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

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

2.3 Ergodicity

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

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

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

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

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

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

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

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

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

---

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

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

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

For any $N\geqslant 0$,

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

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

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

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

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

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

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

Moreover,

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

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

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

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

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

so

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

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

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

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

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

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

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

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

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

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

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

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

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

Proposition 2.15. *Bernoulli shifts are ergodic.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Exercises for Section 2.3

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