\infty$ makes L^∞_μ into a Banach space. For $1 \leqslant p < q \leqslant \infty$ we have $L^p \supseteq L^q$ for any finite measure space, with strict inclusion except in some degenerate cases.

In practice we will more often use $\mathscr{L}^\infty$, which denotes the bounded functions.

An important consequences of the Borel–Cantelli lemma is that norm convergence in L^p forces pointwise convergence along a subsequence.

Corollary A.12. *If (f_n) is a sequence convergent in L^p_μ $(1 \leqslant p \leqslant \infty)$ to f, then there is a subsequence (f_{n_k}) converging pointwise almost everywhere to f.*

Proof. Choose the sequence (n_k) so that

$$\|f_{n_k} - f\|_p^p < \frac{1}{k^{2+p}}$$

for all $k \geqslant 1$. Then

$$\mu\left(\left\{x \in X \,\Big|\, |f_{n_k}(x) - f(x)| > \frac{1}{k}\right\}\right) < \frac{1}{k^2}.$$

It follows by Theorem A.9 that for almost every x, $|f_{n_k}(x) - f(x)| > \frac{1}{k}$ for only finitely many k, so $f_{n_k}(x) \to f(x)$ for almost every x. □

Finally we turn to integration of functions of several variables; a measure space $(X, \mathscr{B}, \mu)$ is called σ-finite if there is a sequence $A_1, A_2, \dots$ of measurable sets with $\mu(A_n) < \infty$ for all $n \geqslant 1$ and with $X = \bigcup_{n \geqslant 1} A_n$.

Theorem A.13 (Fubini–Tonelli[(104)]). *Let f be a non-negative integrable function on the product of two σ-finite measure spaces $(X, \mathscr{B}, \mu)$ and $(Y, \mathscr{C}, \nu)$. Then, for almost every $x \in X$ and $y \in Y$, the functions*

$$h(x) = \int_Y f(x, y)\,\mathrm{d}\nu, \quad g(y) = \int_X f(x, y)\,\mathrm{d}\mu$$

are integrable, and

$$\int_{X \times Y} f\,\mathrm{d}(\mu \times \nu) = \int_X h\,\mathrm{d}\mu = \int_Y g\,\mathrm{d}\nu. \tag{A.2}$$

This may also be written in a more familiar form as

$$\begin{aligned}\int_{X \times Y} f(x, y)\,\mathrm{d}(\mu \times \nu)(x, y) &= \int_X \left(\int_Y f(x, y)\,\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x) \\ &= \int_Y \left(\int_X f(x, y)\,\mathrm{d}\mu(x)\right)\mathrm{d}\nu(y).\end{aligned}$$

We note that integration makes sense for functions taking values in some other spaces as well, and this will be discussed further in Section B.7.

A.4 Radon–Nikodym Derivatives

One of the fundamental ideas in measure theory concerns the properties of a probability measure viewed from the perspective of a given measure. Fix a σ-finite measure space $(X, \mathscr{B}, \mu)$ and some measure ν defined on $\mathscr{B}$.

- The measure ν is *absolutely continuous* with respect to μ, written $\nu \ll \mu$, if $\mu(A) = 0 \implies \nu(A) = 0$ for any $A \in \mathscr{B}$.

- If $\nu \ll \mu$ and $\mu \ll \nu$ then μ and ν are said to be *equivalent*.
- The measures μ and ν are mutually *singular*, written $\mu \perp \nu$, if there exist disjoint sets A and B in $\mathscr{B}$ with $A \cup B = X$ and with $\mu(A) = \nu(B) = 0$.

These notions are related by two important theorems.

Theorem A.14 (Lebesgue decomposition). *Given σ-finite measures μ and ν on $(X, \mathscr{B})$, there are measures ν_0 and ν_1 with the properties that*

(1) $\nu = \nu_0 + \nu_1$;
(2) $\nu_0 \ll \mu$; *and*
(3) $\nu_1 \perp \mu$.

The measures ν_0 and ν_1 are uniquely determined by these properties.

Theorem A.15 (Radon–Nikodym derivative[105]). *If $\nu \ll \mu$ then there is a measurable function $f \geqslant 0$ on X with the property that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu$$

for any set $A \in \mathscr{B}$.

By analogy with the fundamental theorem of calculus (Theorem A.25), the function f is written $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and is called the *Radon–Nikodym derivative* of ν with respect to μ. Notice that for any two measures μ_1, μ_2 we can form a new measure $\mu_1 + \mu_2$ simply by defining $(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A)$ for any measurable set A. Then $\mu_i \ll \mu_1 + \mu_2$, so there is a Radon–Nikodym derivative of μ_i with respect to $\mu_1 + \mu_2$ for $i = 1, 2$.

A.5 Convergence Theorems

The most important distinction between integration on L^p spaces as defined above and Riemann integration on bounded Riemann-integrable functions is that the L^p functions are closed under several natural limiting operations, allowing for the following important convergence theorems.

Theorem A.16 (Monotone Convergence Theorem). *If $f_1 \leqslant f_2 \leqslant \cdots$ is a pointwise increasing sequence of integrable functions on the probability space $(X, \mathscr{B}, \mu)$, then $f = \lim_{n\to\infty} f_n$ satisfies*

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

In particular, if $\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu < \infty$, then f is finite almost everywhere.

Theorem A.17 (Fatou's Lemma). *Let $(f_n)_{n\geqslant 1}$ be a sequence of measurable real-valued functions on a probability space, all bounded below by some integrable function. If $\liminf_{n\to\infty}\int f_n\,\mathrm{d}\mu<\infty$ then $\liminf_{n\to\infty} f_n$ is integrable, and*

$$\int \liminf_{n\to\infty} f_n\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty}\int f_n\,\mathrm{d}\mu.$$

Theorem A.18 (Dominated Convergence Theorem). *If $h : X \to \mathbb{R}$ is an integrable function and $(f_n)_{n\geqslant 1}$ is a sequence of measurable real-valued functions which are dominated by h in the sense that $|f_n| \leqslant h$ for all $n \geqslant 1$, and $\lim_{n\to\infty} f_n = f$ exists almost everywhere, then f is integrable and*

$$\int f\,\mathrm{d}\mu = \lim_{n\to\infty}\int f_n\,\mathrm{d}\mu.$$

A.6 Well-behaved Measure Spaces

It is convenient to slightly extend the notion of a Borel probability space as follows (cf. Definition 5.13).

Definition A.19. Let X be a dense Borel subset of a compact metric space $\overline{X}$, with a probability measure μ defined on the restriction of the Borel σ-algebra $\mathscr{B}$ to X. The resulting probability space $(X,\mathscr{B},\mu)$ is a *Borel probability space**.

For our purposes, this is the most convenient notion of a measure space that is on the one hand sufficiently general for the applications needed, while on the other has enough structure to permit explicit and convenient proofs.

A circle of results called Lusin's theorem [237] (or Luzin's theorem) show that measurable functions are continuous off a small set. These results are true in almost any context where continuity makes sense, but we state a form of the result here in the setting needed.

Theorem A.20 (Lusin). *Let $(X,\mathscr{B},\mu)$ be a Borel probability space and let $f : X \to \mathbb{R}$ be a measurable function. Then, for any $\varepsilon > 0$, there is a continuous function $g : X \to \mathbb{R}$ with the property that*

$$\mu\left(\{x \in X \mid f(x) \neq g(x)\}\right) < \varepsilon.$$

As mentioned in the endnote to Definition 5.13, there is a slightly different formulation of the standard setting for ergodic theory, in terms of Lebesgue spaces.

---

* Commonly the σ-algebra $\mathscr{B}$ is enlarged to its *completion* $\mathscr{B}_\mu$, which is the smallest σ-algebra containing both $\mathscr{B}$ and all subsets of null sets with respect to μ. It is also standard to allow any probability space that is isomorphic to $(X,\mathscr{B}_\mu,\mu)$ in Definition A.19 as a measure space to be called a Lebesgue space.

Definition A.21. A probability space is a *Lebesgue space* if it is isomorphic as a measure space to

$$\left([0,s]\sqcup A, \mathscr{B}, m_{[0,s]} + \sum_{a\in A} p_a\delta_a\right)$$

for some countable set A of *atoms* and numbers $s, p_a > 0$ with

$$s + \sum_{a\in A} p_a = 1,$$

where $\mathscr{B}$ comprises unions of Lebesgue measurable sets in $[0,s]$ and arbitrary subsets of A, $m_{[0,s]}$ is the Lebesgue measure on $[0,s]$, and δ_a is the Dirac measure defined by $\delta_a(B) = \chi_B(a)$.

The next result shows, *inter alia*, that this notion agrees with that used in Definition A.19 (a proof of this may be found in the book of Parthasarathy [280, Chap. V]) up to completion of the measure space (a measure space is complete if all subsets of a null set are measurable and null). We will not use this result here.

Theorem A.22. *A probability space is a Lebesgue space in the sense of Definition* A.21 *if and only if it is isomorphic to $(X, \mathscr{B}_\mu, \mu)$ for some probability measure μ on the completion $\mathscr{B}_\mu$ of the Borel σ-algebra $\mathscr{B}$ of a complete separable metric space X.*

The function spaces from Section A.3 are particularly well-behaved for Lebesgue spaces.

Theorem A.23 (Riesz–Fischer[(106)]). *Let $(X, \mathscr{B}, \mu)$ be a Lebesgue space. For any p, $1 \leqslant p < \infty$, the space L^p_μ is a separable Banach space with respect to the $\|\cdot\|_p$-norm. In particular, L^2_μ is a separable Hilbert space.*

A.7 Lebesgue Density Theorem

The space $\mathbb{R}$ together with the usual metric and Lebesgue measure $m_{\mathbb{R}}$ is a particularly important and well-behaved special case, and here it is possible to say that a set of positive measure is thick in a precise sense.

