Dynamical systems

École Normale Supérieure, 2023-2026

Cyril Houdayer

École Normale Supérieure, Département de Mathématiques et Applications, Université Paris-Saclay, Institut Universitaire de France, 45 rue d'Ulm, 75230 Paris Cedex 05, FRANCE
cyril.houdayer@ens.psl.eu.

ABSTRACT. These are the lecture notes of a master course given at École Normale Supérieure during 2023-2026. In this master course, we give an introduction to topological dynamics and ergodic theory, and we discuss examples of dynamical systems arising from group theory, symbolic dynamics, geometry and homogeneous spaces. Topics include: Topological dynamics (topological transtivity and recurrence, Ramsey theory, topological entropy); Ergodic theory (recurrence, ergodicity, weak mixing, ergodic theorems, random walks, measure entropy); Homogeneous dynamics (locally compact groups, lattices, $\mathrm{SL}_n(\mathbb{Z}) < \mathrm{SL}_n(\mathbb{R})$, Howe–Moore's property, Moore's ergodicity theorem).

Contents

CHAPTER 1

Topological dynamics

In these lecture notes, all topological spaces are assumed to be Hausdorff. For any topological space X, a continuous map $T : X \to X$ is called a *topological dynamical system*. If $T : X \to X$ is moreover a homeomorphism, then $T : X \to X$ induces an action $\mathbb{Z} \times X \to X : (n, x) \mapsto T^n(x)$ by homeomorphisms.

A subset $Y \subset X$ is (forward) T-*invariant* if $T(Y) \subset Y$. When $T : X \to X$ is a homeomorphism, a subset $Y \subset X$ is T-*invariant* if $T(Y) = Y$. For every $x \in X$, define the (positive) T-*orbit* $\mathscr{O}_T^+(x) = \{T^n(x) \mid n \in \mathbb{N}\}$. When $T : X \to X$ is a homeomorphism, for every $x \in X$, define the (negative) T-*orbit* $\mathscr{O}_T^-(x) = \{T^{-n}(x) \mid n \in \mathbb{N}\}$ and the T-*orbit* $\mathscr{O}_T(x) = \mathscr{O}_T^+(x) \cup \mathscr{O}_T^-(x) = \{T^n(x) \mid n \in \mathbb{Z}\}$. We say that $x \in X$ is T-*fixed* if $\mathscr{O}_T^+(x) = \{x\}$ and T-*periodic* if there exists $k \geq 1$ such that $T^k(x) = x$.

For every $i \in \{1, 2\}$, let $T_i : X_i \to X_i$ be a topological dynamical system. We say that T_2 is a *topological factor* of T_1 or that T_1 is a *topological extension* of T_2 if there exists a surjective continuous map $\pi : X_1 \to X_2$ such that $\pi \circ T_1 = T_2 \circ \pi$. We say that T_1 and T_2 are *topologically conjugate* if there exists a homeomorphism $\pi : X_1 \to X_2$ such that $\pi \circ T_1 = T_2 \circ \pi$. Topologically conjugate dynamical systems have identical topological properties. Therefore, all properties and invariants we introduce in this chapter including minimality, topological transitivity, topological recurrence, topological mixing, topological entropy are preserved by topological conjugacy.

For this chapter, we follow the presentation given in [**BS02**, **EW11**].

1. Examples of topological dynamical systems

1.1. Rotations. Denote by $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ the unit circle, which is a compact metrizable group. For every $\alpha \in \mathbb{R}$, define the *rotation* $T_\alpha : \mathbb{T} \to \mathbb{T} : z \mapsto \exp(\mathrm{i}2\pi\alpha)z$ which is a homeomorphism. For every $\alpha \in \mathbb{R}$, we have $(T_\alpha)^{-1} = T_{-\alpha}$. Endow $\mathbb{T}$ with the metric defined by $d : \mathbb{T} \times \mathbb{T} \to [0, 1) : (z_1, z_2) \mapsto \min(\theta, 1 - \theta)$ where $\theta \in [0, 1)$ is the unique element such that $z_2 z_1^{-1} = \exp(\mathrm{i}2\pi\theta)$. Then $T_\alpha : \mathbb{T} \to \mathbb{T}$ is an isometry in the sense that $d(T_\alpha(z_1), T_\alpha(z_2)) = d(z_1, z_2)$ for all $z_1, z_2 \in \mathbb{T}$.

If $\alpha = \frac{p}{q} \in \mathbb{Q}$, then $(T_\alpha)^q = \mathrm{id}_\mathbb{T}$ and so every point $z \in \mathbb{T}$ is T_α-periodic. If $\alpha \notin \mathbb{Q}$, then for every $z \in \mathbb{T}$, the positive T_α-orbit $\mathscr{O}_{T_\alpha}^+(z)$ is dense in $\mathbb{T}$. In that case, we say that T_α is a *minimal* topological dynamical system. Indeed, let $z \in \mathbb{T}$ and $N \geq 2$. Since $(\exp(\mathrm{i}2\pi\alpha n))_n$ are pairwise distinct in

$\mathbb{T}$, the pigeon-hole principle implies that there exist $0 \leq m < n \leq N$ such that $d(\exp(\mathrm{i}2\pi\alpha n), \exp(\mathrm{i}2\pi\alpha m)) \leq \frac{1}{N}$. This means that the rotation T_α^{n-m} has an angle $\theta = d(T_\alpha^{n-m}(1), 1) \leq \frac{1}{N}$. Since $n - m \geq 1$, it follows that the positive T_α-orbit $\mathscr{O}^+_{T_\alpha}(z)$ comes within distance $\frac{1}{N}$ of every point in $\mathbb{T}$. Since this is true for every $N \geq 2$, the positive T_α-orbit $\mathscr{O}^+_{T_\alpha}(z)$ is dense in $\mathbb{T}$.

Rotations on the unit cercle are particular examples of group translations. Let G be a (Hausdorff) topological group. For every $g \in G$, define the homeomorphism $T_g : G \to G : x \mapsto gx$. If G is first countable, that is, if the identity element $e \in G$ admits a countable neighborhood basis, then by Birkhoff–Kalutani's theorem, G possesses a left-invariant compatible metric $d : G \times G \to \mathbb{R}_+$. In that case, $T_g : G \to G$ is an isometry for every $g \in G$.

1.2. Bernoulli shifts. Let Y be a topological space and I an at most countable index set (e.g. $I = \{1, \dots, n\}$ for $n \geq 1$, or $I = \mathbb{N}$, or $I = \mathbb{Z}$). Consider the product space $Y^I = \{(y_i)_{i\in I} \mid \forall i \in I, y_i \in Y\}$ endowed with the product topology. For every nonempty finite subset $\mathscr{F} \subset I$ and every family of open sets $(U_i)_{i\in\mathscr{F}}$ of Y, define the *cylinder* open set $\mathscr{C}(\mathscr{F}, (U_i)_{i\in\mathscr{F}}) = \prod_{i\in I} Z_i$ where $Z_i = U_i$ if $i \in \mathscr{F}$ and $Z_i = Y$ if $i \notin \mathscr{F}$. Then the family $(\mathscr{C}(\mathscr{F}, (U_i)_{i\in\mathscr{F}}))_{\mathscr{F}\subset I, \mathscr{F} \text{ finite}}$ is a basis of open sets for the product topology on Y^I. If Y is a compact space, then so is Y^I by Tychonov's theorem. If Y is a Polish space, meaning that Y is a separable complete metrizable topological space, then so is Y^I.

Consider the product space $Y^{\mathbb{N}}$ endowed with the product topology and define the (noninvertible) *forward Bernoulli shift* $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$, which is a topological dynamical system. Likewise, consider the product space $Y^{\mathbb{Z}}$ endowed with the product topology and define the (invertible) *Bernoulli shift* $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$, which is a homeomorphism.

