Introduction to Ergodic Theory

Summer Term 2023

Lecture by

Dr. Ivan Yaroslavtsev

University of Hamburg

TeXed by

Xiangxiang Michael Zheng

Contents

1 Motivation **1**

2 A bit of Measure Theory **5**

3 Measure-preserving Transformations **13**

4 Invariant Measures **18**

5 Conditional Expectations **22**

6 Ergodic Transforms and Birkhoff's Ergodic Theorem **24**

7 Ergodic Measures for Continuous Transforms **35**

8 Mixing **39**

9 One- and Two-Point Motion **44**

A Solutions to Exercise Sheets **51**

1 Motivation

Let $d \geq 1$, let $\omega\colon \mathbb{R}^d \to \mathbb{R}^d$ has a smooth vector field. Fix $x \in \mathbb{R}^d$. Then the ordinary differential equation (ODE)

$$\begin{cases} \mathrm{d}\, x_t &= \omega(x_t)\,\mathrm{d}\, t, t \geq 0 \\ x_0 &= x \end{cases} \tag{$*$}$$

has a unique solution $x_{(\cdot)}\colon \mathbb{R}_+ \to \mathbb{R}^d$ by the Picard-Lindelöf Theorem.

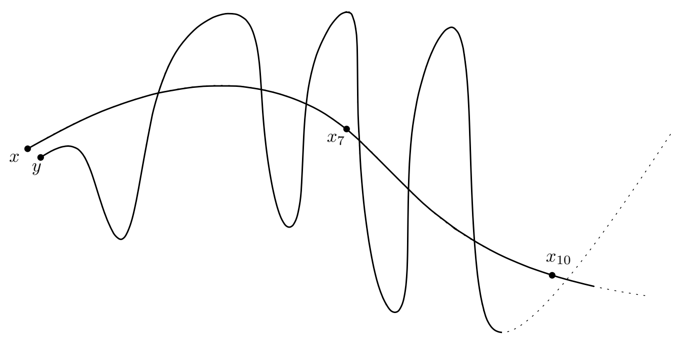


Figure 1: Sample trajectories that vary widely but have similar initial conditions

But in practice, for some ω sufficiently terrible, $(x_t)_{t\geq 0}$ is extremely sensitive to changes of the initial value, so the ODE theory can be (completely) useless in practice.

Example 1.1 (Classical Example by Lorenz in 1963)**.** Let $d = 3, x = (x^1, x^2, x^3)$, and $x_t = (x_t^1, x_t^2, x_t^3), t \geq 0$, solve

$$\begin{cases} \mathrm{d}\, x_t^1 &= \sigma(x_t^2 - x_t^1)\,\mathrm{d}\, t, t \geq 0 \\ \mathrm{d}\, x_t^2 &= (x_t^1(\rho - x_t^3) - x_t^2)\,\mathrm{d}\, t, t \geq 0 \\ \mathrm{d}\, x_t^3 &= (x_t^1 x_t^2 - \beta x_t^3)\,\mathrm{d}\, t, t \geq 0 \\ x_0 &= x. \end{cases} \qquad (\beta, \rho, \sigma \in \mathbb{R})$$

For $\sigma = 10, \rho = 28, c = 8/3$, the solution gets *highly chaotic*, i.e. it changes a lot for minor changes in the initial conditions.

Moreover, instead of the whole trajectory, one is typically interested in its discretization, i.e. $(x_{\varepsilon n})_{n\geq 0}$ for some $\varepsilon > 0$ instead of $(x_t)_{t\geq 0}$. As $(*)$ has a unique solution for all $x \in \mathbb{R}^d$, we set the injective map $T\colon \mathbb{R}^d \to \mathbb{R}^d$ to be $Tx = x_\varepsilon$ for all $x \in \mathbb{R}^d$.
Furthermore, as $(x_t)_{t\geq\varepsilon}$ solves

$$\begin{cases} \mathrm{d}\, x_t &= \omega(x_t)\,\mathrm{d}\, t, t \geq \varepsilon \\ x_\varepsilon &= x_\varepsilon, \end{cases}$$

it follows that $x_{2\varepsilon} = Tx_\varepsilon = T^2 x$, so in general $(x_{\varepsilon n})_{n\geq 0} = (T^n x)_{n\geq 0}$.

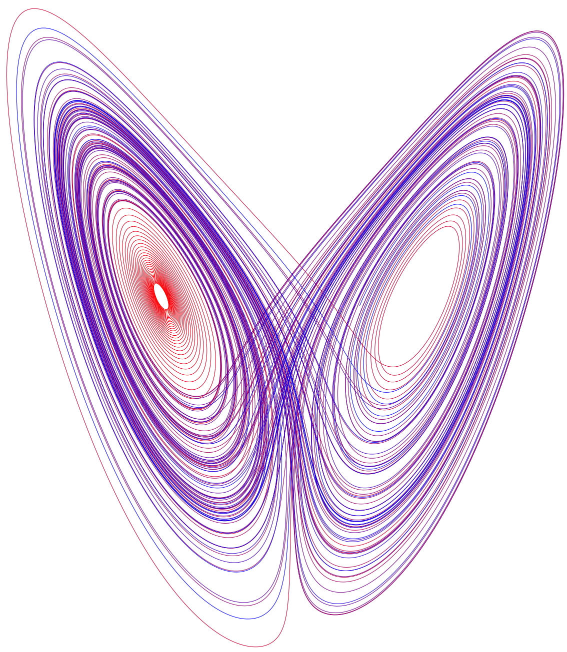
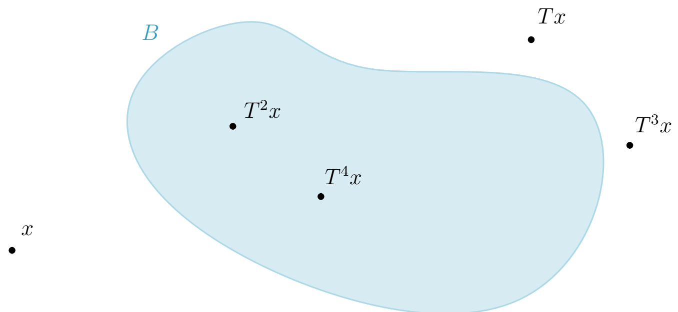


Figure 2: A solution in the *Lorenz attractor* (Source: Wikipedia)

What do we want to solve?

Typically, what is of interest in such chaotic systems is the *statistics* of $(T^n x)_{n\geq 0}$.

1. Fix $B \subseteq \mathbb{R}^d$. How often is $T^n x$ in B?



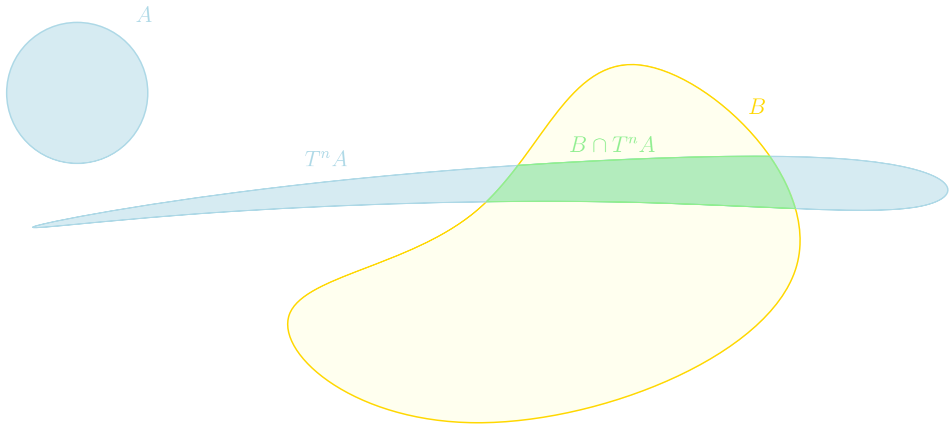
For this reason, one would like to know the *mean time* of $T^n x$ being in B, i.e.

$$\Phi_x(B) \coloneqq \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_B(T^n x) \tag{$*$}$$

if the latter exists. If this limit exists for any Borel set $B \subseteq \mathbb{R}^d$, then $B \mapsto \Phi_x(B)$ is

additive (i.e. $\Phi_x(A \cupdot B) = \Phi_x(A) + \Phi_x(B)$ for disjoint $A, B \subseteq \mathbb{R}^d$), and therefore, highly likely, a *measure*. We will explore for which T we can find such a measure and when Φ_x does *not* depend on x.
We will usually assume that T is *measure-preserving*, i.e. $T^{-1}(B)$ has the same volume as B for all Borel sets $B \subseteq \mathbb{R}^d$. In the case of T defined by $(*)$ this is equivalent to $\operatorname{div} \omega = 0$, but we will discuss this later.

2. Another interesting question is *mixing*. E.g., fix $B \subseteq \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$.



We want to tell something about $B \cap T^n A$, where $T^n A = \{T^n x, x \in A\} \subseteq \mathbb{R}^d$. T is called *mixing* if there exists a measure μ on $\mathbb{R}^d$ such that B and $T^n A$ are *almost μ-independent*, i.e. for sufficiently large n,

$$\mu(B \cap T^n A) \approx \mu(B) \underbrace{\mu(T^n A)}_{(=\mu(A) \text{ usually})}$$

or, if T is invertible and μ-preserving,

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

where $A, B \subseteq \mathbb{R}^d$ are Borel sets.

3. Sometimes instead of mixing, one has the same property, but for the average over all the time steps, i.e.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(B \cap T^n A) = \mu(B)\mu(A).$$

This is called *weak mixing*. I.e., the sequence $(T^n)_{n\geq 1}$ is *μ-independent* if one averages over time.

Example 1.2 (Grasshopper)**.** Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \simeq [0, 2\pi)$ be the one-dimensional torus. Let $\varphi \in [0, 2\pi)$. For all $x \in \mathbb{T}$, set $Tx = x + \varphi$.

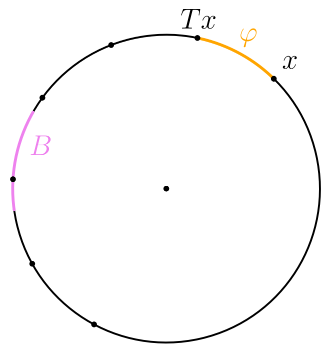


Figure 3: Grasshopper setup

The Question: Let $B \subseteq [0, 2\pi)$ be an interval. What can be shown about $\chi_B(T^n x)$? E.g., what is $\Phi_x(B) = \lim_{N\to\infty} 1/N \cdot \sum_{n=1}^{N} \chi_B(T^n x)$?

The Answer: If φ is 2π-irrational, i.e. $\varphi/2\pi \notin \mathbb{Q}$, then $\Phi_x(B) = \lambda(B)$, i.e. the normalized length of B. If x is 2π-rational, i.e. $\varphi/2\pi \in \mathbb{Q}$, $\Phi_x(B)$ depends on both x and B.

2 A bit of Measure Theory

We will give a brief review of measure theory. For basic proofs and exercises, see *Measure Theory* by Bogachev.

Definition 2.1. Let X be a set. A collection $\mathcal{B}$ of subsets of X is called *algebra* if

(i) $\emptyset \in \mathcal{B}$ and $X \in \mathcal{B}$,

(ii) if $B \in \mathcal{B}$, then its complement $X \setminus B \in \mathcal{B}$,

(iii) for all $B_1, \ldots, B_N \in \mathcal{B}$ one has $\bigcup_{n=1}^{N} B_n \in \mathcal{B}$.

$\mathcal{B}$ is called a *σ-algebra* if additionally

$(iii)'$ for all $(B_n)_{n\geq 1}$ from $\mathcal{B}$ one has $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

Let $\mathcal{C}$ be a collection of subsets of X. Then we denote by $\sigma(\mathcal{C})$ the *σ-algebra generated by $\mathcal{C}$*. Equivalently, $\sigma(\mathcal{C})$ is the intersection of *all σ-algebras containing $\mathcal{C}$*.

Exercise 2.2. Let $(\mathcal{B}_\alpha)_{\alpha\in\Lambda}$ be a family of σ-algebras in X. Prove that $\bigcap_{\sigma\in\Lambda} \mathcal{B}_\alpha$ is a σ-algebra as well. Do we have the same for $\bigcup_{\alpha\in\Lambda} \mathcal{B}_\alpha$?[1]

Definition 2.3. Let X be a topological space (e.g. $\mathbb{R}^d$ with the topology generated by all open balls). Then the σ-algebra generated by all open sets in X is called the *Borel σ-algebra* and is denoted by $\mathcal{B}(X)$.

Exercise 2.4. Show that a Cantor set C is in $\mathcal{B}(\mathbb{R})$ where C has the following form:

$$C = [0,1] \setminus \bigcup_{n=1}^{\infty} I_n,$$

where $(I_n)_{n\geq 1}$ are disjoint open intervals.

Figure 4: The classical Cantor set

The classical Cantor set is an example of an uncountable set with length 0 (with respect to the *Lebesgue measure*).[2]

Definition 2.5. Let X be a set and let $\mathcal{B}$ be a σ-algebra on X. Then $(X, \mathcal{B})$ is called a *measurable space*.

[1] Spoiler: No, not in general.

[2] "You can go rock climbing or think about the Cantor set, both are equally exciting." – Dr. Ivan Yaroslavtsev, 17th of April 2023.

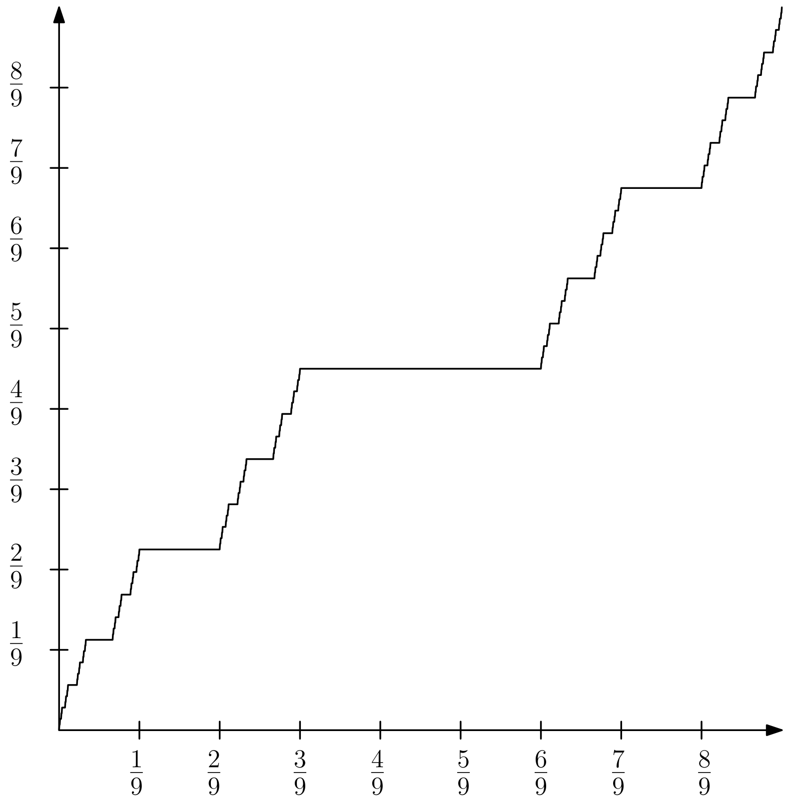


Figure 5: The distribution of the classical Cantor set

Definition 2.6. Let X be a set, $\mathcal{B}$ be an algebra on X. A map $\mu\colon \mathcal{B} \to [0,\infty]$ is called

(i) *additive* if for any disjoint $B_1, \dots, B_n \in \mathcal{B}$ one has

$$\mu\left(\bigcup_{n=1}^{N} B_n\right) = \sum_{n=1}^{N} \mu(B_n).$$

(ii) *countably additive* or *σ-additive* if for all $(B_n)_{n\geq 1}$ from $\mathcal{B}$ that are pairwise disjoint such that $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$ one has

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n).$$

If $\mathcal{B}$ is a σ-algebra and μ is σ-additive, then μ is called a *measure*. If μ takes values in $\mathbb{R}_+$, then μ is called a *finite measure*. In this case, $(X,\mathcal{B},\mu)$ is called a *measure space*.[3] If, additionally, $\mu(X) = 1$, then μ is a *probability* or a *probability measure* and $(X,\mathcal{B},\mu)$

[3]Others might follow the convention that $(X,\mathcal{B},\mu)$ is called a measure space even if μ is only a measure.

is called a *probability space.* $\mu\colon X \to \mathbb{R}$ is called a *finite signed measure* if $\mu = \mu_1 - \mu_2$ for some finite measures μ_1 and μ_2.
Let $(X, \mathcal{B})$ be a measurable space and μ be a measure on $(X, \mathcal{B})$. Then μ is called *σ-finite* if there exists $(B_n)_{n\geq 1}$ in $\mathcal{B}$ such that

$$\bigcup_{n=1}^{\infty} B_n = X, \qquad \forall\, n \in \mathbb{N}\colon \mu(B_n) < \infty.$$

Example 2.7. Let $X = \mathbb{R}, \mathcal{B} = \mathcal{B}(\mathbb{R}), \mu = \lambda$ be the Lebesgue measure, i.e. $\mu([a, b]) = b - a$ for all $b \geq a$, μ is σ-finite as $\mu([n, n+1]) = 1$ for all $n \in \mathbb{Z}$, and $\bigcup_{n\in\mathbb{Z}}[n, n+1] = \mathbb{R}$.

Example 2.8. Let $(X, \mathcal{B}, \mu)$ be a measure space and assume that μ is finite and $\mu(X) > 0$. Let $\hat{\mu} = \mu/\mu(X)$, i.e.

$$\forall\, B \in \mathcal{B}\colon \hat{\mu} = \frac{\mu(B)}{\mu(X)}.$$

Then $\hat{\mu}$ is a probability measure, and hence $(X, \mathcal{B}, \hat{\mu})$ is a probability space.

Exercise 2.9. Let $(X, \mathcal{B})$ be a measurable space and $\mu\colon X \to [0, \infty]$ be additive. Then the following are equivalent:

(*i*) μ is σ-additive.

(*ii*) For any increasing $(B_n)_{n\geq 1}$ in $\mathcal{B}$, i.e. $B_n \subseteq B_m$ for all $n \leq m$, one has

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n).$$

(*iii*) For any decreasing $(B_n)_{n\geq 1}$ in $\mathcal{B}$, i.e. $B_n \supseteq B_m$ for all $n \leq m$, one has

$$\mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n).$$

Typically, one defines a measure not on the whole σ-algebra, but only on some sets. For example, the Lebesgue measure λ on $\mathbb{R}^d$ is first defined on cubes, i.e. we set

$$\lambda\left([a_1, b_1] \times \cdots \times [a_d, b_d]\right) = \prod_{i=1}^{d} (b_i - a_i)$$

for all $b_1 \geq a_1, \ldots, b_d \geq a_d$. Then we can extend it to a measure on $\mathcal{B}(\mathbb{R}^d)$ by the following theorem.[4]

Theorem 2.10 (Carathéodory, without proof)**.** Let X be a set and $\mathcal{A}$ be an algebra on X. Let $\tilde{\mu}\colon \mathcal{A} \to \mathbb{R}_+$ be countably additive. Then there exists a *unique* measure on $\sigma(\mathcal{A})$ such that $\mu\big|_{\mathcal{A}} = \tilde{\mu}$.

Exercise 2.11. Show the existence of the Lebesgue measure λ on $\mathbb{R}^d$ using the theorem above and assuming that $\mathcal{A}$ is generated by all open and closed cubes.

[4] You don't need to know it super in-depth for the exam.

2.1 Measures on $\mathbb{R}$

Let $X = \mathbb{R}$.

Definition 2.12. A right-continuous non-decreasing function $F\colon \mathbb{R} \to \mathbb{R}$ is called a *distribution*. Concretely, F is a distribution if

$$\begin{aligned} &(i) \quad F(x) \le F(y) \text{ for all } x \le y, \\ &(ii) \quad F(x+\varepsilon) \to F(x) \text{ as } \varepsilon \downarrow 0. \end{aligned} \tag{$*$}$$

Note that for any such F one can define a map $\mu\colon \mathcal{J} \to \mathbb{R}_+$ by $\mu((a,b]) = F(b) - F(a)$ for all $b \ge a$, where $\mathcal{J} = \{(a,b]\colon b \ge a\} \subseteq \mathcal{B}(\mathbb{R})$.

Exercise 2.13. Check that μ is σ-additive on the algebra $\mathcal{A}(\mathcal{J})$ generated by $\mathcal{J}$ and conclude that it can be uniquely extended to a measure on $\mathbb{R}$.

For the inverse question ("Does any measure have a distribution?") one needs the following definition:

Definition 2.14. A measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called *Lebesgue-Stieltjes* if $\mu((a,b]) < \infty$ for all $b \ge a$.

Proposition 2.15. Let μ be a Lebesgue-Stieltjes measure. Then it has a distribution.

Proof idea. Set

$$F(x) := \begin{cases} \mu((0,x]), & x \ge 0 \\ -\mu((x,0]), & x < 0. \end{cases}$$

Then $F(b) - F(a) = \mu((a,b])$ for $b \ge a$, i.e. μ has distribution F. □ $(**)$

$(**)$ follows from the construction of F and the additivity of μ. The properties $(*)$ follow from the σ-additivity of μ.

Examples 2.16.

(1) Let $\alpha \in \mathbb{R}, \mu = \delta_\alpha$, i.e.

$$\mu(B) = \delta_\alpha(B) = \begin{cases} 1, & \alpha \in B \\ 0, & \alpha \notin B. \end{cases}$$

In this case we can set

$$F(x) = \begin{cases} 1, & x \ge \alpha \\ 0, & x < \alpha. \end{cases}$$

(2) Let $\mu = \lambda$ be the Lebesgue measure on $\mathbb{R}$. Then we can set $F(x) = x$ for all $x \in \mathbb{R}$.

(3) Let $f\colon \mathbb{R} \to \mathbb{R}_+$ be continuous and assume that

$$\mu((a,b]) = \int_a^b f(x)\, \mathrm{d}x.$$

Then we can set

$$F(x) := \int_0^x f(t)\, \mathrm{d}t.$$

2.2 Measures on Sequence Spaces

Let $k \geq 2$ and let Σ_k^+ be the set of all infinite sequences of k symbols, i.e.

$$\Sigma_k^+ \coloneqq \{a = (a_1, a_2, a_3, \dots) \colon a_1, a_2, \dots \in \{0, \dots, k-1\}\}.$$

Fix $n \geq 1$ and $x_1, \dots, x_n \in \{0, \dots, k-1\}$. Then the set

$$I_{x_1,\dots,x_n} = \left\{a \in \Sigma_k^+ \colon a_1 = x_1, \dots, a_n = x_n\right\}$$

is called a *cylinder*. Let Cyl be the set of all cylinders.

Exercise 2.17. Find $\mathcal{A} = \mathcal{A}(\mathrm{Cyl})$, the algebra generated by all cylinders.

Exercise 2.18. Fix $p_0, \dots, p_{k-1} \in [0,1]$ such that $p_0 + \dots + p_{k-1} = 1$. Let $\mu \colon \mathcal{A} \to \mathbb{R}_+$ be defined by

$$\mu\left(I_{x_1,\dots,x_n}\right) = \prod_{k=1}^{n} p_{x_k}.$$

Show that μ is a probability.

Exercise 2.19. Let $F \colon \Sigma_k^+ \to [0,1]$ be defined by $F(a) = \sum_{k\geq 1} a_n k^{-n}$ for

$$a = (a_1, a_2, \dots) \in \Sigma_k^+.$$

Show that if $p_0 = \dots = p_{k-1} = 1/k$, then for all $x, y \in [0,1], x \leq y$, one has that

$$\mu\left(F^{-1}(x,y]\right) = y - x = \lambda((x,y]).$$

Hint: It is sufficient to show the statement $x = \sum_{n=1}^N x_n k^{-n}$ and $y = \sum_{n=1}^N y_n k^{-n}$ with $N \geq 1, x_1, \dots, x_N, y_1, \dots, y_N \in \{0, \dots, k-1\}$.

Proof. As described in the hint, it suffices to consider $x = \sum_{n=1}^N x_n k^{-n}$ and $y = \sum_{n=1}^N y_n k^{-n}$ with $N \geq 1, x_1, \dots, x_N, y_1, \dots, y_N \in \{0, \dots, k-1\}$ as we can arbitrarily well approximate general x and y by numbers of that form. Now, it directly follows that

$$\begin{aligned}
\mu\left(F^{-1}(x,y]\right) &= \mu\left(\{(a_n)_{n\geq 1} \colon x_1 x_2 \dots x_N < a_1 a_2 \dots a_N \leq y_1 y_2 \dots y_N\}\right) \\
&= \sum_{x_1 x_2 \dots x_N < z_1 z_2 \dots z_N \leq y_1 y_2 \dots y_N} \mu\left(\{(a_n)_{n\geq 1} \colon a_1 = z_1, a_2 = z_2, \dots, a_N = z_N\}\right) \\
&= \frac{y_1 y_2 \dots y_N - x_1 x_2 \dots x_N}{k^N} \\
&= y - x,
\end{aligned}$$

where $x_1 x_2 \dots x_N$ is shorthand notation for $\sum_{i=1}^N x_i \cdot k^{N-i}$. □

Example 2.20 (Cantor set)**.** Set $k = 3, p_0 = p_2 = 1/2, p_1 = 0$. In such a way we can introduce a probability measure on the Cantor set.

