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.

M. Einsiedler, T. Ward, *Ergodic Theory*, Graduate Texts in Mathematics 259, DOI 10.1007/978-0-85729-021-2_2,

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 Sects. 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.

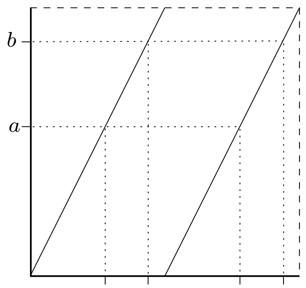

Fig. 2.1 The pre-image of $[a,b)$ under the circle-doubling map

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 Fig. 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 Sect. 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 (2.1) holds for $f \in L^1_\mu$.

Proof. If (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 (2.1) holds for any function of the form χ_B and hence for any simple function (see Sect. 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 Sect. 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 (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

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\phi} \\ Y & \xrightarrow[S]{} & Y \end{array}$$

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 Sect. 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 Sect. 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 Sect. 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 Sect. 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 Sect. 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 $x \in \widetilde{X}$;
- $\widetilde{\mu}\big(\{x \in \widetilde{X} \mid x_0 \in A\}\big) = \mu(A)$ 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 Sect. 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 Chap. 7.

Exercises for Sect. 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 (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 (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. $\square$

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,

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

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 Sects. 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 Sect. 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 Sect. 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 (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 $f(t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} nt}$ (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 $c_n = c_n \mathrm{e}^{2\pi \mathrm{i} n\alpha}$ for all $n \in \mathbb{Z}$. Since α is irrational, $\mathrm{e}^{2\pi \mathrm{i} n\alpha}$ 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 $g(t) = \mathrm{e}^{2\pi \mathrm{i} qt}$ 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 $f(t) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} nt}$ with

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

By invariance under T_2,

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

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 (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 Sect. 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 Sect. 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 (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

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

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

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

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 (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 Sect. 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 Sect. 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*

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

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 $\left(A, \mathscr{B}\big|_A, \frac{1}{\mu(A)}\mu\big|_A, T\big|_A\right)$. □

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*

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

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

$$r_1 \geqslant r_2 \geqslant \cdots \geqslant r_K.$$

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

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

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

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

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*

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

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 (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 (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 (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 $\alpha\mu(E^g_{\alpha,M}) \leqslant 3\|g\|_1$, 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 (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 (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 (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 (2.25), (2.26) and (2.27) together we see that $f^*=f'\in\mathscr{L}^1_\mu$ and that the convergence in (2.25) also happens in L^1_μ. Finally we also get

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

□

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

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

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

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

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

Exercises for Sect. 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 $f : X \to \mathbb{R}$,

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

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 Sects. 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

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

as

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

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

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 Chap. 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 Sect. 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 nth 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 Sect. 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

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

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

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

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 $T : \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} y \\ x+y \end{smallmatrix}\right)$.
(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* $a_n \underset{n\notin J}{\longrightarrow} 0$;
(3) $\displaystyle\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} a_j^2 = 0$.

Proof. (1) $\implies$ (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 (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 (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 (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) $\iff$ (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 $a_n = \left|\mu\left(A \cap T^{-n}B\right) - \mu(A)\mu(B)\right|$.

(6) $\implies$ (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) $\implies$ (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) $\implies$ (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 $A_2, B_2 \in \mathscr{B}_Y$,

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

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

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

The second term in (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 (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) $\Longrightarrow$ (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) $\Longrightarrow$ (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

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

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 (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 (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 Sect. 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 Sect. 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

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

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* Fig. 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.

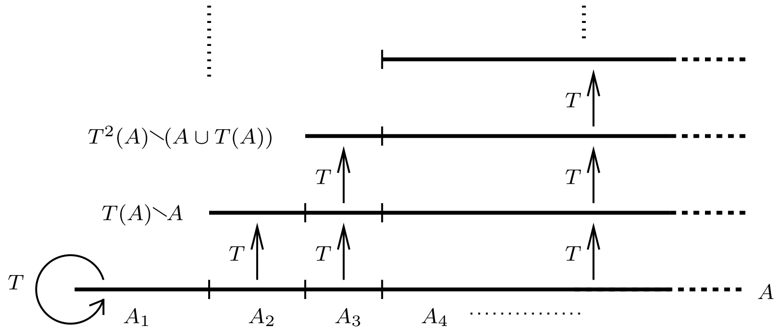

Fig. 2.2 The induced transformation T_A

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

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

by (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 Fig. 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 Fig. 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 Sect. 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 Chap. 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 [374, 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 [370] 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 Sect. 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 [395]. The issue has also been discussed by Ulam [365] 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, 391, 392].

(20)(Page 38) The maximal ergodic theorem is due to Wiener [382] and was also proved by Yosida and Kakutani [392].

(21)(Page 40) Covering lemmas of this sort were introduced by Vitali [369], 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

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

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 $0.123456789101112131415\dots$ 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, 390] 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) $\implies$ (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 [389], in which Kac's theorem is extended to measurable transformations.
(36)(Page 64) 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 $a_1, \dots, a_n$.

http://www.springer.com/978-0-85729-020-5

Ergodic Theory
with a view towards Number Theory
Einsiedler, M.; Ward, Th.
2011, XVII, 481 p., Hardcover
ISBN: 978-0-85729-020-5