1.3. Toral automorphisms. Let $d \geq 1$ and consider the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, which is a compact metrizable group. For every $A \in \mathrm{GL}_d(\mathbb{Z})$, since $A(\mathbb{Z}^d) \subset \mathbb{Z}^d$, we may consider the continuous automorphism

$$T_A : \mathbb{T}^d \to \mathbb{T}^d : x + \mathbb{Z}^d \mapsto Ax + \mathbb{Z}^d.$$

Then $T_A : \mathbb{T}^d \to \mathbb{T}^d$ naturally preserves the Haar (Lebesgue) probability measure on $\mathbb{T}^d$.

Toral automorphisms are particular examples of compact group automorphisms. Let G be a compact metrizable group and denote by m_G its unique Haar probability measure (see Chapter 3). Let $T \in \mathrm{Aut}(G)$ be a continuous group automorphism. Then $T_* m_G = m_G$ and so the topological dynamical system $T : G \to G$ preserves the Haar probability measure $m_G \in \mathrm{Prob}(G)$.

1.4. Homogeneous dynamical systems. Let G be a locally compact second countable group and $\Gamma < G$ a lattice, meaning that $\Gamma < G$ is a discrete subgroup for which the homogeneous locally compact second countable space G/Γ carries a (unique) G-invariant Borel probability measure $\nu \in \mathrm{Prob}(G/\Gamma)$ (see Chapter 3). Let $g \in G$ and consider the left translation homeomorphism $T_g : G/\Gamma \to G/\Gamma : h\Gamma \mapsto gh\Gamma$. Then ${T_g}_*\nu = \nu$ and so the topological dynamical system $T_g : G/\Gamma \to G/\Gamma$ preserves the Borel probability measure $\nu \in \mathrm{Prob}(G/\Gamma)$.

2. Topological transitivity, recurrence, minimality, mixing

Let X be a topological space and $T : X \to X$ a topological dynamical system. For every $x \in X$, the ω-*limit set* of x with respect to T is the closed subset

$$\omega_T(x) = \bigcap_{n\in\mathbb{N}} \overline{\{T^i(x) \mid i \geq n\}}.$$

If X is compact, then $\omega_T(x) \neq \emptyset$. For every $y \in X$, we have $y \in \omega_T(x)$ if and only if there exists a net $(n_i)_{i\in I}$ such that $\lim_{i\to\infty} n_i = +\infty$ and $y = \lim_{i\to\infty} T^{n_i}(x)$. Note that $\omega_T(x)$ is T-invariant. We say that a point $x \in X$ is (positively) T-*recurrent* if $x \in \omega_T(x)$. For every $x \in X$, we have that x is T-recurrent if and only if there exists a net $(n_i)_{i\in I}$ such that $\lim_{i\to\infty} n_i = +\infty$ and $x = \lim_{i\to\infty} T^{n_i}(x)$. The set of all T-recurrent points is T-invariant. Any T-periodic point is T-recurrent.

Definition 1.1. Let X be a topological space and $T : X \to X$ a topological dynamical system. We say that T is *topologically transitive* if there exists $x \in X$ such that $\mathscr{O}_T^+(x)$ is dense in X.

Let X be a topological space. We say that X is *locally compact* if every point $x \in X$ possesses a compact neighborhood. We say that X is *second countable* if there exists a countable family of open sets $(V_i)_{i\in\mathbb{N}}$ that generates the topology of X. We say that X satisfies the *Baire property* if any countable intersection of dense open subsets is dense. Examples of topological spaces with the Baire property include Polish spaces and locally compact topological spaces. We record the following useful sufficient condition that implies topological transitivity.

Proposition 1.2. *Let X be a second countable topological space with the Baire property and $T : X \to X$ a topological dynamical system. Assume that for any nonempty open sets $U, V \subset X$, there exists $n \in \mathbb{N}$ such that $T^{-n}(U) \cap V \neq \emptyset$. Then T is topologically transitive.*

Proof. By assumption, for every nonempty open set $V \subset X$, the open set $\bigcup_{n\in\mathbb{N}} T^{-n}(V)$ intersects any nonempty open set $U \subset X$ and so $\bigcup_{n\in\mathbb{N}} T^{-n}(V)$ is dense in X. Choose a countable family of open sets $(V_i)_{i\in\mathbb{N}}$ that generates the topology of X. Since X satisfies the Baire property, the intersection $\bigcap_{i\in\mathbb{N}}(\bigcup_{n\in\mathbb{N}} T^{-n}(V_i))$ is not empty. Choose a point $y \in \bigcap_{i\in\mathbb{N}}(\bigcup_{n\in\mathbb{N}} T^{-n}(V_i))$. Then for every $i \in \mathbb{N}$, we have $\mathscr{O}_T^+(y) \cap V_i \neq \emptyset$, which implies that $\mathscr{O}_T^+(y)$ is dense in X. Thus, T is topologically transitive. $\square$

In compact metrizable spaces without isolated points, we show that the existence of a dense orbit implies the existence of a dense forward orbit.

PROPOSITION 1.3. *Let X be a compact metrizable space without isolated points and $T : X \to X$ a homeomorphism. Assume that there exists $x \in X$ such that $\mathscr{O}_T(x)$ is dense in X. Then there exists $y \in X$ such that $\mathscr{O}_T^+(y)$ is dense in X, that is, T is topologically transitive.*

PROOF. Fix a compatible metric $d : X \times X \to \mathbb{R}_+$. Since X has no isolated points, for every $k \geq 1$, we may choose $n_k \in \mathbb{Z}$ such that $T^{n_k}(x) \in B(x, \frac{1}{k})$ and $|n_k| \to +\infty$. Note that for every $\ell \in \mathbb{Z}$, we have $T^{n_k+\ell}(x) \to T^\ell(x)$. Firstly, assume that there are infinitely many $k \in \mathbb{N}$ such that $n_k > 0$. Then we have $\mathscr{O}(x) \subset \overline{\mathscr{O}_T^+(x)}$ and so $\mathscr{O}_T^+(x)$ is dense in X. Then we are done. Secondly, assume that there are infinitely many $k \in \mathbb{N}$ such that $n_k < 0$. Then we have $\mathscr{O}(x) \subset \overline{\mathscr{O}_T^-(x)}$ and so $\mathscr{O}_T^-(x)$ is dense in X. Since X has no isolated points, this implies that for any nonempty open sets $U, V \subset X$, we can find integers $i < j < 0$ such that $T^i(x) \in U$ and $T^j(x) \in V$. This implies that $T^j(x) \in T^{j-i}(U) \cap V \neq \emptyset$. By Proposition 1.2, there exists $y \in X$ such that $\mathscr{O}_T^+(y)$ is dense in X. We are done. $\square$

Next, we discuss a strengthening of topological transitivity.

DEFINITION 1.4. A closed nonempty forward T-invariant subset $Y \subset X$ is a *minimal* set for T if Y contains no proper closed nonempty forward T-invariant subset. If X itself is a minimal set for T, then we say that $T : X \to X$ is a *minimal* topological dynamical system.

Any minimal topological dynamical system is topologically transitive. In the following proposition, we record some useful properties of minimality in the setting of topological dynamical systems defined on compact topological spaces.

PROPOSITION 1.5. *Let X be a compact topological space and $T : X \to X$ a topological dynamical system. The following assertions hold:*

(i) *There exists a minimal set $Y \subset X$ for T.*
(ii) *Let $Y \subset X$ be a closed nonempty forward T-invariant subset. Then Y is minimal for T if and only if $\omega_T(y) = Y$ for every $y \in Y$.*
(iii) *The topological dynamical system $T : X \to X$ is minimal if and only if for every $x \in X$, the positive T-orbit $\mathscr{O}_T^+(x)$ is dense in X.*

PROOF. (i) Denote by $\mathscr{Y}$ the collection of all closed nonempty forward T-invariant subsets $Y \subset X$ with the partial ordering given by inclusion. Then for all $Y_1, Y_2 \in \mathscr{Y}$, we have $Y_1 \leq Y_2$ if and only if $Y_2 \subset Y_1$. Note that $\mathscr{Y}$ is not empty since $X \in \mathscr{Y}$. Let $\mathscr{K} \subset \mathscr{Y}$ be a totally ordered subset. For any nonempty finite subset $\mathscr{F} \subset \mathscr{K}$, we have $\emptyset \neq \bigcap_{Y \in \mathscr{F}} Y \in \mathscr{K}$. Since X is compact, the finite intersection property implies that $\bigcap_{Y \in \mathscr{K}} Y \neq \emptyset$ and so $\bigcap_{Y \in \mathscr{K}} Y$ is un upper bound for $\mathscr{K}$. By Zorn's lemma, $\mathscr{C}$ contains a maximal element $Y \in \mathscr{C}$, which is a minimal set for T.