Theorem A.24 (Lebesgue[(107)]). *If $A \subseteq \mathbb{R}$ is a measurable set, then*

$$\lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} m_{\mathbb{R}}\left(A \cap (a-\varepsilon, a+\varepsilon)\right) = 1$$

for $m_{\mathbb{R}}$-almost every $a \in A$.

A point a with this property is said to be a *Lebesgue density point* or a *point with Lebesgue density* 1. An equivalent and more familiar formulation of the result is a form of the fundamental theorem of calculus.

Theorem A.25. *If $f : \mathbb{R} \to \mathbb{R}$ is an integrable function then*

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_s^{s+\varepsilon} f(t)\,\mathrm{d}t = f(s)$$

for $m_{\mathbb{R}}$-almost every $s \in [0, \infty)$.

The equivalence of Theorem A.24 and A.25 may be seen by approximating an integrable function with simple functions.

A.8 Substitution Rule

Let $O \subseteq \mathbb{R}^n$ be an open set, and let $\phi : O \to \mathbb{R}^n$ be a C^1-map with Jacobian $J_\phi = |\det \mathrm{D}\,\phi|$. Then for any measurable function $f \geqslant 0$ (or for any integrable function f) defined on $\phi(O) \subseteq \mathbb{R}^n$ we have[(108)].

$$\int_O f(\phi(\mathbf{x})) J_\phi(\mathbf{x})\,\mathrm{d}m_{\mathbb{R}^n}(\mathbf{x}) = \int_{\phi(O)} f(\mathbf{y})\,\mathrm{d}m_{\mathbb{R}^n}(\mathbf{y}). \tag{A.3}$$

We recall the definition of the push-forward of a measure. Let $(X, \mathscr{B}_X)$ and $(Y, \mathscr{B}_Y)$ be two spaces equipped with σ-algebras. Let μ be a measure on X defined on $\mathscr{B}_X$, and let $\phi : X \to Y$ be measurable. Then the push-forward $\phi_*\mu$ is the measure on $(Y, \mathscr{B}_Y)$ defined by $(\phi_*\mu)(B) = \mu(\phi^{-1}(B))$ for all $B \in \mathscr{B}_Y$.

The substitution rule allows us to calculate the push-forward of the Lebesgue measure under smooth maps as follows.

Lemma A.26. *Let $O \subseteq \mathbb{R}^n$ be open, let $\phi : O \to \mathbb{R}^n$ be a smooth injective map with non-vanishing Jacobian $J_\phi = |\det \mathrm{D}\,\phi|$. Then the push-forward $\phi_* m_O$ of the Lebesgue measure $m_O = m_{\mathbb{R}^n}|_O$ restricted to O is absolutely continuous with respect to $m_{\mathbb{R}^n}$ and is given by*

$$\mathrm{d}\phi_* m_O = J_\phi^{-1} \circ \phi^{-1}\,\mathrm{d}m_{\phi(O)}.$$

Moreover, if we consider a measure $\mathrm{d}\mu = F\,\mathrm{d}m_O$ absolutely continuous with respect to m_O, then similarly

$$\mathrm{d}\phi_*\mu = F \circ \phi^{-1} J_\phi^{-1} \circ \phi^{-1}\,\mathrm{d}m_{\phi(O)}.$$

Proof. Recall that under the assumptions of the lemma, ϕ^{-1} is smooth and $J_{\phi^{-1}} = J_\phi^{-1} \circ \phi^{-1}$. Therefore, by equation (A.3) and the definition of the push-forward,

$$\begin{aligned}\int_{\phi(O)} f(x) J_\phi^{-1}\left(\phi^{-1}(x)\right) \,\mathrm{d}m_{\mathbb{R}^n}(x) &= \int_{\phi(O)} f\left(\phi(\phi^{-1}(x))\right) J_{\phi^{-1}}(x) \,\mathrm{d}m_{\mathbb{R}^n}(x)\\ &= \int_O f(\phi(y)) \,\mathrm{d}m_{\mathbb{R}^n}(y)\\ &= \int_{\phi(O)} f(x) \,\mathrm{d}\phi_* m_O(x)\end{aligned}$$

for any characteristic function $f = \chi_B$ of a measurable set $B \subseteq \phi(O)$. This implies the first claim. Moreover, for any measurable functions $f \geqslant 0, F \geqslant 0$ defined on $\phi(O), O$ respectively,

$$\int_{\phi(O)} f(x)F(\phi^{-1}(x)) J_\phi^{-1}(\phi^{-1}(x)) \,\mathrm{d}m_{\mathbb{R}^n}(x) = \int_O f(\phi(y))F(y) \,\mathrm{d}m_{\mathbb{R}^n},$$

which implies the second claim. □

Notes to Appendix A

(102)(Page 405) This result was stated by Borel [40, p. 252] for independent events as part of his study of normal numbers, but as pointed out by Barone and Novikoff [18] there are some problems with the proofs. Cantelli [46] noticed that half of the theorem holds without independence; this had also been noted by Hausdorff [142] in a special case. Erdős and Rényi [84] showed that the result holds under the much weaker assumption of pairwise independence.

(103)(Page 406) This is shown, for example, in Parthasarathy [280, Th. 1.2]: defining a Borel set A to be regular if, for any $\varepsilon > 0$, there is an open set O_ε and a closed set C_ε with $C_\varepsilon \subseteq A \subseteq O_\varepsilon$ and $\mu(O_\varepsilon \smallsetminus C_\varepsilon) < \varepsilon$, it may be shown that the collection of all regular sets forms a σ-algebra and contains the closed sets.

(104)(Page 409) A form of this theorem goes back to Cauchy for continuous functions on the reals, and this was extended by Lebesgue [220] to bounded measurable functions. Fubini [97] extended this to integrable functions, showing that if $f : [a,b] \times [c,d] \to \mathbb{R}$ is integrable then $y \mapsto f(x,y)$ is integrable for almost every x, and proving equation (A.2). Tonelli [362] gave the formulation here, for non-negative functions on products of σ-finite spaces. Complete proofs may be found in Royden [320] or Lieb and Loss [229, Th. 1.12]. While the result is robust and of central importance, some hypotheses are needed: if the function is not integrable or the spaces are not σ-finite, the integrals may have different values. A detailed treatment of the minimal hypotheses needed for a theorem of Fubini type, along with counterexamples and applications, is given by Fremlin [96, Sect. 252].

(105)(Page 410) This result is due to Radon [297] when μ is Lebesgue measure on $\mathbb{R}^n$, and to Nikodym [272] in the general case.

(106)(Page 412) This result emerged in several notes of Riesz and two notes of Fischer [91], [92], with a full treatment of the result that $L^2(\mathbb{R})$ is complete appearing in a paper of Riesz [311].

(107)(Page 412) This is due to Lebesgue [220], and a convenient source for the proof is the monograph of Oxtoby [276]. Notice that Theorem A.24 expresses how constrained measurable sets are: it is impossible, for example, to find a measurable subset A of $[0,1]$ with the property that $m_{\mathbb{R}}(A \cap [a,b]) = \frac{1}{2}(b-a)$ for all $b > a$. While a measurable subset of measure $\frac{1}{2}$ may have an intricate structure, it cannot occupy only half of the space on all possible scales.

(108)(Page 413) The usual hypotheses are that the map ϕ is injective and the Jacobian non-vanishing; these may be relaxed considerably, and the theorem holds in very general settings both measurable (see Hewitt and Stromberg [152]) and smooth (see Spivak [349]).

Appendix B: Functional Analysis

Functional analysis abstracts the basic ideas of real and complex analysis in order to study spaces of functions and operators between them[109]. A *normed space* is a vector space E over a field $\mathbb{F}$ (either $\mathbb{R}$ or $\mathbb{C}$) equipped with a map $\|\cdot\|$ from $E \to \mathbb{R}$ satisfying the properties

- $\|x\| \geqslant 0$ for all $x \in E$ and $\|x\| = 0$ if and only if $x = 0$;
- $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in E$ and $\lambda \in \mathbb{F}$; and
- $\|x + y\| \leqslant \|x\| + \|y\|$.

If $(E, \|\cdot\|)$ is a normed space, then $\mathsf{d}(x, y) = \|x - y\|$ defines a metric on E. A *semi-norm* is a map with the first property weakened to

- $\|x\| \geqslant 0$ for all $x \in E$.

A normed space is a *Banach space* if it is complete as a metric space: that is, the condition that the sequence (x_n) is Cauchy (for all $\varepsilon > 0$ there is some N for which $m > n > N$ implies $\|x_m - x_n\| < \varepsilon$) is equivalent to the condition that the sequence (x_n) converges (there is some $y \in E$ with the property that for all $\varepsilon > 0$ there is some N for which $n > N$ implies $\|x_n - y\| < \varepsilon$).

As discussed in Section A.3, for any probability space $(X, \mathscr{B}, \mu)$, the norm $\|\cdot\|_p$ makes the space L^p_μ into a Banach space.

B.1 Sequence Spaces

For $1 \leqslant p < \infty$ and a countable set Γ (in practice this will be $\mathbb{N}$ or $\mathbb{Z}$) we denote by $\ell^p(\Gamma)$ the space

$$\{x = (x_\gamma) \in \mathbb{R}^\Gamma \mid \sum_{\gamma \in \Gamma} |x_\gamma|^p < \infty\},$$

and for $p = \infty$ write

$$\ell^\infty(\Gamma) = \{x = (x_\gamma) \in \mathbb{R}^\Gamma \mid \sup_{\gamma\in\Gamma} |x_\gamma| < \infty\}.$$

The norms $\|x\|_p = \left(\sum_{\gamma\in\Gamma} |x_\gamma|^p\right)^{1/p}$ and $\|x\|_\infty = \sup_{\gamma\in\Gamma} |x_\gamma|$ make $\ell_p(\Gamma)$ into a complete space for $1 \leqslant p \leqslant \infty$.