Keep in mind that, when thinking about such sequences, you can equivalently think about $[0,1]$.

2.3 Lebesgue Integration

Let $(X, \mathcal{B}, \mu)$ be a measure space. $f\colon X \to \mathbb{R}$ is called *measurable* if for all $D \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(D) = \{x \in X \colon f(x) \in D\} \in \mathcal{B}.$$

Equivalently, f is measurable if

$$f^{-1}((c, \infty)) \in \mathcal{B}$$

for all $c \in \mathbb{R}$. A function $f\colon X \to \mathbb{C}$ is called *measurable* if both $\mathrm{Re}\,(f)$ and $\mathrm{Im}\,(f)$ are measurable. For all $B \in \mathcal{B}$, we define the characteristic function $\chi_B\colon X \to \mathbb{R}$ by

$$\chi_B(x) = \begin{cases} 1, & x \in B \\ 0, & x \notin B. \end{cases}$$

A function $f\colon X \to \mathbb{R}$ is called *simple* if there exists $N \geq 1, c_1, \ldots, c_N \in \mathbb{R}$ and $B_1, \ldots, B_N \in \mathcal{B}$ such that

$$f = \sum_{n=1}^{N} c_n \chi_{B_n}. \tag{$\star$}$$

For simplicity sake, assume that μ is a probability even though most of the following results hold for σ-finite μ. Then for f of the form $(\star)$ we can define $\int_X f \; \mathrm{d}\,\mu$ by

$$\int_X f \; \mathrm{d}\,\mu \coloneqq \sum_{n=1}^{N} c_n \mu(B_n).$$

Now, let $f\colon X \to [0, \infty]$ be a general positive measurable function.
For each $n \geq 1$, we set

$$f_n(x) \coloneqq \begin{cases} \frac{i-1}{2^n}, & \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}, i = 1, \ldots, n \cdot 2^n, \\ n, & f(x) \geq n. \end{cases}$$

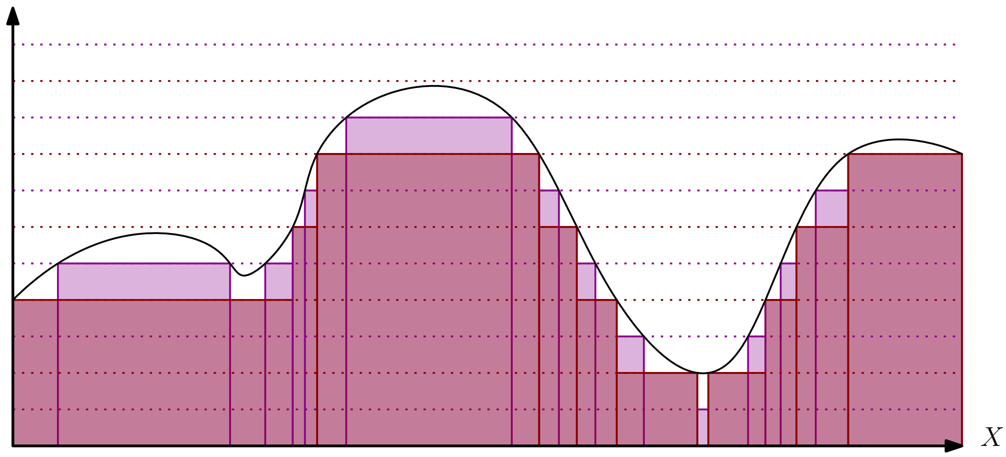


Figure 6: Approximation of f through f_n and f_{n+1}

Exercise 2.21. Show that $f_m \geq f_n$ for all $m \geq n$.

As $(f_n)_{n\geq 1}$ are simple,

$$\left(\int_X f_n \,\mathrm{d}\,\mu\right)_{n\geq 1}$$

is a non-decreasing sequence. Then we set

$$\int_X f \,\mathrm{d}\,\mu \coloneqq \lim_{n\to\infty} \int_X f_n \,\mathrm{d}\,\mu \in [0,\infty].$$

The latter is called *the Lebesgue integral of f with respect to μ.* If $\int_X f \,\mathrm{d}\,\mu < \infty$, then f is called *integrable*. A measurable function $f\colon X \to \mathbb{R}$ is called *integrable* if both $f^+ = \max\{0, f\}$ and $f^- = -\min\{0, f\}$ are integrable.
In this case, as $f = f^+ - f^-$, we set

$$\int_X f \,\mathrm{d}\,\mu = \int_X f^+ \,\mathrm{d}\,\mu - \int_X f^- \,\mathrm{d}\,\mu.$$

2.4 Properties of the Lebesgue Integral

Let $(X, \mathcal{B}, \mu)$ be a measure space.

1. For all integrable f, g and all $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ is integrable and

$$\int_X \alpha f + \beta g \,\mathrm{d}\,\mu = \alpha \int_X f \,\mathrm{d}\,\mu + \beta \int_X g \,\mathrm{d}\,\mu.$$

2. For all integrable f and $B \in \mathcal{B}$ one has that $\chi_B \cdot f$ is integrable and one can define

$$\int_B f \,\mathrm{d}\,\mu = \int_X \chi_B \cdot f \,\mathrm{d}\,\mu.$$

Furthermore, if $\mu(B) = 0$, then $\int_B f \,\mathrm{d}\,\mu = 0$. Therefore, the integral does not change if f has different values on a set B of measure $\mu(B) = 0$.
Hence, we can give the following definition:

Definition 2.22. For all $f, g\colon X \to \mathbb{R}$ we say that $f = g$ *almost everywhere* (or short $f = g$ a.e.) if

$$\mu(\{x \in X \colon f(x) \neq g(x)\}) = 0.$$

If in addition μ is a probability, then we say that $f = g$ *almost surely* (short $f = g$ a.s.).

Definition 2.23. Let $L^1(X, \mu)$ (short $L^1(X)$) be the linear space of all $f\colon X \to \mathbb{R}$ integrable, where $f = g$ in $L^1(X)$ if $f = g$ a.e.

3. For all $f \in L^1(X)$ with $f \geq 0$

$$\int_X f \,\mathrm{d}\,\mu = 0 \iff f = 0 \text{ a.s.}$$

4. For all $f, g \in L^1(X)$ such that $f \le g$ a.e. (i.e. $\mu(\{x \in X : f(x) > g(x)\} = 0)$, one has
$$\int_X f \; \mathrm{d}\,\mu \le \int_X g \; \mathrm{d}\,\mu.$$

5. If $f \in L^1(X)$, then $|f| \in L^1(X)$. Indeed, in this case for $f^+ = \max\{0, f\}, f^- = -\min\{0, f\}$, one has $|f| = f^+ + f^-$, so
$$\int_X f \; \mathrm{d}\,\mu = \underbrace{\int_X f^+ \; \mathrm{d}\,\mu}_{<\infty} + \underbrace{\int_X f^- \; \mathrm{d}\,\mu}_{<\infty} < \infty$$
as $f \in L^1(X)$. Thus,
$$\begin{aligned} |f| \in L^1(X) &\iff \int_X |f| \; \mathrm{d}\,\mu < \infty \\ &\iff \left(\int_X f^+ \; \mathrm{d}\,\mu < \infty \wedge \int_X f^- \; \mathrm{d}\,\mu < \infty\right) \\ &\iff f \in L^1(X). \end{aligned}$$
So $f \in L^1(X)$ if and only if $\int_X |f| \; \mathrm{d}\,\mu < \infty$.

6. (Monotone convergence theorem) Let $f_1, f_2, \ldots : X \to \mathbb{R}$ be a pointwise increasing sequence (i.e. $f_m(x) \ge f_n(x)$ for all $x \in X$ and $m \ge n$) of integrable functions. Then there exists $f\colon X \to (-\infty, +\infty]$ such that $f(x) = \lim_{n\to\infty} f_n(x)$, which is measurable. Moreover, $f^+ = \max\{0, f\}$ and $f^- = -\min\{0, f\}$ one has that
$$\int_X f \; \mathrm{d}\,\mu = \int_X f^+ \; \mathrm{d}\,\mu - \int_X f^- \; \mathrm{d}\,\mu = \lim_{n\to\infty} \int_X f_n \; \mathrm{d}\,\mu \in (-\infty, \infty],$$
i.e. f^+ may not be integrable, but f^- is always integrable.

7. (Fatou's lemma) Let $f_1, f_2, \ldots : X \to \mathbb{R}$ be non-negative and measurable. If $\liminf_{n\to\infty} \int_X f_n \; \mathrm{d}\,\mu < \infty$, then there exists $f\colon X \to [0, +\infty]$ such that $f(x) = \liminf_{n\to\infty} f_n(x)$ for all $x \in X$, $f \in L^1(X)$ and
$$\int_X f \; \mathrm{d}\,\mu \le \liminf_{n\to\infty} \int_X f_n \; \mathrm{d}\,\mu.$$

8. (Dominated convergence theorem) Let $f_1, f_2, \cdots \in L^1(X)$ and assume that there exists non-negative $g \in L^1(X)$ such that $|f_n| \le g$ a.e. Assume that there exists $f\colon X \to \mathbb{R}$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ for a.e. $x \in X$. Then $f \in L^1(X)$ and
$$\int_X f \; \mathrm{d}\,\mu = \lim_{n\to\infty} \int_X f_n \; \mathrm{d}\,\mu.$$

Proof of these properties. Exercise or read any book on measure theory. □

3 Measure-preserving Transformations

Definition 3.1. Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be two probability spaces. A map $T\colon X_1 \to X_2$ is called *measurable* if for all $B \in \mathcal{B}_2$

$$T^{-1}(B) \coloneqq \{x \in X_1 \colon Tx \in B\} \subseteq X_1$$

is in $\mathcal{B}_1$. T is called *measure-preserving* if additionally one has $\mu_2(B) = \mu_1(T^{-1}(B))$ for all $B \in \mathcal{B}_2$. T is called an *invertible measure-preserving transformation* if it is a bijection with both T and $T^{-1}\colon X_2 \to X_1$ being measure-preserving.

Remarks 3.2.

1. Let $(X_1, \mathcal{B}_1, \mu_1), (X_2, \mathcal{B}_2, \mu_2)$ and $(X_3, \mathcal{B}_3, \mu_3)$ be probability spaces, $T\colon X_1 \to X_2$ and $S\colon X_2 \to X_3$ be measure-preserving. Then $ST\colon X_1 \to X_3$ is also measure-preserving. In order to show this, fix $B \in \mathcal{B}_3$. Then

$$\mu_3(B) = \mu_2\left(S^{-1}(B)\right) = \mu_1\left(T^{-1}S^{-1}(B)\right) = \mu_1\left((ST)^{-1}(B)\right).$$

2. Typically, $(X_1, \mathcal{B}_1, \mu_1) = (X_2, \mathcal{B}_2, \mu_2)$, so $T\colon X_1 \to X_2$ is an *automorphism*, i.e. a mapping from a set to itself.

Example 3.3.

1. Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \simeq [0, 2\pi)$ be a torus with a normalized Lebesgue measure λ, let $\alpha \in [0, 2\pi)$, and let $T\colon \mathbb{T} \to \mathbb{T}$ be such that $Tx = x + \alpha$. Then T is an invertible measure-preserving transformation.[5]
2. Let $X = [0,1)^2, \mathcal{B} = \mathcal{B}(X)$ and let λ be the Lebesgue measure on X. Let $a, b, c, d \in \mathbb{Z}$ be such that $ad - bc = 1$. For each $x = (x^1, x^2)^\top \in \mathbb{R}^2$ set

$$Tx \coloneqq \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix}\right)_{\text{mod } 1} = \begin{pmatrix} \{ax^1 + bx^2\} \\ \{cx^1 + dx^2\} \end{pmatrix},$$

where for $y \in \mathbb{R}$ $\{y\} \in [0,1)$ denotes the *fractional part* of y, i.e. the unique number $\{y\} \in [0,1)$ such that $y - \{y\} \in \mathbb{Z}$.

Exercise 3.4. Show that T is measurable.

Let us show that T is a bijection. Note that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

as

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad - bc & bd - bd \\ -ac + ac & ad - bc \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

[5] Note that this is the same example as Example 1.2 at the beginning of the lecture.

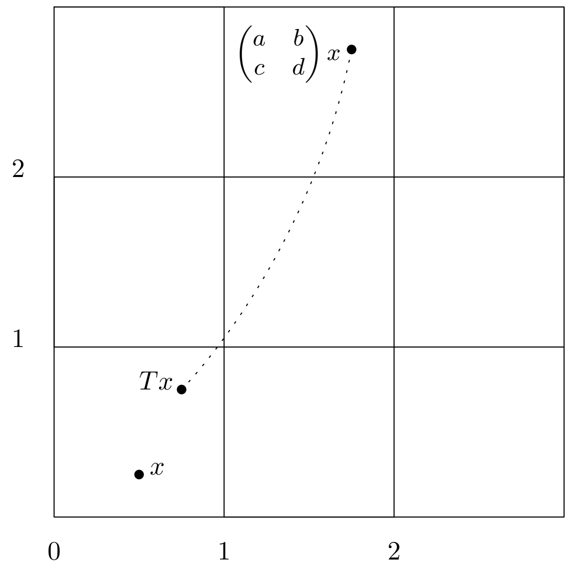


Figure 7: Action of T in Example 3.3.2

Thus, if for some $x, y \in X$ one has $Tx = y$, then

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} Tx = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} y \,.$$

In particular,

$$\begin{aligned} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} y &= \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} \{ax^1 + bx^2\} \\ \{cx^1 + dx^2\} \end{pmatrix} \\ &= \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} z \end{aligned}$$

for some $z \in \mathbb{Z}^2$. The latter equals $x + \tilde{z}$ for some $\tilde{z} \in \mathbb{Z}^2$. Therefore,

$$x = \left(\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} y \right)_{\bmod 1} = \begin{pmatrix} \{dy^1 - by^2\} \\ \{-cy^1 + ay^2\} \end{pmatrix}$$

The fact that then additionally $Tx = y$ follows analogously.

3. Let $X = \mathbb{T}, \mathcal{B} = \mathcal{B}(\mathbb{T}), \mu$ be the normalized Lebesgue measure. Fix an integer $n \geq 1$. For any $x \in [0, 2\pi) \simeq \mathbb{T}$ set $Tx = (n \cdot x)_{\bmod 2\pi} = \{nx/(2\pi)\} \cdot 2\pi$.[6]

Exercise 3.5. Show that T is measure-preserving. Is T bijective?

[6] It's the same as considering $z \mapsto z^n$ on the unit circle in $\mathbb{C}$.

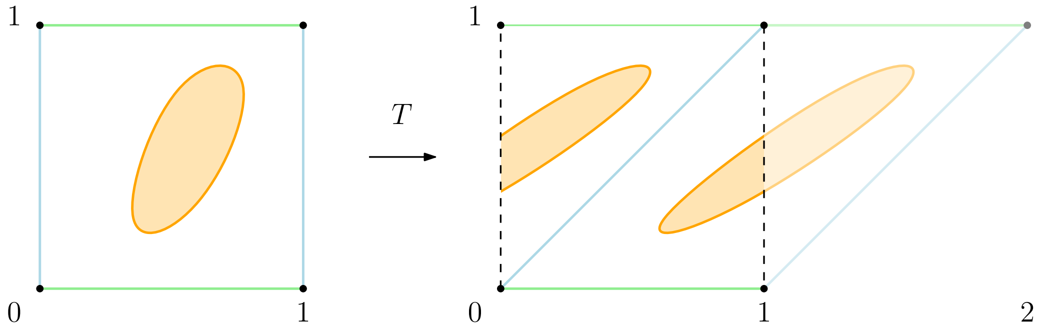


Figure 8: Action of the transformation $Tx = \left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x\right)_{\mathrm{mod}\ 1}$

We will start with the following important statistical property which measure-preserving transforms enjoy.

Theorem 3.6 (Poincaré's Recurrence Theorem)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T\colon X \to X$ be measure-preserving and let $E \in \mathcal{B}$ be such that $\mu(E) > 0$. Then μ-almost all points of E return infinitely often to E under the iteration of T. I.e. there exists $F \subseteq E$ such that $\mu(E \setminus F) = 0$ and

$$\forall\, x \in F\colon \exists\, n_1 < n_2 < \ldots \in \mathbb{N}\colon T^{n_1}x, T^{n_2}x, \ldots \in E.$$

Proof. Let B be the set of points in E that never return to E, i.e.

$$B = \left\{x \in E \,\middle|\, \forall\, k \geq 1\colon T^k x \notin E\right\} = E \setminus \left(\bigcup_{k\geq 1} T^{-k}(E)\right).$$

As $E \in \mathcal{B}$ and T is measurable, $T^{-k}(E) \in \mathcal{B}$, so $B \in \mathcal{B}$ by the definition of σ-algebras. Furthermore, by definition, $T^{-n}(B)$ is the set of all points that are in E after $n \geq 0$ steps but then never return to E. Let us show that for all $m > n \geq 0$ $T^{-m}(B)$ and $T^{-n}(B)$ have the same measure and are disjoint. The first follows from the fact that T is measure-preserving, so T^{m-n} is so as well, see Remark 3.2, and the following identity:

$$T^{-m}(B) = T^{-(m-n)}\left(T^{-n}(B)\right) \implies \mu\left(T^{-m}(B)\right) = \mu(T^{-n}(B)).$$

Let us now show that $T^{-m}(B) \cap T^{-n}(B) = \emptyset$. Assume the converse, i.e. there exists $x \in T^{-m}(B) \cap T^{-n}(B)$. Then

$$T^n x \in T^{-(m-n)}(B) \cap B \subseteq T^{-(m-n)}(E) \cap B = \emptyset$$

by the definition of B. That's why $(T^{-n}(B))_{n\geq 0}$ are disjoint, so

$$1 = \mu(X) \geq \mu\left(\bigcup_{n\geq 0} T^{-n}(B)\right) = \sum_{n\geq 0} \mu(T^{-n}(B)) = \sum_{n\geq 0} \mu(B) \implies \mu(B) = 0.$$

In particular, as $\mu(E \setminus B) = \mu(E)$, *almost every point in E returns to E*. Now, in order to show that a.e. point returns infinitely many times, we set for all $n \geq 1$

$$\begin{aligned}\tilde{B}_n &= \left\{x \in E \,\middle|\, T^n x \in E \text{ but } T^k x \notin E \;\forall\, k > n\right\} \\ &= (E \cap T^{-n}(E)) \setminus \Big(\bigcup_{k>n} T^{-k}(E) \Big).\end{aligned}$$

In other words, $\tilde{B}_n$ are the set of points in E that are the last time in E after n steps. Let us show that $\mu(\tilde{B}_n) = 0$. To this end, note that

$$T^n(\tilde{B}_n) = (T^n(E) \cap E) \setminus \Big(\bigcup_{k\geq 1} T^{-k}(E) \Big) \subseteq E \setminus \Big(\bigcup_{k\geq 1} T^{-k}(E) \Big) = B,$$

i.e. $\tilde{B}_n \subseteq T^{-n}(T^n(\tilde{B}_n)) \subseteq T^{-n}(B)$, so $\mu(\tilde{B}_n) \leq \mu(T^{-n}(B)) = \mu(B) = 0$. Now we let $\tilde{B} = \bigcup_{n\geq 0} \tilde{B}_n$, where $\tilde{B}_0 = B$, be the set of points in E that only return to E finitely many times and let $F = E \setminus \tilde{B}$. Then for all $x \in F$, $(T^n x)_{n\geq 1}$ is infinitely often in E, and $\mu(E \setminus F) = \mu(\tilde{B}) = 0$, so F is a desired set. $\square$

We will show for a concrete example the statement of the theorem (without using it).

Example 3.7 (Grasshopper revisited)**.** Let $\mathbb{T} \simeq [0, 2\pi)$ be the torus, $\alpha \in (0, 2\pi)$, and $T\colon \mathbb{T} \to \mathbb{T}$ be defined by $Tx = x + \alpha$ for $x \in \mathbb{T}$.

1. Assume that α is 2π-rational, i.e. $\alpha/(2\pi) \in \mathbb{Q}$. Then $\alpha/(2\pi) = m/l$ for some coprime $m, l \in \mathbb{N}$, i.e. m and l have no common prime factor.

Exercise 3.8. Show that $nm/l \in \mathbb{N}$ for $n \in \mathbb{N}$ if and only if $n/l \in \mathbb{N}$.

Then $T^n x = x + n\alpha = x + (2\pi nm)/l$ for all $n \in \mathbb{N}$. By the exercise above,

$$\begin{aligned} T^n x = x \iff & \Big(\exists\, \tilde{k} \in \mathbb{N}\colon \frac{2\pi nm}{l} = 2\pi\tilde{k}\Big) \\ \iff & \Big(\exists\, \tilde{k} \in \mathbb{N}\colon \frac{nm}{l} = \tilde{k}\Big) \\ \iff & \Big(\exists\, k \in \mathbb{N}\colon n = lk\Big) \end{aligned}$$

Thus, T is *periodic with period l*. In particular, *every* point in a subset $E \in \mathcal{B}(\mathbb{T})$ returns infinitely often to E.

2. Assume that α is 2π-irrational, i.e. $\alpha/2\pi \notin \mathbb{Q}$.

Lemma 3.9. Let $\alpha \in (0, 2\pi)$. There exists $n \in \mathbb{N}$ such that $(n\alpha)_{\bmod 2\pi} \leq \min(\alpha/2, \pi - \alpha/2))$ or $(2\pi - n\alpha)_{\bmod 2\pi} \leq \min(\alpha/2, \pi - \alpha/2))$.

Proof. W.l.o.g. assume that $\alpha < \pi$, otherwise consider $2\pi - \alpha$. Let $k \geq 1$ be such that $k\alpha \geq 2\pi$ and $(k-1)\alpha < 2\pi$. If $k\alpha - 2\pi > \alpha/2$, then

$$2\pi - (k-1)\alpha = 2\pi - k\alpha + \alpha < \alpha - \frac{\alpha}{2} = \frac{\alpha}{2},$$

so we can set $n = k - 1$. Otherwise $n = k$ is as desired. □

Corollary 3.9.1. For any $\alpha \in (0, 2\pi)$ and $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that either $\alpha n < \varepsilon \pmod{2\pi}$ or $2\pi - \alpha n \leq \varepsilon \pmod{2\pi}$. In other words, $\alpha n \in (-\varepsilon, \varepsilon)$ $\pmod{2\pi}$.

Proof. By Lemma 3.9, there exists $n_1 \in \mathbb{N}$ such that $\alpha_1 := \alpha n_1 \in (-\alpha/2, \alpha/2)$ $\pmod{2\pi}$. Further, there exists $n_2 \in \mathbb{N}$ such that $\alpha_1 n_2 = \alpha n_1 n_2 \in (-\alpha/4, \alpha/4)$ $\pmod{2\pi}$. By induction, there exists for any $k \in \mathbb{N}$ $n_1, \dots, n_k \in \mathbb{N}$ such that

$$\alpha n_1 \dots n_k \in \left(-\frac{\alpha}{2^k}, \frac{\alpha}{2^k}\right) \pmod{2\pi}.$$

Hence, choose $k \in \mathbb{N}$ such that $k > \log_2(\alpha/\varepsilon)$. Then $2^k > \alpha/\varepsilon$, so $\alpha/2^k < \varepsilon$. Thus, $\alpha n \in (-\varepsilon, \varepsilon)$ for $n = n_1 \dots n_k$. □

Corollary 3.9.2. Let $\alpha \in (0, 2\pi)$ be 2π-irrational, $I \subseteq \mathbb{T}$ be an open interval. Then for all $x \in \mathbb{T}$, $(T^n x)_{n \geq 1}$ occurs in I infinitely many times.

As every non-empty open set is the union of open intervals, the corollary also holds for any non-empty open set. In particular, the orbit of x is *dense in* $\mathbb{T}$.

Proof. Let $\varepsilon > 0$ be smaller then the length of I. W.l.o.g. $x = 0, I = (a, b), 0 \leq a < b < 2\pi$.[7] By Corollary 3.9.1, there exists $n \geq 1$ such that $\alpha n \in (-\varepsilon, \varepsilon) \pmod{2\pi}$. As α is 2π-irrational, $\alpha n \neq 0 \pmod{2\pi}$, so either $\alpha n \in (-\varepsilon, 0) \pmod{2\pi}$ or $\alpha n \in (0, \varepsilon) \pmod{2\pi}$. Assume the latter (the first case is similar). If $m \geq 1$ is such that $\alpha n(m-1) \leq a \pmod{2\pi}$ and $\alpha n m > a \pmod{2\pi}$, then

$$\alpha m n = \underbrace{\alpha n}_{\leq \varepsilon} + \underbrace{(m-1)\alpha n}_{\leq a} \leq a + \varepsilon < b,$$

as $\varepsilon < b - a$. Hence, $\alpha m n \in (a, b) = I \pmod{2\pi}$, so $T^{mn} x \in I$. As there are infinitely many points $m \geq 1$ such that

$$\alpha(m-1)n \leq a < \alpha m n \pmod{2\pi},$$

and as $T^{mn} x \in I$ for such m, $(T^n x)_{n \geq 1}$ occurs in I infinitely many times. □

[7] Even if your interval "wraps back around" at 0, you can always restrict to a subinterval of the desired form.