(ii) Assume that Y is minimal for T. For every $y \in Y$, since $\omega_T(y) \subset X$ is a closed nonempty forward T-invariant set and since $\omega_T(y) \subset Y$, it follows that $\omega_T(y) = Y$. Conversely, assume that for every $y \in Y$, we have $\omega_T(y) = Y$. Let $Z \subset Y$ be a closed nonempty forward T-invariant subset. Choose $z \in Z \subset Y$. Since $\omega_T(z) \subset Z \subset Y$ and $\omega_T(z) = Y$, it follows that $Z = Y$.

(iii) Assume that $T : X \to X$ is minimal. For every $x \in X$, since $\overline{\mathscr{O}_T^+(x)} \subset X$ is a closed nonempty forward T-invariant, we have $\overline{\mathscr{O}_T^+(x)} = X$. Conversely, assume that for every $x \in X$, the positive T-orbit $\mathscr{O}_T^+(x)$ is dense in X. Let $Y \subset X$ be a closed nonempty forward T-invariant subset. Choose $y \in Y$. Since $\mathscr{O}_T^+(y) \subset Y \subset X$ and since $\overline{\mathscr{O}_T^+(y)} = X$, it follows that $Y = X$. □

As an application of Proposition 1.5, we derive Birkhoff's recurrence theorem.

Corollary 1.6. *Let X be a compact topological space and $T : X \to X$ a topological dynamical system. Then there exists a T-recurrent point $x \in X$.*

Remark 1.7. We may also define the notion of a minimal set for a homeomorphism $T : X \to X$. A closed nonempty T-invariant subset $Y \subset X$ is a *minimal* set for T if Y contains no proper closed nonempty T-invariant subset. If X itself is a minimal set for T, then we say that $T : X \to X$ is a *minimal* homeomorphism. If $T : X \to X$ is a homeomorphism and is minimal as a topological dynamical system, then it is minimal as a homeomorphism. The converse does not hold in general (consider the map $T : \mathbb{Z} \to \mathbb{Z} : n \mapsto n+1$). However, when X is a compact topological space and $T : X \to X$ is a homeomorphism, then T is minimal as a topological dynamical system if and only if T is minimal as a homeomorphism.

We say that a subset $A \subset \mathbb{N}$ is *syndetic* if there exists $k > 0$ such that for every $n \in \mathbb{N}$, we have $\{n, n+1, \dots, n+k\} \cap A \neq \emptyset$. For every $k > 0$, observe that the periodic set $\{km \mid m \in \mathbb{N}\}$ is syndetic.

We say that $x \in X$ is T-*almost periodic* if for every neighborhood $U \subset X$ of $x \in X$, the set $A_U = \left\{i \in \mathbb{N} \mid T^i(x) \in U\right\}$ is syndetic. Observe that any T-periodic point is T-almost periodic.

Proposition 1.8. *Let X be a compact topological space, $T : X \to X$ a topological dynamical system and $x \in X$. Then $\overline{\mathscr{O}_T^+(x)}$ is minimal for T if and only if x is T-almost periodic.*

Proof. Assume that x is T-almost periodic. Let $y \in \overline{\mathscr{O}_T^+(x)}$. We need to show that $x \in \overline{\mathscr{O}_T^+(y)}$. Let $U \subset X$ be a neighborhood of $x \in X$. We show that $\mathscr{O}_T^+(y) \cap U \neq \emptyset$. We may choose an open set $U_0 \subset U$ such that $x \in U_0$ and an open set $V \subset X \times X$ such that $\Delta_X = \{(x,x) \mid x \in X\} \subset V$ and such that whenever $x_1 \in U_0$ and $(x_1, x_2) \in V$, we have $x_2 \in U$. Since x is T-almost periodic, we may choose $k > 0$ with the property that for every $n \in \mathbb{N}$, there exists $0 \le j \le k$ such that $T^{n+j}(x) \in U_0$. Set $V_0 = \bigcap_{j=0}^{k}(T \times T)^{-j}(V)$ and observe that $V_0 \subset X \times X$ is an open set such that $\Delta_X \subset V_0$. Choose an open set $W \subset X$ such that $y \in W$ and $W \times W \subset V_0$. Since $y \in \overline{\mathscr{O}_T^+(x)}$, there exists $n \in \mathbb{N}$ such that $T^n(x) \in W$. Choose $0 \le j \le k$ such that $T^{n+j}(x) \in U_0$. Then $(T^n(x), y) \in W \times W \subset V_0$ and so $(T^{n+j}(x), T^j(y)) \in V$. Since $T^{n+j}(x) \in U_0$, we have $T^j(y) \in U$. This shows that $\overline{\mathscr{O}_T^+(x)}$ is minimal for T.

Conversely, assume that x is not T-almost periodic. Then there exists a neighborhood $U \subset X$ of $x \in X$ such that the set $A_U = \{i \in \mathbb{N} \mid T^i(x) \in U\}$ is not syndetic meaning that for every $k > 0$, there exists $n_k \in \mathbb{N}$ such that $\{n_k, n_k + 1, \dots, n_k + k\} \cap A_U = \emptyset$. Choose $y \in \bigcap_{k>0} \overline{\{T^{n_j}(x) \mid j \ge k\}}$ and note that $y \in \overline{\mathscr{O}_T^+(x)}$. Then there exists a net $(k_i)_{i \in I}$ in $\mathbb{N}$ such that $k_i \to \infty$ and $T^{n_{k_i}}(x) \to y$. For every $\ell \in \mathbb{N}$, we have $T^{n_{k_i}+\ell}(x) \to T^\ell(y)$. For every $\ell \in \mathbb{N}$, there exists $i_\ell \in I$ such that for every $i \ge i_\ell$, we have $\ell \le k_i$ and so $T^{n_{k_i}+\ell}(x) \notin U$. This implies that $T^\ell(y) \notin U$ for every $\ell \in \mathbb{N}$ and so $x \notin \overline{\mathscr{O}_T^+(y)}$. This shows that $\overline{\mathscr{O}_T^+(x)}$ is not minimal for T. $\square$

Next, we discuss yet another strengthening of topological transitivity.

Definition 1.9. Let X be a topological space and $T : X \to X$ a topological dynamical system. We say that T is *topologically mixing* if for any nonempty open sets $U, V \subset X$, there exists $n_0 \in \mathbb{N}$ such that for every $n \ge n_0$, we have $T^{-n}(U) \cap V \neq \emptyset$.

By Proposition 1.2, for any second countable topological space X with the Baire property, any topologically mixing dynamical system is topologically transitive. The converse is not true. An (irrational) rotation is not topologically mixing. More generally, for any metric space (X, d) that is not a singleton and any isometry $T : (X, d) \to (X, d)$, the topological dynamical system T is not topologically mixing. Indeed, by contradiction, assume that T is topologically mixing. Let $x_1, x_2 \in X$ such that $x_1 \neq x_2$ and set $\kappa = d(x_1, x_2) > 0$. Define the open sets $V_1 = B(x_1, \frac{1}{4}\kappa)$ and $V_2 = B(x_2, \frac{1}{4}\kappa)$ and set $U = V_1$. Then we can find $n \in \mathbb{N}$ large enough such that $T^{-n}(U) \cap V_1 \neq \emptyset$ and $T^{-n}(U) \cap V_2 \neq \emptyset$. Choose $y_1 \in V_1$ and $y_2 \in V_2$ such that $T^n(y_1), T^n(y_2) \in U = V_1$. Then we have

$$\frac{1}{2}\kappa \le d(y_1, y_2) = d(T^n(y_1), T^n(y_2)) < \frac{1}{2}\kappa.$$

This is a contradiction.