B.2 Linear Functionals

A vector space V over a normed field $\mathbb{F}$, equipped with a topology τ, and with the property that each point of V is closed and the vector space operations (addition of vectors and multiplication by scalars) are continuous is called a *topological vector space*. Any topological vector space is Hausdorff. If $0 \in V$ has an open neighborhood with compact closure, then V is said to be locally compact.

Let $\lambda : V \to W$ be a linear map between topological vector spaces. Then the following properties are equivalent:

(1) λ is continuous;
(2) λ is continuous at $0 \in V$;
(3) λ is uniformly continuous in the sense that for any neighborhood O_W of $0 \in W$ there is a neighborhood O_V of $0 \in V$ for which $v - v' \in O_V$ implies $\lambda(v) - \lambda(v') \in O_W$ for all $v, v' \in V$.

Of particular importance are linear maps into the ground field. For a linear map $\lambda : V \to \mathbb{F}$, the following properties are equivalent:

(1) λ is continuous;
(2) the kernel $\ker(\lambda) = \{v \in V \mid \lambda(v) = 0\}$ is a closed subset of V;
(3) $\ker(\lambda)$ is not dense in V;
(4) λ is bounded on some neighborhood of $0 \in V$.

Continuous linear maps $\lambda : V \to \mathbb{F}$ are particularly important: they are called *linear functionals* and the collection of all linear functionals is denoted V^*. If V has a norm $\|\cdot\|$ defining the topology τ, then V^* is a normed space under the norm

$$\|\lambda\|_{\text{operator}} = \sup_{\|v\|=1} \{|\lambda(v)|_{\mathbb{F}}\}$$

where $|\cdot|_{\mathbb{F}}$ is the norm on the ground field $\mathbb{F}$. The normed space V^* is complete if $\mathbb{F}$ is complete. The next result asserts that there are many linear functionals, and allows them to be constructed in a flexible and controlled way.

Theorem B.1 (Hahn–Banach[(110)]). *Let $\lambda : U \to \mathbb{F}$ be a linear functional defined on a subspace $U \subseteq V$ of a normed linear space and let*

$$p : V \to \mathbb{R}_{\geqslant 0}$$

be a semi-norm. If $|f(u)| \leqslant p(u)$ for $u \in U$, then there is a linear functional $\lambda' : V \to \mathbb{F}$ that extends λ in the sense that $\lambda'(u) = \lambda(u)$ for all $u \in U$, and $|\lambda'(v)| \leqslant p(v)$ for all $v \in V$.

B.3 Linear Operators

It is conventional to call maps between normed spaces *operators*, because in many of the applications the elements of the normed spaces are themselves functions. A map $f : E \to F$ between normed vector spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ is *continuous at a* if for any $\varepsilon > 0$ there is some $\delta > 0$ for which

$$\|x - a\|_E < \delta \implies \|f(x) - f(a)\|_F < \varepsilon,$$

is *continuous* if it is continuous at every point, and is *bounded* if there is some R with $\|f(x)\|_F \leqslant R\|x\|_E$ for all $x \in E$. If $f : E \to F$ is linear, then the following are equivalent:

- f is continuous;
- f is bounded;
- f is continuous at $0 \in E$.

A linear map $f : E \to F$ is an *isometry* if $\|f(x)\|_F = \|x\|_E$ for all $x \in E$, and is an *isomorphism of normed spaces* if f is a bijection and both f and f^{-1} are continuous.

Norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on E are *equivalent* if the identity map

$$(E, \|\cdot\|_1) \to (E, \|\cdot\|_2)$$

is an isomorphism of normed spaces; equivalently, if there are positive constants r, R for which

$$r\|x\|_1 \leqslant \|x\|_2 \leqslant R\|x\|_1$$

for all $x \in E$. If E, F are finite-dimensional, then all norms on E are equivalent and all linear maps $E \to F$ are continuous.

Theorem B.2 (Open Mapping Theorem). *If $f : E \to F$ is a continuous bijection of Banach spaces, then f is an isomorphism.*

The space of all bounded linear maps from E to F is denoted $B(E, F)$; this is clearly a vector space. Defining

$$\|f\|_{\text{operator}} = \sup_{\|x\|_E \leqslant 1} \{\|f(x)\|_F\}$$

makes $B(E, F)$ into a normed space, and if F is a Banach space then $B(E, F)$ is a Banach space. An important special case is the space of linear functionals, $E^* = B(E, \mathbb{F})$.

Assume now that E and F are Banach spaces. An operator $f : E \to F$ is *compact* if the image $f(U)$ of the open unit ball $U = \{x \in E \mid \|x\|_E < 1\}$ has compact closure in F. Equivalently, an operator is compact if and only if every bounded sequence (x_n) in E contains a subsequence (x_{n_j}) with the property that $\big(f(x_{n_j})\big)$ converges in F. Many operators that arise naturally in the study of integral equations, for example the Hilbert–Schmidt integral operators T defined on $L^2_\mu(X)$ by

$$(Tf)(s) = \int_X K(s,t)\,\mathrm{d}\mu(t)$$

for some *kernel* $K \in L^2_{\mu\times\mu}(X \times X)$, are compact operators.

Now assume that E is a Banach space. Then $B(E) = B(E,E)$ is not only a Banach space but also an algebra: if $S, T \in B(E)$ then $ST \in B(E)$ where $(ST)(x) = S(T(x))$, and $\|ST\| \leqslant \|S\|\|T\|$. Write I for the identity operator, and define the *spectrum* of an operator $T \in B(E)$ to be

$$\sigma_{\text{operator}}(T) = \{\lambda \in \mathbb{F} \mid (T - \lambda I) \text{ does not have a continuous inverse}\}.$$

Theorem B.3. *Let E and F be Banach spaces.*

(1) *If $T \in B(E,E)$ is compact and $\lambda \neq 0$, then the kernel of $T - \lambda I$ is finite-dimensional.*
(2) *If E is not finite-dimensional and $T \in B(E)$ is compact, then $\sigma_{\text{operator}}(T)$ contains 0.*
(3) *If $S, T \in B(E)$ and T is compact, then ST and TS are compact.*

Functional analysis on Hilbert space is particularly useful in ergodic theory, because each measure-preserving system $(X, \mathscr{B}, \mu, T)$ has an associated Koopman operator $U_T : L^2_\mu \to L^2_\mu$ defined by $U_T(f) = f \circ T$.

An invertible measure-preserving transformation T is said to have *continuous spectrum* if 1 is the only eigenvalue of U_T and any eigenfunction of U_T is a constant.

Theorem B.4 (Spectral Theorem). *Let U be a unitary operator on a complex Hilbert space $\mathscr{H}$.*

(1) *For each element $f \in \mathscr{H}$ there is a unique finite Borel measure μ_f on $\mathbb{S}^1$ with the property that*

$$\langle U^n f, f\rangle = \int_{\mathbb{S}^1} z^n \,\mathrm{d}\mu_f(z) \tag{B.1}$$

for all $n \in \mathbb{Z}$.
(2) *The map*

$$\sum_{n=-N}^{N} c_n z^n \mapsto \sum_{n=-N}^{N} c_n U^n f$$

extends by continuity to a unitary isomorphism between $L^2(\mathbb{S}^1, \mu_f)$ and the smallest U-invariant subspace in $\mathscr{H}$ containing f.

(3) *If T has continuous spectrum and $f \in L^2_\mu$ has $\int_X f \,\mathrm{d}\mu = 0$, then the spectral measure μ_f associated to the unitary operator U_T is non-atomic.*

We will also need two fundamental compactness results due to Alaoglu, Banach and Tychonoff[111].

Theorem B.5 (Tychonoff). *If $\{X_\gamma\}_{\gamma\in\Gamma}$ is a collection of compact topological spaces, then the product space $\prod_{\gamma\in\Gamma} X_\gamma$ endowed with the product topology is itself a compact space.*

Theorem B.6 (Alaoglu). *Let X be a topological vector space with U a neighborhood of 0 in X. Then the set of linear operators $x^* : X \to \mathbb{R}$ with $\sup_{x\in U} |x^*(x)| \leqslant 1$ is weak*-compact.*

B.4 Continuous Functions

Let (X, d) be a compact metric space. The space $C_{\mathbb{C}}(X)$ of continuous functions $f : X \to \mathbb{C}$ is a metric space with respect to the uniform metric

$$\mathsf{d}(f, g) = \sup_{x\in X} |f(x) - g(x)|;$$

defining $\|f\|_\infty = \sup_{x\in X} |f(x)|$ makes $C_{\mathbb{C}}(X)$ into a normed space.

It is often important to know when a subspace of a normed space of functions is dense.

Theorem B.7 (Stone–Weierstrass Theorem[112]). *Let (X, d) be a compact metric space, and let $\mathscr{A} \subseteq C_{\mathbb{C}}(X)$ be a linear subspace with the following properties:*

- *$\mathscr{A}$ is closed under multiplication (that is, $\mathscr{A}$ is a subalgebra);*
- *$\mathscr{A}$ contains the constant functions;*
- *$\mathscr{A}$ separates points (for $x \neq y$ there is a function $f \in \mathscr{A}$ with $f(x) \neq f(y)$); and*
- *for any $f \in \mathscr{A}$, the complex conjugate $\overline{f} \in \mathscr{A}$.*

Then $\mathscr{A}$ is dense in $C_{\mathbb{C}}(X)$.

Lemma B.8. *The spaces $C_{\mathbb{C}}(X)$ and $C(X)$ are separable metric spaces with respect to the metric induced by the uniform norm.*

Proof. Let $\{x_1, x_2, \dots\}$ be a dense set in X, and define a set

$$F = \{f_1, f_2, \dots\}$$

of continuous functions by $f_n(x) = \mathsf{d}(x, x_n)$ where d is the given metric on X. The set F separates points since the set $\{x_1, x_2, \dots\}$ is dense. It follows that the algebra generated by F is dense in $C(X)$ by the Stone–Weierstrass Theorem (Theorem B.7). The same holds for the $\mathbb{Q}$-algebra generated by F (that is, for the set of finite linear combinations $\sum_{i=1}^{m} c_i h_i$ with $c_i \in \mathbb{Q}$ and $h_i = \prod_{k=1}^{K_i} g_{k,i}$ with $g_{k,i} \in F$ and $K_i \in \mathbb{N}$). However, this set is countable, which shows the lemma for real-valued functions. The same argument using the $\mathbb{Q}(i)$-algebra gives the complex case. □

The next lemma is a simple instance of a more general result of Urysohn that characterizes normal spaces(113).