4 Invariant Measures

In this section, we are concerned with the situation where our measure *isn't given.* In other words, we are interested in the following question:

Given a measurable space $(X, \mathcal{B})$ and a measurable transform $T\colon X \to X$, can one find a probability measure μ on $(X, \mathcal{B})$ such that T is μ-preserving?

Such μ is then also called *T-invariant.*

Example 4.1. Let $(X, \mathcal{B})$ be a measurable space, let $T\colon X \to X$ be such that $T^n x = x$ for all $x \in X$, i.e. T is n-periodic. Let $a_1, \dots, a_n \in X$ be an orbit, i.e. $a_1, \dots, a_n$ are distinct, $Ta_i = a_{i+1}$ for all $1 \le i \le n-1$, and $Ta_n = a_1$. Assume that $\{a_1\}, \dots, \{a_n\} \in \mathcal{B}$. Fix some $p_1, \dots, p_n \in [0,1]$ such that $\sum_{k=1}^n p_k = 1$ and set $\mu = \sum_{k=1}^n p_k \delta_{a_k}$.

Exercise 4.2. Show that such μ is T-invariant if and only if $p_1 = \cdots = p_n = 1/n$.

Recall that X is a *metric space* if there exists $\rho\colon X \to \mathbb{R}_+$ such that

1. $\rho(x,y) = 0$ if and only if $x = y$ for all $x, y \in X$,
2. $\rho(x,y) = \rho(y,x)$ for all $x, y \in X$,
3. $\rho(x,y) + \rho(y,z) \ge \rho(x,z)$ for all $x, y, z \in X$.

X is called *compact* if for all $(x_n)_{n\ge1} \in X^{\mathbb{N}}$ there exists $x \in X$ and increasing $(n_i)_{i\ge1} \in \mathbb{N}^{\mathbb{N}}$ such that $x_{n_i} \to x$ as $i \to \infty$, i.e. $\rho(x, x_{n_i}) \to 0$ as $i \to \infty$. In other words, X is compact if every sequence has a convergent subsequence with limit in X.
The following theorem is important as it gives a positive answer to our question for a wide class of spaces and transforms.

Theorem 4.3 (Krylov-Bogolyubov)**.** Let X be a compact metric space, $T\colon X \to X$ be continuous.[8] Then there exists a probability μ on $(X, \mathcal{B})$ where $\mathcal{B} = \mathcal{B}(X)$ such that T is μ-preserving or, equivalently, μ is T-invariant.

The proof is very long, so we will go step-by-step. Firstly, we will start with the following definition.

Definition 4.4. Let $(X, \mathcal{B})$ be a measurable space. We denote by $M(X)$ the set of all probability measures on $(X, \mathcal{B})$. Note that $M(X)$ is convex, i.e. for all $\mu, \nu \in M(X)$ and $p \in [0,1]$ one has that $p\mu + (1-p)\nu \in M(X)$. Further, let X be a metric space. We endow $M(X)$ with the *weak*-topology*: A sequence $(\mu_n)_{n\ge1} \in M(X)^{\mathbb{N}}$ *converges to* $\mu \in M(X)$ *weakly** if

$$\forall\, f \in C(X)\colon \int_X f \, \mathrm{d}\,\mu_n \to \int_X f \, \mathrm{d}\,\mu, n \to \infty.$$

Here $C(X)$ denotes the set of continuous, bounded[9] functions on X.

[8] Here, T is called *continuous* in a pointwise sense, i.e. if $Tx_n \to Tx$ if $x_n \to x$ as $n \to \infty$.
[9] Note that any continuous function is also bounded since X is compact.

Lemma 4.5 (Without proof)**.** Let X be a compact metric space. Then $M(X)$ endowed with the weak*-topology is compact, i.e. for all $(\mu_n)_{n\geq 1} \in M(X)^{\mathbb{N}}$ there exists $\mu \in M(X)$ and increasing $(n_i)_{i\geq 1} \in \mathbb{N}^{\mathbb{N}}$ such that $\mu_{n_i} \to \mu$ weakly* as $i \to \infty$.

Definition 4.6. Let $T\colon X \to X$ be measurable. Then we define $\tilde{T}\colon M(X) \to M(X)$ by

$$(\tilde{T}\mu)(B) = \mu\left(T^{-1}(B)\right), \mu \in M(X), B \in \mathcal{B}.$$

In other words, $\tilde{T}$ maps a probability measure to the corresponding image measure.

From now on, let X be a compact metric space. We are interested in properties of $\tilde{T}$.

Lemma 4.7. For all $\mu \in M(X)$ and $f \in C(X)$ one has

$$\int_X f \ \mathrm{d}\tilde{T}\mu = \int_X f \circ T \ \mathrm{d}\mu \left(= \int_X f(T(x)) \ \mathrm{d}\mu(x)\right).$$

Proof. First, assume that $f = \chi_B$ for some $B \in \mathcal{B}$. Then

$$\begin{aligned}\int_X \chi_B \ \mathrm{d}\tilde{T}\mu = (\tilde{T}\mu)(B) \qquad &= \mu(T^{-1}(B)) \\ = \int_X \chi_{T^{-1}(B)}(x) \ \mathrm{d}\mu(x) &= \int_X \chi_B(Tx) \ \mathrm{d}\mu(x).\end{aligned}$$

By linearity, the same can be shown for any step function. The desired follows from the fact that any continuous function can be approximated by step functions.[10] □

Lemma 4.8. Let $T\colon X \to X$ be continuous. Then $\tilde{T}\colon M(X) \to M(X)$ is weak*-continuous and affine, i.e. for all $\mu, \nu \in M(X)$ and $p \in [0,1]$ one has that

$$\tilde{T}(p\mu + (1-p)\nu) = p\tilde{T}(\mu) + (1-p)\tilde{T}(\nu). \tag{$*$}$$

Proof. Let us start by showing that $\tilde{T}$ is weak*-continuous. Let $\mu \in M(X), (\mu_n)_{n\geq 1} \in M(X)^{\mathbb{N}}$ be such that μ_n converges to μ in the sense of the weak*-topology, i.e.

$$\forall f \in C(X)\colon \int_X f \ \mathrm{d}\mu_n \to \int_X f \ \mathrm{d}\mu$$

as $n \to \infty$. Let us show that $\int_X f \ \mathrm{d}\tilde{T}\mu_n \to \int_X f \ \mathrm{d}\tilde{T}\mu$ as $n \to \infty$. This follows from Lemma 4.7 and the fact that $f \circ T$ is continuous and bounded as both f and T are continuous and f is bounded, so

$$\int_X f \ \mathrm{d}\tilde{T}\mu_n = \int_X (f \circ T)(x) \ \mathrm{d}\mu_n(x) \overset{n\to\infty}{\to} \int_X (f \circ T)(x) \ \mathrm{d}\mu(x) = \int_X f \ \mathrm{d}\tilde{T}\mu.$$

$(*)$ follows from the fact that for all $B \in \mathcal{B}$

$$\begin{aligned}(p\tilde{T}\mu + (1-p)\tilde{T}\nu)(B) &= p\mu\left(T^{-1}(B)\right) + (1-p)\nu\left(T^{-1}(B)\right) \\ &= (p\mu + (1-p)\nu)\left(T^{-1}(B)\right) \\ &= \left(\tilde{T}(p\mu + (1-p)\nu)\right)(B).\end{aligned}$$

□

[10] In other words, we use that the set of step functions is dense in in $C(X)$.

For measurable $T\colon X \to X$, let

$$M(X,T) \coloneqq \left\{\mu \in M(X)\colon \mu = \tilde{T}\mu\right\}$$

denote the set of T-invariant probability measures on X. Like $M(X)$, $M(X,T)$ is convex. We would like to show that $M(X,T) \neq \emptyset$. Let us start with the following theorem.

Theorem 4.9. Let T be continuous. Then we have $\mu \in M(X,T)$ if and only if $\int_X f \circ T \, \mathrm{d}\,\mu = \int_X f \, \mathrm{d}\,\mu$ for all $f \in C(X)$.

Proof. The statement follows from Lemma 4.7 and the fact that for $\mu, \nu \in M(X)$ one has that $\mu = \nu$ if and only if

$$\forall\, f \in C(X)\colon \int_X f \, \mathrm{d}\,\mu = \int_X f \, \mathrm{d}\,\nu.$$[11] □

Theorem 4.10. Let $T\colon X \to X$ be continuous and $(\sigma_n)_{n\geq 1} \in M(X)^{\mathbb{N}}$. For any $n \geq 1$, set

$$\mu_n \coloneqq \frac{1}{n}\sum_{i=1}^{n} \tilde{T}^i \sigma_n \in M(X).$$

Then *any* limit point of $(\mu_n)_{n\geq 1}$ is in $M(X,T)$, i.e. T-invariant.

Proof. Let μ be a limit point of $(\mu_n)_{n\geq 1}$, i.e. there exists an increasing sequence $(n_j)_{j\geq 1}$ such that $\mu_{n_j} \to \mu$ weakly* as $j \to \infty$. Then for every $f \in C(X)$ one has that $f \circ T \in C(X)$, so

$$\begin{aligned}
&\left|\int_X f \circ T \, \mathrm{d}\,\mu - \int_X f \, \mathrm{d}\,\mu\right| \\
&= \lim_{j\to\infty} \left|\int_X f \circ T \, \mathrm{d}\,\mu_{n_j} - \int_X f \, \mathrm{d}\,\mu_{n_j}\right| \\
&= \lim_{j\to\infty} \frac{1}{n_j}\left|\sum_{i=1}^{n_j}\int_X \left(f \circ T^{i+1}\right) - \left(f \circ T^{i}\right) \, \mathrm{d}\,\sigma_{n_j}\right| && \text{(Def. of } \mu_n\text{, Lem. 4.7)} \\
&= \lim_{j\to\infty} \frac{1}{n_j}\left|\int_X \left(f \circ T^{n_j+1}\right) - (f \circ T) \, \mathrm{d}\,\sigma_{n_j}\right| && \text{(Telescoping sum)} \\
&\leq \limsup_{j\to\infty} \frac{2\left\|f\right\|_\infty}{n_j} && \text{(Triangle inequality)} \\
&= 0,
\end{aligned}$$

hence $\mu \in M(X,T)$ by Theorem 4.9. □

Finally, we are ready to prove the Krylov-Bogolyubov theorem.

Proof of Theorem 4.3. Fix any $(\sigma_n)_{n\geq 1} \in M(X)^{\mathbb{N}}$. Then $(\mu_n)_{n\geq 1}$ from the previous theorem has at least one limit point $\mu \in M(X)$ by the weak*-compactness of $M(X)$, see Lemma 4.5. So, by Theorem 4.10, μ is the desired measure. □

[11] For this fact, see any book on measure theory.

Remark. Note that such a sequence $(\sigma_n)_{n\geq 1}$ exists as every singleton set can be written as an intersection of balls with fixed centre and decreasing radii. In particular, all singleton sets are contained in our σ-algebra. Hence, we always have Dirac measures in $M(X)$.

5 Conditional Expectations

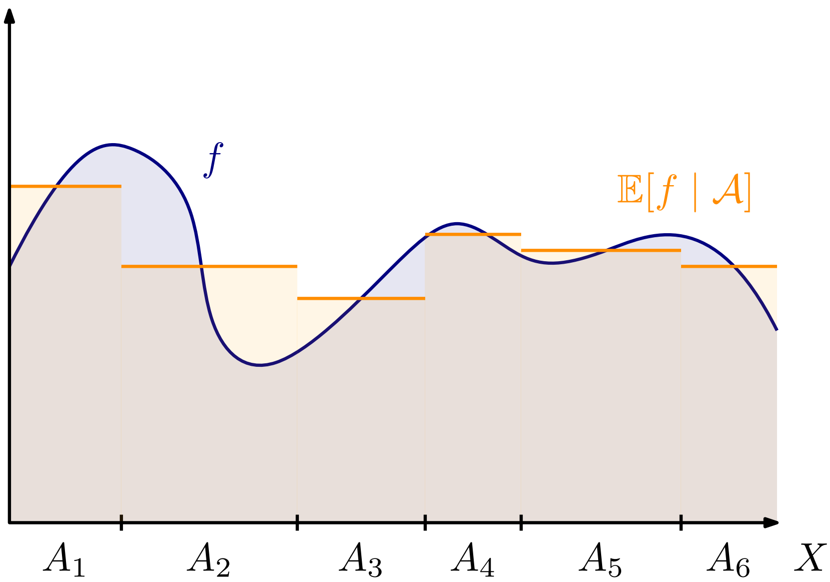
Definition 5.1. Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathcal{A} \subseteq \mathcal{B}$ be a sub-σ-algebra. Let $f\colon X \to \mathbb{R}$ be integrable and $\mathcal{B}$-measurable. Then there exists a *unique*[12] $\mathcal{A}$-measurable function $h \in L^1(X)$ such that

$$\int_A h \, \mathrm{d}\,\mu = \int_A f \, \mathrm{d}\,\mu$$

for all $A \in \mathcal{A}$. Such h is called *the conditional expectation of f with respect to $\mathcal{A}$* and is denoted by $\mathbb{E}\,[\,f \mid \mathcal{A}\,]$.

The existence and uniqueness is proven in any measure theory book, e.g. *Measure Theory* by Bogachev.

Examples 5.2.



1. Let $A_1, \ldots, A_n \in \mathcal{B}$ be disjoint, $A_1 \cup \cdots \cup A_n = X$, and let $\mathcal{A} = \sigma(\{A_1, \ldots, A_n\})$. Assume additionally that $\mu(A_i) > 0, i = 1, \ldots, n$. Then

$$\mathbb{E}\,[\,f \mid \mathcal{A}\,] = \sum_{i=1}^{n} \frac{\int_{A_i} f \, \mathrm{d}\,\mu}{\mu(A_i)} \chi_{A_i}.$$

2. Let $\mathcal{A} = \{\emptyset, X\}$. Then $\mathbb{E}\,[\,f \mid \mathcal{A}\,]$ must be a constant as it is $\mathcal{A}$-measurable, and thus

$$\mathbb{E}\,[\,f \mid \mathcal{A}\,] = \int_X \mathbb{E}\,[\,f \mid \mathcal{A}\,] \, \mathrm{d}\,\mu = \int_X f \, \mathrm{d}\,\mu = \mathbb{E}\,[\,f\,].$$

[12] Up to "almost everywhere" equality.

Properties of the conditional expectation

1. For all $f, g \in L^1(X), a, b \in \mathbb{R}$ one has that

$$\mathbb{E}[\, af + bg \mid \mathcal{A}\,] = a\mathbb{E}[\, f \mid \mathcal{A}\,] + b\mathbb{E}[\, g \mid \mathcal{A}\,]$$

almost surely. In other words, $\mathbb{E}[\,\cdot \mid \mathcal{A}\,]$ is linear.

2. If $f \leq g$, then $\mathbb{E}[\, f \mid \mathcal{A}\,] \leq \mathbb{E}[\, g \mid \mathcal{A}\,]$. In other words, $\mathbb{E}[\,\cdot \mid \mathcal{A}\,]$ is monotone.
Hint: First show that if for two $\mathcal{A}$-measurable functions $f_1, f_2 \in L^1(X)$ one has

$$\int_A f_1 \;\mathrm{d}\,\mu \geq \int_A f_2 \;\mathrm{d}\,\mu$$

for all $A \in \mathcal{A}$, then $f_1 \geq f_2$ almost surely.

3. As $X \in \mathcal{A}$,

$$\int_X \mathbb{E}[\, f \mid \mathcal{A}\,] \;\mathrm{d}\,\mu = \int_X f \;\mathrm{d}\,\mu.$$

4. For any $\mathcal{A}$-measurable $f \in L^1(X)$ one has that $\mathbb{E}[\, f \mid \mathcal{A}\,] = f$ almost surely.

5. Two sets A and B from $\mathcal{B}$ are called *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$. Two σ-algebras $\mathcal{F}, \mathcal{G} \subseteq \mathcal{B}$ on $(X, \mathcal{B})$ are called independent if for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$ one has that A and B are independent. We say that $L^1(X) \ni f \colon X \to \mathbb{R}$ is *independent of the σ-algebra* $\mathcal{A}$ if $\sigma(f) := \sigma(\{f^{-1}(B) \colon B \in \mathcal{B}(\mathbb{R})\})$ and $\mathcal{A}$ are independent. In this case,

$$\mathbb{E}[\, f \mid \mathcal{A}\,] = \int_X f \;\mathrm{d}\,\mu$$

almost surely.

Proof. Exercise for the reader. □

Hint: Use that $\int_X fg \;\mathrm{d}\,\mu = \int_X f \;\mathrm{d}\,\mu \int_X g \;\mathrm{d}\,\mu$ for independent $f, g \colon X \to \mathbb{R}$ where $f, g, fg \in L^1(X)$.

6. Let $f \in L^1(X)$, where f is not necessarily $\mathcal{A}$-measurable, and $g \in L^\infty(X)$ be $\mathcal{A}$-measurable. Then

$$\mathbb{E}[\, fg \mid \mathcal{A}\,] = g\mathbb{E}[\, f \mid \mathcal{A}\,]$$

almost surely.

7. Let $\mathcal{A}_2 \subseteq \mathcal{A}_1 \subseteq \mathcal{B}$. Then for every $f \in L^1(X)$ we have

$$\mathbb{E}[\, \mathbb{E}[\, f \mid \mathcal{A}_1\,] \mid \mathcal{A}_2\,] = \mathbb{E}[\, f \mid \mathcal{A}_2\,] \qquad (\textit{Tower property})$$

almost surely.

8. Let $f \colon X \to \mathbb{R}$ be measurable and $\mathcal{A}$-independent and let $g \colon X \to \mathbb{R}$ be $\mathcal{A}$-measurable. Let $h \colon \mathbb{R}^2 \to \mathbb{R}$ be a Borel-measurable function such that $h(f, g) \in L^1(X)$. Then for μ-almost all $y \in X$ one has

$$\mathbb{E}[\, h(f, g) \mid \mathcal{A}\,](y) = \int_X h(f(x), g(y)) \;\mathrm{d}\,\mu(x).$$

6 Ergodic Transforms and Birkhoff's Ergodic Theorem

Let us now return to transformations: Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ be measurable and μ-preserving.

Definition 6.1 (Ergodicity)**.** T is called *ergodic* if for any $A \in \mathcal{B}$ such that $T^{-1}(A) = A$ one has $\mu(A) \in \{0, 1\}$. A set $A \in \mathcal{B}$ is called T*-invariant* if $T^{-1}(A) = A$. A measurable function $f \to \mathbb{R}$[13] is called T*-invariant* if $f \circ T = f$ almost surely.

Recall that the *symmetric difference* of $A, B \in \mathcal{B}$ is defined by

$$A \,\Delta\, B \coloneqq (A \setminus B) \cup (B \setminus A).$$

Proposition 6.2. The following statements are equivalent:

- (i) T is ergodic.
- (ii) For any $A \in \mathcal{B}$ such that $\mu(A \,\Delta\, T^{-1}(A)) = 0$, one has $\mu(A) \in \{0, 1\}$.[14]
- (iii) Every T-invariant function is constant.

Note that the second condition is really a natural relaxation of our definition above for ergodicity as it states that measurable sets still have measure one or zero if they are T-invariant up to a null set. This will play in nicely to show that (ii) implies (iii) since the T-invariance of a function only gives equality almost everywhere.

Proof.

$(i) \implies (ii)$: Let T be ergodic and let $A \in \mathcal{B}$ be such that $\mu(A \,\Delta\, T^{-1}(A)) = 0$. Let $\tilde{A} = \bigcap_{n \geq 0} \bigcup_{i \geq n} T^{-i}(A)$.[15] Let us show that $\tilde{A}$ is T-invariant. To this end, start with $T^{-1}(\tilde{A}) \subseteq \tilde{A}$: Fix $x \in \tilde{A}$. For $n \geq 0$ there exists $j(n) \in \mathbb{N}, j(n) \geq n$, such that $x \in T^{-j(n)}(A)$. Then

$$T^{-1}(x) \subseteq T^{-(j(n)+1)}(A) \subseteq \bigcup_{i \geq n} T^{-i}(A).$$

As this holds for every $n \geq 0$,

$$T^{-1}(x) \subseteq \bigcap_{n \geq 0} \bigcup_{i \geq n} T^{-i}(A) = \tilde{A},$$

so the desired follows. Let us show that $\tilde{A} \subseteq T^{-1}(\tilde{A})$. This holds as

$$T^{-1}(\tilde{A}) = T^{-1}\left(\bigcap_{n \geq 0} \bigcup_{i \geq n} T^{-i}(A)\right)$$

[13] One may also replace $\mathbb{R}$ by the compactification $[-\infty, \infty]$.

[14] I.e. every measurable set A that is apart from a null set T-invariant must have measure zero or one.

[15] I.e. $\tilde{A}$ is the set of $x \in X$ for which $T^i(x) \in A$ for infinitely many $i \in \mathbb{N}_0$.

$$\begin{aligned}
&= \bigcap_{n\geq 0} \bigcup_{i\geq n} T^{-(i+1)}(A) \\
&= \bigcap_{n\geq 1} \bigcup_{i\geq n} T^{-i}(A) \\
&\supseteq \tilde{A}.
\end{aligned}$$

As $\tilde{A}$ is T-invariant, $\mu(\tilde{A}) \in \{0,1\}$. It remains to show that $\mu(A \,\Delta\, \tilde{A}) = 0$. We start with noticing that[16]

$$\begin{aligned}
\mu(T^{-n}(A) \,\Delta\, A) &\leq \mu\left(\bigcup_{i=1}^{n} T^{-i}(A) \,\Delta\, T^{-(i-1)}(A)\right) \\
&\leq \sum_{i=1}^{n} \mu\left(T^{-i}(A) \,\Delta\, T^{-(i-1)}(A)\right) \\
&= \sum_{i=1}^{n} \mu\left(T^{-(i-1)}\left(T^{-1}(A) \,\Delta\, A\right)\right) \\
&= \sum_{i=1}^{n} \mu\left(T^{-1}(A) \,\Delta\, A\right) \\
&= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu\left(\left(\bigcup_{i\geq n} T^{-i}(A)\right) \,\Delta\, A\right) &\leq \mu\left(\bigcup_{i\geq n}\left(T^{-i}(A) \,\Delta\, A\right)\right) \\
&\leq \sum_{i\geq n} \mu\left(T^{-i}(A) \,\Delta\, A\right) \\
&= 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mu\left(\tilde{A} \,\Delta\, A\right) &= \mu\left(\left(\bigcap_{n\geq 0} A_n\right) \,\Delta\, A\right) && \leq \mu\left(\bigcup_{n\geq 0}(A_n \,\Delta\, A)\right) \\
&= \sum_{n\geq 0} \mu(A_n \,\Delta\, A) && = 0,
\end{aligned}$$

where $A_n = \bigcup_{i\geq n} T^{-i}(A)$.

$(ii) \implies (iii)$: Let $f\colon X \to \mathbb{R}$ be T-invariant. Let $B \in \mathcal{B}(\mathbb{R}), A = f^{-1}(B)$. First, note that

$$\mu\left((f\circ T)^{-1}(B) \,\Delta\, f^{-1}(B)\right) \leq \mu(f\circ T \neq f) = 0.$$

[16] As $A \,\Delta\, C \subseteq ((A \,\Delta\, B) \cup (B \,\Delta\, C))$ for any sets A, B, C.

Next, notice that

$$T^{-1}(A) = T^{-1}\left(f^{-1}(B)\right) = (f \circ T)^{-1}(B),$$

so $\mu(A \,\Delta\, T^{-1}(A)) = 0$, i.e. $\mu(A) \in \{0, 1\}$. Thus,

$$\mu\left(f^{-1}(B)\right) \in \{0, 1\}$$

for all $B \in \mathcal{B}(\mathbb{R})$.

Exercise 6.3. Conclude that f is constant a.s.