We show that Bernoulli shifts are topologically mixing.

Proposition 1.10. *Let Y be a topological space. Then both the forward Bernoulli shift $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$ and the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$ are topologically mixing.*

Proof. We only prove that the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$ is topological mixing. The proof that the forward Bernoulli shift is topologically mixing is completely analogous. Since the family of cylinder open sets forms a basis of open sets for the product topology, we may assume that U and V are cylinder open sets. Let $n_1, n_2 \in \mathbb{N}$, $(U_j)_{|j| \leq n_1}$ and $(V_k)_{|k| \leq n_2}$ be families of nonempty open sets in Y such that $\mathscr{U} = \mathscr{C}(\{-n_1, \dots, n_1\}, U_{-n_1}, \dots, U_{n_1})$ and $\mathscr{V} = \mathscr{C}(\{-n_2, \dots, n_2\}, V_{-n_2}, \dots, V_{n_2})$. For every $n \geq n_1 + n_2 + 1$, we have

$$T^{-n}(\mathscr{U}) = \mathscr{C}(\{n - n_1, \dots, n + n_1\}, U_{-n_1}, \dots, U_{n_1})$$

and so $T^{-n}(\mathscr{U}) \cap \mathscr{V} \neq \emptyset$. This shows that the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$ is topological mixing. $\square$

3. Applications to combinatorial number theory

The main result of this section is Furstenberg–Weiss' multiple recurrence theorem, which strengthens Birkhoff's recurrence theorem (see Corollary 1.6). Throughout, we assume that X is a compact metrizable space.

THEOREM 1.11 (Furstenberg–Weiss). *Let $T : X \to X$ be a homeomorphism. Then for every $\ell \geq 1$, there exist $x \in X$ and a sequence $(n_k)_k$ in $\mathbb{N}$ such that $n_k \to +\infty$ and for every $1 \leq j \leq \ell$, we have $T^{jn_k}(x) \to x$.*

Before proving Theorem 1.11, let us derive van der Waerden's theorem, which is a Ramsey-type result in combinatorial number theory.

COROLLARY 1.12 (van der Waerden). *Let $r \geq 1$ and $\mathbb{Z} = \bigsqcup_{i=1}^{r} C_i$ be a partition of the integers into r nonempty subsets. Then for every $\ell \geq 1$, there exists $1 \leq j \leq r$ such that C_j contains an arithmetic progression of length $\ell + 1$. In particular, one of the sets C_i contains arbitrarily long arithmetic progressions.*

PROOF. Let $\ell \geq 1$. Set $Y = \{1, \dots, r\}$ and consider the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$. Define the compatible metric $d : Y^{\mathbb{Z}} \times Y^{\mathbb{Z}} \to \mathbb{R}_+$ by the formula $d((y_n)_n, (z_n)_n) = 2^{-k}$ where $k = \min\{|n| \in \mathbb{N} \mid y_n \neq z_n\}$. Observe that for all $(y_n)_n, (z_n)_n \in Y^{\mathbb{Z}}$, we have $d((y_n)_n, (z_n)_n) < 1$ if and only if $y_0 = z_0$.

Define the point $y \in Y^{\mathbb{Z}}$ by the formula $y_n = i$ for every $1 \leq i \leq r$ and every $n \in C_i$, and set $X = \overline{\mathscr{O}_T(y)} \subset Y^{\mathbb{Z}}$. Consider the homeomorphism $T : X \to X$. By Theorem 1.11, there exists $x \in X$ and $n \in \mathbb{N}$ large enough such that for every $1 \leq j \leq \ell$, we have $d(T^{jn}(x), x) < 1$. By definition of $T : X \to X$, this implies that $x_0 = x_n = \cdots = x_{\ell n}$ and we denote this common value by $j \in \{1, \dots, r\}$. Since $X = \overline{\mathscr{O}_T(y)}$, by definition of the metric d, there exists $m \in \mathbb{Z}$ large enough such that $T^m(y)$ and x agree on the interval $[-\ell n, \ell n]$. In particular, we have $y_m = y_{m+n} = \cdots = y_{m+\ell n} = x_0 = j$. By definition of $y \in Y^{\mathbb{Z}}$, this implies that $m, m+n, \dots, m+\ell n \in C_j$. Thus, by pigeon-hole principle, it follows that one the of sets C_i contains arbitrarily long arithmetic progressions. $\square$

Before proving Theorem 1.11, we need to introduce some further terminology. Fix a compatible metric $d : X \times X \to \mathbb{R}_+$. Let $T : X \to X$ be a topological dynamical system and $Y \subset X$ a nonempty closed subset. Denote by $\mathrm{Homeo}_T(X)$ the group of all homeomorphisms of X that commute with T. We say that

- Y is *T-recurrent* if for every $\varepsilon > 0$, there exist $n \geq 1$ and $(x, y) \in Y \times Y$ such that $d(T^n(x), y) < \varepsilon$.
- Y is *T-homogeneous* if there exists a subgroup $G < \mathrm{Homeo}_T(X)$ such that $S(Y) = Y$ for every $S \in G$ and the action $G \curvearrowright Y$ is minimal.

Firstly, we prove a preliminary result on recurrent homogeneous sets.

LEMMA 1.13. *Let $T : X \to X$ be a topological dynamical system and $Y \subset X$ a nonempty T-recurrent T-homogeneous closed subset. Then for every $\varepsilon > 0$, there exist $n \geq 1$ and $z \in Y$ such that $d(T^n(z), z) < \varepsilon$.*

PROOF. Let $Y \subset X$ be a nonempty T-recurrent T-homogeneous closed subset. Denote by $G < \mathrm{Homeo}_T(X)$ the corresponding subgroup. We start by proving the following claim.

CLAIM 1.14. For every $y \in Y$ and every $\varepsilon > 0$, there exist $n \geq 1$ and $x \in Y$ such that $d(T^n(x), y) < \varepsilon$.

Indeed, consider the subset $Z \subset Y$ of all the elements $z \in Y$ with the property that for every $\varepsilon > 0$, there exist $n \geq 1$ and $x \in Y$ such that $d(T^n(x), z) < \varepsilon$. Since $d(T^n(S(x)), S(z)) = d(S(T^n(x)), S(z))$ for every $S \in G$ and every $n \in \mathbb{N}$, it follows that $Z \subset Y$ is a closed G-invariant subset. Since Y is T-recurrent, we may find sequences $(r_n)_n$ in $\mathbb{N}^*$ and $(x_n)_n, (y_n)_n$ in Y such that $\lim_n d(T^{r_n}(x_n), y_n) = 0$. Since Y is compact, upon passing to a subsequence, we may assume that there exists $y \in Y$ such that $y_n \to y$. Then $y \in Z$ and so $Z \neq \emptyset$. Since Y is T-homogeneous, it follows that $Z = Y$. This finishes the proof of Claim 1.14.