Theorem B.9 (Tietze–Urysohn extension). *Any continuous real-valued function on a closed subspace of a normal topological space may be extended to a continuous real-valued function on the entire space.*

We will only need this in the metric setting, and any metric space is normal as a topological space.

Corollary B.10. *If (X, d) is a metric space, then for any non-empty closed sets $A, B \subseteq X$ with $A \cap B = \varnothing$, there is a continuous function $f : X \to [0, 1]$ with $f(A) = \{0\}$ and $f(B) = \{1\}$.*

B.5 Measures on Compact Metric Spaces

The material in this section deals with measures and linear operators. It is standard; a convenient source is Parthasarathy [280].

Let (X, d) be a compact metric space, with Borel σ-algebra $\mathscr{B}$. Denote by $\mathscr{M}(X)$ the space of Borel probability measures on X. The dual space $C(X)^*$ of continuous real functionals on the space $C(X)$ of continuous functions $X \to \mathbb{R}$ can be naturally identified with the space of finite signed measures on X. A functional $F : C(X) \to \mathbb{C}$ is called *positive* if $f \geqslant 0$ implies that $F(f) \geqslant 0$, and the *Riesz representation theorem* states that any continuous positive functional F is defined by a unique measure $\mu \in \mathscr{M}(X)$ via

$$F(f) = \int_X f \, \mathrm{d}\mu.$$

The main properties of $\mathscr{M}(X)$ needed are the following. Recall that a set $\mathscr{M}$ of measures is said to be *convex* if the convex combination

$$s\mu_1 + (1 - s)\mu_2$$

lies in $\mathscr{M}$ for any $\mu_1, \mu_2 \in \mathscr{M}(X)$ and $s \in [0, 1]$.

Theorem B.11. (1) *$\mathscr{M}(X)$ is convex.*

(2) *For* $\mu_1, \mu_2 \in \mathscr{M}(X)$,
$$\int f \,\mathrm{d}\mu_1 = \int f \,\mathrm{d}\mu_2 \tag{B.2}$$
for all $f \in C(X)$ *if and only if* $\mu_1 = \mu_2$.

(3) *The weak*-topology on* $\mathscr{M}(X)$ *is the weakest topology making each of the evaluation maps*
$$\mu \mapsto \int f \,\mathrm{d}\mu$$
continuous for any $f \in C(X)$*; this topology is metrizable and in this topology* $\mathscr{M}(X)$ *is compact.*

(4) *In the weak*-topology,* $\mu_n \to \mu$ *if and only if any of the following conditions hold:*
- $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ *for every* $f \in C(X)$*;*
- *for every closed set* $C \subseteq X$, $\limsup_{n\to\infty} \mu_n(C) \leqslant \mu(C)$*;*
- *for every open set* $O \subseteq X$, $\liminf_{n\to\infty} \mu_n(O) \geqslant \mu(O)$*;*
- *for every Borel set* A *with* $\mu(\partial(A)) = 0$, $\mu_n(A) \to \mu(A)$.

Proof of part (3). Recall that by the Riesz representation theorem the dual space $C(X)^*$ of continuous linear real functionals $C(X) \to \mathbb{R}$ with the operator norm coincides with the space of finite signed measures, with the functional being given by integration with respect to the measure. Moreover, by the Banach–Alaoglu theorem the unit ball B_1 in $C(X)^*$ is compact in the weak*-topology. It follows that
$$\mathscr{M}(X) = \left\{\mu \in C(X) \mid \int 1 \,\mathrm{d}\mu = 1, \int f \,\mathrm{d}\mu \geqslant 0 \text{ for } f \in C(X) \text{ with } f \geqslant 0\right\}$$
is a weak*-closed subset of B_1 and is therefore compact in the weak*-topology.

To show that the weak*-topology is metrizable on $\mathscr{M}(X)$ we use the fact that $C(X)$ is separable by Lemma B.8. Suppose that $\{f_1, f_2, \dots\}$ is a dense set in $C(X)$. Then the weak*-topology on $\mathscr{M}(X)$ is generated by the intersections of the open neighborhoods of $\mu \in \mathscr{M}(X)$ defined by
$$V_{\varepsilon,n}(\mu) = \left\{\nu \in \mathscr{M}(X) \mid \left|\int f_n \,\mathrm{d}\nu - \int f_n \,\mathrm{d}\mu\right| < \varepsilon\right\}.$$
This holds since for any $f \in C(X)$ and neighborhood
$$V_{\varepsilon,f}(\mu) = \left\{\nu \in \mathscr{M}(X) \mid \left|\int f \,\mathrm{d}\nu - \int f \,\mathrm{d}\mu\right| < \varepsilon\right\}$$
we can find some n with $\|f_n - f\| < \frac{\varepsilon}{3}$ and it is easily checked that
$$V_{\varepsilon/3,n}(\mu) \subseteq V_{\varepsilon,f}(\mu).$$
Define

$$\mathsf{d}_{\mathscr{M}}(\mu,\nu)=\sum_{n=1}^{\infty}\frac{1}{2^n}\frac{|\int f_n\,\mathrm{d}\mu-\int f_n\,\mathrm{d}\nu|}{1+|\int f_n\,\mathrm{d}\mu-\int f_n\,\mathrm{d}\nu|}\tag{B.3}$$

for $\mu,\nu\in\mathscr{M}(X)$. A calculation shows that $\mathsf{d}_{\mathscr{M}}$ is a metric on $\mathscr{M}(X)$.

We finish the proof by comparing the metric neighborhoods $B_\delta(\mu)$ defined by $\mathsf{d}_{\mathscr{M}}$ with the neighborhoods $V_{\varepsilon,n}(\mu)$. Fix $\delta>0$ and choose K such that $\sum_{n=K+1}^{\infty}\frac{1}{2^n}<\frac{\delta}{2}$. Then, for sufficiently small $\varepsilon>0$, any measure

$$\nu\in V_{\varepsilon,f_1}(\mu)\cap\cdots\cap V_{\varepsilon,f_K}(\mu)$$

will satisfy

$$\sum_{n=1}^{K}\frac{1}{2^n}\frac{|\int f_n\,\mathrm{d}\mu-\int f_n\,\mathrm{d}\nu|}{1+|\int f_n\,\mathrm{d}\mu-\int f_n\,\mathrm{d}\nu|}<\frac{\delta}{2},$$

showing that $\nu\in B_\delta(\mu)$. Similarly, if $n\geqslant 1$ and $\varepsilon>0$ are given, we may choose δ small enough to ensure that $\frac{1}{2^n}\frac{s}{1+s}<\delta$ implies that $s<\varepsilon$. Then for any $\nu\in B_\delta(\mu)$ we will have $\nu\in V_{\varepsilon,n}$. It follows that the metric neighborhoods give the weak*-topology. □

A continuous map $T:X\to X$ induces a map $T_*:\mathscr{M}(X)\to\mathscr{M}(X)$ defined by $T_*(\mu)(A)=\mu(T^{-1}A)$ for any Borel set $A\subseteq X$. Each $x\in X$ defines a measure δ_x by

$$\delta_x(A)=\begin{cases}1 \text{ if } x\in A;\\ 0 \text{ if } x\notin A.\end{cases},$$

and $T_*(\delta_x)=\delta_{T(x)}$ for any $x\in X$.

For $f\geqslant 0$ a measurable map and $\mu\in\mathscr{M}(X)$,

$$\int f\,\mathrm{d}T_*\mu=\int f\circ T\,\mathrm{d}\mu.\tag{B.4}$$

This may be seen by the argument used in the first part of the proof of Lemma 2.6. In particular, equation (B.4) holds for all $f\in C(X)$, and from this it is easy to check that the map $T_*:\mathscr{M}(X)\to\mathscr{M}(X)$ is continuous with respect to the weak*-topology on $\mathscr{M}(X)$.

Lemma B.12. *Let μ be a measure in $\mathscr{M}(X)$. Then $\mu\in\mathscr{M}^T(X)$ if and only if $\int f\circ T\,\mathrm{d}\mu=\int f\,\mathrm{d}\mu$ for all $f\in C(X)$.*

The map T_* is continuous and affine, so the set $\mathscr{M}^T(X)$ of T-invariant measures is a closed convex subset of $\mathscr{M}(X)$.

B.6 Measures on Other Spaces

Our emphasis is on compact metric spaces and finite measure spaces, but we are sometimes forced to consider larger spaces. As mentioned in Definition A.5, a measure space is called σ-finite if it is a countable union of measurable sets with finite measure. Similarly, a metric space is called σ-compact if it is a countable union of compact subsets. A measure defined on the Borel sets of a metric space is called *locally finite* if every point of the space has an open neighborhood of finite measure.

Theorem B.13. *Let μ be a locally finite measure on the Borel sets of a σ-compact metric space. Then μ is* regular*, meaning that*

$$\mu(B) = \sup\{\mu(K) \mid K \subseteq B, K \textit{ compact}\} = \inf\{U \mid B \subseteq U, U \textit{ open}\}$$

for any Borel set B.