Proof. For each $k \in \mathbb{Z}$ and $n \in \mathbb{N}$ we define

$$A_n^k := \left\{x \in X : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\}.$$

$(A_n^k)_{k\in\mathbb{Z}}$ forms a partition for every fixed $n \in \mathbb{N}$ and since f is T-invariant so is A_n^k. As the preimage of a Borel set of f, $\mu(A_n^k)$ equals 0 or 1. Hence, there exists exactly one $k =: k_n$ such that $\mu(A_n^{k_n}) = 1$. Thus,

$$\tilde{X} := \bigcap_{n\in\mathbb{N}} A_n^{k_n}$$

has measure 1 and f is constant on $\tilde{X}$ with value $\lim_{n\to\infty} k_n/2^n$. $\square$

$(iii) \implies (i)$: Assume that every T-invariant function is constant a.s. Fix T-invariant $A \in \mathcal{B}$. Let $f = \chi_A$. Then

$$f \circ T = \chi_{T^{-1}(A)} = \chi_A,$$

so f is T-invariant. As f is constant a.s., $\mu(A) \in \{0, 1\}$. $\square$

Theorem 6.4 (Birkhoff's ergodic theorem)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T \colon X \to X$ be μ-preserving. Furthermore, let $f \in L^1(X)$. Then, one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \mathbb{E}\left[\, f \mid \mathcal{G}\,\right](x) \tag{$*$}$$

for almost every $x \in X$, where the left hand side of $(*)$ is well-defined and

$$\mathcal{G} := \left\{A \in \mathcal{B} \colon T^{-1}(A) = A\right\}$$

is the σ-algebra generated by / of all T-invariant sets.

Exercise 6.5. Show that any $A \in \mathcal{G}$ is T-invariant.[17] Conclude that any $\mathcal{G}$-measurable $f \colon X \to \mathbb{R}$ is T-invariant.[18]

[17] In other words, show that the set of all T-invariant sets is a σ-algebra.

[18] Just plug in the definitions.

Hint: Show that countable unions, intersections, and complements of T-invariant sets are T-invariant.

Corollary 6.5.1. T is ergodic if and only if for every $f \in L^1(X)$ one has

$$\mathbb{E}\left[\, f \mid \mathcal{G} \,\right] = \int_X f \; \mathrm{d}\,\mu$$

almost surely. In particular, for almost all $x \in X$, one has

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f \; \mathrm{d}\,\mu$$

almost surely. That is, T *is ergodic if and only if the time average converges to the space average.*

Proof. Let T be ergodic and let $f \in L^1(X)$. We need to show that

$$\mathbb{E}\left[\, f \mid \mathcal{G} \,\right] = \int_X f \; \mathrm{d}\,\mu$$

almost surely. As $\mathcal{G}$ is the σ-algebra of all T-invariant sets (see Exercise 6.5), the sets

$$\begin{aligned} A_+ &= \left\{ \mathbb{E}\left[\, f \mid \mathcal{G} \,\right] > \int_X f \; \mathrm{d}\,\mu \right\}, \\ A_- &= \left\{ \mathbb{E}\left[\, f \mid \mathcal{G} \,\right] < \int_X f \; \mathrm{d}\,\mu \right\}, \\ A_0 &= \left\{ \mathbb{E}\left[\, f \mid \mathcal{G} \,\right] = \int_X f \; \mathrm{d}\,\mu \right\} \end{aligned}$$

are T-invariant. In particular, their measure is 0 or 1. Now, as $A_+ \mathbin{\dot\cup} A_- \mathbin{\dot\cup} A_0 = X$, one of them has measure one and the remaining two have measure zero. If $\mu(A_+) = 1$, then

$$\begin{aligned} 0 &= \int_{A_+} f \; \mathrm{d}\,\mu - \int_X f \; \mathrm{d}\,\mu && = \int_{A_+} \mathbb{E}\left[\, f \mid \mathcal{G} \,\right] \; \mathrm{d}\,\mu - \int_{A_+} \int_X f \; \mathrm{d}\,\mu \; \mathrm{d}\,\mu \\ &= \int_{A_+} \underbrace{\left[\mathbb{E}\left[\, f \mid \mathcal{G} \,\right] - \int_X f \; \mathrm{d}\,\mu \right]}_{>0} \mathrm{d}\,\mu && > 0. \; \lightning \end{aligned}$$

So, $\mu(A_+) = 0$. Similarly, $\mu(A_-) = 0$ and hence $\mu(A_0) = 1$.
Now, let T be not ergodic. Let us find $f \in L^1(X)$ such that $\mathbb{E}\left[\, f \mid \mathcal{G} \,\right]$ is almost surely not a constant. As T is not ergodic, there exists T-invariant $A \in \mathcal{B}$ such that $0 < \mu(A) < 1$. Set $f \coloneqq \chi_A$. Then f is $\mathcal{G}$-measurable as $A \in \mathcal{G}$. Therefore,

$$\mathbb{E}\left[\, f \mid \mathcal{G} \,\right] = f = \chi_A$$

almost surely, which is almost surely not constant. ∎

Corollary 6.5.2 (Borel's theorem on Normal Numbers)**.** Almost any number $x \in [0,1)$ has the same frequency of ones and zeros in its binary code, i.e. if

$$x = \sum_{k \in \mathbb{N}} a_k 2^{-k}$$

with $a_1, a_2, \ldots \in \{0,1\}$[19], then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \underbrace{\chi_1(a_n)}_{=a_n} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \underbrace{\chi_0(a_n)}_{=1-a_n} = \frac{1}{2}.$$

In this case we say that x is *normal to base 2*.

Remark. This theorem can be generalized to an arbitrary base $\mathbb{N} \ni b > 1$. Also, since the frequency of ones and zeros of any real number is dominated by its fractional part, the theorem above actually holds for almost all real numbers.

Proof. Let $T\colon [0,1) \to [0,1)$ be defined by $Tx = 2x \pmod 1, x \in [0,1)$.

Exercise 6.6. Show that such T is measure-preserving (with respect to the Lebesgue measure).

Let us show that T is ergodic. By Proposition 6.2, it is sufficient to show that for every T-invariant $f\colon [0,1) \to \mathbb{R}$, i.e. $f \circ T = f$ almost surely, one has that f is constant almost surely. Fix such a T-invariant $f\colon [0,1) \to \mathbb{R}$. W.l.o.g. we may assume that $f \in L^\infty([0,1))$ as we can reset for every $C > 0$

$$\tilde{f}(x) = \begin{cases} C, & f(x) > C \\ f(x), & -C \le f(x) \le C \\ -C, & f(x) < -C, \end{cases}$$

which is also T-invariant. As $f \in L^\infty([0,1))$, then in particular $f \in L^2([0,1))$, so it has a Fourier decomposition

$$f(x) = \sum_{z \in \mathbb{Z}} b_z \,\mathrm{e}^{2\pi \,\mathrm{i}\, zx}\,.$$

As T is measure-preserving, we have in an L^2-sense

$$\begin{aligned}(f \circ T)(x) &= f(Tx) \\ &= \sum_{z \in \mathbb{Z}} b_z \,\mathrm{e}^{2\pi \,\mathrm{i}\, z\cdot(2x \pmod 1)} \\ &= \sum_{z \in \mathbb{Z}} b_z \,\mathrm{e}^{2\pi \,\mathrm{i}\, z\cdot 2x} \\ &= \sum_{z \in \mathbb{Z}} b_z \,\mathrm{e}^{2\pi \,\mathrm{i}(2z)x}\end{aligned}$$

[19] Note that the coefficients $a_1, a_2, \ldots$ are for almost any number unique.

$$\stackrel{!}{=} f(x). \qquad (T \text{ is } \lambda\text{-preserving.})$$

Thus, $b_z = b_{2z} = b_{4z} = \dots$ for all $z \in \mathbb{Z}$. If $b_z \neq 0$ for some $z \neq 0$, then $\sum_{z\in\mathbb{Z}} |b_z|^2 = \infty$, which by Parseval's theorem contradicts the fact that f is in $L^2([0,1))$. So, $b_z = 0$ for $z \neq 0$ and f is almost surely a constant. Thus, T is ergodic. Set $f = \chi_{[1/2,1)}$. Then by Corollary 6.5.1, for almost every

$$x = \sum_{k\geq 1} a_k 2^{-k} \in [0,1)$$

we have that

$$\begin{aligned}
\frac{1}{2} &= \int_0^1 f(y) \ \mathrm{d}\, y \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(2^n x \pmod 1) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(\sum_{k\geq 1} a_k 2^{n-k} \pmod 1\right) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(\sum_{l\geq 1} a_{l+n} 2^{-l}\right) && (l = k-n) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} a_{n+1} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n,
\end{aligned}$$

which is the desired result. □

Let us now show Birkhoff's theorem / Theorem 6.4. In order to do so we will need the following lemma:

Lemma 6.7 (Maximal ergodic theorem)**.** In the setting of Theorem 6.4, set for every $x \in X$ and $N \geq 1$

$$S_N(x) = \sum_{n=0}^{N-1} f(T^n x), \qquad M_N(x) \coloneqq \max\left\{S_0(x), \dots, S_N(x)\right\},$$

where $S_0(x) \coloneqq 0$. Then

$$\int_{\{M_N > 0\}} f \ \mathrm{d}\,\mu \geq 0$$

for every $N \geq 1$.

Proof. For any $0 \le k \le N$ and $x \in X$ one has $M_N(Tx) \ge S_k(Tx)$ and so

$$\begin{aligned} f(x) + M_N(Tx) &\ge f(x) + S_k(Tx) \\ &= f(x) + \sum_{n=0}^{k-1} f(T^{n+1}x) \\ &= \sum_{n=0}^{k} f(T^n x) \\ &= S_{k+1}(x). \end{aligned}$$

Therefore,

$$f(x) \ge \max\{S_1(x), \dots, S_N(x)\} - M_N(Tx). \tag{$*$}$$

Next, for $x \in \{M_N > 0\}$, one has that $0 = S_0(x) < M_N(x)$, so

$$M_N(x) = \max\{S_1(x), \dots, S_N(x)\}$$

and hence by $(*)$

$$f(x) \ge M_N(x) - (M_N \circ T)(x).$$

Thus,

$$\begin{aligned} \int_{\{M_N>0\}} f \,\mathrm{d}\mu &\ge \int_{\{M_N>0\}} M_N - M_N \circ T \,\mathrm{d}\mu \\ &= \int_X M_N \,\mathrm{d}\mu - \int_{\{M_N>0\}} M_N \circ T \,\mathrm{d}\mu \end{aligned} \tag{$**$}$$

as

$$\int_{\{M_N \le 0\}} M_N \,\mathrm{d}\mu = 0$$

since M_N is nonnegative, so $M_N = 0$ on $\{M_N \le 0\}$. Now,

$$\begin{aligned} \int_{\{M_N>0\}} M_N \circ T \,\mathrm{d}\mu &= \int_X \chi_{\{M_N>0\}} \cdot M_N \circ T \,\mathrm{d}\mu \\ &\le \int_X M_N \circ T \,\mathrm{d}\mu \\ &= \int_X M_N \,\mathrm{d}\mu \end{aligned} \tag{$***$}$$

as T is μ-preserving.[20] Hence, by $(**)$ and $(***)$

$$\int_{\{M_N>0\}} f \,\mathrm{d}\mu \ge 0. \qquad \square$$

[20] This fact is left as an exercise for the reader.

Proof of Theorem 6.4. First, note that it is sufficient to show that

$$\begin{aligned} 0 &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left(f(T^n x) - \mathbb{E}\left[\, f \mid \mathcal{G}\,\right](x)\right) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left[f - \mathbb{E}\left[\, f \mid \mathcal{G}\,\right]\right](T^n x), \end{aligned}$$

where $\mathbb{E}[\, f \mid \mathcal{G}\,](x) = \mathbb{E}[\, f \mid \mathcal{G}\,](T^n x)$ almost surely as $\mathbb{E}[\, f \mid \mathcal{G}\,]$ is $\mathcal{G}$-measurable, and hence T-invariant (see Exercise 6.5). By replacing f by $f - \mathbb{E}[\, f \mid \mathcal{G}\,]$, we may assume that $\mathbb{E}[\, f \mid \mathcal{G}\,] = 0$. Assume additionally that f is bounded as bounded functions are dense in $L^1(X)$ and as each of the mappings

$$\begin{aligned} f &\mapsto \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n, \\ f &\mapsto \mathbb{E}\left[\, f \mid \mathcal{G}\,\right] \end{aligned}$$

are continuous linear operators on $L^1(X)$ with norm 1. Define

$$\begin{aligned} \bar{S} &\coloneqq \limsup_{n\to\infty} \frac{S_n}{n} \\ \underline{S} &\coloneqq \liminf_{n\to\infty} \frac{S_n}{n}, \end{aligned}$$

where S_n is defined as in Lemma 6.7. We want to show that $\bar{S} = \underline{S} = 0$ almost surely. To this end it is enough to show that $\bar{S} \leq 0$ almost surely, as then by considering $-f$ one gets $\underline{S} \geq 0$ almost surely, so $0 \leq \underline{S} \leq \bar{S} \leq 0$, hence $\bar{S} = \underline{S} = 0$.

Let us show that almost surely $\bar{S} \leq 0$. We start by noticing that

$$\begin{aligned} (\bar{S} \circ T)(x) &= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{n+1} x) \\ &= \limsup_{N\to\infty} \left[\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) + \underbrace{\frac{1}{N} \Big(f(T^N x) - f(x) \Big)}_{\leq \frac{2\|f\|_\infty}{N} \to 0, N\to\infty} \right] \\ &= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \\ &= \bar{S}(x). \end{aligned}$$

So, $\bar{S} \circ T = \bar{S}$. Hence $\bar{S}$ is $\mathcal{G}$-measurable. Fix $\varepsilon > 0$ and set

$$A^\varepsilon \coloneqq \Big\{ x \in X \colon \bar{S}(x) > \varepsilon \Big\}.$$

Then $A^\varepsilon \in \mathcal{G}$ as $\bar{S}$ is $\mathcal{G}$-measurable. We want to show that

$$\mu(A^\varepsilon) = 0.$$

The idea: If there is a such a set with positive measure, we could take our original f and reduce it by something smaller or equal to ε. Call this new function f^ε. By definition that would also reduce $\bar{S}$ by ε. But since A^ε is $\mathcal{G}$-measurable we already have

$$\int_{A^\varepsilon} f \;\mathrm{d}\,\mu = \int_{A^\varepsilon} \mathbb{E}\left[\, f \mid \mathcal{G}\,\right] \;\mathrm{d}\,\mu = 0.$$

If we could then show that

$$0 \leq \int_{A^\varepsilon} f^\varepsilon \;\mathrm{d}\,\mu$$

we would be left with

$$0 \leq -\varepsilon\mu(A^\varepsilon).$$

The proof (*continued*): Set $f^\varepsilon := (f - \varepsilon)\chi_{A^\varepsilon}$. As $A^\varepsilon \in \mathcal{G}$, $T^{-1}(A^\varepsilon) = A^\varepsilon$, so $\chi_{A^\varepsilon} \circ T^n = \chi_{A^\varepsilon}$, and

$$\begin{aligned} S_N^\varepsilon(x) &:= \sum_{n=0}^{N-1} f^\varepsilon(T^n x) \qquad\qquad = \left(S_N(x) - N\varepsilon\right)\chi_{A^\varepsilon}(x),\\ M_N^\varepsilon(x) &:= \max\left\{S_0^\varepsilon(x), \ldots, S_N^\varepsilon(x)\right\}. \end{aligned}$$

In particular,

$$\frac{S_N^\varepsilon(x)}{N} = \begin{cases} 0, & \bar{S}(x) \leq \varepsilon \\ \frac{S_N(x)}{N} - \varepsilon, & \text{otherwise.} \end{cases} \tag{$*$}$$

The sequence of sets $\{M_N^\varepsilon > 0\}$ increases as $N \mapsto M_N^\varepsilon$ increases pointwise. Moreover,

$$B^\varepsilon := \bigcup_{N\geq 1}\{M_N^\varepsilon > 0\} = \left\{\sup_{N\in\mathbb{N}} S_N^\varepsilon > 0\right\} = \left\{\sup_{N\in\mathbb{N}} \frac{S_N^\varepsilon}{N} > 0\right\}. \tag{$**$}$$

By $(*)$, the following equivalences hold:

$$\sup_{N\in\mathbb{N}} \frac{S_N^\varepsilon(x)}{N} > 0 \iff \left(\exists\, N \geq 1\colon \frac{S_N(x)}{N} - \varepsilon > 0\right) \iff \bar{S}(x) > \varepsilon.$$

Therefore,

$$B^\varepsilon = \left\{\sup_{N\in\mathbb{N}} \frac{S_N^\varepsilon}{N} > 0\right\} = \left\{\bar{S} > \varepsilon\right\} = A^\varepsilon.$$

As $f \in L^1(X)$, we have $f^\varepsilon \in L^1(X)$ as well, so by the dominated convergence theorem, $(**)$ and the previous lemma one has

$$\begin{aligned} 0 &\leq \lim_{N\to\infty} \int_{\{M_N^\varepsilon > 0\}} f^\varepsilon \;\mathrm{d}\,\mu \\ &= \int_{\bigcup_{N\in\mathbb{N}}\{M_N^\varepsilon > 0\}} f^\varepsilon \;\mathrm{d}\,\mu \\ &= \int_{B^\varepsilon} f^\varepsilon \;\mathrm{d}\,\mu \end{aligned}$$

$$
\begin{aligned}
&= \int_{A^\varepsilon} f^\varepsilon \, \mathrm{d}\,\mu \\
&= \int_{A^\varepsilon} f \, \mathrm{d}\,\mu - \varepsilon\mu(A^\varepsilon).
\end{aligned}
$$

As $A^\varepsilon \in \mathcal{G}$, the latter equals

$$\int_{A^\varepsilon} \mathbb{E}\left[\, f \mid \mathcal{G} \,\right] \, \mathrm{d}\,\mu - \varepsilon\mu(A^\varepsilon) = -\varepsilon\mu(A^\varepsilon),$$

where we used that $\mathbb{E}\left[\, f \mid \mathcal{G} \,\right] = 0$. Thus, $\mu(A^\varepsilon) = 0$ for all $\varepsilon > 0$, so $\bar{S} \leq 0$ almost surely and the desired follows. $\square$

Corollary 6.7.1. Let $(X, \mathcal{B}, \mu)$ be a probability space, $T \colon X \to X$ be μ-preserving. Then T is ergodic if and only if for any $A, B \in \mathcal{B}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B). \tag{$\star$}$$

In other words, the event that a point is in B (if following a probability distribution according to μ) and the event that such a point ends up in A after n steps are in a time average sense almost independent for large n.

Proof. Let T be ergodic. Fix $A, B \in \mathcal{B}$. Then by Theorem 6.4,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \chi_A(T^n x) = \int_X \chi_A \, \mathrm{d}\,\mu = \mu(A)$$

for almost every $x \in X$. Hence,

$$\lim_{N\to\infty} \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} \chi_A(T^n x)\chi_B(x)}_{\leq \chi_B(x)} = \mu(A)\chi_B(x)$$

for almost every $x \in X$. Thus, by the dominated convergence theorem, we get

$$
\begin{aligned}
\mu(A)\mu(B) &= \int_X \mu(A)\chi_B \, \mathrm{d}\,\mu \\
&= \lim_{N\to\infty} \int_X \frac{1}{N} \sum_{n=0}^{N-1} \chi_A(T^n x)\chi_B(x) \, \mathrm{d}\,\mu(x) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X \chi_{T^{-n}(A)}(x)\chi_B(x) \, \mathrm{d}\,\mu(x) \qquad (T^n x \in A \iff x \in T^{-n}(A)) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}(A) \cap B),
\end{aligned}
$$

which is ($\star$).

For the other direction, let ($\star$) holds for any $A, B \in \mathcal{B}$. Fix $A \in \mathcal{B}$ such that $T^{-1}(A) = A$. Then by ($\star$)

$$(\mu(A))^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}(A) \cap A) = \mu(A).$$

So, $\mu(A) \in \{0, 1\}$ and hence T is ergodic. $\square$

7 Ergodic Measures for Continuous Transforms

Recall the notation used in Section 4, i.e.

- X is a compact metric space,
- $\mathcal{B} = \mathcal{B}(X)$ is the Borel σ-algebra generated by all balls in X,
- $T\colon X \to X$ is continuous,
- $M(X)$ is the set of all probability measures on $(X, \mathcal{B})$ and endowed with a weak*-topology, i.e. $\mu_n \to \mu$ in $M(X)$ if for all $f \in C(X)$
$$\int_X f \,\mathrm{d}\mu_n \to \int_X f \,\mathrm{d}\mu, n \to \infty,$$
- $M(X,T) \subseteq M(X)$ is the set of all T-invariant probabilistic measures, i.e. $\mu \in M(X,T)$ if $\mu(T^{-1}(A)) = \mu(A)$ for any $A \in \mathcal{B}$. By Section 4, $M(X,T)$ is non-empty.

Our goal in this chapter is to study the structure of $M(X,T)$ further and, in particular, establish that there always exists a measure such that T is ergodic under the above conditions.

Definition 7.1 (Ergodicity, absolute continuity, mutual singularity)**.** We say that $\mu \in M(X,T)$ is called *ergodic* if T is ergodic on $(X, \mathcal{B}, \mu)$. For $\mu, \nu \in M(X)$ we say that μ is *absolutely continuous* with respect to ν if for all $A \in \mathcal{B}$ $\nu(A) = 0$ implies $\mu(A) = 0$. We denote this by $\mu \ll \nu$. μ and ν are *mutually singular* if there exists $A \in \mathcal{B}$ such that $\mu(A) = 0$ and $\nu(A) = 1$.

Lemma 7.2. Let $\mu, \nu \in M(X,T)$ be ergodic. If $\mu \ll \nu$, then $\mu = \nu$.

Proof. Fix a measurable and bounded function $f\colon X \to \mathbb{R}$. By Theorem 6.4, one has
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f \,\mathrm{d}\nu \tag{$*$}$$
for ν-a.e. $x \in X$. Let $\Omega \subseteq X$ be such that $\nu(\Omega) = 1$ and $(*)$ holds pointwise on Ω. Then $\nu(X \setminus \Omega) = 0$, so $\mu(X \setminus \Omega) = 0, \mu(\Omega) = 1$. Thus,
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f \,\mathrm{d}\nu$$
for μ-a.e. $x \in X$. But by Theorem 6.4, it equals
$$\int_X f \,\mathrm{d}\mu$$
μ-a.e. Hence,
$$\int_X f \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\nu.$$
By taking $f = \chi_A, A \in \mathcal{B}$, one gets $\mu(A) = \nu(A)$, so $\mu = \nu$. ☐

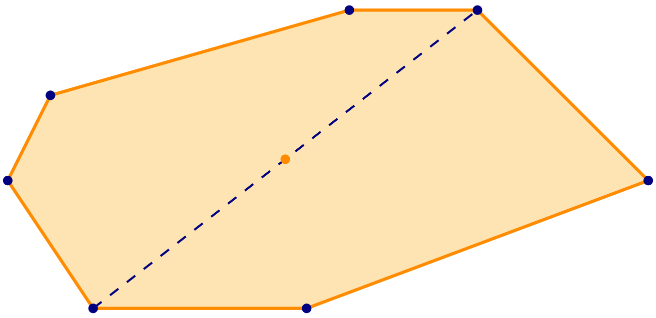


Figure 9: Extremal and non-extremal points of a convex polygon

We say that $\mu \in M(X,T)$ is an *extremal point* of $M(X,T)$ if there is no $\nu, \eta \in M(X,T)$ and $p \in (0,1)$ such that $\mu = p\nu + (1-p)\eta$, $\nu \neq \mu, \eta \neq \mu$. In other words, μ is extremal if it can't be written as the convex combination of two other points.

It is known that in a space of dimension $d \in \mathbb{N}$, any non-extremal point can be written as the convex combination of d extremal points. The same doesn't generally not hold in an infinite dimensional space.

Proposition 7.3. $\mu \in M(X,T)$ is extremal if and only if it is ergodic.

Proof.

"$\implies$": Let μ be extremal. Assume that it is not ergodic. Let $A \in \mathcal{B}$ be such that $0 < \mu(A) < 1$ and $T^{-1}(A) = A$. Let for any $B \in \mathcal{B}$

$$\mu_A(B) := \frac{\mu(A \cap B)}{\mu(A)}, \qquad \mu_{X \setminus A}(B) := \frac{\mu(B \cap (X \setminus A))}{\mu(X \setminus A)}.$$

By Exercise Sheet 3, $\mu_A, \mu_{X\setminus A}$ are in $M(X,T)$. Furthermore, $\mu = \mu(A) \cdot \mu_A + \mu(X \setminus A) \cdot \mu_{X\setminus A}$. As $0 < \mu(A) < 1$ and as $\mu_A \neq \mu, \mu \neq \mu_{X\setminus}$, μ is not an extremal point. ↯ So, μ is ergodic.

"$\impliedby$": Let μ be ergodic. Assume that μ is not extremal. Then there exists $\nu, \eta \in M(X,T), p \in (0,1)$ such that $\nu \neq \eta, \eta \neq \mu$, and $\mu = p\nu + (1-p)\eta$. If for $A \in \mathcal{B}$ one has $\mu(A) = 0$, then $0 = p\nu(A) + (1-p)\eta(A)$, so as $0 < 1-p, p < 1$ and $\nu(A), \eta(A)$ are nonnegative, $\nu(A) = \eta(A) = 0$. Thus, $\nu \ll \mu$ and $\eta \ll \mu$. Similarly, if $\mu(A) = 1$, then $1 = p\nu(A) + (1-p)\eta(A)$, so $\nu(A) = \eta(A) = 1$.