Let $\varepsilon > 0$ and fix $z_0 \in Y$. Letting $\varepsilon_1 = \frac{\varepsilon}{2}$, using Claim 1.14, there exist $n_1 \geq 1$ and $z_1 \in Y$ such that $d(T^{n_1}(z_1), z_0) < \varepsilon_1$. By continuity, there exists $\varepsilon_2 < \varepsilon_1$ such that for every $z \in Y$ that satisfies $d(z, z_1) < \varepsilon_2$, we have $d(T^{n_1}(z), z_0) < \varepsilon_1$. Using Claim 1.14, there exist $n_2 \geq 1$ and $z_2 \in Y$ such that $d(T^{n_2}(z_2), z_1) < \varepsilon_2$. Then we also have $d(T^{n_2+n_1}(z_2), z_0) < \varepsilon_1$. Proceeding by induction, if $d(T^{n_k}(z_k), z_{k-1}) < \varepsilon_k$ for $k \geq 1$, by continuity, there exists $\varepsilon_{k+1} < \varepsilon_k$ such that for every $z \in Y$ that satisfies $d(z, z_k) < \varepsilon_{k+1}$, we have $d(T^{n_k}(z), z_{k-1}) < \varepsilon_k$. Using Claim 1.14, there exist $n_{k+1} \geq 1$ and $z_{k+1} \in Y$ such that $d(T^{n_{k+1}}(z_{k+1}), z_k) < \varepsilon_{k+1}$. Then for all $i < j$, we have $d(T^{n_j+\cdots+n_{i+1}}(z_j), z_i) < \varepsilon_{i+1}$. Since Y is compact, there exist $i < j$ such that $d(z_j, z_i) < \varepsilon_1$. Then we obtain

$$d(T^{n_j+\cdots+n_{i+1}}(z_j), z_j) \leq d(T^{n_j+\cdots+n_{i+1}}(z_j), z_i) + d(z_i, z_j) < \varepsilon_{i+1} + \varepsilon_1 \leq \varepsilon.$$

Letting $z = z_j$, we are done. $\square$

Secondly, we prove the following key result regarding the existence of recurrent points.

THEOREM 1.15. *Let $T : X \to X$ be a topological dynamical system and $Y \subset X$ a nonempty T-recurrent T-homogeneous closed subset. Then there exists a T-recurrent point $y \in Y$.*

PROOF. Let $Y \subset X$ be a nonempty T-recurrent T-homogeneous closed subset. Denote by $G < \mathrm{Homeo}_T(X)$ the corresponding subgroup. Define the function $F : X \to \mathbb{R}_+ : x \mapsto \inf_{n\geq 1} d(T^n(x), x)$. Observe that $x \in X$ is T-recurrent if and only if $F(x) = 0$. By Lemma 1.13, we have $\inf_{x\in Y} F(x) = 0$. By construction, the function F is upper semicontinuous in the sense that for every $x \in X$ and every sequence $(x_n)_n$ in X such that $x_n \to x$, we have $\limsup_n F(x_n) \leq F(x)$. Consider the restriction $F|_Y : Y \to \mathbb{R}_+$. We claim that $F|_Y$ has a point of continuity. Indeed, denote by $\mathscr{D}$ the set of all $y \in Y$ where $F|_Y$ is not continuous at y. If $y \in \mathscr{D}$, then there exist $r \in \mathbb{Q}$ and a sequence $(y_n)_n$ in Y such that $y_n \to y$ and $F(y_n) < r < F(y)$ for every $n \in \mathbb{N}$. Define the closed subset $\mathscr{F}_r = F^{-1}([r, +\infty)) \cap Y$. Then we have $y \in \partial\mathscr{F}_r = \mathscr{F}_r \setminus \mathrm{Int}(\mathscr{F}_r)$. This shows that $\mathscr{D} \subset \bigcup_{r\in\mathbb{Q}} \partial\mathscr{F}_r$. Since Y has the Baire property and since $\partial\mathscr{F}_r$ has empty interior for every $r \in \mathbb{Q}$, it follows that $\mathscr{D}$ has empty interior and so $Y \setminus \mathscr{D}$ is not empty.

Let $y \in Y$ be a point of continuity for $F|_Y$. We claim that $F(y) = 0$. By contradiction, assume that $F(y) > 0$. Then there exist $\varepsilon > 0$ and an open set $U \subset Y$ such that $y \in U$ and $F(x) \geq \varepsilon$ for every $x \in U$. Since Y is T-homogeneous, the nonempty G-invariant open subset $\bigcup_{S\in G} S(U)$ is necessarily equal to Y. By compactness, there exist $S_1, \dots, S_r \in G$ such that $\bigcup_{i=1}^r S_i(U) = Y$. We may choose $\delta > 0$ such that for all $z_1, z_2 \in X$, if $d(z_1, z_2) < \delta$, then $d(S_i^{-1}(z_1), S_i^{-1}(z_2)) < \varepsilon$ for every $1 \leq i \leq r$. This further implies that for every $x \in X$ and every $1 \leq i \leq r$, if $F(x) < \delta$, then $F(S_i^{-1}(x)) = \inf_{n\geq 1} d(S_i^{-1}(T^n(x)), S_i^{-1}(x)) < \varepsilon$. By the choice of $\varepsilon > 0$ and since $Y = \bigcup_{i=1}^r S_i(U)$, it follows that $F(x) \geq \delta$ for every $x \in Y$. This contradicts the fact that $\inf_{x\in Y} F(x) = 0$. Therefore, we have $F(y) = 0$ and so $y \in Y$ is a T-recurrent point. $\square$

We are now ready to prove Theorem 1.11. We will actually prove the following slightly more general result which implies Theorem 1.11.

THEOREM 1.16. *Let $\ell \geq 1$ and $(T_j : X \to X)_{1\leq j\leq \ell}$ be a family of pairwise commuting homeomorphisms. Then there exist $x \in X$ and a sequence $(n_k)_k$ in $\mathbb{N}$ such that $n_k \to +\infty$ and for every $1 \leq j \leq \ell$, we have $T_j^{n_k}(x) \to x$.*

PROOF. We proceed by induction on $\ell \geq 1$. For $\ell = 1$, the result follows from Birkhoff's recurrence theorem (see Corollary 1.6). Assume that the result holds for $\ell \geq 1$ and let us prove that it holds for $\ell + 1$.

Let $(T_j : X \to X)_{1\leq j\leq \ell+1}$ be a family of pairwise commuting homeomorphisms. Denote by G the abelian group generated by $T_1, \dots, T_{\ell+1}$. Upon passing to a nonempty closed G-invariant subset, we may assume that the action $G \curvearrowright X$ is minimal. Set $\widehat{X} = X^{\ell+1}$, $\Delta = \{(x, \dots, x) \in \widehat{X} \mid x \in X\}$ and for every $S \in G$, set $\widehat{S} = \widehat{X} \to \widehat{X} : (x_j)_j \mapsto (S(x_j))_j$. Then we may consider the natural action $G \curvearrowright \widehat{X}$. Observe that $\Delta \subset \widehat{X}$ is a closed G-invariant subset and that the action $G \curvearrowright \Delta$ is minimal. Define $\widehat{T} = T_1 \times \cdots \times T_{\ell+1} : \widehat{X} \to \widehat{X}$. For every $S \in G$, we have $\widehat{S} \circ \widehat{T} = \widehat{T} \circ \widehat{S}$. Thus, the closed subset $\Delta \subset \widehat{X}$ is $\widehat{T}$-homogeneous. Next, we show that $\Delta \subset \widehat{X}$ is $\widehat{T}$-recurrent. Applying the induction hypothesis to the family $(S_j = T_j \circ T_{\ell+1}^{-1} : X \to X)_{1 \leq j \leq \ell}$, there exist $x \in X$ and a sequence $(n_k)_k$ in $\mathbb{N}$ such that $n_k \to +\infty$ and for every $1 \leq j \leq \ell$, we have $S_j^{n_k}(x) \to x$. Therefore, for every $\varepsilon > 0$, there exists $n \geq 1$ such that the points $(x, \dots, x) \in \Delta \subset \widehat{X}$ and $\widehat{T}^n(T_{\ell+1}^{-n}(x), \dots, T_{\ell+1}^{-n}(x)) \in \widehat{X}$ are within distance ε of one another. Since $(x, \dots, x) \in \Delta$ and $(T_{\ell+1}^{-n}(x), \dots, T_{\ell+1}^{-n}(x)) \in \Delta$, it follows that the closed subset $\Delta \subset \widehat{X}$ is $\widehat{T}$-recurrent. By Theorem 1.15, there exists a $\widehat{T}$-recurrent point $(x, \dots, x) \in \Delta$. This implies that the result holds for $\ell+1$ and finishes the proof of Theorem 1.16. ☐

4. Topological entropy

In this section, we introduce the notion of *topological entropy*. It is a topological invariant that measures the complexity of the orbit structure of a dynamical system. Topological entropy is analogous to measure entropy we will introduce in the next chapter. Throughout, we assume that (X, d) is a compact metric space. Fix a topological dynamical system $T : X \to X$.