B.7 Vector-valued Integration

It is often useful to integrate functions taking values in the space of measures (for example, in Theorem 6.2, in Section 6.5, and in Theorem 8.10). It is also useful to integrate functions $f : X \to V$ defined on a measure space $(X, \mathscr{B}, \mu)$ and taking values in a topological vector space V. The goal is to define $\int_X f \,\mathrm{d}\mu$ as an element of V that behaves like an integral: for example, if $\lambda : V \to \mathbb{R}$ is a continuous linear functional on V, then we would like

$$\lambda\left(\int_X f \,\mathrm{d}\mu\right) = \int_X (\lambda f) \,\mathrm{d}\mu \tag{B.5}$$

to hold whenever $\int_X f \,\mathrm{d}\mu$ is defined. One (of many[(114)]) approaches to defining integration in this setting is to use the property in equation (B.5) to *characterize* the integral; in order for this to work we need to restrict attention to topological vector spaces in which there are enough functionals. We say that V^* separates points in V if for any $v \neq v'$ in V there is a $\lambda \in V^*$ with $\lambda(v) \neq \lambda(v')$.

Definition B.14. Let V be a topological vector space on which V^* separates points, and let $f : X \to V$ be a function defined on a measure space $(X, \mathscr{B}, \mu)$ with the property that the scalar functions $\lambda(f) : X \to \mathbb{F}$ lie in $L^1_\mu(X)$ for every $\lambda \in V^*$. If there is an element $v \in V$ for which

$$\lambda(v) = \int_X (\lambda f) \,\mathrm{d}\mu$$

for every $\lambda \in V^*$, then we define

$$\int_X f \,\mathrm{d}\mu = v.$$

We start with the simplest example of integration for functions taking values in a Hilbert space.

Example B.15. If V is a Hilbert space $\mathscr{H}$ then the characterization in Definition B.14 takes the form

$$\left\langle \int_X f \,\mathrm{d}\mu, h \right\rangle = \int_X \langle f(x), h\rangle \,\mathrm{d}\mu(x) \tag{B.6}$$

for all $h \in V$. Note that in this setting the right-hand side of equation (B.6) defines a continuous linear functional on $\mathscr{H}$. It follows that the integral $\int_X f \,\mathrm{d}\mu$ exists by the Riesz representation theorem (see p. 422).

We now describe two more situations in which the existence of the integral can be established quite easily.

Example B.16. Let $V = L^p_\nu(Y)$ for a probability space (Y, ν) with $1 \leqslant p < \infty$, and let $F : X \times Y \to \mathbb{C}$ be an element of $L^p_{\mu\times\nu}(X \times Y)$. In this case we define

$$f : (X, \mu) \to V$$

by defining $f(x)$ to be the equivalence class of the function

$$F(x, \cdot) : y \longmapsto F(x, y).$$

We claim that $v = \int_X f \,\mathrm{d}\mu$ exists and is given by the equivalence class of

$$\nu(y) = \int F(x, y) \,\mathrm{d}\mu(x),$$

which is well-defined by the Fubini–Tonelli Theorem (Theorem A.13), since

$$L^p_{\mu\times\nu}(X \times Y) \subseteq L^1_{\mu\times\nu}(X \times Y).$$

To see this claim, recall that $V^* = L^q_\nu(Y)$ where $\frac{1}{p}+\frac{1}{q} = 1$, and let $w \in L^q_\nu(Y)$. Then $Fw \in L^1_{\mu\times\nu}(X \times Y)$ and so

$$\begin{aligned}\int_X \langle f(x), w)\rangle \,\mathrm{d}\mu &= \int_X \int_Y F(x, y)w(y) \,\mathrm{d}\nu \,\mathrm{d}\mu \\ &= \int_Y \int_X F(x, y) \,\mathrm{d}\mu \cdot w(y) \,\mathrm{d}\nu = \langle v, w\rangle\end{aligned}$$

by Fubini, as required (notice that the last equation also implies that v lies in $L^p_\nu(Y)$).

Example B.17. Suppose now that V is a Banach space, and that $f : X \to V^*$ takes values in the dual space V^* of V. Assume moreover that $\|f(x)\|$ is integrable and for any $v \in V$ the map $x \mapsto \langle v, f(x) \rangle$ is measurable (and hence automatically integrable, since $|\langle v, f(x) \rangle| \leqslant \|v\| \cdot \|f(x)\|$). Then

$$\int_X f(x)\,\mathrm{d}\mu(x) \in V^*$$

exists if we equip V^* with the weak*-topology: In fact, we may let $\int_X f\,\mathrm{d}\mu$ be the map

$$V \ni v \longmapsto \int_X \langle v, f(x) \rangle\,\mathrm{d}\mu,$$

which depends linearly and continuously on v. Moreover, with respect to the weak*-topology on V^* all continuous functionals on V^* are evaluation maps on V.

The last example includes (and generalizes) the first two examples above, but also includes another important case. A similar construction is used in Section 5.3, in the construction of conditional measures.

Example B.18. Let $V = C(Y)$ for a compact metric space Y, so that V^* is the space of signed finite measures on Y. Hence, for any probability-valued function

$$\Theta : X \to \mathscr{M}(Y)$$

with the property that $\int f(y)\,\mathrm{d}\Theta_x(y)$ depends measurably on $x \in X$, there exists a measure $\int_X \Theta_x\,\mathrm{d}\mu(x)$ on Y.

The next result gives a general criterion that guarantees existence of integrals in this sense (see Folland [94, App. A]).

Theorem B.19. *If $(X, \mathscr{B}, \mu)$ is a Borel probability space, V^* separates points of V, $f : X \to V$ is measurable, and the smallest closed convex subset I of V containing $f(X)$ is compact, then the integral $\int_X f\,\mathrm{d}\mu$ in the sense of Definition* B.14 *exists, and lies in I.*

A second approach is to generalize Riemann integration to allow continuous functions defined on a compact metric space equipped with a Borel probability measure and taking values in a Banach space. If V is a Banach space with norm $\|\cdot\|$, (X, d) is a compact metric space with a finite Borel measure μ, and $f : X \to V$ is continuous, then f is uniformly continuous since X is compact. Given a finite partition ξ of X into Borel sets and a choice $x_P \in X$ of a point $x_P \in P \in \xi$ for each atom P of ξ, define the associated Riemann sum

$$R_\xi(f) = \sum_{P \in \xi} f(g_P)\mu(P).$$

It is readily checked that $R_\xi(f)$ converges as

$$\mathrm{diam}(\xi) = \max_{P\in\xi} \mathrm{diam}(P) \to 0,$$

and we define

$$\int_X f \,\mathrm{d}\mu = \lim_{\mathrm{diam}(\xi)\to 0} R_\xi(f)$$

to be the (Riemann) integral of f with respect to μ. It is clear from the definition that

$$\left\| \int_X f \,\mathrm{d}\mu \right\| \leqslant \int_X \|f\| \,\mathrm{d}\mu,$$

where the integral on the right-hand side has the same definition for the continuous function $x \mapsto \|f(x)\|$ taking values in $\mathbb{R}$ (and therefore coincides with the Lebesgue integral).

Notes to Appendix B

(109) (Page 417) Convenient sources for most of the material described here include Rudin [321] and Folland [94]; many of the ideas go back to Banach's monograph [17], originally published in 1932.

(110) (Page 418) The Hahn–Banach theorem is usually proved using the Axiom of Choice (though it is not equivalent to it), and is often the most convenient form of the Axiom of Choice for functional analysis arguments. Significant special cases were found by Riesz [312], [313] in connection with extending linear functionals on L^q, and by Helly [147] who gave a more abstract formulation in terms of operators on normed sequence spaces. Hahn [132] and Banach [16] formulated the theorem as it is used today, using transfinite induction in a way that became a central tool in analysis.

(111) (Page 421) Tychonoff's original proof appeared in 1929 [363]; the result requires and implies the Axiom of Choice. Alaoglu's theorem appeared in 1940 [4], clarifying the treatment of weak topologies by Banach [17].

(112) (Page 421) Weierstrass proved that the polynomials are dense in $C[a,b]$ (corresponding to the algebra of real functions generated by the constants and the function $f(t) = t$). Stone [355] proved the result in great generality.

(113) (Page 422) Urysohn [365] shows that a topological space is *normal* (that is, Hausdorff and with the property that disjoint closed sets have disjoint open neighborhoods) if and only if it has the extension property in Theorem B.9. A simple example of a non-normal topological space is the space of all functions $\mathbb{R} \to \mathbb{R}$ with the topology of pointwise convergence. Earlier, Tietze [361] had shown the same extension theorem for metric spaces, and in particular Corollary B.10, which for normal spaces is usually called Urysohn's lemma.

(114) (Page 425) Integration can also be defined by emulating the real-valued case using partitions of the domain to produce a theory of vector-valued Riemann integration, or by using the Borel σ-algebra in V to produce a theory of vector-valued Lebesgue integration: the article of Hildebrandt [153] gives an overview.

Appendix C: Topological Groups

Many groups arising naturally in mathematics have a topology with respect to which the group operations are continuous. Abstracting this observation has given rise to the important theory described here. We give a brief overview, but note that most of the discussions and examples in this volume concern concrete groups, so knowledge of the general theory summarized in this appendix is useful but often not strictly necessary.

C.1 General Definitions

Definition C.1. A *topological group* is a group G that carries a topology with respect to which the maps $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous as maps $G \times G \to G$ and $G \to G$ respectively.

Any topological group can be viewed as a uniform space in two ways: the *left uniformity* renders each left multiplication $g \mapsto hg$ into a uniformly continuous map while the *right uniformity* renders each right multiplication $g \mapsto gh$ into a uniformly continuous map. As a uniform space, any topological group is completely regular, and hence[(115)] is Hausdorff if it is T_0.

Since the topological groups we need usually have a natural metric giving the topology, we will not need to develop this further.

The topological and algebraic structure on a topological group interact in many ways. For example, in any topological group G:

- the connected component of the identity is a closed normal subgroup;
- the inverse map $g \mapsto g^{-1}$ is a homeomorphism;
- for any $h \in G$ the left multiplication map $g \mapsto hg$ and the right multiplication map $g \mapsto gh$ are homeomorphisms;
- if H is a subgroup of G then the closure of H is also a subgroup;
- if H is a normal subgroup of G, then the closure of H is also a normal subgroup.