Fix T-invariant A. Then $\mu(A) \in \{0,1\}$. So by the derivations above,

$$\nu(A), \eta(A) \in \{0,1\},$$

thus ν and η are ergodic. By Lemma 7.2, $\nu = \eta = \mu$. ↯ So, μ is extremal. □

Proposition 7.4. $M(X,T)$ is convex compact with respect to the weak*-topology.

Proof. $M(X,T)$ is convex as by our remarks in Definition 4.4. Let us show that it is closed, i.e. for all $(\mu_n)_{n\geq 1} \in M(X,T)^{\mathbb{N}}$ such that $\mu_n \to \mu$ weakly* one has that $\mu \in M(X,T)$. To this end, note that, by the definition of weak*-convergence, for all $f \in C(X)$

$$\begin{aligned}\int_X f \,\mathrm{d}\mu &= \lim_{n\to\infty} \int_X f \,\mathrm{d}\mu_n \\ &= \lim_{n\to\infty} \int_X f \circ T \,\mathrm{d}\mu_n && \text{(Theorem 4.9)} \\ &\overset{f\circ T \in C(X)}{=} \int_X f \circ T \,\mathrm{d}\mu. && (*)\end{aligned}$$

By Theorem 4.9 and $(*)$, we conclude that $\mu \in M(X,T)$. Thus, $M(X,T)$ is closed. As $M(X,T)$ is closed and as $M(X)$ is compact by Lemma 4.5, $M(X,T) \subseteq M(X)$ is compact. ☐

Theorem 7.5. There exists at least one ergodic measure $\mu \in M(X,T)$.

Proof. By the *Krein-Milman Theorem*[21], any convex compact set in $M(X)$ is a closed convex hull of its extremal points, i.e. the smallest closed convex set containing extremal points. In particular, as $M(X,T)$ is non-empty (see Theorem 4.3), it has extremal points, which are ergodic measures by Proposition 7.4. ☐

Example 7.6. If $M(X,T)$ contains only one measure, then this measure is ergodic. E.g. $X = \mathbb{T}$, $Tx = x + \alpha$, $x \in \mathbb{T}$, where α is 2π-irrational, see Exercise Sheet 2.

Proposition 7.7. Let $\mu, \nu \in M(X,T)$ be ergodic and $\mu \neq \nu$. Then μ and ν are mutually singular.

Proof. By the definition of mutual singularity, we need to find some $A \in \mathcal{B}$ such that $\mu(A) = 1$ and $\nu(A) = 0$. By Lemma 7.2, as $\mu \neq \nu$ and as μ and ν are ergodic, they are not absolutely continuous to each other. Therefore, there exists $B \in \mathcal{B}$ such that $\mu(B) > 0$ ad $\nu(B) = 0$. Consider the set of points that return to B infinitely often, i.e.

$$A \coloneqq \bigcap_{n\geq 0} \bigcup_{i \geq n} T^{-i}(B).$$

We have previously shown that $T^{-1}(A) = A$. Furthermore,

$$\begin{aligned}\nu(A) &= \nu\left(\bigcap_{n\geq 0} \bigcup_{i\geq n} T^{-i}(B)\right) \\ &\leq \sum_{i\geq 0} \nu(T^{-i}(B)) \\ &= \sum_{i\geq 0} \nu(B)\end{aligned}$$

[21] See any reference on functional analysis, as we *won't* state it here.

$$= 0.$$

As A is T-invariant and μ is ergodic, $\mu(A) \in \{0, 1\}$. It remains to show that $\mu(A) > 0$. To this end, note that

$$\mu\left(\bigcup_{i\geq n} T^{-i}(B)\right) \geq \mu\left(T^{-n}(B)\right) = \mu(B),$$

i.e. the measure of those sets is bounded from below. So,

$$\mu\left(\bigcap_{n\geq 0}\bigcup_{i\geq n} T^{-1}(B)\right) = \lim_{n\to\infty} \mu\left(\bigcup_{i\geq n} T^{-1}(B)\right) \geq \mu(B) > 0$$

since $\left(\bigcup_{i\geq n} T^{-1}(B)\right)_{n\geq 0}$ is a decreasing sequence of sets. □

8 Mixing

Definition 8.1. Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ be μ-preserving. The T is called *(strongly) mixing* if for all $A, B \in \mathcal{B}$

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

T is called *weakly mixing*, if

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

In other words, strong mixing means that $T^{-n}(A)$ and B are almost independent for n large enough, while weak mixing means the same but in the time-average.

Remark (Why use $-n$ in the exponent instead of n?)**.** In our introductory chapter we have discussed that mixing means in a literal sense that applying T n times to a certain portion of space $A \in \mathcal{B}$ should give for large n a sort of mixing behaviour. But here, we use $-n$ instead of n. Formally, T^{-1} "behaves" better when working in measure theory. But there are other reasons: If T is an invertible measure-preserving transformation, then $\mu(T^{-n}(A) \cap B) = \mu(A \cap T^n(B))$ so one may think of applying the transformation a bunch of times to B instead A to achieve this mixing behaviour. So, our original idea of mixing agrees with our new notion for suitable T. For general T, one can also get a different natural interpretation using probabilities: $\mu(T^{-n}(A) \cap B)$ can be seen as the probability that a particle following the probability distribution μ is in B and after n steps in A. Strong mixing states that for large n the two sub-events are almost independent, i.e. the probability is asymptotically the same as the product of the two probabilities.

Remark. The conditions of the definition can be checked only for $A, B \in \mathcal{A}$, where $\mathcal{A}$ is an algebra generating $\mathcal{B}$, i.e. $\sigma(\mathcal{A}) = \mathcal{B}$, see Theorem 1.17 in Walters' *An Introduction to Ergodic Theory*.

Proposition 8.2. Strong mixing implies weak mixing, weak mixing implies ergodicity.

Proof. For the former implication, consider the real-valued sequence $(a_n)_{n\geq 0}$ defined by

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

Since T is by assumption strongly mixing, it means that $a_n \to \mu(A)\mu(B) =: a, n \to \infty$. As $|a_n - a|$ is non-negative, it suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n - a| \leq 0.$$

For that, fix $\varepsilon > 0$. Then there exists $N_0 \in \mathbb{N}$ such that $|a_n - a| \leq \varepsilon$ for all $n \geq N_0$. It follows that

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}|a_n - a| &= \frac{1}{N}\left(\sum_{n=0}^{N_0-1}|a_n - a| + \sum_{n=N_0}^{N-1}|a_n - a|\right)\\ &\leq \frac{1}{N}\left(N_0 \max_{0\leq n\leq N_0-1}|a_n - a| + (N - N_0)\cdot\varepsilon\right)\\ &\to \varepsilon, N \to \infty.\end{aligned}$$

Let us now show the latter implication. Fix $A \in \mathcal{B}$ such that $T^{-1}(A) = A$. By the definition of weak mixing,

$$0 = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(T^{-n}(A)\cap A) - \mu(A)^2\right| = \left|\mu(A) - \mu(A)^2\right|,$$

so $\mu(A) \in \{0, 1\}$, so T is ergodic. □

The question is now if for some of these notions we have equality, which turns out to be false for both cases.

Remark ("Ergodicity $\neq$ weak mixing")**.** Indeed, let $X = \mathbb{T} \simeq [0, 2\pi)$, $\alpha \in [0, 2\pi)$ be 2π-irrational, $Tx = x + \alpha$ for all $x \in X$. Then T is ergodic[22] (see Exercise Sheet 3). Let us show that T is not weakly mixing. W.l.o.g. $\alpha \in (0, \pi)$, otherwise we reset $\alpha' := 2\pi - \alpha$. Let $A = B = (0, \varepsilon)$ for some $\varepsilon < \alpha/2$. Then if for some $n \geq 0$ one has that $T^{-n}(A) \cap B \neq \emptyset$, then $T^{-(n+1)}(A) \cap B = \emptyset$. Thus,

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}|\mu(T^{-n}(A)\cap B) - \mu(A)\mu(B)| \geq \liminf_{N\to\infty}\frac{1}{N}\frac{N}{2}\mu(A)\mu(B) = \frac{\mu(A)\mu(B)}{2}.$$

So T is not weakly mixing.

Remark ("Strong mixing $\neq$ weak mixing")**.** For such an example, have a look at Walters' *An Introduction to Ergodic Theory* and the references therein.

Let $f\colon X \to \mathbb{R}$ be $\mathcal{B}$-measurable. Set $U_T f := f \circ T$. Note that $\|U_T f\|_{L^p} = \|f\|_{L^p}$.

Theorem 8.3. Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ be μ-preserving.

(a) The following are equivalent:

- (i) T is ergodic.
- (ii) For all $f, g \in L^2(X)$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_X U_T^n f\cdot g \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu\int_X g \,\mathrm{d}\mu.$$

[22] With respect to the normalized Lebesgue measure.

(*iii*) For all $f \in L^2(X)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X U_T^n f \cdot f \;\mathrm{d}\,\mu = \left(\int_X f \;\mathrm{d}\,\mu\right)^2.$$

(b) The following are equivalent:

(*i*) T is weakly mixing.

(*ii*) For all $f, g \in L^2(X)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot g \;\mathrm{d}\,\mu - \int_X f \;\mathrm{d}\,\mu \int_X g \;\mathrm{d}\,\mu \right| = 0.$$

(*iii*) For all $f \in L^2(X)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot f \;\mathrm{d}\,\mu - \left(\int_X f \;\mathrm{d}\,\mu\right)^2 \right| = 0.$$

(*iv*) For all $f \in L^2(X)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot f \;\mathrm{d}\,\mu - \left(\int_X f \;\mathrm{d}\,\mu\right)^2 \right|^2 = 0.$$

(c) The following are equivalent:

(*i*) T is strongly mixing.

(*ii*) For all $f, g \in L^2(X)$

$$\lim_{n\to\infty} \int_X U_T^n f \cdot g \;\mathrm{d}\,\mu = \int_X f \;\mathrm{d}\,\mu \int_X g \;\mathrm{d}\,\mu.$$

(*iii*) For all $f \in L^2(X)$

$$\lim_{n\to\infty} \int_X U_T^n f \cdot f \;\mathrm{d}\,\mu = \left(\int_X f \;\mathrm{d}\,\mu\right)^2.$$

In particular, T is strongly mixing if for any $f, g \in L^2(X)$ $U_T^n f$ and g are almost independent for n large enough.

Proof. We will only show (c), (a) and (b) can be shown similarly.

(*ii*) $\implies$ (*i*): Let $A, B \in \mathcal{B}, f = \chi_A, g = \chi_B$. Then

$$\begin{aligned}
\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) &= \lim_{n\to\infty} \int_X \chi_{T^{-n}(A)} \cdot \chi_B \;\mathrm{d}\,\mu \\
&= \lim_{n\to\infty} \int_X \chi_A \circ T^n \cdot \chi_B \;\mathrm{d}\,\mu \\
&= \lim_{n\to\infty} \int_X U_T^n f \cdot g \;\mathrm{d}\,\mu \\
&\overset{(ii)}{=} \int_X f \;\mathrm{d}\,\mu \int_X g \;\mathrm{d}\,\mu \\
&= \mu(A)\mu(B).
\end{aligned}$$

$(i) \implies (ii)$: By (i), one has that the equality in (ii) holds for any $A, B \in \mathcal{B}, f = \chi_A, g = \chi_B$. By the bilinearity of that equality, we can extend it to all simple functions. Fix any $f, g \in L^2(X), \varepsilon > 0$. Then there exists simple $\tilde{f}, \tilde{g} \in L^2(X)$ such that

$$\left\| f - \tilde{f} \right\|_{L^2} < \varepsilon, \qquad \|g - \tilde{g}\|_{L^2} < \varepsilon.$$

Then for all $n \geq 0$ we have with the triangle inequality that

$$\begin{aligned}
&\left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right| \\
\leq &\left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X U_T^n f \cdot \tilde{g} \, \mathrm{d}\mu \right| && (\star) \\
&+ \left| \int_X U_T^n f \cdot \tilde{g} \, \mathrm{d}\mu - \int_X U_T^n \tilde{f} \cdot \tilde{g} \, \mathrm{d}\mu \right| && (\star\star) \\
&+ \left| \int_X U_T^n \tilde{f} \cdot \tilde{g} \, \mathrm{d}\mu - \int_X \tilde{f} \, \mathrm{d}\mu \int_X \tilde{g} \, \mathrm{d}\mu \right| && (\star\star\star) \\
&+ \left| \int_X \tilde{f} \, \mathrm{d}\mu \int_X \tilde{g} \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X \tilde{g} \, \mathrm{d}\mu \right| && (\star\star\star\star) \\
&+ \left| \int_X f \, \mathrm{d}\mu \int_X \tilde{g} \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right|. && (\star\star\star\star\star)
\end{aligned}$$

- By the Cauchy-Schwarz inequality, $(\star)$ is bounded from above by

  $$\|U_T^n f\|_{L^2} \|g - \tilde{g}\|_{L^2} \leq \varepsilon \|f\|_{L^2}$$

  as $\|U_T^n f\|_{L^2} = \|f\|_{L^2}$.
- $(\star\star)$ is bounded from above by

  $$\left\| U_T^n (f - \tilde{f}) \right\|_{L^2} \|\tilde{g}\|_{L^2} \leq \varepsilon \left(\|g\|_{L^2} + \varepsilon \right).$$
- $(\star\star\star)$ goes to zero as n goes to infinity as $\tilde{f}, \tilde{g}$ are simple.
- $(\star\star\star\star)$ is bounded from above by

  $$\|\tilde{g}\|_{L^2} \left\| f - \tilde{f} \right\|_{L^2} \leq \varepsilon \left(\|g\|_{L^2} + \varepsilon \right).$$
- $(\star\star\star\star\star)$ is bounded from above by

  $$\|f\|_{L^2} \|g - \tilde{g}\|_{L^2} \leq \varepsilon \|f\|_{L^2}.$$

As $\varepsilon > 0$ was arbitrary,

$$\left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right| \to 0, n \to \infty.$$

$(ii) \implies (iii)$: Obvious.

$(iii) \implies (ii)$: Fix $f \in L^2(X)$. Let

$$F_f \coloneqq \left\{ g \in L^2(X) \colon \int_X U_T^n f \cdot g \, \mathrm{d}\,\mu \to \int_X f \, \mathrm{d}\,\mu \int_X g \, \mathrm{d}\,\mu, n \to \infty \right\}.$$

We want to show that $F_f = L^2(X)$. Note that F_f is linear, contains f by (iii) and contains all constants as

$$\int_X U_T^n f \, \mathrm{d}\,\mu = \int_X f \, \mathrm{d}\,\mu.$$

F_f is closed, which can be shown analogously as in the "$(i) \implies (ii)$" step of the proof. Further, F_f is U_T-invariant, i.e. if $g \in F_f$ then $U_T g \in F_f$ because

$$\begin{aligned}
\lim_{n\to\infty} \int_X U_T^n f \cdot U_T g \, \mathrm{d}\,\mu &= \lim_{n\to\infty} \int_X U_T^{n-1} f \cdot g \, \mathrm{d}\,\mu \\
&= \int_X f \, \mathrm{d}\,\mu \int_X g \, \mathrm{d}\,\mu \\
&= \int_X f \, \mathrm{d}\,\mu \int_X U_T g \, \mathrm{d}\,\mu,
\end{aligned}$$

where the first and last equality follows from T being measure-preserving. Now, let $g \perp F_f$. Then, as $U_T^n f \in F_f$, $\int_X U_T^n f \cdot g \, \mathrm{d}\,\mu = 0$. Further, as all constants are in F_f, $\int_X g \, \mathrm{d}\,\mu = 0$. Thus,

$$\lim_{n\to\infty} \int_X U_T^n f \cdot g \, \mathrm{d}\,\mu = 0 = \int_X f \, \mathrm{d}\,\mu \int_X g \, \mathrm{d}\,\mu,$$

so $g \in F_f$. Hence, $g = 0$ and $F_f^\perp = \{0\}$. Thus, $F_f = L^2(X)$. $\square$

9 One- and Two-Point Motion

Let $(X, \mathcal{B}, \mu)$ be a probability space, $T\colon X \to X$ be μ-preserving. A *two-point* motion $T \times T\colon X \times X \to X \times X$ is defined by $(T \times T)(x, y) = (Tx, Ty)$ for $x, y \in X$. Let $\mathcal{B} \otimes \mathcal{B}$ be the σ-algebra generated by $\{A \times B \subseteq X \times X\colon A, B \in \mathcal{B}\}$. Let the measure $\mu \otimes \mu$ on $(X \times X, \mathcal{B} \otimes \mathcal{B})$ be the product measure defined by $(\mu \otimes \mu)(A \times B) = \mu(A)\mu(B)$ for $A, B \in \mathcal{B}$: $\mu \otimes \mu$ can be uniquely extended to the algebra $\mathcal{A}$ generated by $A \times B, A, B \in \mathcal{B}$ adn then extended to $\mathcal{B} \otimes \mathcal{B}$ by Theorem 2.10.

Exercise 9.1. Show that $T \times T$ is $(\mathcal{B} \otimes \mathcal{B})$-measurable and $(\mu \otimes \mu)$-preserving.

There are two strong connections between the properties of T and $T \times T$. We start with the following:

Theorem 9.2. *T is strongly mixing if and only if $T \times T$ is strongly mixing.*

Proof. See Problem 4 on Exercise Sheet 5. □

Our main goal is to get the same result, but for weak mixing. To this end, let $L^2(X)$ contain all $\mathbb{C}$-valued L^2 functions and let $L^2_0 := \{f \in L^2(X)\colon \int_X f \, \mathrm{d}\,\mu = 0\}$.

Definition 9.3. Let $U_T\colon L^2(X) \to L^2(X)$ be defined by $U_T f := f \circ T$. Then T is called to have *continuous spectrum* if the only eigenvalue of U_T is 1 and the only eigenfunctions are constants, i.e. the equation

$$U_T f = \lambda f \qquad (\lambda \in \mathbb{C}, f \in L^2(X))$$

has only the solution $\lambda = 1$ and $f = c$ for some $c \in \mathbb{C} \setminus \{0\}$. Equivalently, T has continuous spectrum if U_T defined on L^2_0 has no eigenvalues.

Remark. Let λ be an eigenvalue of U_T, i.e. $U_T f = \lambda f$ for some $f \in L^2(X), f \neq 0$. Then

$$\|f\|_2^2 = \langle f, f \rangle \overset{(*)}{=} \langle U_T f, U_T f \rangle = \langle \lambda f, \lambda f \rangle = |\lambda|^2 \langle f, f \rangle \implies |\lambda| = 1,$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $L^2(X)$ defined by

$$\langle f, g \rangle := \int_X f \bar{g} \, \mathrm{d}\,\mu,$$

and where $(*)$ follows from the fact that T is μ-preserving.

Theorem 9.4 (Mixing theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space, $T\colon X \to X$ be an invertible μ-preserving transformation. Then the following are equivalent:*

- (i) T is weakly mixing.
- (ii) $T \times T$ is weakly mixing.
- (iii) $T \times T$ is ergodic.

(*iv*) T has continuous spectrum.

Further, if T is not invertible, then $(i) \iff (ii) \iff (iii)$.

Lemma 9.5 (Cesàro product)**.** Let $(a_n)_{n\geq 0}, (b_n)_{n\geq 0}$ be real-valued (or complex-valued) bounded sequences and let $a, b \in \mathbb{R}$ (or $a, b \in \mathbb{C}$) such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n - a| = 0, \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |b_n - b| = 0,$$

then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n b_n - ab| = 0.$$

Proof. It is sufficient to note that

$$|a_n b_n - ab| \leq |a_n| \, |b_n - b| + |b| \, |a_n - a| \leq \tilde{a} \, |b_n - b| + \tilde{b} \, |a_n - a| \, ,$$

where $\tilde{a} = \sup_{n\geq 0} |a_n| \vee |a| \, , \tilde{b} = \sup_{n\geq 0} |b_n| \vee |b|$. ☐

Lemma 9.6. $T \times T$ is weakly mixing $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu)$ if and only if for all $A, B, C, D \in \mathcal{B}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| (\mu \otimes \mu) \left((T \times T)^{-n} (A \times B) \cap (C \times D) \right) - (\mu \otimes \mu)(A \times B)(\mu \otimes \mu)(C \times D) \right| = 0. \qquad (\star)$$

In other words, it suffices to look at product sets to show that $T \times T$ is weakly mixing.

Proof. Let $\mathcal{A} \subseteq \mathcal{B} \otimes \mathcal{B}$ be the algebra generated by $\mathcal{A}_0 \coloneqq \{A \times B \colon A, B \in \mathcal{B}\}$. Note that any $E \in \mathcal{A}$ can be represented as $E = E_1 \cup \cdots \cup E_k$ for some disjoint $E_1, \ldots, E_k \in \mathcal{A}_0$.[23] So, for all $E, F \in \mathcal{A}$ there exist disjoint $E_1, \ldots, E_k \in \mathcal{A}_0$ and disjoint $F_1, \ldots, F_m \in \mathcal{A}_0$ such that $E = E_1 \cup \cdots \cup E_k, F = F_1 \cup \cdots \cup F_m$. Therefore, we get

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| (\mu \otimes \mu) \left((T \times T)^{-n} (E) \cap F \right) - (\mu \otimes \mu)(E)(\mu \otimes \mu)(F) \right| \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \sum_{i=1}^{k} \sum_{j=1}^{m} \left((\mu \otimes \mu) \left((T \times T)^{-n} (E_i) \cap F_j \right) - (\mu \otimes \mu)(E_i)(\mu \otimes \mu)(F_j) \right) \right| \\
&\leq \sum_{i=1}^{k} \sum_{j=1}^{m} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| (\mu \otimes \mu) \left((T \times T)^{-n} (E_i) \cap F_j \right) - (\mu \otimes \mu)(E_i)(\mu \otimes \mu)(F_j) \right| \\
&= 0
\end{aligned}$$

by $(\star)$ and the triangle inequality. As $\mathcal{B} \otimes \mathcal{B} = \sigma(\mathcal{A})$, the desired follows from a remark in the last chapter. ☐

[23] This fact can be proved inductively and is left as an exercise for the reader.

Proposition 9.7. Let $\mathcal{H}$ be a ($\mathbb{C}$-valued) Hilbert space, $U\colon \mathcal{H} \to \mathcal{H}$ be a unitary operator, i.e. $UU^* = U^*U = \mathrm{id}_{\mathcal{H}}$. Then for any $f \in \mathcal{H}$ there exists a probability measure E_f on $\mathbb{S}^1 \coloneqq \{z \in \mathbb{C}\colon |z| = 1\}$ such that

$$\forall\, n \in \mathbb{Z}\colon \langle U^n f, f\rangle = \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\, E_f(\lambda). \tag{$\star\star$}$$

In particular, if $(X, \mathcal{B}, \mu)$ is a probability space and T an invertible μ-preserving transformation, then U_T is unitary on $L^2_0(X)$, has continuous spectrum, so for all $f \in L^2_0(X)$, $E_\lambda(f)$ from ($\star\star$) has no atoms,[24] i.e. $E_f(\{\lambda\}) = 0$ for all $\lambda \in \mathbb{S}^1$, and ($\star\star$) becomes

$$\langle U_T^n f, f\rangle_{L^2} = \int_X U_T^n f \cdot \bar{f} \,\mathrm{d}\,\mu = \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\, E_f(\lambda).$$

Proof. See Theorem 1.25 from Walters' *An Introduction to Ergodic Theory* or Halmos' *Lectures on Ergodic Theory.* ☐

Proof idea. Let E be the spectral measure of U_T on $L^2_0(x)$. This means that for all measurable $K \subseteq \mathbb{S}^1$ $E(K)$ is a projection, and that $\int_{\mathbb{S}^1} \lambda \,\mathrm{d}\, E(\lambda) = U_T$ in a weak sense, i.e. for all $n \geq 0, f, g \in L^2_0(X)$, we have

$$\langle U_T^n f, g\rangle = \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\,\langle E(\lambda) f, g\rangle.$$

As U_T has continuous spectrum, $E(\{\lambda\}) = 0$ for all $\lambda \in \mathbb{S}^1$, so $E_f \coloneqq \langle Ef, f\rangle$ does not have atoms as well. ☐

Proof of Theorem 9.4. First, let T be an invertible μ-preserving transformation:

(i) $\implies$ *(ii)*: Let T be weakly mixing. Let us show that $T \times T$ is weakly mixing. By Lemma 9.6, it is sufficient to show ($\star$) for all $A, B, C, D \in \mathcal{B}$, which follows from the fact that

$$\begin{aligned}
&\frac{1}{N}\sum_{n=0}^{N-1} \left|(\mu\otimes\mu)\left((T\times T)^{-n}(A\times B)\cap(C\times D)\right) - (\mu\otimes\mu)(A\times B)(\mu\otimes\mu)(C\times D)\right| \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \left|\mu(T^{-n}(A)\cap C)\mu(T^{-n}(B)\cap D) - (\mu(A)\mu(C))\,(\mu(B)\mu(D))\right| \\
&\to 0, N \to \infty
\end{aligned}$$

by Lemma 9.5 and the fact that T is weakly mixing.