Firstly, we recall the following elementary lemma on subadditive sequences.

LEMMA 1.17 (Fekete). *Let $(a_n)_{n \geq 1}$ be a subadditive sequence in $\mathbb{R}$, meaning that $a_{m+n} \leq a_m + a_n$ for all $m, n \geq 1$. Then the sequence $(\frac{a_n}{n})_n$ is convergent in $[-\infty, +\infty)$ and we have*

$$\lim_n \frac{a_n}{n} = \inf_{n \geq 1} \frac{a_n}{n}.$$

PROOF. Set $\ell = \inf_{n \geq 1} \frac{a_n}{n}$. Let $\ell_0 > \ell$ and choose $m \geq 1$ such that $\frac{a_m}{m} < \ell_0$. Set $a_0 = 0$. For every $n \geq m$, write $n = qm + r$ with $q \geq 1$ and $0 \leq r < m$. By subadditivity, we have

$$\frac{a_n}{n} \leq \frac{q a_m}{n} + \frac{a_r}{n} \leq \frac{a_m}{m} + \frac{1}{n} \max\{a_0, \dots, a_{m-1}\}.$$

Next, choose $n_0 \geq m$ large enough such that for all $n \geq n_0$, we have $\frac{1}{n} \max\{a_0, \dots, a_{m-1}\} \leq \ell_0 - \frac{a_m}{m}$. Then for all $n \geq n_0$, we have $\ell \leq \frac{a_n}{n} \leq \ell_0$. Therefore, $\lim_n \frac{a_n}{n} = \ell$. ☐

We define the topological entropy of $T : X \to X$ using open covers. We say that a set $\mathscr{U}$ consisting of open subsets of X is an *open cover* if $X = \bigcup_{U \in \mathscr{U}} U$. For every open cover $\mathscr{U}$ and every $j \in \mathbb{N}$, define the open cover $T^{-j}(\mathscr{U}) = \{T^{-j}(U) \mid U \in \mathscr{U}\}$. For all open covers $\mathscr{U}, \mathscr{V}$, define the *join* open cover $\mathscr{U} \vee \mathscr{V} = \{U \cap V \mid U \in \mathscr{U}, V \in \mathscr{V}\}$. For every open cover $\mathscr{U}$, using compactness, we may define $N(\mathscr{U}) \in \mathbb{N}^*$ as the minimum cardinality of a finite subcover. It is plain to see that for all open covers $\mathscr{U}, \mathscr{V}$, we have $N(\mathscr{U} \vee \mathscr{V}) \leq N(\mathscr{U})N(\mathscr{V})$. For every $n \geq 1$, define the open cover $\mathscr{U}_n = \mathscr{U} \vee \cdots \vee T^{-n+1}(\mathscr{U})$. Then the sequence $(\log(N(\mathscr{U}_n)))_n$ is subadditive and by using Lemma 1.17, we may define

$$h_{\mathrm{top}}(T, \mathscr{U}) = \lim_n \frac{1}{n} \log(N(\mathscr{U}_n)) = \inf_{n \geq 1} \frac{1}{n} \log(N(\mathscr{U}_n)) \geq 0.$$

The topological entropy of $T : X \to X$ is defined by the formula

$$h_{\mathrm{top}}(T) = \sup \{h_{\mathrm{top}}(T, \mathscr{U}) \mid \mathscr{U} \text{ open cover}\}.$$

Next, we define the topological entropy of $T : X \to X$ as the exponential growth rate of the number of essentially different orbit segments of length $n \in \mathbb{N}$. We show that the two notions coincide. This second definition of topological entropy will be very convenient to work with. For every $n \geq 1$, define the metric $d_n : X \times X \to \mathbb{R}_+ : (x, y) \mapsto \max\{d(T^k(x), T^k(y)) \mid 0 \leq k \leq n-1\}$. For every $(x, y) \in X \times X$ and every $n \geq 1$, $d_n(x, y)$ measures the maximum distance between the first n iterates of x and y. We claim that d is equivalent to d_n for every $n \geq 1$. Firstly, observe that $d(x, y) \leq d_n(x, y)$ for every $n \geq 1$ and every $(x, y) \in X \times X$. Secondly, let $(x_j)_{j \in \mathbb{N}}$ be a sequence in X and $x \in X$ such that $\lim_j d(x_j, x) = 0$. Fix $n \geq 1$. For every $0 \leq k \leq n-1$, since $T^k : X \to X$ is continuous, we have $\lim_j d(T^k(x_j), T^k(x)) = 0$ and so $\lim_j d_n(x_j, x) = 0$. Thus, for every $n \geq 1$, the metrics d and d_n induce the same topology on X. We denote by $B_n(x, r) = \{y \in X \mid d_n(x, y) < r\}$ the open ball, with respect to the metric d_n, of center $x \in X$ and radius $r > 0$ in X.

Fix $n \geq 1$ and $\varepsilon > 0$. A subset $\mathscr{F} \subset X$ is said to be (n, ε)-*spanning* if $X = \bigcup_{x \in \mathscr{F}} B_n(x, \varepsilon)$. By compactness, there exists a finite (n, ε)-spanning subset $\mathscr{F} \subset X$. We then denote by $\mathrm{span}(n, \varepsilon, T)$ the minimum cardinality of an (n, ε)-spanning set. Likewise, we denote by $\mathrm{cov}(n, \varepsilon, T)$ the minimum cardinality of a covering of X by sets of d_n-diameter less than ε. Again by compactness, we have $\mathrm{cov}(n, \varepsilon, T) < +\infty$. A subset $\mathscr{F} \subset X$ is said to be (n, ε)-*separated* if for any $x, y \in \mathscr{F}$ such that $x \neq y$, we have $d_n(x, y) \geq \varepsilon$. By compactness, any (n, ε)-separating set is finite.

LEMMA 1.18. *Keep the same notation as above. Let $n \geq 1$ and $\varepsilon > 0$. The following assertions hold:*

(i) *For every (n, ε)-separating set $\mathscr{F} \subset X$, we have $|\mathscr{F}| \leq \mathrm{cov}(n, \varepsilon, T)$. We may then denote by $\mathrm{sep}(n, \varepsilon, T)$ the maximum cardinality of an (n, ε)-separating set.*

(ii) *We have*

$$\mathrm{cov}(n, 2\varepsilon, T) \leq \mathrm{span}(n, \varepsilon, T) \leq \mathrm{sep}(n, \varepsilon, T) \leq \mathrm{cov}(n, \varepsilon, T).$$

PROOF. (i) Assume that $\mathscr{F}$ is an (n, ε)-separating set and $\mathscr{V}$ is a covering of X of minimal cardinality by sets of d_n-diameter less than ε. For every $x \in \mathscr{F}$, denote by $\mathscr{V}_x$ the nonempty set of all sets $V \in \mathscr{V}$ such that $x \in V$.

Since $\mathscr{F}$ is an (n,ε)-separating set and since each $V \in \mathscr{V}$ has d_n-diameter less than ε, it follows that the sets $(\mathscr{V}_x)_{x\in\mathscr{F}}$ are pairwise disjoint. Therefore, we have $|\mathscr{F}| = \sum_{x\in\mathscr{F}} 1 \leq \sum_{x\in\mathscr{F}} |\mathscr{V}_x| \leq |\mathscr{V}| = \operatorname{cov}(n,\varepsilon,T)$. Then we have $\operatorname{sep}(n,\varepsilon,T) \leq \operatorname{cov}(n,\varepsilon,T)$.