A topological group is called *monothetic* if it is Hausdorff and has a dense cyclic subgroup; a monothetic group is automatically abelian. Any generator of a dense subgroup is called a *topological generator*. Monothetic groups arise in many parts of dynamics.

A subgroup of a topological group is itself a topological group in the subspace topology. If H is a subgroup of a topological group G then the set of left (or right) cosets G/H (or $H\backslash G$) is a topological space in the quotient topology (the smallest topology which makes the natural projection $g \mapsto gH$ or Hg continuous). The quotient map is always open. If H is a normal subgroup of G, then the quotient group becomes a topological group. However, if H is not closed in G, then the quotient group will not be T_0 even if G is. It is therefore natural to restrict attention to the category of Hausdorff topological groups, continuous homomorphisms and closed subgroups, which is closed under many natural group-theoretic operations.

If the topology on a topological group is metrizable[(116)], then there is a compatible metric defining the topology that is invariant under each of the maps $g \mapsto hg$ (a left-invariant metric) and there is similarly a right-invariant metric.

Lemma C.2. *If G is compact and metrizable, then G has a compatible metric invariant under all translations (that is, a bi-invariant metric).*

PROOF. Choose a basis $\{U_n\}_{n\geqslant 1}$ of open neighborhoods of the identity $e \in G$, with $\cap_{n\geqslant 1} U_n = \{e\}$, and for each $n \geqslant 1$ choose (by Theorem B.9) a continuous function $f_n : G \to [0,1]$ with $\|f_n\| = 1$, $f_n(e) = 1$ and $f_n(G\smallsetminus U_n) = \{0\}$. Let

$$f(g) = \sum_{n=1}^{\infty} f_n(g)/2^n,$$

so that f is continuous, $f^{-1}(\{1\}) = e$, and define

$$\mathsf{d}(x,y) = \sup_{a,b\in G} \{|f(axb) - f(ayb)|\}.$$

Then d is bi-invariant and compatible with the topology on G. □

Example C.3. [(117)] The group $\mathrm{GL}_n(\mathbb{C})$ carries a natural norm

$$\|x\| = \max\left\{ \left(\sum_{i=1}^{n} \Big| \sum_{j=1}^{n} x_{ij} v_j \Big|^2 \right)^{1/2} \mid \sum_{i=1}^{n} |v_i|^2 = 1 \right\}$$

from viewing a matrix $x = (x_{ij})_{1\leqslant i,j\leqslant n} \in \mathrm{GL}_n(\mathbb{C})$ as a linear operator on $\mathbb{C}^n$. Then the function

$$\mathsf{d}(x,y) = \log\left(1 + \|x^{-1}y - I_n\| + \|y^{-1}x - I_n\|\right)$$

is a left-invariant metric compatible with the topology. For $n \geqslant 2$, there is no bi-invariant metric on $\mathrm{GL}_n(\mathbb{C})$. To see this, notice that for such a metric conjugation would be an isometry, while

$$\begin{pmatrix} m & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{m} & \frac{1}{m^2} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 + \frac{1}{m} \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

as $m \to \infty$, and

$$\begin{pmatrix} \frac{1}{m} & \frac{1}{m^2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} m & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{m} + \frac{1}{m^2} \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as $m \to \infty$.

C.2 Haar Measure on Locally Compact Groups

Further specializing to locally compact topological groups (that is, topological groups in which every point has a neighborhood containing a compact neighborhood) produces a class of particular importance in ergodic theory for the following reason.

Theorem C.4 (Haar[118]). *Let G be a locally compact group.*

(1) *There is a measure m_G defined on the Borel subsets of G that is invariant under left translation, is positive on non-empty open sets, and is finite on compact sets.*
(2) *The measure m_G is unique in the following sense: if μ is any measure with the properties of* (1) *then there is a constant C with $\mu(A) = Cm_G(A)$ for all Borel sets A.*
(3) *$m_G(G) < \infty$ if and only if G is compact.*

The measure m_G is called (a) *left Haar measure* on G; if G is compact it is usually normalized to have $m_G(G) = 1$. There is a similar right Haar measure. If m_G is a left Haar measure on G, then for any $g \in G$ the measure defined by $A \mapsto m_G(Ag)$ is also a left Haar measure. By Theorem C.4, there must therefore be a unique function mod, called the *modular function* or *modular character* with the property that

$$m_G(Ag) = \mathrm{mod}(g) m_G(A)$$

for all Borel sets A. The modular function is the continuous homomorphism $\mathrm{mod} : G \to \mathbb{R}_{>0}$. A group in which the left and right Haar measures coincide (equivalently, whose modular function is identically 1) is called *unimodular*: examples include all abelian groups, all compact groups (since there are no non-trivial compact subgroups of $\mathbb{R}_{>0}$), and semi-simple Lie groups.

There are several different proofs of Theorem C.4. For compact groups, it may be shown using fixed-point theorems from functional analysis. A particularly intuitive construction, due to von Neumann, starts by assigning measure one to some fixed compact set K with non-empty interior, then uses translates of some small open set to efficiently cover K and any other compact set L. The Haar measure of L is then approximately the number of translates needed to cover L divided by the number needed to cover K (see Section 8.3 for more details). Remarkably, Theorem C.4 has a converse: under some technical hypotheses, a group with a Haar measure must be locally compact(119).

Haar measure produces an important class of examples for ergodic theory: if $\phi : G \to G$ is a surjective homomorphism and G is compact, then ϕ preserves(120) the Haar measure on G. Haar measure also connects(121) the topology and the algebraic structure of locally compact groups.

Example C.5. In many situations, the Haar measure is readily described.

(1) The Lebesgue measure λ on $\mathbb{R}^n$, characterized by the property that

$$\lambda\left([a_1, b_1] \times \cdots \times [a_n, b_n]\right) = \prod_{i=1}^{n} (b_i - a_i)$$

for $a_i < b_i$, is translation invariant and so is a Haar measure for $\mathbb{R}^n$ (unique up to multiplication by a scalar).

(2) The Lebesgue measure λ on $\mathbb{T}^n$, characterized in the same way by the measure it gives to rectangles, is a Haar measure (unique if we choose to normalize so that the measure of the whole group $\mathbb{T}^n$ is 1).

As we have seen, a measure can be described in terms of how it integrates integrable functions. For the remaining examples, we will describe a Haar measure m_G by giving a 'formula' for $\int f \, \mathrm{d}m_G$. Thus the statement about the Haar measure $m_{\mathbb{R}^n}$ in (1) above could be written somewhat cryptically as

$$\int_{\mathbb{R}^n} f(\mathbf{x}) \, \mathrm{d}m_{\mathbb{R}^n}(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}) \, \mathrm{d}x_1 \dots \mathrm{d}x_n$$

for all functions f for which the right-hand side is finite. Evaluating a Haar measure on a group with explicit coordinates often amounts to computing a Jacobian.

(3) Let $G = \mathbb{R} \smallsetminus \{0\} = \mathrm{GL}_1(\mathbb{R})$, the real multiplicative group. The transformation $x \mapsto ax$ has Jacobian a: it can be readily checked that

$$\int f(ax) \frac{\mathrm{d}x}{|x|} = \int f(x) \frac{\mathrm{d}x}{|x|}$$

for any integrable f and $a \neq 0$. Hence a Haar measure m_G is defined by

$$\int_G f(x)\,\mathrm{d}m_G(x) = \int_G \frac{f(x)}{|x|}\,\mathrm{d}x$$

for any integrable f. Similarly, if $G = \mathbb{C}\smallsetminus\{0\} = \mathrm{GL}_1(\mathbb{C})$, then

$$\int_{\mathbb{C}\smallsetminus\{0\}} f(z)\,\mathrm{d}m_G(z) = \iint_{\mathbb{R}^2\smallsetminus\{(0,0)\}} \frac{f(x+\mathrm{i}y)}{x^2+y^2}\,\mathrm{d}x\,\mathrm{d}y.$$

(4) Let $G = \left\{\begin{pmatrix} a & b \\ 0 & 1\end{pmatrix} \mid a \in \mathbb{R}\smallsetminus\{0\}, b \in \mathbb{R}\right\}$, and identify elements of G with pairs (a,b). Then

$$\int_G f(a,b)\,\mathrm{d}m_G^{(\ell)} = \int_{\mathbb{R}}\int_{\mathbb{R}\smallsetminus\{0\}} \frac{f(a,b)}{a^2}\,\mathrm{d}a\,\mathrm{d}b$$

defines a left Haar measure, while

$$\int_G f(a,b)\,\mathrm{d}m_G^{(r)} = \int_{\mathbb{R}}\int_{\mathbb{R}\smallsetminus\{0\}} \frac{f(a,b)}{|a|}\,\mathrm{d}a\,\mathrm{d}b$$

defines a right Haar measure. As G is isomorphic to the group of affine transformations $x \mapsto ax+b$ under composition, it is called the '$ax+b$' group. It is an example of a non-unimodular group, with $\mathrm{mod}(a,b) = \frac{1}{|a|}$.

(5) Let $G = \mathrm{GL}_2(\mathbb{R})$, and identify the element $(x_{ij})_{1\leqslant i,j\leqslant 2}$ with

$$(x_{11}, x_{12}, x_{21}, x_{22}) \in A = \{\mathbf{x} \in \mathbb{R}^4 \mid x_{11}x_{22} - x_{12}x_{21} \neq 0\}.$$

Then

$$\int_G f\,\mathrm{d}m_G = \iiiint_A \frac{f(x_{11},x_{12},x_{21},x_{22})}{(x_{11}x_{22}-x_{12}x_{21})^2}\,\mathrm{d}x_{11}\,\mathrm{d}x_{12}\,\mathrm{d}x_{21}\,\mathrm{d}x_{22}$$

defines a left and a right Haar measure on G, which is therefore unimodular.