(ii) $\implies$ *(iii)*: See Proposition 8.2.

(iii) $\implies$ *(iv)*: Let $T \times T$ be ergodic. Assume that T is without continuous spectrum, i.e. there exists $\lambda \in \mathbb{C}$ and nonconstant $f \in L^2(X)$ such that $U_T f = f \circ T = \lambda f$. Set

$$\tilde{f}(x, y) = f(x)\bar{f}(y)$$

[24] See StackExchange.

for $x, y \in X$. Then $\tilde{f} \in L^2(X \times X)$ and for all $x, y \in X$

$$\begin{aligned} U_{T\times T}\tilde{f}(x,y) &= (\tilde{f} \circ (T \times T))(x,y) \\ &= \tilde{f}(Tx, Ty) \\ &= f(Tx)\bar{f}(Ty) \\ &= \lambda\bar{\lambda} f(x)\bar{f}(y) \\ &= \tilde{f}(x,y), \end{aligned}$$

where $|\lambda| = 1$ by the remark above.

As $\tilde{f}$ is not constant but $(T \times T)$-invariant, $T \times T$ is not ergodic by Proposition 6.2. ↯ Thus, T has continuous spectrum.

$(iv) \implies (i)$: Assume that T has continuous spectrum. Let us show that T is weakly mixing. By Theorem 8.3, it is sufficient to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot \bar{f} \,\mathrm{d}\,\mu - \left| \int_X f \,\mathrm{d}\,\mu \right|^2 \right|^2 = 0$$

for all $f \in L^2(X)$. Clearly, equality holds if f is constant a.e. Hence, w.l.o.g., by setting $\tilde{f} := f - \int_X f \,\mathrm{d}\,\mu$, we may assume that $f \in L_0^2(X)$, so we need to show that

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot \bar{f} \,\mathrm{d}\,\mu \right|^2 = 0 \\ \iff &\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\,E_f(\lambda) \right|^2 = 0, \end{aligned}$$

where the equivalence follows from Proposition 9.7. By Fubini's theorem,

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\,E_f(\lambda) \right|^2 &= \frac{1}{N} \sum_{n=0}^{N-1} \int_{\mathbb{S}^1} \lambda^n \,\mathrm{d}\,E_f(\lambda) \int_{\mathbb{S}^1} \bar{\lambda}^n \,\mathrm{d}\,\bar{E}_f(\lambda) \\ &= \int_{\mathbb{S}^1\times\mathbb{S}^1} \frac{1}{N} \sum_{n=0}^{N-1} (\lambda\bar{\nu})^n \,\mathrm{d}\,(E_f \otimes \bar{E}_f)(\lambda, \nu). \end{aligned}$$

Recall that if $x \neq 1$, then

$$\sum_{n=0}^{N-1} x^n = \frac{x^N - 1}{x - 1},$$

so if $\lambda\bar{\nu} \neq 1$, then

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} (\lambda\bar{\nu})^n \right| = \frac{1}{N} \left| \frac{(\lambda\bar{\nu})^N - 1}{\lambda\bar{\nu} - 1} \right| \leq \frac{2}{N} \left| \frac{1}{\lambda\bar{\nu} - 1} \right| \to 0, N \to \infty.$$

If $\lambda\bar{\nu} = 1$, then

$$\frac{1}{N}\sum_{n=0}^{N-1}(\lambda\bar{\nu})^n = 1.$$

By the dominated convergence theorem, we thus get that

$$\begin{aligned}
&\int_{\mathbb{S}^1\times\mathbb{S}^1}\frac{1}{N}\sum_{n=0}^{N-1}(\lambda\bar{\nu})^n \;\mathrm{d}\,(E_f\otimes\bar{E}_f)(\lambda,\nu)\\
\overset{N\to\infty}{\to}\quad &\int_{\mathbb{S}^1\times\mathbb{S}^1}\chi_{\lambda\bar{\nu}=1}\;\mathrm{d}\,(E_f\otimes\bar{E}_f)(\lambda,\nu)\\
=\quad &\int_{\mathbb{S}^1\times\mathbb{S}^1}\chi_{\lambda=\nu}\;\mathrm{d}\,(E_f\otimes\bar{E}_f)(\lambda,\nu)\\
=\quad &(E_f\otimes\bar{E}_f)(D),
\end{aligned}$$

where D is the diagonal of $\mathbb{S}^1\times\mathbb{S}^1$. The latter equals zero due to Fubini's theorem as E_f has no atoms:

$$(E_f\otimes\bar{E}_f)(D) = \int_{\mathbb{S}^1}E_f(\{\lambda\})\;\mathrm{d}\,\bar{E}_f(\lambda) = 0.$$

Let us now show the second part of the theorem: While proving $(i) \implies (ii) \implies (iii)$, we did not use the invertibility of T, so we only need to show that $(iii) \implies (i)$. So, let $T\times T$ be ergodic. Fix $A, B\in\mathcal{B}$. Analogously to Theorem 8.3 (or Exercise Sheet 5) it is enough to show that

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

in order to prove (i). First note that

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu(T^{-n}(A)\cap B) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}(\mu\otimes\mu)\left((T\times T)^{-n}(A\times X)\cap(B\times X)\right)\\
&= (\mu\otimes\mu)(A\times X)(\mu\otimes\mu)(B\times X)\\
&= \mu(A)\mu(B),
\end{aligned}$$

where the second equality follows from $T\times T$ being ergodic. Similarly,

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

Thus,

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(T^{-n}(A)\cap B)-\mu(A)\mu(B)\right|^2\\
&=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\Big(\left(\mu\left(T^{-n}(A)\cap B\right)\right)^2-2\mu(A)\mu(B)\mu\left(T^{-n}(A)\cap B\right)+\left(\mu(A)\right)^2\left(\mu(B)\right)^2\Big)\\
&=0.
\end{aligned}$$

This completes the proof. $\square$

Example 9.8 ("Mixing should change distances."). Let (X,d) be a metric space $\mathcal{B}=\mathcal{B}(X)$ be the Borel σ-algebra, μ be a probability measure on $(X,\mathcal{B})$. Let $T\colon X\to X$ be μ-preserving such that T preserves small distances, i.e. there exists $\delta>0$ such that $d(x,y)<\delta \implies d(Tx,Ty)<\delta$ for all $x,y\in X$. Assume that δ is neither too small or too big, i.e. for all $x\in X$ the ball

$$B_\delta(x) \coloneqq \{y\in X\colon d(x,y)<\delta\}$$

does not have measure 0 or 1. Then T is not weakly mixing. Indeed, let

$$A\coloneqq\{(x,y)\in X\times X\colon d(x,y)<\delta\}\subseteq X\times X.$$

Then

$$\begin{aligned}
(T\times T)^{-1}(A)&=\{(x,y)\in X\times X\colon d(Tx,Ty)<\delta\}\\
&\supseteq\{(x,y)\in X\times X\colon d(x,y)<\delta\}\\
&=A.
\end{aligned}$$

As T is μ-preserving, $T\times T$ is $(\mu\otimes\mu)$-preserving, so

$$\mu\Big((T\times T)^{-1}(A)\Big)=(\mu\otimes\mu)(A),$$

so we get

$$\begin{aligned}
(\mu\otimes\mu)\Big((T\times T)^{-1}(A)\,\Delta\,A\Big)&=(\mu\otimes\mu)\Big((T\times T)^{-1}(A)\setminus A\Big)\\
&=(\mu\otimes\mu)\Big((T\times T)^{-1}(A)\Big)-(\mu\otimes\mu)(A)\\
&=0.
\end{aligned}$$

On the other hand, by Fubini's theorem,

$$\begin{aligned}
(\mu\otimes\mu)(A)&=\int_X\int_X\chi_{d(x,y)<\delta}\ \mathrm{d}\,\mu(y)\ \mathrm{d}\,\mu(x)\\
&=\int_X\mu\left(B_\delta(x)\right)\ \mathrm{d}\,\mu(x).
\end{aligned}$$

As $\mu(B_\delta(x)) \in (0,1)$ for all $x \in X$, the latter integral is within $(0,1)$ as well, so

$$0 < (\mu \otimes \mu)(A) < 1.$$

Hence, by Proposition 6.2, $T \times T$ is not ergodic, so by Theorem 9.4, T is not weakly mixing.

Remark. Similarly, if there is some notion of angles in your space, one may show that if the angle is sufficiently preserved by the transformation, then one can show that the transformation can't be weakly mixing by a similar ad-hoc argument.

A Solutions to Exercise Sheets

Exercise Sheet 1

1. Show that there exist measurable spaces $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$, a measurable transform $T\colon X_1 \to X_2$ and a set $A \in \mathcal{B}_1$ such that $T(A) = \{Tx\colon x \in A\} \subseteq X_2$ is not in $\mathcal{B}_2$.

Solution. Take $X_1 = X_2 = \{0,1\} =: X, T = \mathrm{id}_X, \mathcal{B}_1 = \sigma(\{0\}, \{1\}) = 2^X$ and $\mathcal{B}_2 = \{\emptyset, X_2\}$. Clearly, T is measurable as $\{T^{-1}(\emptyset), T^{-1}(X)\} = \{\emptyset, X\} \subseteq \mathcal{B}_1$. However, $T(A) = \{0\} \notin \mathcal{B}_2$ for $A = \{0\} \in \mathcal{B}_1$. □

2. (Exercise 3.5) Let $X = \mathbb{T} \simeq [0, 2\pi)$ be the torus, $\mathcal{B} = \mathcal{B}(\mathbb{T})$, λ be the normalised Lebesgue measure on $\mathbb{T}$, i.e. $\forall\, 0 \le a \le b < 2\pi$ we set $\lambda([a,b]) = (b-a)/(2\pi)$. Let $n \ge 1$ be an integer and let $T\colon X \to X$ be defined by

$$Tx = (nx) \mod 2\pi := 2\pi \left\{ \frac{nx}{2\pi} \right\}, \quad x \in [0, 2\pi).$$

(a) T is measurable and measure-preserving as a map from $(X, \mathcal{B}, \lambda)$ to itself.

Proof. Let $0 \le a < b \le 2\pi$. We see that

$$\begin{aligned} T^{-1}([a,b)) &= \{x \in X \mid \exists\, k \in \{0, \dots, n-1\} : a \le nx - 2\pi k < b\} \\ &= \dot{\bigcup_{k=0}^{n-1}} \left[\frac{a + 2\pi k}{n}, \frac{b + 2\pi k}{n} \right). \end{aligned}$$

So the preimage of all half-open intervals that are subsets of X are all in $\mathcal{B}$. Furthermore, this shows that

$$\lambda\left(T^{-1}([a,b))\right) = \sum_{k=0}^{n-1} \lambda\left(\left[\frac{a+2\pi k}{n}, \frac{b+2\pi k}{n}\right)\right) = \frac{b-a}{2\pi} = \lambda([a,b)).$$

As the half-open intervals generate $\mathcal{B}$ and λ is σ-finite, T is measurable and measure-preserving by Theorem 2.10. □

(b) If $n \ne 1$, then T is not bijective. Furthermore, there exists a segment $A \subseteq [0, 2\pi)$ such that $\lambda(A) \ne \lambda(T(A))$.

Proof. For the first part of the claim, consider 0 and $2\pi/n$. Obviously, those are two distinct numbers with $T(0) = 0 = T(2\pi/n)$. So T is not injective and in particular not bijective.
For the second part, consider $A = [0, 2\pi/n)$. Then we have

$$\lambda(A) = \frac{1}{n} \ne 1 = \lambda(T(A))$$

since $n \ne 1$. □

3. Let $a, b, c, d \in \mathbb{Z}$ be such that $ad - bc = 1$.

(a) Let $h \in \mathbb{R}^2$. Let $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$Sx := \begin{pmatrix} a & b \\ c & d \end{pmatrix} x - h, x \in \mathbb{R}^2.$$

Then S is a bijection and its inverse given by

$$S^{-1}(y) := \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} (y + h).$$

Furthermore, S preserves the Lebesgue measure.

Proof. We first show that S is a bijection with the given inverse: If $y = Sx$, then, due to $ad - bc = 1$, it follows that

$$S^{-1}(y) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} (Sx + h) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} x = x.$$

Similarly, one can show that if $x = S^{-1}(y)$, then $Sx = y$. Thus, S is a bijection with the given inverse.
To show that S preserves the Lebesgue measure, note that S is a diffeomorphism. Furthermore, we have for all $y \in \mathbb{R}^2$ that

$$DS^{-1}(x) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \implies \left|\det\left(DS^{-1}(x)\right)\right| = ad - bc = 1.$$

So, by a change of variables we get for all $A \in \mathcal{B}(\mathbb{R}^2)$

$$\lambda\left(S^{-1}(A)\right) = \int_A \left|\det\left(DS^{-1}(x)\right)\right| \,\mathrm{d}\, x = \lambda(A).$$

Thus, S is measure-preserving (with respect to the Lebesgue measure). □

(b) Let $X := [0, 1)^2$ and let $\mathcal{B}$ and $T\colon X \to X$ be defined as in Example 3.3.2. Then T preserves the Lebesgue measure.

Proof. As we have already shown, T is a bijection with inverse

$$T^{-1}(y) \left(\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} y\right)_{\text{mod } 1} = \begin{pmatrix} \{dy^1 - by^2\} \\ \{-cy^1 + ay^2\} \end{pmatrix}, y = \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \in X.$$

We first show that T is measurable in the first place, see Exercise 3.4. For that, it suffices to show that the mapping

$$F\colon \mathbb{R}^2 \to X, \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \mapsto \begin{pmatrix} \{x^1\} \\ \{x^2\} \end{pmatrix}$$

is measurable.[25] For that, fix $A \in \mathcal{B}$. We then have

$$F^{-1}(A) = \bigcup_{v \in \mathbb{Z}^2} (A + v).$$

As translations are measurable and $\mathbb{Z}^2$ countable, $F^{-1}(A)$ is the countable union of measurable sets, so $F^{-1}(A) \in \mathcal{B}$. Thus, T is measurable.
To show that T is measure-preserving, we use the following lemma:

Lemma A.1. Let $X := [0,1)^2, T\colon X \to X$ be a measurable bijection. Assume additionally that there exists an open subset $B \subseteq X$ of full Lebesgue measure λ such that T is continuous on B. Then T preserves the Lebesgue measure if and only if for any $x \in B$ there exists $\delta > 0$ such that for any measurable $A \subseteq X$ within the ball of radius δ with the centre in Tx one has that $\lambda(A) = \lambda(T^{-1}(A))$.

To use this lemma, note that T is continuous for $x = (x^1, x^2)^\top \in X$ if none of $ax^1 + bx^2$ and $cx^1 + dx^2$ is an integer. Indeed, in a small neighborhood around x, where $a\tilde{x}^1 + b\tilde{x}^2$ and $c\tilde{x}^1 + d\tilde{x}^2$ are also not integers for all $\tilde{x}$ in that neighborhood, we have that T then acts just like a linear transformation, which in particular is continuous.
Next, let $B_1 \subseteq X$ be the subset of all such $x = (x^1, x^2)^\top \in X$ where none of $ax^1 + bx^2$ and $cx^1 + dx^2$ is an integer. Clearly, B_1 is open, as one can always (as we implicitly used before) find a neighborhood around $x \in B_1$ that is also completely contained in B_1. By our previous observation, B_1 is thus an open set on which T is continuous on.
To show that B_1 is also of full measure, note that

$$T(X \setminus B_1) = \left\{x \in X\colon x^1 = 0\right\} \cup \left\{x \in X\colon x^2 = 0\right\},$$

i.e. it's the union of two finite line segments with one endpoint at the origin. Let us now think of each step when applying T^{-1} on both sides. Applying a linear transformation would preserve them being lines with one endpoint at the origin and applying F would turn them into a union of finite affine line segments. Thus, $\lambda(X \setminus B_1) = 0$, so B_1 is of full measure.
Furthermore, $B_2 = (0,1)^2$ is also open and of full measure. Thus, $B = B_1 \cap B_2$ is an open subset of full measure on which T is continuous.
Now we can apply the lemma: Let $x \in B$. Choose $\delta > 0$ in such a way that the ball of radius δ with centre Tx is contained in $T(B)$. Let $A \in \mathcal{B}$ be within the ball of radius δ around Tx. As the Borel σ-algebra is generated by open sets, we may assume w.l.o.g. that A is open. Then $T^{-1}(x) \in (0,1)^2$ for all $x \in A$ by definition of B_2. This implies that none of $dx^1 - bx^2$ and $-cx^1 + ax^2$ is an integer, meaning that T^{-1} is continuous on A. More

[25] If one were pedantic, we should really consider $F\big|_X$, but as $\mathcal{B}$ is the "relative σ-algebra" on X induced by $\mathcal{B}(\mathbb{R}^2)$, it doesn't matter.

specifically, T^{-1} is – restricted on A – a diffeomorphism with

$$DT^{-1}(x) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \implies \left|\det\left(DT^{-1}(x)\right)\right| = ad - bc = 1.$$

In particular, we see again by a change of variables that

$$\lambda\left(T^{-1}(A)\right) = \int_A \left|\det\left(DT^{-1}(x)\right)\right| \,\mathrm{d}\,x = \lambda(A).$$

Thus, by Lemma A.1, T is measure-preserving. $\square$

Remark. Alternatively, one may argue that T^{-1} restricted on A looks like a function as in (a) for a suitable h. Hence, it immediately follows that T^{-1} is measure-preserving when restricted onto A.

Exercise Sheet 2

The goal in this sheet is to show that the only probability measure for which the 2π-irrational rotations on $\mathbb{T}$ are measure-preserving, is the normalized Lebesgue measure.

1. Let $\mathbb{T} \simeq [0, 2\pi)$ be the torus and let μ be a probability measure on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Assume that there exists $f\colon [0, 2\pi) \to [0, 1]$ such that for any $0 \le a \le b < 2\pi$ one has $\mu([a, b]) = f(b - a)$.

(a) f is non-decreasing, right-continuous, and $f(0) = 0$. Furthermore, for any $a, b \in [0, 2\pi), a + b < 2\pi$, we have

$$f(a + b) = f(a) + f(b).$$

Proof. f is non-decreasing since for any $0 \le a \le b < 2\pi$ we have

$$f(a) = \mu([0, a]) \le \mu([0, a]) + \mu((a, b]) = \mu([0, b]) = f(b).$$

Next, let $(x_n)_{n\in\mathbb{N}} \in \mathbb{T}^{\mathbb{N}}$ be a decreasing sequence such that $x_n \downarrow x \in \mathbb{T}$ as $n \uparrow \infty$. Define $B_n = [0, x_n] \in \mathcal{B}(\mathbb{T})$ for all $n \in \mathbb{N}$. Clearly, $(B_n)_{n\in\mathbb{N}}$ is then decreasing with

$$B := \bigcap_{n\in\mathbb{N}} B_n = [0, x].$$

Indeed, clearly $[0, x] \subseteq B$ as $x \le x_n$ for all $x \in \mathbb{N}$. On the other hand, if $\mathbb{T} \ni y > x$ there must be $n_0 \in \mathbb{N}$ such that $y > x_n$ for all $n \ge n_0$, hence $y \notin B$. So, by Exercise 2.9,

$$\lim_{n\uparrow\infty} f(x_n) = \lim_{n\uparrow\infty} \mu(B_n) = \mu(B) = f(x).$$

Thus, f is right-continuous. Next, for the sake of contradiction, assume that $f(0) > 0$. That means that

$$\mu(\{x\}) = f(0) > 0$$

for all $x \in \mathbb{T}$. However, that would imply that the set $M = \{\pi/n\colon n \in \mathbb{N}\}$ has infinite measure, since, as μ is σ-additive, we have

$$\mu(M) = \sum_{n=1}^{\infty} \mu\left(\left\{\frac{\pi}{n}\right\}\right) = \sum_{n=1}^{\infty} f(0) = \infty.$$

This contradicts the fact that μ is a finite measure. Hence $f(0) = 0$. Lastly, for $a, b \in [0, 2\pi), a + b < 2\pi$, we get

$$\begin{aligned} f(a + b) &= \mu([0, a + b]) \\ &= \mu([0, a) \,\dot\cup\, [a, a + b]) \\ &= \mu([0, a)) + \mu([a, a + b]) \end{aligned}$$

$$= (\mu([0,a]) - \mu(\{a\})) + f(b)$$
$$= f(a) + f(b)$$

since μ is additive and $\mu(\{a\}) = 0$. $\square$

(b) μ is the normalized Lebesgue measure.

Proof. By Theorem 2.10, it suffices to show that for any $0 \le a \le b \le 2\pi$

$$\mu([a,b)) = \frac{b-a}{2\pi}.$$

Consider integer sequences $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ such that $a_k/2^k \uparrow a/(2\pi)$ and $b_k/2^k \downarrow b/(2\pi)$ as $k \uparrow \infty$. These sequences can be constructed by considering the binary expansions of $a/2\pi$ and $1 - b/(2\pi)$. In particular,

$$\bigcap_{k\in\mathbb{N}} \left[\frac{2\pi a_k}{2^k}, \frac{2\pi b_k}{2^k}\right) = [a,b),$$

where $[a_k/2^k, b_k/2^k)$ is descreasing in k. Now, note that for all $n \in \mathbb{N}$

$$\begin{aligned}
1 &= \mu(\mathbb{T}) \\
&= \mu\left(\dot{\bigcup_{m=0}^{n-1}} \left[\frac{2\pi m}{n}, \frac{2\pi(m+1)}{n}\right)\right) \\
&= \sum_{m=0}^{n-1} \mu\left(\left[\frac{2\pi m}{n}, \frac{2\pi(m+1)}{n}\right)\right) \\
&= \sum_{m=0}^{n-1} f\left(\frac{2\pi}{n}\right) \\
&= nf\left(\frac{2\pi}{n}\right).
\end{aligned}$$

Thus, we get by Exercise 2.9

$$\begin{aligned}
\mu([a,b)) &= \lim_{k\to\infty} \mu\left(\left[\frac{2\pi a_k}{2^k}, \frac{2\pi b_k}{2^k}\right)\right) \\
&= \lim_{k\to\infty} \sum_{m=a_k}^{b_k-1} \mu\left(\left[\frac{2\pi m}{n}, \frac{2\pi(m+1)}{n}\right)\right) \\
&= \lim_{k\to\infty} \sum_{m=a_k}^{b_k-1} f\left(\frac{2\pi}{2^k}\right) \\
&= \lim_{k\to\infty} \frac{b_k - a_k}{2^k} \\
&= \frac{b-a}{2\pi}.
\end{aligned}$$

$\square$

2. Let $\mathbb{T} \simeq [0, 2\pi)$ be the torus and let $\alpha \in [0, 2\pi)$ be 2π-irrational. Set $T\colon \mathbb{T} \to \mathbb{T}$ to be $Tx\colon x + \alpha$ for $x \in \mathbb{T}$. Let $\mathcal{B} = \mathcal{B}(\mathbb{T})$ be the corresponding Borel σ-algebra and let μ be a T-invariant probability measure on $(\mathbb{T}, \mathcal{B})$, i.e. for any $B \in \mathcal{B}$ one has $\mu(T^{-1}(B)) = \mu(B)$.

(a) T is a bijection and $T^{-1}x = x - \alpha$ for any $x \in \mathbb{T}$.

Proof. Consider $x, y \in \mathbb{T}$ with $Tx = Ty$. That means that $x + \alpha \equiv y + \alpha \pmod{2\pi}$. So,

$$(y + \alpha) - (x + \alpha) = y - x$$

is an integer multiple of 2π. But since $0 \leq x, y < 2\pi$, this implies $x = y$. So, T is injective.
On the other hand, for $y \in \mathbb{T}$ there exists suitable $k \in \mathbb{N}$ such that $x := y - \alpha + 2\pi k \in [0, 2\pi)$. In particular, $Tx \equiv y \pmod{2\pi}$. So, T is surjective. Lastly, as we are in $\mathbb{T}$ always doing arithmetic in Modulo 2π, it follows from the above that $T^{-1}x = x - \alpha$. □

(b) For any $x \in \mathbb{T}$ one has that $T^n x \neq T^m x$ for any $n, m \in \mathbb{Z}$. Furthermore, $\mu(\{x\}) = 0$.

Proof. Suppose there exist distinct $m, n \in \mathbb{Z}$ and $x \in \mathbb{T}$ such that that were the case. Then,

$$x + \alpha m \equiv T^m x = T^n x \equiv x + \alpha n \pmod{2\pi},$$

so $\alpha(m - n) \in 2\pi\mathbb{Z}$. But since $m \neq n$, that means that $\alpha \in 2\pi\mathbb{Q}$. ↯
As T is bijective, the fact that μ is T-invariant implies that

$$\mu(T^z(B)) = \mu(B)$$

for all $z \in \mathbb{Z}$ and $B \in \mathcal{B}$. Consider the orbit $B = \{T^z x\colon z \in \mathbb{Z}\}$. As a collection of countably many points, B must be measurable. Note that due to the first part of the claim, B can be written as the disjoint, countable union of the singleton sets containing $T^z x$, $z \in \mathbb{Z}$. So, using the monotonicity of μ and the fact that $T^z x$, we get

$$1 = \mu(\mathbb{T}) \geq \mu(B) = \sum_{z \in \mathbb{Z}} \mu(\{T^z x\}) = \sum_{z \in \mathbb{Z}} \mu(\{x\}).$$

This chain of inequalities can only hold if $\mu(\{x\}) = 0$. □

(c) Fix $0 \leq a < b < 2\pi$. Let $I = [a, b]$. Set $c := b - a$ and $I_0 := [0, c]$. Then for any $\varepsilon > 0$ there exists $n \geq 0$ such that $T^{-n}(I) \,\Delta\, I_0 \subseteq (-\varepsilon, \varepsilon) \cup (c - \varepsilon, c + \varepsilon)$.[26]

[26] Technically, it was written with closed brackets, but it's correct as stated.