(ii) Assume that $\mathscr{F}$ is an (n,ε)-spanning set of minimal cardinality. Then we have $X = \bigcup_{x\in\mathscr{F}} B_n(x,\varepsilon)$. By compactness, there exists $\delta < \varepsilon$ such that we still have $X = \bigcup_{x\in\mathscr{F}} B_n(x,\delta)$. Then $(B_n(x,\delta))_{x\in\mathscr{F}}$ is a covering of X by subsets of d_n-diameter at most $2\delta < 2\varepsilon$. It follows that $\operatorname{cov}(n,2\varepsilon,T) \leq \operatorname{span}(n,\varepsilon,T)$.

Assume that $\mathscr{G}$ is an (n,ε)-separating set of maximal cardinality. Then by maximality, we have $X = \bigcup_{x\in\mathscr{G}} B_n(x,\varepsilon)$ and so $\mathscr{G}$ is an (n,ε)-spanning set. It follows that $\operatorname{span}(n,\varepsilon,T) \leq \operatorname{sep}(n,\varepsilon,T)$. □

Lemma 1.19. *For every $\varepsilon > 0$, the sequence $(\log(\operatorname{cov}(n,\varepsilon,T)))_n$ is subadditive.*

Proof. Let $m,n \geq 1$ and $\varepsilon > 0$. Let $\mathscr{U}$ (resp. $\mathscr{V}$) be a finite covering of X of minimal cardinality by elements of d_m-diameter (resp. d_n-diameter) less than ε. Then $\mathscr{W} = \{U \cap T^{-m}(V) \mid U \in \mathscr{U}, V \in \mathscr{V}\}$ is a covering of X. Moreover, for every $U \in \mathscr{U}$ and every $V \in \mathscr{V}$, the set $U \cap T^{-m}(V)$ has d_{m+n}-diameter less than ε. Thus, we infer that

$$\operatorname{cov}(m+n,\varepsilon,T) \leq \operatorname{cov}(m,\varepsilon,T)\operatorname{cov}(n,\varepsilon,T).$$

This finishes the proof of the lemma. □

A combination of Lemmas 1.17 and 1.19 implies that the quantity

$$h_\varepsilon(T) = \lim_n \frac{1}{n}\log(\operatorname{cov}(n,\varepsilon,T)) = \inf_{n\geq 1} \frac{1}{n}\log(\operatorname{cov}(n,\varepsilon,T)) \geq 0$$

exists and is finite. Moreover, for every $n \geq 1$, the function $(0,+\infty) \to [0,+\infty) : \varepsilon \mapsto \log(\operatorname{cov}(n,\varepsilon,T))$ is non-increasing. This implies that the function $(0,+\infty) \to [0,+\infty) : \varepsilon \mapsto h_\varepsilon(T)$ is non-increasing. Then we may define the *entropy* of the topological dynamical system $T : X \to X$ by the formula

$$h(T) = \lim_{\varepsilon\to 0^+} h_\varepsilon(T) = \sup_{\varepsilon>0} h_\varepsilon(T) \in [0,+\infty].$$

It follows from Lemma 1.18 that

$$\begin{aligned} h(T) &= \lim_{\varepsilon\to 0^+} \lim_n \frac{1}{n}\log(\operatorname{cov}(n,\varepsilon,T)) \\ &= \lim_{\varepsilon\to 0^+} \limsup_n \frac{1}{n}\log(\operatorname{span}(n,\varepsilon,T)) = \lim_{\varepsilon\to 0^+} \liminf_n \frac{1}{n}\log(\operatorname{span}(n,\varepsilon,T)) \\ &= \lim_{\varepsilon\to 0^+} \limsup_n \frac{1}{n}\log(\operatorname{sep}(n,\varepsilon,T)) = \lim_{\varepsilon\to 0^+} \liminf_n \frac{1}{n}\log(\operatorname{sep}(n,\varepsilon,T)). \end{aligned}$$

We now prove that the two notions of topological entropy coincide.

Theorem 1.20. *For every topological dynamical system $T : X \to X$, we have $h_{\mathrm{top}}(T) = h(T)$.*

PROOF. Firstly, we show that $h(T) \leq h_{\text{top}}(T)$. Let $\varepsilon > 0$ and $n \geq 1$. Let $\mathscr{U}$ be an open cover whose all elements have d-diameter less than ε. Then $\mathscr{U}_n$ is an open cover whose all elements have d_n-diameter less than ε and so $\operatorname{cov}(n, \varepsilon, T) \leq N(\mathscr{U}_n)$. This implies that

$$\begin{aligned} h_\varepsilon(T) &= \lim_n \frac{1}{n} \log(\operatorname{cov}(n, \varepsilon, T)) \\ &\leq \lim_n \frac{1}{n} \log(N(\mathscr{U}_n)) = h_{\text{top}}(T, \mathscr{U}) \leq h_{\text{top}}(T). \end{aligned}$$

By taking the limit as $\varepsilon \to 0^+$, this further implies that $h(T) \leq h_{\text{top}}(T)$.

Secondly, we show that $h_{\text{top}}(T) \leq h(T)$. Let $\mathscr{U}$ be an open cover of X. We may then choose $\varepsilon > 0$ such that the Lebesgue number of $\mathscr{U}$ with respect to d is at least 2ε, that is, for all $x \in X$, there exists $U_x \in \mathscr{U}$ such that $B(x, \varepsilon) \subset U_x$. For every $n \geq 1$, $\mathscr{U}_n$ is an open cover of X whose Lebesgue number with respect to d_n is at least 2ε. Choose an (n, ε)-spanning set $\mathscr{F} \subset X$ of minimal cardinality. For every $x \in \mathscr{F}$, we may choose $U_{n,x} \in \mathscr{U}_n$ such that $B_n(x, \varepsilon) \subset U_{n,x}$. For every $y \in X$, there exists $x \in \mathscr{F}$ such that $d_n(x, y) < \varepsilon$. It follows that $(U_{n,x})_{x \in \mathscr{F}}$ is an open subcover of X and so $N(\mathscr{U}_n) \leq |\mathscr{F}| = \operatorname{span}(n, \varepsilon, T)$. This implies that

$$\begin{aligned} h_{\text{top}}(T, \mathscr{U}) &= \lim_n \frac{1}{n} \log(N(\mathscr{U}_n)) \\ &\leq \limsup_n \frac{1}{n} \log(\operatorname{span}(n, \varepsilon, T)) \leq h_\varepsilon(T) \leq h(T). \end{aligned}$$

By taking the supremum over all open covers of X, this further implies that $h_{\text{top}}(T) \leq h(T)$. This finishes the proof. □

Theorem 1.20 implies that the topological entropy $h(T) = h_{\text{top}}(T)$ of the topological dynamical system $T : X \to X$ only depends on the topology on X and does not depend on the compatible metric d on X. In particular, the topological entropy is an invariant of topological conjugacy.

EXAMPLE 1.21. Let $T : (X, d) \to (X, d)$ be an isometry. Then we have $h(T) = 0$. Indeed, let $\varepsilon > 0$. For every $n \geq 1$, we have $d_n = d$ and so $\operatorname{cov}(n, \varepsilon, T) = \operatorname{cov}(1, \varepsilon, T)$. Then $h_\varepsilon(T) = \lim_n \frac{1}{n} \log(\operatorname{cov}(n, \varepsilon, T)) = 0$. Therefore, we have $h(T) = \lim_{\varepsilon \to 0^+} h_\varepsilon(T) = 0$.

We collect some useful properties of topological entropy.

PROPOSITION 1.22. *Let (X, d) be a compact metric space and $T : X \to X$ a topological dynamical system.*

(i) *For every $m \in \mathbb{N}$, we have $h(T^m) = m\, h(T)$.*
(ii) *If $T : X \to X$ is a homeomorphism, then $h(T^{-1}) = h(T)$. Thus, for every $m \in \mathbb{Z}$, we have $h(T^m) = |m|\, h(T)$.*

For every $i \in \{1, 2\}$, let (X_i, d^i) be a compact metric space and $T_i : X_i \to X_i$ a topological dynamical system.