C.3 Pontryagin Duality

Specializing yet further brings us to the class of *locally compact abelian groups* (LCA groups) which have a very powerful theory[(122)] generalizing Fourier analysis on the circle. Throughout this section, $L^p(G)$ denotes $L^p_{m_G}(G)$ for some Haar measure m_G on G.

A *character* on a LCA group G is a continuous homomorphism

$$\chi : G \to \mathbb{S}^1 = \{z \in \mathbb{C} \mid |z| = 1\}.$$

The set of all continuous characters on G forms a group under pointwise multiplication, denoted $\widehat{G}$ (this means the operation on $\widehat{G}$ is defined by

$$(\chi_1 + \chi_2)(g) = \chi_1(g)\chi_2(g)$$

for all $g \in G$, and the trivial character $\chi(g) = 1$ is the identity). The image of $g \in G$ under $\chi \in \widehat{G}$ will also be written $\langle g, \chi \rangle$ to emphasize that this is a pairing between G and $\widehat{G}$. For compact $K \subseteq G$ and $\varepsilon > 0$ the sets

$$N(K, \varepsilon) = \{\chi \mid |\chi(g) - 1| < \varepsilon \text{ for } g \in K\}$$

and their translates form a basis for a topology on $\widehat{G}$, the topology of *uniform convergence on compact sets.*

Theorem C.6. *In the topology described above, the character group of a LCA group is itself a LCA group. A subgroup of the character group that separates points is dense.*

A subset $E \subseteq \widehat{G}$ is said to *separate points* if for $g \neq h$ in G there is some $\chi \in E$ with $\chi(g) \neq \chi(h)$.

Using the Haar measure on G the usual L^p function spaces may be defined. For $f \in L^1(G)$ the *Fourier transform of* f, denoted $\widehat{f}$ is the function on $\widehat{G}$ given by

$$\widehat{f}(\chi) = \int_G f(g)\overline{\langle g, \chi \rangle}\, \mathrm{d}m_G.$$

Some of the basic properties of the Fourier transform are as follows.

- The image of the map $f \mapsto \widehat{f}$ is a separating self-adjoint algebra in $C_0(\widehat{G})$ (the continuous complex functions vanishing at infinity) and hence is dense in $C_0(\widehat{G})$ in the uniform metric.
- The Fourier transform of the convolution $f * g$ is the product $\widehat{f} \cdot \widehat{g}$.
- The Fourier transform satisfies $\|\widehat{f}\|_\infty \leqslant \|f\|_1$ and so is a continuous operator from $L^1(G)$ to $L^\infty(\widehat{G})$.

Lemma C.7. *If G is discrete, then $\widehat{G}$ is compact, and if G is compact then $\widehat{G}$ is discrete.*

We prove the second part of this lemma to illustrate how Fourier analysis may be used to study these groups. Assume that G is compact, so that the constant function $\chi_0 \equiv 1$ is in $L^1(G)$.

Also under the assumption of compactness of G, we have the following orthogonality relations. Let $\chi \neq \eta$ be characters on G. Then we may find an element $h \in G$ with $(\chi\eta^{-1})(h) \neq 1$. On the other hand,

$$\int_G (\chi\eta^{-1})(g)\, \mathrm{d}m_G = \int_G (\chi\eta^{-1})(g + h)\, \mathrm{d}m_G = (\chi\eta^{-1})(h) \int (\chi\eta^{-1})(g)\, \mathrm{d}m_G,$$

so $\int_G (\chi\eta^{-1})(g)\,\mathrm{d}m_G = 0$ and the characters χ and η are orthogonal with respect to the inner-product

$$\langle f_1, f_2 \rangle = \int_G f_1 \overline{f_2}\,\mathrm{d}m_G$$

on $\widehat{G}$. Thus distinct characters are orthogonal as elements of $L^2(G)$.

Finally, note that the Fourier transform of any L^1 function is continuous on the dual group, and the orthogonality relations mean that $\widehat{\chi_0}(\chi) = 1$ if χ is the trivial character χ_0, and $\widehat{\chi_0}(\chi) = 0$ if not. It follows that $\{\chi_0\}$ is an open subset of $\widehat{G}$, so $\widehat{G}$ is discrete.

The Fourier transform is defined on $L^1(G) \cap L^2(G)$, and maps into a dense linear subspace of $L^2(\widehat{G})$ as an L^2 isometry. It therefore extends uniquely to an isometry $L^2(G) \to L^2(\widehat{G})$, known as the *Fourier* or *Plancherel transform* and also denoted by $f \mapsto \widehat{f}$. We note that this map is surjective.

Recall that there is a natural inner-product structure on $L^2(G)$.

Theorem C.8 (Parseval Formula). *Let f and g be functions in $L^2(G)$. Then*

$$\langle f, g \rangle_G = \int_G f(x)\overline{g(x)}\,\mathrm{d}m_G = \int_{\widehat{G}} \widehat{f}(\chi)\overline{\widehat{g}(\chi)}\,\mathrm{d}m_{\widehat{G}} = \langle\, \widehat{f}, \widehat{g}\, \rangle_{\widehat{G}}\,.$$

Given a finite Borel measure μ on the dual group $\widehat{G}$ of a locally compact abelian group G, the inverse Fourier transform of μ is the function $\check{\mu} : G \to \mathbb{C}$ defined by

$$\check{\mu}(x) = \int_{\widehat{G}} \chi(x)\,\mathrm{d}\mu(\chi).$$

A function $f : G \to \mathbb{C}$ is called *positive-definite* if for any $a_1, \dots, a_r \in \mathbb{C}$ and $x_1, \dots, x_r \in G$,

$$\sum_{i=1}^{r} \sum_{j=1}^{r} a_i \overline{a_j} f(x_i x_j^{-1}) \geqslant 0. \tag{C.1}$$

Theorem C.9 (Herglotz–Bochner(123)). *Let G be an abelian locally compact group. A function $f : G \to \mathbb{C}$ is positive-definite if and only if it is the Fourier transform of a finite positive Borel measure.*

Denote by $B(G)$ the set of all functions f on G which have a representation in the form

$$f(x) = \int_{\widehat{G}} \langle x, \chi \rangle\; \mathrm{d}\mu(\chi)$$

for $x \in G$ and a finite positive Borel measure μ on $\widehat{G}$. A consequence of the Herglotz–Bochner theorem (Theorem C.9) is that $B(G)$ coincides with the set of finite linear combinations of continuous positive-definite functions on G (see equation (C.1)).

Theorem C.10 (Inversion Theorem). *Let G be a locally compact group. If $f \in L^1(G) \cap B(G)$, then $\widehat{f} \in L^1(\widehat{G})$. Having chosen a Haar measure on G, the Haar measure $m_{\widehat{G}}$ on $\widehat{G}$ may be normalized to make*

$$f(g) = \int_{\widehat{G}} \widehat{f}(\chi)\langle g, \chi\rangle \,\mathrm{d}m_{\widehat{G}} \tag{C.2}$$

for $g \in G$ and any $f \in L^1(G) \cap B(G)$.

We will usually use Theorem C.10 for a compact metric abelian group G. In this case the Haar measure is normalized to make $m(G) = 1$, and the measure on the discrete countable group $\widehat{G}$ is simply counting measure, so that the right-hand side of equation (C.2) is a series.

In particular, for the case $G = \mathbb{T} = \mathbb{R}/\mathbb{Z}$ we find $\widehat{G} = \{\chi_k \mid k \in \mathbb{Z}\}$ where $\chi_k(t) = \mathrm{e}^{2\pi \mathrm{i} k t}$. Theorem C.10 then says that for any $f \in L^2(\mathbb{T})$ we have the Fourier expansion

$$f(t) = \sum_{k \in \mathbb{Z}} \widehat{f}(\chi_k)\mathrm{e}^{2\pi \mathrm{i} k t}$$

for almost every t.

Similarly, for any compact G, the set of characters of G forms an orthonormal basis of $L^2(G)$. We already showed the orthonormality property in the discussion after Lemma C.7; here we indicate briefly how the completeness of the set of characters can be established, both for concrete groups and in general.

Let $\mathscr{A}$ denote the set of finite linear combinations of the form

$$p(g) = \sum_{i=1}^{n} c_i \chi_i$$

with $c_i \in \mathbb{C}$. Then $\mathscr{A}$ is a subalgebra of $C_{\mathbb{C}}(X)$ which is closed under conjugation. If we know in addition that $\mathscr{A}$ separates points in G, then by the Stone–Weierstrass Theorem (Theorem B.7) we have that $\mathscr{A}$ is dense in $C_{\mathbb{C}}(X)$. Moreover, in that case $\mathscr{A}$ is also dense in $L^2(G)$, and so the set of characters forms an orthonormal basis for $L^2(G)$. That $\mathscr{A}$ separates points can be checked explicitly for many compact abelian groups; in particular for $G = \mathbb{R}^d/\mathbb{Z}^d$ the characters are of the form

$$\chi_{\mathbf{n}}(\mathbf{x}) = \mathrm{e}^{2\pi \mathrm{i}(n_1 x_1 + \cdots + n_d x_d)} \tag{C.3}$$

with $\mathbf{n} \in \mathbb{Z}^d$, and this explicit presentation may be used to show that the set of characters separates points on the d-torus. In general, one can prove that $\widehat{G}$ separates points by showing that the functions in $B(G)$ separate points, and then applying the Herglotz–Bochner theorem (Theorem C.9).

Theorem C.11. *For any compact abelian group G, the set of characters separates points and therefore forms a complete orthonormal basis for $L^2(G)$.*

The highlight of this theory is *Pontryagin duality*, which directly links the algebraic structure of LCA groups to their (Fourier-)analytic structure. If G is an LCA group, then $\Gamma = \widehat{G}$ is also an LCA group, which therefore has a character group $\widehat{\Gamma}$, which is again LCA. Any element $g \in G$ defines a character $\chi \mapsto \chi(g)$ on Γ.

Theorem C.12 (Pontryagin Duality). *The map $\alpha : G \to \widehat{\Gamma}$ defined by*

$$\langle g, \chi \rangle = \langle \chi, \alpha(g) \rangle$$

is a continuous isomorphism of LCA groups.