Proof. Let $\varepsilon > 0$ be arbitrary. First, as $-\alpha$ is also 2π-irrational, we have by (a) and Corollary 3.9.2 that there exists $n \in \mathbb{N}$ such that $T^{-n}a \in (-\varepsilon, \varepsilon)$. For that n, we explicitly get

$$\begin{aligned} T^{-n}(I) &= \{T^{-n}x \colon x \in I\} = \{T^{-n}(a+y) \colon y \in I_0\} \\ &= T^{-n}a + I_0 \qquad \subseteq (-\varepsilon, \varepsilon + c). \end{aligned}$$

Thus, we get that

$$\begin{aligned} T^{-n}(I) \,\Delta\, I_0 &= (T^{-n}(I) \setminus I_0) \cup (I_0 \setminus T^{-n}(I)) \\ &\subseteq ((-\varepsilon, 0) \cup (c, c+\varepsilon)) \cup ([0, \varepsilon) \cup (c-\varepsilon, c]) \\ &= (-\varepsilon, \varepsilon) \cup (c-\varepsilon, c+\varepsilon). \end{aligned}$$ □

(d) For any $\varepsilon > 0$ one has that

$$|\mu(I) - \mu(I_0)| \leq \mu(A_\varepsilon),$$

where $A_\varepsilon \coloneqq [-\varepsilon, \varepsilon] \cup [c-\varepsilon, c+\varepsilon]$. In particular, $\mu(I) = \mu(I_0)$.

Proof. Let $\varepsilon > 0$ be arbitrary. Choose $n \geq 0$ as in (c). Consider the function $f(x) = \mathbb{1}_{T^{-n}(I)}(x) - \mathbb{1}_{I_0}(x)$. Clearly, this function is measurable and, as μ is a probability measure, integrable. Further, it is clear that $f(x) \neq 0$ if and only if $x \in T^{-n}(x) \,\Delta\, I_0$. So, $|f| \leq \mathbb{1}_{A_\varepsilon}$ by (c). So, as μ is T-invariant, we get

$$\begin{aligned} |\mu(I) - \mu(I_0)| &= |\mu(T^{-n}(I) - \mu(I_0)| \\ &= \left| \int_{\mathbb{T}} f \;\mathrm{d}\,\mu \right| \\ &\leq \int_{\mathbb{T}} |f| \;\mathrm{d}\,\mu \\ &\leq \int_{\mathbb{T}} \mathbb{1}_{A_\varepsilon} \;\mathrm{d}\,\mu \\ &= \mu(A_\varepsilon). \end{aligned}$$

This shows the first part of the claim. Now, $(A_{1/k})_{k\geq 1}$ is a decreasing sequence with $\bigcap_{k\in\mathbb{N}} A_{1/k} = \{0, c\}$. So, by Exercise 2.9 and (b), we get that

$$\lim_{k\to\infty} \mu\left(A_{\frac{1}{k}}\right) = \mu(\{0\}) + \mu(\{c\}) = 0.$$

This shows that $\mu(I) = \mu(I_0)$. □

(e) μ is the normalized Lebesgue measure.

Proof. Let $f \colon [0, 2\pi) \to [0,1], x \mapsto \mu([0,x])$. By (d), this function satisfies the property that for any $0 \leq a \leq b < 2\pi$ one has $\mu([a,b]) = \mu([0, b-a]) = f(b-a)$. So, by Exercise 1 (b), it follows that μ must be the normalized Lebesgue measure. □

Exercise Sheet 3

1. Let $(X, \mathcal{B})$ be a measurable space, $T: X \to X$ be measurable.

(a) For any $A, B \in \mathcal{B}$ we have $T^{-1}(A) \cap T^{-1}(B) = T^{-1}(A \cap B)$.

Proof. Via elementary equivalences, we get

$$\begin{aligned} x \in T^{-1}(A) \cap T^{-1}(B) &\iff \big(x \in T^{-1}(A) \wedge x \in T^{-1}(B)\big) \\ &\iff (Tx \in A \wedge Tx \in B) \\ &\iff (Tx \in A \cap B) \\ &\iff x \in T^{-1}(A \cap B). \end{aligned}$$

So, we're done. □

(b) Let μ be a T-invariant probability measure on $(X, \mathcal{B})$. Assume that there exists $A \in \mathcal{B}$ such that $T^{-1}(A) = A$ and $\mu(A) > 0$. Then the measure ν defined by $\nu(B) := \mu(A \cap B)/\mu(A)$ is a T-invariant probability measure.

Proof. Clearly, ν is well-defined since $\mu(A) > 0$ and is a measure since it only differs from μ by a constant factor. It is a probability measure since $\nu(X) = \mu(A)/\mu(A) = 1$. To show that ν is T-invariant, let $B \in \mathcal{B}$ be arbitrary. We get

$$\begin{aligned} \nu(T^{-1}(B)) &= \frac{\mu(A \cap T^{-1}(B))}{\mu(A)} \\ &= \frac{\mu(T^{-1}(A) \cap T^{-1}(B))}{\mu(A)} \\ &\overset{\text{(a)}}{=} \frac{\mu(T^{-1}(A \cap B))}{\mu(A)} \\ &= \frac{\mu(A \cap B)}{\mu(A)} && (T\text{-invariance of } \mu) \\ &= \nu(B). \end{aligned}$$

This completes the proof. □

2. Let $(X, \mathcal{B}, \mu)$ be a probability space and $T: X \to X$ be measurable, μ-preserving, and *not ergodic*. Then there exists another probability measure ν on $(X, \mathcal{B})$ such that $\mu \neq \nu$ and T is ν-preserving.

Proof. Since T is not ergodic, there exists $A \in \mathcal{B}$ such that $T^{-1}(A) = A$ but $\mu(A) \in (0, 1)$. Define ν as in Exercise 1 (b). It suffices to show that $\mu \neq \nu$. This is clear however, as $\mu(X \setminus A) = 1 - \mu(A) \in (0, 1)$ but $\nu(X \setminus A) = 0$. □

3. $\mathbb{T} \simeq [0, 2\pi)$ be the torus and let $\alpha \in [0, 2\pi)$. Set $T\colon \mathbb{T} \to \mathbb{T}$ to be $Tx := x+\alpha, x \in \mathbb{T}$. Let $\mathcal{B} = \mathcal{B}(\mathbb{T})$ be the corresponding Borel σ-algebra and let μ be the normalized Lebesgue measure on $\mathbb{T}$.

(a) If α is 2π-irrational, then T is ergodic.

Proof. Assume otherwise. If T is not ergodic, then Exercise 2 would give us that there are at least two different probability measures on $(\mathbb{T}, \mathcal{B})$. But by Exercise 2 from Exercise Sheet 2, we know that the only T-invariant probability measure on $(\mathbb{T}, \mathcal{B})$ is μ, a contradiction. □

(b) If α is 2π-rational, then T is not ergodic.

Proof. W.l.o.g. $\alpha \neq 0$ as otherwise $T = \mathrm{id}_{\mathbb{T}}$. Let

$$\frac{\alpha}{2\pi} = \frac{m}{n}$$

for some coprime $m, n \in \mathbb{N}$. We then get

$$T^{-n}x = x - \alpha n = x - 2\pi m \equiv x \pmod{2\pi}.$$

Let $0 < \varepsilon < 2\pi/n$ be arbitrary and let $I = [0, \varepsilon]$ and

$$A := \dot{\bigcup_{k=0}^{n-1}} T^{-k}(I).$$

Note that the $T^{-k}(I)$ are pairwise disjoint by our choice of ε. Thus we get

$$\begin{aligned} T^{-1}(A) &= \dot{\bigcup_{k=0}^{n-1}} \left(T^{-1} \circ T^{-k}\right)(I) \\ &= \dot{\bigcup_{k=0}^{n-1}} T^{-(k+1)}(I) \\ &= \dot{\bigcup_{k=1}^{n}} T^{-k}(I) \\ &= \dot{\bigcup_{k=0}^{n-1}} T^{-k}(I) && (T^{-n}(I) = I) \\ &= A. \end{aligned}$$

Furthermore, we have $\mu(A) = n\mu(I) = n\varepsilon/(2\pi) \in (0, 1)$ by our choice of ε. So, T is by definition not ergodic. □

(c) If $\alpha \neq 0$ is 2π-rational, i.e. $\alpha/2\pi = m/n$ for some coprime $m, n \in \mathbb{N}$, then $A \in \mathcal{B}$ is T-invariant if and only if A is preserved under the rotation $2\pi/n$.

Proof. Let $0 \leq k < l < n$ be arbitrary. As m and n are coprime and in particular $0 \leq l-k < n$, we have that n doesn't divide $m(l-k)$. It follows that $(l-k)\alpha \not\equiv 0 \pmod{2\pi}$. Thus, we have for any $x \in \mathbb{T}$ that

$$T^{-k}x = x - k\alpha \not\equiv x - l\alpha = T^{-l}x \pmod{2\pi}.$$

We also have for any $x \in \mathbb{T}$ and $0 \leq k < n$ that $T^{-k}x = x - k\alpha = x - 2\pi l/n$ for some $l \in \mathbb{N}_0$ where, because we are calculating Modulo 2π, we may assume that $l \in \{0, \ldots, n-1\}$. Together, these two facts imply that the (periodic) orbit of x is

$$\left\{x, T^{-1}x, \ldots, T^{-(n-1)}x\right\} = \left\{x, x - \frac{2\pi}{n}, \ldots, x - \frac{2\pi(n-1)}{n}\right\}.$$

Now, let $A \in \mathcal{B}$ be T-invariant. By definition, this means that for any $x \in A$, $T^{-1}x$ is also in A. Inductively and due to $T^{-n}x = x$, we see that A is T-invariant, if and only if for any $x \in A$, we have

$$\left\{T^{-1}x, \ldots, T^{-(n-1)}x\right\} = \left\{x - \frac{2\pi}{n}, \ldots, x - \frac{2\pi(n-1)}{n}\right\} \in A.$$

So, $A \in \mathcal{B}$ is T-invariant if and only if A is preserved under a $2\pi/n$ rotation. $\square$

Exercise Sheet 4

1. Let $(Y, \mathcal{F}, \nu)$ be a probability space and let $X = Y^{\mathbb{Z}}$. For each $k \geq 1$ we endow Y^k with the σ-algebra $\mathcal{F}^{\otimes k}$ generated by all sets of the form $A_1 \times \ldots \times A_k \subseteq Y^k, A_1, \ldots, A_k \in \mathcal{F}$. A set $C \subseteq X$ is called a *cylinder* if there exist $k \geq 1$, $n_1, \ldots, n_k \in \mathbb{Z}$, and $A \in \mathcal{F}^{\otimes k}$ such that

$$C = \left\{ x = (y_n)_{n\in\mathbb{Z}} \in X \colon\ (y_{n_i})_{i=1}^k \in A \right\}. \tag{1}$$

Let $\mathcal{C}$ be the set of all cylinders in X.

(a) $\mathcal{C}$ is an algebra.

Proof. Clearly, $\emptyset \in \mathcal{C}$ since one can look at the cylinder with $k = 1, n_1 = 0$ and $A_0 = \emptyset$. Similarly, $X \in \mathcal{C}$ by choosing $k = 1, n_1 = 0$ and $A_0 = Y$. Next, we show that $\mathcal{C}$ is closed under taking complements. For that, let C be described by $k \in \mathbb{N}, n_1, \ldots, n_k \in \mathbb{Z}, A \in \mathcal{F}^{\otimes k}$. As $\mathcal{F}^{\otimes k}$ is a σ-algebra, we get that $Y^k \setminus A \in \mathcal{F}^{\otimes k}$ and as the complement is just given by

$$X \setminus C = \left\{ x = (y_n)_{n\in\mathbb{Z}} \in X \colon\ (y_{n_i})_{i=1}^k \in Y^k \setminus A \right\},$$

this shows $X \setminus C \in \mathcal{C}$. Lastly, consider $C_1, C_2 \in \mathcal{C}$ formed by $k_1, k_2 \in \mathbb{N}, n_1^1, \ldots, n_{k_1}^1, n_1^2, \ldots, n_{k_2}^2 \in \mathbb{Z}, A_1 \in \mathcal{F}^{\otimes k_1}, A_2 \in \mathcal{F}^{\otimes k_2}$ respectively. By appropriately extending A_1, A_2 by $\times Y$ in the corresponding components, we may assume that $k_i = 2m+1, n_1^i = -m, \ldots, n_{k_i}^i = m$ for all $i \in \{1, 2\}$, where m may be chosen as

$$m = \max \left\{ \left| n_j^i \right| : i \in \{1, 2\}, 1 \leq j \leq k_i \right\}.$$

But then $C_1 \cup C_2$ is a cylinder formed by $2m+1 \in \mathbb{N}, n_1 = -m, \ldots, n_{2m+1} = m$, and $A = A_1 \cup A_2 \in \mathcal{F}^{\otimes 2m+1}$. □

Now, let $\mathcal{B} = \sigma(\mathcal{C})$. For each cylinder $C \in \mathcal{C}$ of the form (1) with $A = A_1 \times \ldots \times A_k \subseteq Y^k$ we set

$$\mu(C) = \prod_{i=1}^{k} \nu(A_i).$$

Then μ can be extended to a probability measure on $(X, \mathcal{B})$. Let $T \colon X \to X$ be the *right shift*, i.e. for any $x = (y_n)_{n\in\mathbb{Z}}$ and $n \in \mathbb{Z}$ one has that $(Tx)_n = y_{n-1}$, so $Tx = (y_{n-1})_{n\in\mathbb{Z}}$.

(b) T is $\mathcal{B}$-measurable, invertible, and μ-preserving.

Proof. Clearly, T is invertible with $\left(T^{-1}(x)\right)_n = y_{n+1}$ for any $x = (y_n)_{n\in\mathbb{Z}}$:

$$\left(\left(T^{-1} \circ T\right)(x)\right)_n = \left(T^{-1}\left((y_{m-1})_{m\in\mathbb{Z}}\right)\right)_n = y_{(n-1)+1},$$

$$\left(\left(T \circ T^{-1}\right)(x)\right)_n = \left(T\left((y_{m+1})_{m\in\mathbb{Z}}\right)\right)_n \qquad = y_{(n+1)-1}.$$

As $\mathcal{B}$ is generated by $\mathcal{C}$, it is sufficient to prove measurability and the μ-preserving property of the sets from $\mathcal{C}$. In particular, by the definition of a cylinder, it is clear that $\mathcal{B}$ is already generated by cylinders of the form (1) with $A = A_1 \times \cdots \times A_k$. Thus, it suffices to show the two claims for those particular cylinders. So, let $C \in \mathcal{C}$ be a cylinder as in (1) with $A = A_1 \times \cdots \times A_k$. We can see that $T^{-1}(C) \in \mathcal{B}$ is the cylinder given by the same k as C, the same $A_1, \ldots, A_k$ and $n_i' = n_i - 1$ for all $i \in \{1, \ldots, k\}$. In particular, we get that $\mu(T^{-1}(C)) = \mu(C)$ and are done. □

(c) T is ergodic.

Proof. Let $A \subseteq X$ in $\mathcal{B}$ such that $T^{-1}(A) = A$. We need to show that $\mu(A) \in \{0, 1\}$. We fix $\varepsilon > 0$. Recall the following lemma:

Lemma A.2. Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{A} \subseteq \mathcal{B}$ be an algebra such that $\mathcal{B} = \sigma(\mathcal{A})$. Then for any $B \in \mathcal{B}$ and any $\varepsilon > 0$ there exists $A \in \mathcal{A}$ such that $\mu(A \,\Delta\, B) < \varepsilon$.

Since $\mathcal{C}$ is an algebra that generates $\mathcal{B}$, we can find some C[27] in $\mathcal{C}$ such that $\mu(A \,\Delta\, C) < \varepsilon$ and therefore also $|\mu(A) - \mu(C)| < \varepsilon$. Our next goal will be to show that $|\mu(A) - \mu(C)^2| < 2\varepsilon$, which for $\varepsilon \downarrow 0$ would imply that $\mu(A) = \mu(A)^2$, giving us $\mu(A) \in \{0, 1\}$.

For that, we want to start of by showing that $\mu(T^{-n}(C) \cap C) = \mu(C)^2$ for some n large enough. For that we note that for

$$C = \{x = (y_m)_{m\in\mathbb{Z}} \in X \colon (y_{m_i})_{i=1}^k \in A\}$$

we can choose $n > \max_{1\le i\le k} |m_i|$ to get

$$\{m_i \colon 1 \le i \le k\} \cap \{m_i - n \colon 1 \le i \le k\} = \emptyset.$$

Thus, $T^{-n}(C) \cap C$ is a cylinder of measure

$$\begin{aligned}\mu(T^{-n}(C) \cap C) &= \mu(\{x = (y_m)_{m\in\mathbb{Z}} \colon (y_{m_i})_{i=1}^k \in A \wedge (y_{m_i - n})_{i=1}^k \in A\}) \\ &= \mu(C)^2.\end{aligned}$$

Next we can use the fact that

$$(T^{-n}(C) \cap C) \,\Delta\, (T^{-n}(A) \cap A) \subseteq (T^{-n}(C) \,\Delta\, (T^{-n}(A)) \cup (C \,\Delta\, A)$$

to deduce

$$\mu((T^{-n}(C) \cap C) \,\Delta\, (T^{-n}(A) \cap A)) \le \mu(T^{-n}(C) \,\Delta\, (T^{-n}(A)) + \mu(C \,\Delta\, A)$$
$$< 2\varepsilon$$

[27] Described by $k \in \mathbb{N}, m_1, \ldots, m_k \in \mathbb{Z}, A \in \mathcal{F}^{\otimes k}$.

since T is μ-preserving. Thus, we also get by using the above statement and the T invariance of A that

$$\begin{aligned}\left|\mu(C)^2 - \mu(A)\right| &= |\mu(T^{-n}(C) \cap C) - \mu(T^{-n}(A) \cap A)| \\ &< 2\varepsilon.\end{aligned}$$

As outlined, to conclude that $\mu(A) \in \{0, 1\}$, we consider the case when $\varepsilon \downarrow 0$. More formally, let $\varepsilon_l = 1/l$ for all $l \in \mathbb{N}$ and let $C_l \in \mathcal{C}$ be the cylinder with

$$\mu(A \,\Delta\, C_l) < \varepsilon_l.$$

Consider then $x := \limsup_{l\to\infty} \mu(C_l)$. By construction, $x \in [0, 1]$ and

$$x^2 = \mu(A) = x.$$

In particular, x must satisfy $x^2 - x = x\cdot(x-1) = 0$. Thus, $\mu(A) \in \{0, 1\}$. □

(d) T is *strong mixing*, i.e. for any $A, B \in \mathcal{B}$ one has that

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

Proof. Let $C, C' \in \mathcal{C}$ be such that for some $\varepsilon > 0$ fixed $\mu(A \,\Delta\, C) < \varepsilon$ and $\mu(B \,\Delta\, C') < \varepsilon$. Then, as before, we get

$$\begin{aligned}\mu((T^{-n}(A) \cap B) \,\Delta\, (T^{-n}(C) \cap C')) &\le \mu(T^{-n}(A) \,\Delta\, T^{-n}(C)) + \mu(B \,\Delta\, C') \\ &< 2\varepsilon.\end{aligned}$$

However, as C, C' are in $\mathcal{C}$, we can use the above argument to find a large enough $n_0 \in \mathbb{N}$ such that $\mu(T^{-n}(C) \cap C') = \mu(C)\mu(C')$. Combining all that we have for all $\varepsilon > 0$ and $n \ge n_0$ with the triangle inequality

$$\begin{aligned}&|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| \\ &\le |\mu(T^{-n}(A) \cap B) - \mu(T^{-n}(C) \cap C')| + |\mu(C)\mu(C') - \mu(A)\mu(B)| \\ &< 2\varepsilon + |\mu(C')(\mu(C) - \mu(A)) + \mu(A)(\mu(C') - \mu(B))| \\ &< 2\varepsilon + |\mu(C')(\mu(C) - \mu(A))| + |\mu(A)(\mu(C') - \mu(B))| \\ &< 4\varepsilon.\end{aligned}$$

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

2. Consider the *Baker's Transformation* $S\colon [0, 1)^2 \to [0, 1)^2$ defined by

$$S(x, y) := \begin{cases} \left(2x, \frac{y}{2}\right), & x \in [0, \frac{1}{2}), y \in [0, 1), \\ \left(2x - 1, \frac{y+1}{2}\right), & x \in [\frac{1}{2}, 1), y \in [0, 1). \end{cases}$$

(a) Let $[0,1)^2$ be endowed with the Borel σ-algebra and the Lebesgue measure. Then S is isomorphic to a right shift. In particular, S is ergodic and strong-mixing.

Proof. Consider $(X, \mathcal{B}, \mu)$ as in Exercise 1 with

$$Y = \{0,1\}, \mathcal{F} = \sigma(\{0\}, \{1\}), \nu(\{0\}) = \nu(\{1\}) = \frac{1}{2}.$$

Let $\tilde{Z} \subseteq [0,1)^2$ be the set of all numbers where both components have unique binary decompositions. As there are only countably many with non-unique binary decomposition, it follows that $\tilde{Z}$ is of full measure. Now, as each component is nonnegative and smaller than 1, the binary decomposition of $(x,y) \in \tilde{Z}$ is given by

$$x = \sum_{n\in\mathbb{N}} a_n \cdot 2^{-n}, \qquad (a_1, \ldots \in Y)$$
$$y = \sum_{n\in\mathbb{N}} b_n \cdot 2^{-n}. \qquad (b_1, \ldots \in Y)$$

For the sake of clarity, we alternatively express x and y in their binary representations $0.a_1a_2\ldots, 0.b_1b_2\ldots$ respectively. Thus, we may consider the mapping $V\colon \tilde{Z} \to \tilde{X}$ defined by

$$V(x,y)_n = \begin{cases} a_{-n}, & n < 0 \\ b_{n+1}, & n \geq 0, \end{cases}$$

where we take $\tilde{X}$ to be the image of V.

n	$\ldots$	-2	-1	0	1	2	$\ldots$
n-th term	$\ldots$	a_2	a_1	b_1	b_2	b_3	$\ldots$

Table 1: Schematic representation of $V(x,y)$

Note that by construction, $\tilde{X}$ is the set of all sequences whose terms for both $n \to \infty$ and $n \to -\infty$ don't become constant. Hence, $\tilde{X}$ is also of full measure. In fact, V is a bijection[28] by the construction of $\tilde{Z}$ that lets one interchange between the measure spaces, i.e. for measurable $A \subseteq \tilde{Z}$ and measurable $B \subseteq \tilde{X}$ we have

$$\mu(V(A)) = \lambda^2(A) \qquad \lambda^2(V^{-1}(B)) = \mu(B).$$

We will sketch the proof of this fact now. A formal treatment of a similar statement is given in Exercise 2.19: Consider a cylinder C as in (1) where $A = A_1 \times \cdots \times A_k$. In particular, let all A_i be singleton sets. To what

[28] With V^{-1} defined in the obvious way.

set in $\tilde{Z}$ does C correspond to? Well, A basically specifies in the binary decomposition certain digits that then either have to be 0 or 1. As such, one can deduce that C corresponds to a countable union of cubes. Each such specification due to the n_i's halves the area of the corresponding union of cubes and thus for those sets the measures are equivalent. Those cylinders in the form of C already generate the whole σ-algebra as Y is discrete. So, we are done in one direction. In the other direction, one can show that using countably many unions and intersections of those cylinders like C, one can describe any half-cube in $\tilde{Z}$. This then completes the claim.

Now, by simply plugging all the definitions, we see that for $x = (y_m)_{m\in\mathbb{Z}} \in \tilde{X}$ we get for $n \in \mathbb{Z}$

$$\begin{aligned}
&\left(\left(VSV^{-1}\right)x\right)_n \\
=\ &\left((VS)\left(0.y_{-1}y_{-2}y_{-3}\ldots, 0.y_0y_1y_2\ldots\right)\right)_n \\
=\ &\left(V\left(y_{-1}.y_{-2}y_{-3}\ldots - y_{-1}, 0.0y_0y_1y_2\ldots + 0.y_{-1}\right)\right)_n \\
=\ &\left(V\left(0.y_{-2}y_{-3}\ldots, 0.y_{-1}y_0y_1y_2\ldots\right)\right)_n \\
=\ &\left(V\left(\sum_{m\in\mathbb{N}} y_{-m-1}\cdot 2^{-m}, \sum_{m\in\mathbb{N}} y_{(m-1)-1}\cdot 2^{-m}\right)\right)_n \\
=\ &y_{n-1}
\end{aligned}$$

Thus, VSV^{-1} is the right shift on $(X, \mathcal{B}, \mu)$.

Even though this shows that S is isomorphic to a right shift and thus inherits all the above shown properties, we show each of the properties explicitly for completeness: We have for measurable $A \subseteq \tilde{Z}$ with $S^{-1}(A) = A$ that $V(A)$ satisfies $(VSV^{-1})^{-1}(V(A)) = V(A)$, so

$$\lambda^2(A) = \mu\left(V(A)\right) \in \{0, 1\}$$

as VSV^{-1} is ergodic by the first exercise. In a similar vein, using the fact that VSV^{-1} is strong mixing, for measurable $A, B \subseteq \tilde{Z}$ we have that

$$\begin{aligned}
\lambda^2(S^{-n}(A) \cap B) &= \mu(V(S^{-n}(A) \cap B)) \\
&= \mu\left(\left(VSV^{-1}\right)^{-n}(V(A)) \cap V(B)\right) \\
&\overset{n\to\infty}{\to} \mu(V(A))\mu(V(B)) \\
&= \lambda^2(A)\lambda^2(B).
\end{aligned}$$

Thus, S is ergodic and strong mixing. □

Exercise Sheet 5

1. This problem will deal with one particular equivalence in Theorem 8.3.

(a) For any bounded sequence $(a_n)_{n\geq 0} \in \mathbb{R}^{\mathbb{N}}$ the following are equivalent:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n| = 0 \iff \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n|^2 = 0.$$

Proof. Let us first assume that the limit on the right hand side holds. Then the Cauchy-Schwarz inequality implies

$$\begin{aligned}\frac{1}{N} \sum_{n=0}^{N-1} |a_n| = \sum_{n=0}^{N-1} \frac{1}{N} \cdot |a_n| \quad &\leq \left(\sum_{n=0}^{N-1} \frac{1}{N^2}\right)^{\frac{1}{2}} \cdot \left(\sum_{n=0}^{N-1} |a_n|^2\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{N} \sum_{n=0}^{N-1} |a_n|^2\right)^{\frac{1}{2}} \to 0, N \to \infty.\end{aligned}$$

Note that we didn't use the fact that $(a_n)_{n\geq 0}$ is bounded.

For the other direction, choose $M > 0$ such that $|a_n| \leq M$ for all $n \in \mathbb{N}$. In particular, we have that $|a_n|/M \leq 1$ for all $n \in \mathbb{N}$, so

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{|a_n|^2}{M} \leq \frac{1}{N} \sum_{n=0}^{N-1} |a_n| \to 0, N \to \infty.$$

Hence, as $1/M$ is just a constant the desired follows. □

(b) Let $(X, \mathcal{B}, \mu)$ be a probability space. For any measure-preserving transform $T\colon X \to X$ and any $f, g \in L^2(X)$, the following are equivalent:

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right| = 0 \\ \iff &\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right|^2 = 0.\end{aligned}$$

Proof. Set $a_n := \int_X U_T^n f \cdot g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu$ for all $n \in \mathbb{N}$. As $f, g \in L^2(X)$, we get using the triangle inequality and Cauchy-Schwarz inequality

$$\begin{aligned}|a_n| &\leq \left| \int_X U_T^n f \cdot g \, \mathrm{d}\mu \right| + \left| \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right| \\ &\leq \|f\|_{L^2} \|g\|_{L^2} + \|f\|_{L^1} \|g\|_{L^1} \\ &< \infty\end{aligned}$$

since T is μ-preserving. So the sequence $(a_n)_{n\geq 0}$ is bounded and we are done by (a). □

2. Let $(X, \mathcal{B}, \mu)$ be a probability space, $T \colon X \to X$ be μ-preserving, and let $(e_k)_{k \geq 1}$ be an orthonormal basis of $L^2(X)$ (possibly, $\mathbb{C}$-valued). Then T is strongly mixing if and only if for any $k, m \geq 1$

$$\lim_{n \to \infty} \int_X U_T^n e_k \bar{e}_m \, \mathrm{d}\,\mu = \int_X e_k \, \mathrm{d}\,\mu \int_X \bar{e}_m \, \mathrm{d}\,\mu,$$

where $\bar{e}_m$ is the complex conjugate of e_m.

Proof. Recall the following lemma which can be derived from Theorem 8.3:

Lemma A.3. Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T \colon X \to X$ be μ-preserving. Then T is strongly mixing if and only if for any $f, g \in L^2(X)$ (possibly, $\mathbb{C}$-valued) one has that

$$\lim_{n \to \infty} \int_X U_T^n f \bar{g} \, \mathrm{d}\,\mu = \int_X f \, \mathrm{d}\,\mu \int_X g \, \mathrm{d}\,\mu,$$

where $\bar{g}$ is the complex conjugate of g.

Using the equivalence, it is clear that if T is strongly mixing, we do get

$$\lim_{n \to \infty} \int_X U_T^n e_k \bar{e}_m \, \mathrm{d}\,\mu = \int_X e_k \, \mathrm{d}\,\mu \int_X \bar{e}_m \, \mathrm{d}\,\mu$$

for all $k, m \geq 1$. For the other direction, we will do an approximation argument: Recall that every function $f \in L^2(X)$ can be uniquely represented as $\sum_{j \geq 1} f_j e_j$ for some l^2-sequence $(f_j)_{j \geq 1}$. So, fix any $f, g \in L^2(X), \varepsilon > 0$ and let $f = \sum_{j \geq 1} f_j e_j$ and $g = \sum_{l \geq 1} g_l e_l$ be their respective representations. For $k \in \mathbb{N}$, let $f^{(k)}$ be the partial sum formed by the first k terms in the representation of f, and define $g^{(m)}$ similarly for $m \in \mathbb{N}$. Since $(f_j)_{j \geq 1}$ is an l^2-sequence, we have

$$\left\| f - f^{(k)} \right\|_{L^2} = \left(\sum_{j \geq k+1} f_j^2 \right)^{\frac{1}{2}} \to 0, k \to \infty$$

and the same holds for g and $g^{(m)}$. Fix $k, m \in \mathbb{N}$ such that

$$\left\| f - f^{(k)} \right\|_{L^2} < \varepsilon, \qquad \left\| g - g^{(m)} \right\|_{L^2} < \varepsilon.$$

For all $n \geq 0$ we have that

$$\begin{aligned} &\left| \int_X U_T^n f \cdot \bar{g} \, \mathrm{d}\,\mu - \int_X f \, \mathrm{d}\,\mu \int_X \bar{g} \, \mathrm{d}\,\mu \right| \\ \leq &\left| \int_X U_T^n f \cdot \bar{g} \, \mathrm{d}\,\mu - \int_X U_T^n f \cdot \bar{g}^{(m)} \, \mathrm{d}\,\mu \right| && (\star) \\ &+ \left| \int_X U_T^n f \cdot \bar{g}^{(m)} \, \mathrm{d}\,\mu - \int_X U_T^n f^{(k)} \cdot \bar{g}^{(m)} \, \mathrm{d}\,\mu \right| && (\star\star) \end{aligned}$$

$$+\left|\int_X U_T^n f^{(k)} \cdot \bar{g}^{(m)} \, \mathrm{d}\,\mu - \int_X f^{(k)} \, \mathrm{d}\,\mu \int_X \bar{g}^{(m)} \, \mathrm{d}\,\mu\right| \qquad (\star\star\star)$$
$$+\left|\int_X f^{(k)} \, \mathrm{d}\,\mu \int_X \bar{g}^{(m)} \, \mathrm{d}\,\mu - \int_X f \, \mathrm{d}\,\mu \int_X \bar{g}^{(m)} \, \mathrm{d}\,\mu\right| \qquad (\star\star\star\star)$$
$$+\left|\int_X f \, \mathrm{d}\,\mu \int_X \bar{g}^{(m)} \, \mathrm{d}\,\mu - \int_X f \, \mathrm{d}\,\mu \int_X \bar{g} \, \mathrm{d}\,\mu\right|. \qquad (\star\star\star\star\star)$$

We will use the Cauchy-Schwarz inequality multiple times:

- $(\star)$ is bounded from above by
$$\|U_T^n f\|_{L^2} \left\|\bar{g} - \bar{g}^{(m)}\right\|_{L^2} = \|U_T^n f\|_{L^2} \left\|g - g^{(m)}\right\|_{L^2} \leq \varepsilon \|f\|_{L^2}$$
as $\|U_T^n f\|_{L^2} = \|f\|_{L^2}$.
- $(\star\star)$ is bounded from above by
$$\left\|U_T^n \left(f - f^{(k)}\right)\right\|_{L^2} \left\|\bar{g}^{(m)}\right\|_{L^2} \leq \varepsilon \left(\|\bar{g}\|_{L^2} + \left\|\bar{g}^{(m)} - \bar{g}\right\|_{L^2}\right) \leq \varepsilon \left(\|g\|_{L^2} + \varepsilon\right).$$
- $(\star\star\star)$ goes to zero as n goes to infinity since by bilinearity, we get
$$\begin{aligned}\int_X U_T^n f^{(k)} \cdot \bar{g}^{(m)} \, \mathrm{d}\,\mu &= \sum_{j=1}^{k} \sum_{l=1}^{m} f_j \bar{g}_l \int_X U_T^n e_j \cdot \bar{e}_l \, \mathrm{d}\,\mu \\ &\overset{n\to\infty}{\to} \sum_{j=1}^{k} \sum_{l=1}^{m} f_j \bar{g}_l \int_X e_j \, \mathrm{d}\,\mu \int_X \bar{e}_l \, \mathrm{d}\,\mu \\ &= \int_X f^{(k)} \, \mathrm{d}\,\mu \int_X \bar{g}^{(m)} \, \mathrm{d}\,\mu.\end{aligned}$$
- $(\star\star\star\star)$ is bounded from above by
$$\left\|\bar{g}^{(m)}\right\|_{L^2} \left\|f - f^{(k)}\right\|_{L^2} = \left\|g^{(m)}\right\|_{L^2} \left\|f - f^{(k)}\right\|_{L^2} \leq \varepsilon \left(\|g\|_{L^2} + \varepsilon\right).$$
- $(\star\star\star\star\star)$ is bounded from above by
$$\|f\|_{L^2} \left\|\bar{g} - \bar{g}^{(m)}\right\|_{L^2} = \|f\|_{L^2} \left\|g - g^{(m)}\right\|_{L^2} \leq \varepsilon \|f\|_{L^2}.$$

As $\varepsilon > 0$ was arbitrary,
$$\left|\int_X U_T^n f \cdot g \, \mathrm{d}\,\mu - \int_X f \, \mathrm{d}\,\mu \int_X g \, \mathrm{d}\,\mu\right| \to 0, n \to \infty.$$

Hence, T is strongly mixing. □

3. Let $X \coloneqq [0,1)^2$. *Arnold's cat map* $T \colon X \to X$ is defined by
$$Tx \coloneqq \begin{pmatrix} \{x^1 + x^2\} \\ \{x^1 + 2x^2\} \end{pmatrix}, x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \in \mathbb{R}^2,$$
where $\{\,\cdot\,\}$ is the fractional part of a real number. The goal of this exercise is to show that T is strongly mixing. Recall that T is measure-preserving Exercise Sheet 1 for the Lebesgue measure which we will denote by μ.

(a) We have for all $n \in \mathbb{N}$ that

$$T^n x = \begin{pmatrix} \{a_{2n-2}x^1 + a_{2n-1}x^2\} \\ \{a_{2n-1}x^1 + a_{2n}x^2\} \end{pmatrix}, x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \in \mathbb{R}^2,$$

where $(a_n)_{n \geq 0}$ is the Fibonacci sequence which is defined recursively as $a_0 = a_1 = 1$ and $a_n = a_{n-1} + a_{n-2}$ for $n \geq 2$.

Proof. First, note that we can rewrite T as

$$Tx = \left(\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} x \right)_{\mathrm{mod}\ 1}.$$

To simplify the problem, note the following: If $\tilde{x} = x + h$ for some $h \in \mathbb{Z}^2$, then $T\tilde{x} = Tx$ since multiplying h by that matrix gives again a vector in $\mathbb{Z}^2$. So, we inductively have

$$T^n x = \left(\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^n x \right)_{\mathrm{mod}\ 1}.$$

for all $n \in \mathbb{N}$. For the rest of the proof, we will proceed by induction on $n \in \mathbb{N}$ to show that

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^n = \begin{pmatrix} a_{2n-2} & a_{2n-1} \\ a_{2n-1} & a_{2n} \end{pmatrix}.$$

Then the claim follows directly by the above.

Now, the induction base is clear since $a_2 = 2$. For the induction step $n \rightsquigarrow n+1$, using the induction hypothesis, we see that

$$\begin{aligned}
\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{n+1} &= \begin{pmatrix} a_{2n-2} & a_{2n-1} \\ a_{2n-1} & a_{2n} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \\
&= \begin{pmatrix} a_{2n-2} + a_{2n-1} & a_{2n-2} + 2a_{2n-1} \\ a_{2n-1} + a_{2n} & a_{2n-1} + 2a_{2n} \end{pmatrix} \\
&= \begin{pmatrix} a_{2n} & a_{2n-1} + (a_{2n-2} + a_{2n-1}) \\ a_{2n+1} & a_{2n} + (a_{2n-1} + a_{2n}) \end{pmatrix} \\
&= \begin{pmatrix} a_{2n} & a_{2n+1} \\ a_{2n+1} & a_{2n+2} \end{pmatrix} \\
&= \begin{pmatrix} a_{2(n+1)-2} & a_{2(n+1)-1} \\ a_{2(n+1)-1} & a_{2(n+1)} \end{pmatrix}.
\end{aligned}$$

This completes the proof. □

(b) For each $k, l \in \mathbb{Z}$ let $e_{k,l}\colon X \to \mathbb{C}$ be defined by

$$e_{k,l}(x) = \mathrm{e}^{2\pi\,\mathrm{i}(kx^1+lx^2)}, x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \in \mathbb{R}^2.$$

Then we have for every $n \in \mathbb{N}$ and every $p, q \in \mathbb{Z}$

$$\int_X U_T^n e_{k,l} \cdot \bar{e}_{p,q} \, \mathrm{d}\,\mu = \begin{cases} 1, & ka_{2n-2} + la_{2n-1} - p = ka_{2n-1} + la_{2n} - q = 0 \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $\int_X U_T^n e_{k,l} \cdot \bar{e}_{p,q} \, \mathrm{d}\,\mu$ does not converge to zero if and only if there exists an increasing sequence $(n_j)_{j\geq 1} \in \mathbb{N}^{\mathbb{N}}$ such that

$$ka_{2n_j-2} + la_{2n_j-1} - p = ka_{2n_j-1} + la_{2n_j} - q = 0 \tag{2}$$

for any $j \geq 1$.

Proof. Fix $k, l, p, q \in \mathbb{Z}$ and $n \in \mathbb{N}$. Recall that for $m \in \mathbb{Z}, m \neq 0$, we have

$$\int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,m\cdot z} \, \mathrm{d}\,z = \left[\frac{1}{2\pi\,\mathrm{i}\,m} \mathrm{e}^{2\pi\,\mathrm{i}\,m\cdot z}\right]_0^1 = 0.$$

Also note that the integrand is 1 and thus the integral 1 if $m = 0$. So, using Fubini's theorem and (a), we get

$$\begin{aligned}
&\int_X U_T^n e_{k,l} \cdot \bar{e}_{p,q} \, \mathrm{d}\,\mu \\
&= \int_0^1 \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\left(k\left\{a_{2n-2}x^1+a_{2n-1}x^2\right\}+l\left\{a_{2n-1}x^1+a_{2n}x^2\right\}\right)} \mathrm{e}^{-2\pi\,\mathrm{i}(px^1+qx^2)} \, \mathrm{d}\,x^1 \, \mathrm{d}\,x^2 \\
&= \int_0^1 \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\left(k\left(a_{2n-2}x^1+a_{2n-1}x^2\right)+l\left(a_{2n-1}x^1+a_{2n}x^2\right)\right)} \mathrm{e}^{-2\pi\,\mathrm{i}(px^1+qx^2)} \, \mathrm{d}\,x^1 \, \mathrm{d}\,x^2 \\
&= \int_0^1 \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,x^1(ka_{2n-2}+la_{2n-1}-p)+2\pi\,\mathrm{i}\,x^2(ka_{2n-1}+la_{2n}-q)} \, \mathrm{d}\,x^1 \, \mathrm{d}\,x^2 \\
&= \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,x^1(ka_{2n-2}+la_{2n-1}-p)} \, \mathrm{d}\,x^1 \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,x^2(ka_{2n-1}+la_{2n}-q)} \, \mathrm{d}\,x^2 \\
&= \begin{cases} 1, & ka_{2n-2} + la_{2n-1} - p = ka_{2n-1} + la_{2n} - q = 0 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

The second equality holds since $\mathrm{e}^{2\pi\,\mathrm{i}(x+h)} = \mathrm{e}^{2\pi\,\mathrm{i}\,x}$ for any $x \in \mathbb{R}, h \in \mathbb{Z}$. □

(c) If $\int_X U_T^n e_{k,l} \cdot \bar{e}_{p,q} \, \mathrm{d}\,\mu$ does not converge to zero as $n \to \infty$, then using (2) and the fact that

$$\lim_{n\to\infty} \frac{a_{n-1}}{a_n} = \frac{2}{1+\sqrt{5}} \wedge \lim_{n\to\infty} a_n = +\infty,$$

we can conclude that $k = l = p = q = 0$. In particular, T is strongly mixing.

Proof. Just for fun, let's first show the facts: The latter is trivial as the Fibonacci sequence is from the second term onwards strictly increasing by at least one. For the former, one can check that the sequence of ratios $(a_n/a_{n-1})_{n\in\mathbb{N}}$ is converging to some positive number φ satisfying

$$\varphi = \lim_{n\to\infty} \frac{a_n}{a_{n-1}} = 1 + \lim_{n\to\infty} \frac{1}{\frac{a_{n-1}}{a_{n-2}}} = 1 + \frac{1}{\varphi}.$$

In particular, $\varphi^2 = \varphi + 1$, so

$$\varphi \in \left\{ \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2} \right\}$$

and thus $\varphi = (1+\sqrt{5})/2$ as φ has to be non-negative.
Now, if $\int_X U_T^n e_{k,l} \cdot \bar{e}_{p,q} \,\mathrm{d}\,\mu$ does not converge to zero as $n \to \infty$ for some $k, l, p, q \in \mathbb{Z}$, then there must exist an increasing sequence $(n_j)_{j\geq 1}$ satisfying (2). So, in particular, we have

$$k \cdot \frac{a_{2n_j-2}}{a_{2n_j-1}} + l = \frac{p}{a_{2n_j-1}} \wedge k \cdot \frac{a_{2n_j-1}}{a_{2n_j}} + l = \frac{q}{a_{2n_j}} \tag{3}$$

for any $j \geq 1$. For $j \to \infty$, the first identity of (3) gives

$$\frac{k}{\varphi} + l = \lim_{j\to\infty} \left(k \cdot \frac{a_{2n_j-2}}{a_{2n_j-1}} + l \right) = 0 \implies k = -\varphi l.$$

But φ is irrational, so the equality can only hold for $k = l = 0$. This immediately gives us that $p = q = 0$ from (3). Thus, $k = l = p = q = 0$.
This actually shows that for all $k, l, p, q \in \mathbb{Z}$

$$\lim_{n\to\infty} \int_X U_T^n e_{k,l} \bar{e}_{p,q} \,\mathrm{d}\,\mu = \int_X e_{k,l} \,\mathrm{d}\,\mu \int_X \bar{e}_{p,q} \,\mathrm{d}\,\mu,$$

since, by a similar calculation as before, we have

$$\begin{aligned} \int_X e_{k,l} \,\mathrm{d}\,\mu &= \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,kx^1} \,\mathrm{d}\,x^1 \int_0^1 \mathrm{e}^{2\pi\,\mathrm{i}\,lx^2} \,\mathrm{d}\,x^2 \\ &= \begin{cases} 1, & k = l = 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

and the same for $\bar{e}_{k,l}$ for all $k, l \in \mathbb{Z}$. Thus, as $(e_{k,l})_{k,l\in\mathbb{Z}}$ is an orthonormal basis of $L^2(X)$, it follows that T is strongly mixing by Exercise 2. □

It is not clear when exactly $A \in \mathrm{SL}_2(\mathbb{Z})$ induces a strong mixing transformation. In the notation in Exercise Sheet 1, we conjecture the following:

Conjecture A.4 (Ivan's conjecture)**.** A induces a strong mixing transformation if and only if $b \neq 0$, $c \neq 0$ and at most one of a, d is zero.

Maybe someone is able to prove it, perhaps *A Combinatorial Formula for Powers of 2 × 2 Matrices* by Konvalina would be useful.

4. Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T\colon X \to X$ be μ-preserving. Let $T \times T\colon X \times X \to X \times X$ be defined by $(T \times T)(x, y) = (Tx, Ty)$ for $x, y \in X$. Let $\mathcal{B} \otimes \mathcal{B} := \sigma(\{A \times B\colon A, B \in \mathcal{B}\})$ be the product σ-algebra on $X \times X$ and let $\mu \otimes \mu$, defined by $(\mu \otimes \mu)(A \times B) = \mu(A)\mu(B)$ for $A, B \in \mathcal{B}$, be the product measure on $(X \times X, \mathcal{B} \otimes \mathcal{B})$. Then T is strongly mixing if and only if $T \times T$ is strongly mixing on $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu)$.

Proof. For the one direction, let $T \times T$ be strongly mixing. Fix $A, B \in \mathcal{B}$. Then $A \times X, B \times X \in \mathcal{B} \otimes \mathcal{B}$. Thus, on one hand, we get

$$\begin{aligned}(\mu \otimes \mu)\left((T \times T)^{-n}(A \times X) \cap (B \times X)\right) &= (\mu \otimes \mu)\left((T^{-n}(A) \times X) \cap (B \times X)\right)\\ &= (\mu \otimes \mu)\left((T^{-n}(A) \cap B) \times X\right)\\ &= \mu\left(T^{-n}(A) \cap B\right)\mu(X)\\ &= \mu\left(T^{-n}(A) \cap B\right)\end{aligned}$$

for all $n \in \mathbb{N}$ due to $T^{-n}(X) = X$. On the other hand, we have

$$\begin{aligned}\lim_{n\to\infty}(\mu \otimes \mu)\left((T \times T)^{-n}(A \times X) \cap (B \times X)\right) &= (\mu \otimes \mu)(A \times X)(\mu \otimes \mu)(B \times X)\\ &= \mu(A)\mu(B)\mu(X)^2\\ &= \mu(A)\mu(B).\end{aligned}$$

Thus, $\lim_{n\to\infty} \mu\left(T^{-n}(A) \cap B\right) = \mu(A)\mu(B)$ and T is strongly mixing.

For the other direction, let T be strongly mixing. We will use Exercise 2: Let $(e_k)_{k\geq 1}$ be an orthonormal basis of $L^2(X)$. Then, $(e_{k,l})_{k,l\geq 1}$, defined by $e_{k,l} := e_k(x)e_l(y)$ for $x, y \in X$, form an orthonormal basis of $L^2(X \times X)$. Using Fubini's theorem and T being strongly mixing, we get

$$\begin{aligned}&\lim_{n\to\infty}\int_{X\times X} U^n_{T\times T}e_{k,l} \cdot \bar{e}_{p,q}\ \mathrm{d}\,(\mu \otimes \mu)\\ =\ &\lim_{n\to\infty}\int_{X\times X} e_k(T^n x)e_l(T^n y) \cdot \bar{e}_p(x)\bar{e}_q(y)\ \mathrm{d}\,(\mu \otimes \mu)(x, y)\\ =\ &\lim_{n\to\infty}\left(\int_X e_k(T^n x) \cdot \bar{e}_p(x)\ \mathrm{d}\,\mu(x)\int_X e_l(T^n y) \cdot \bar{e}_q(y)\ \mathrm{d}\,\mu(y)\right)\\ =\ &\lim_{n\to\infty}\left(\int_X U^n_T e_k \cdot \bar{e}_p\ \mathrm{d}\,\mu\int_X U^n_T e_l \cdot \bar{e}_q\ \mathrm{d}\,\mu\right)\\ =\ &\left(\int_X e_k\ \mathrm{d}\,\mu\int_X \bar{e}_p\ \mathrm{d}\,\mu\right)\left(\int_X e_l\ \mathrm{d}\,\mu\int_X \bar{e}_q\ \mathrm{d}\,\mu\right)\\ =\ &\int_{X\times X} e_{k,l}\ \mathrm{d}\,(\mu \otimes \mu)\int_{X\times X} \bar{e}_{p,q}\ \mathrm{d}\,(\mu \otimes \mu)\end{aligned}$$

for all $k, l, p, q \in \mathbb{N}$. Thus, by Exercise 2, $T \times T$ is strongly mixing. □