The Pontryagin duality theorem relates to the subgroup structure of an LCA group as follows.

Theorem C.13. *If $H \subseteq G$ is a closed subgroup, then G/H is also an LCA group. The set*

$$H^{\perp} = \{\chi \in \widehat{G} \mid \chi(h) = 1 \text{ for all } h \in H\},$$

the annihilator *of H, is a closed subgroup of $\widehat{G}$. Moreover,*

- $\widehat{G/H} \cong H^{\perp}$*;*
- $\widehat{G}/H^{\perp} \cong \widehat{H}$*;*
- *if H_1, H_2 are closed subgroups of G then*

$$H_1^{\perp} + H_2^{\perp} \cong \widehat{X}$$

where $X = G/(H_1 \cap H_2)$;
- $H^{\perp\perp} \cong H$.

The dual of a continuous homomorphism $\theta : G \to H$ is a homomorphism

$$\widehat{\theta} : \widehat{H} \to \widehat{G}$$

defined by $\widehat{\theta}(\chi)(g) = \chi(\theta(g))$. There are simple dualities for homomorphisms, for example θ has dense image if and only if $\widehat{\theta}$ is injective.

Pontryagin duality expresses topological properties in algebraic terms. For example, if G is compact then $\widehat{G}$ is torsion if and only if G is zero-dimensional (that is, has a basis for the topology comprising sets that are both closed and open), and $\widehat{G}$ is torsion-free if and only if G is connected. Duality also gives a description of monothetic groups: if G is a compact abelian group with a countable basis for its topology then G is monothetic if and only if the dual group $\widehat{G}$ is isomorphic as an abstract group to a countable subgroup of $\mathbb{S}^1$. If G is monothetic, then any such isomorphism is given by choosing a topological generator $g \in G$ and then sending $\chi \in \widehat{G}$ to $\chi(g) \in \mathbb{S}^1$.

Example C.14. As in the case of Haar measure in Example C.5, the character group of many groups can be written down in a simple way.

(1) If $G = \mathbb{Z}$ with the discrete topology, then any character $\chi \in \widehat{\mathbb{Z}}$ is determined by the value $\chi(1) \in \mathbb{S}^1$, and any choice of $\chi(1)$ defines a character. It follows that the map $z \mapsto \chi_z$, where χ is the unique character on $\mathbb{Z}$ with $\chi_z(1) = z$, is an isomorphism $\mathbb{S}^1 \to \widehat{\mathbb{Z}}$.
(2) Consider the group $\mathbb{R}$ with the usual topology. Then for any $s \in \mathbb{R}$ the map $\chi_s : t \mapsto \mathrm{e}^{ist}$ is a character on $\mathbb{R}$, and any character has this form. In other words, the map $s \mapsto \chi_s$ is an isomorphism $\mathbb{R} \to \widehat{\mathbb{R}}$.
(3) More generally, let $\mathbb{K}$ be any locally compact non-discrete field, and assume that $\chi_0 : \mathbb{K} \to \mathbb{S}^1$ is a non-trivial character on the additive group structure of $\mathbb{K}$. Then the map $a \mapsto \chi_a$, where $\chi_a(x) = \chi_0(ax)$, defines an isomorphism $\mathbb{K} \to \widehat{\mathbb{K}}$.
(4) An important example of (3) concerns the field of p-adic numbers $\mathbb{Q}_p$. For each prime number p, the field $\mathbb{Q}_p$ is the set of formal power series $\sum_{n\geqslant k} a_n p^n$ where $a_n \in \{0, 1, \dots, p-1\}$ and $k \in \mathbb{Z}$ and we always choose $a_k \neq 0$, with the usual addition and multiplication. The metric $\mathsf{d}(x, y) = |x - y|_p$, where $|\sum_{n\geqslant k} a_n p^n|_p = p^{-k}$ and $|0|_p = 0$, makes $\mathbb{Q}_p$ into a non-discrete locally compact field. By (3) an isomorphism $\widehat{\mathbb{Q}_p} \to \mathbb{Q}_p$ is determined by any non-trivial character on $\mathbb{Q}_p$, for example the map

$$\sum_{n\geqslant k} a_n p^n \mapsto \exp\Bigl(2\pi\mathrm{i} \sum_{n=k}^{-1} a_n p^{-n} \Bigr).$$

(5) Consider the additive group $\mathbb{Q}$ with the discrete topology. Then the group of characters is compact. Any element of $\widehat{\mathbb{R}}$ restricts to a character of $\mathbb{Q}$, so there is an embedding $\mathbb{R} \hookrightarrow \widehat{\mathbb{Q}}$ (injective because a continuous character on $\mathbb{R}$ is defined by its values on the dense set $\mathbb{Q}$). The group $\widehat{\mathbb{Q}}$ is an example of a *solenoid*, and there is a detailed account of its structure in terms of adeles in the monograph of Weil [377].

Lemma C.15 (Riemmann–Lebesgue[(124)]). *Let G be a locally compact abelian group, and let μ be a measure on G absolutely continuous with respect to Haar measure m_G. Then*

$$\widehat{\mu}(\chi) = \int_G \chi(g)\,\mathrm{d}\mu(t) \to 0$$

*as $\chi \to \infty$**.

---

* A sequence $\chi_n \to \infty$ if for any compact set $K \subseteq \widehat{G}$ there exists $N = N(K)$ for which

$$n \geqslant N \implies \chi_n \notin K.$$

The Riemann–Lebesgue lemma generalizes to absolutely continuous measures with respect to any sufficiently smooth measure.

Lemma C.16. *Let ν be a finite measure on $\mathbb{S}^1$, and assume that*

$$\int \mathrm{e}^{2\pi \mathrm{i} n t}\,\mathrm{d}\nu(t) \to 0$$

as $|n| \to \infty$. Then for any finite measure μ that is absolutely continuous with respect to ν,

$$\int \mathrm{e}^{2\pi \mathrm{i} n t}\,\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\,\mathrm{d}\nu(t) \to 0$$

as $|n| \to \infty$.

Notes to Appendix C

(115)(Page 429) Given a topological space $(X, \mathscr{T})$, points x and y are said to be topologically indistinguishable if for any open set $U \in \mathscr{T}$ we have $x \in U$ if and only if $y \in U$ (they have the same neighborhoods). The space is said to be T_0 or Kolmogorov if distinct points are always topologically distinguishable. This is the weakest of a hierarchy of topological separation axioms; for topological groups many of these collapse to the following natural property: the space is T_2 or Hausdorff if distinct points always have some distinct neighborhoods.

(116)(Page 430) A topological group is metrizable if and only if every point has a countable basis of neighborhoods (this was shown by Kakutani [170] and Birkhoff [34]) and has a metric invariant under all translations if there is a countable basis $\{V_n\}$ at the identity with $xV_nx^{-1} = V_n$ for all n (see Hewitt and Ross [151, p. 79]).

(117)(Page 430) This explicit construction of a left-invariant metric on $\mathrm{GL}_n(\mathbb{C})$ is due to Kakutani [170] and von Dantzig [64].

(118)(Page 431) Haar's original proof appears in his paper [130]; more accessible treatments may be found in the books of Folland [94], Weil [376] or Hewitt and Ross [151]. The important lecture notes of von Neumann from 1940-41, when he developed much of the theory from a new perspective, have now been edited and made available by the American Mathematical Society [269].

(119)(Page 432) This result was announced in part in a note by Weil [375] and then complete proofs were given by Kodaira [206]; these results were later sharpened by Mackey [239].

(120)(Page 432) This observation is due to Halmos [134], who determined when Haar measure is ergodic, and accounts for the special role of compact group automorphisms as distinguished examples of measure-preserving transformations in ergodic theory. The proof is straightforward: the measure defined by $\mu(A) = m_G(\phi^{-1}A)$ is also a translation-invariant probability measure defined on the Borel sets, so $\mu = m_G$.

(121)(Page 432) For example, if G and H are locally compact groups and G has a countable basis for its topology then any measurable homomorphism $\phi : H \to G$ is continuous (Mackey [240]); in any locally compact group, for any compact set A with positive Haar measure, the set AA^{-1} contains a neighborhood of the identity; if $H \subseteq G$ is closed under multiplication and conull then $H = G$.

(122)(Page 433) The theory described in this section is normally called Pontryagin duality or Pontryagin–von Kampen duality; the original sources are the book of Pontryagin [293] and the papers of van Kampen [181]. More accessible treatments may be found in Folland [94], Weil [376], Rudin [322] or Hewitt and Ross [151].

(123)(Page 435) This result is due to Herglotz [148] for functions on $\mathbb{Z}$, to Bochner [37] for $\mathbb{R}$, and to Weil [376] for locally compact abelian groups; accessible sources include the later translation [38] and Folland [94].

(124)(Page 438) Riemann [310] proved that the Fourier coefficients of a Riemann integrable periodic function converge to zero, and this was extended by Lebesgue [219]. The finite Borel measures on $\mathbb{T}$ with $\widehat{\mu}(n) \to 0$ as $|n| \to \infty$ are the *Rajchman* measures; all absolutely continuous measures are Rajchman measures but not conversely. Menshov, in his construction of a Lebesgue null set of multiplicity, constructed a singular Rajchman measure in 1916 by modifying the natural measure on the Cantor middle-third set (though notice that the Cantor–Lebesgue measure ν on the middle-third Cantor set has $\widehat{\nu}(n) = \widehat{\nu}(3n)$, so is a continuous measure that is not Rajchman). Riesz raised the question of whether a Rajchman measure must be continuous, and this was proved by Neder in 1920. Wiener gave a complete characterization of continuous measures by showing that ν is continuous if and only if $\frac{1}{2n+1}\sum_{k=-n}^{n} |\widehat{\mu}(k)| \to 0$ as $n \to \infty$. A convenient account is the survey by Lyons [238].