(iii) *We have $h(T_1 \times T_2) = h(T_1) + h(T_2)$.*

(iv) *If* $T_2 : X_2 \to X_2$ *is a topological factor of* $T_1 : X_1 \to X_1$*, then we have* $h(T_2) \le h(T_1)$*.*

Proof. (i) Let $m \in \mathbb{N}$, $n \ge 1$ and $\varepsilon > 0$. For all $x, y \in X$, we have

$$\max_{0 \le i \le n-1} \left\{ d(T^{mi}(x), T^{mi}(y)) \right\} \le \max_{0 \le j \le mn-1} \left\{ d(T^{j}(x), T^{j}(x)) \right\}.$$

Let $Y \subset X$ be an (mn, ε)-spanning set for T of minimal cardinality. Then $Y \subset X$ is also an (n, ε)-spanning set for T^m. Thus, $\operatorname{span}(n, \varepsilon, T^m) \le \operatorname{span}(mn, \varepsilon, T)$. This further implies that

$$\begin{aligned} h(T^m) &= \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{span}(n, \varepsilon, T^m)) \\ &\le m \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{mn} \log(\operatorname{span}(mn, \varepsilon, T)) \le m\, h(T). \end{aligned}$$

Conversely, for every $\varepsilon > 0$, set

$$\delta(\varepsilon) = \sup \left\{ d(T^i(x), T^i(y)) \mid (x, y) \in X \times X,\, 0 \le i \le m-1,\, d(x, y) \le \varepsilon \right\}.$$

Then we have $\delta(\varepsilon) \to 0$ when $\varepsilon \to 0^+$. Let $Z \subset X$ be an (n, ε)-spanning set for T^m of minimal cardinality. Then by construction, $Z \subset X$ is also an $(mn, \delta(\varepsilon))$-spanning set for T. Thus, $\operatorname{span}(mn, \delta(\varepsilon), T) \le \operatorname{span}(n, \varepsilon, T^m)$. This further implies that

$$\begin{aligned} m\, h(T) &\le m \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{mn} \log(\operatorname{span}(mn, \delta(\varepsilon), T)) \\ &\le \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{n} \log(\operatorname{span}(n, \varepsilon, T^m)) = h(T^m). \end{aligned}$$

Therefore, we have $h(T^m) = m\, h(T)$.

(ii) Assume that $T : X \to X$ is a homeomorphism. For every $n \ge 1$ and every $\varepsilon > 0$, $Y \subset X$ is an (n, ε)-separating set for T if and only if $T^{(n-1)}(Y) \subset X$ is an (n, ε)-separating set for T^{-1}. Thus, $\operatorname{sep}(n, \varepsilon, T) = \operatorname{sep}(n, \varepsilon, T^{-1})$. This further implies that

$$\begin{aligned} h(T) &= \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{sep}(n, \varepsilon, T)) \\ &= \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{sep}(n, \varepsilon, T^{-1})) = h(T^{-1}). \end{aligned}$$

For every $i \in \{1, 2\}$, let (X_i, d^i) be a compact metric space and $T_i : X_i \to X_i$ a topological dynamical system.

(iii) Set $X = X_1 \times X_2$ and $T = T_1 \times T_2 : X \to X$. Define the compatible metric $d : X \times X \to \mathbb{R}_+$ by the formula

$$\forall x = (x_1, x_2), y = (y_1, y_2) \in X, \quad d(x, y) = \max \left\{ d^1(x_1, y_1), d^2(x_2, y_2) \right\}.$$

Moreover, for every $n \ge 1$, we have

$$\forall x = (x_1, x_2), y = (y_1, y_2) \in X, \quad d_n(x, y) = \max \left\{ d^1_n(x_1, y_1), d^2_n(x_2, y_2) \right\}.$$

Let $n \geq 1$ and $\varepsilon > 0$. If $U_1 \subset X_1$ has d^1_n-diameter less than ε and $U_2 \subset X_2$ has d^2_n-diameter less than ε, then $U = U_1 \times U_2 \subset X$ has d_n-diameter less than ε. This implies that

$$\operatorname{cov}(n, \varepsilon, T) \leq \operatorname{cov}(n, \varepsilon, T_1) \operatorname{cov}(n, \varepsilon, T_2).$$

This further implies that

$$\begin{aligned}
h(T) &= \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{cov}(n, \varepsilon, T)) \\
&\leq \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \left(\log(\operatorname{cov}(n, \varepsilon, T_1)) + \log(\operatorname{cov}(n, \varepsilon, T_2))\right) \\
&\leq \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{cov}(n, \varepsilon, T_1)) + \lim_{\varepsilon \to 0^+} \limsup_n \frac{1}{n} \log(\operatorname{cov}(n, \varepsilon, T_2)) \\
&= h(T_1) + h(T_2).
\end{aligned}$$

Conversely, let $n \geq 1$ and $\varepsilon > 0$. If $Y_1 \subset X_1$ is an (n, ε)-separating set for T_1 and $Y_2 \subset X_2$ is an (n, ε)-separating set for T_2, then $Y = Y_1 \times Y_2 \subset X$ is an (n, ε)-separating set for T. This implies that

$$\operatorname{sep}(n, \varepsilon, T_1) \operatorname{sep}(n, \varepsilon, T_2) \leq \operatorname{sep}(n, \varepsilon, T).$$

This further implies that

$$\begin{aligned}
&h(T_1) + h(T_2) \\
&= \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{n} \log(\operatorname{sep}(n, \varepsilon, T_1)) + \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{n} \log(\operatorname{sep}(n, \varepsilon, T_2)) \\
&\leq \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{n} \left(\log(\operatorname{sep}(n, \varepsilon, T_1)) + \log(\operatorname{sep}(n, \varepsilon, T_2))\right) \\
&\leq \lim_{\varepsilon \to 0^+} \liminf_n \frac{1}{n} \log(\operatorname{sep}(n, \varepsilon, T)) \\
&= h(T).
\end{aligned}$$

Therefore, we have $h(T) = h(T_1) + h(T_2)$.

(iv) Assume that $T_2 : X_2 \to X_2$ is a topological factor of $T_1 : X_1 \to X_1$. Then there exists a surjective continuous map $\pi : X_1 \to X_2$ such that $T_2 \circ \pi = \pi \circ T_1$. For every $\varepsilon > 0$, set

$$\delta(\varepsilon) = \sup \left\{ d^2(\pi(x), \pi(y)) \mid (x, y) \in X_1 \times X_1,\, d^1(x, y) \leq \varepsilon \right\}.$$

Then we have $\delta(\varepsilon) \to 0$ when $\varepsilon \to 0^+$. Choose a decreasing sequence $(\varepsilon_k)_k$ in $\mathbb{R}^*_+$ such that $(\delta(\varepsilon_k))_k$ is decreasing and $\lim_k \varepsilon_k = \lim_k \delta(\varepsilon_k) = 0$. For every $n \geq 1$, choose a covering $\mathscr{U}_{n,k}$ of X_1 of minimal cardinality by elements with d^1_n-diameter less than ε_{k+1}. Since $\pi : X_1 \to X_2$ is surjective and since $X_1 = \bigcup_{U \in \mathscr{U}_{n,k}} U$, we have $X_2 = \pi(X_1) = \bigcup_{U \in \mathscr{U}_{n,k}} \pi(U)$. Thus, $\pi(\mathscr{U}_{k,n})$ is a covering of X_2. Moreover by construction, for every $U \in \mathscr{U}_{k,n}$, the d^2_n-diameter of $\pi(U)$ is at most $\delta(\varepsilon_{k+1}) < \delta(\varepsilon_k)$. This implies that $\operatorname{cov}(n, \delta(\varepsilon_k), T_2) \leq \operatorname{cov}(n, \varepsilon_{k+1}, T_1)$. This further implies that

$$h(T_2) = \lim_k \lim_n \frac{1}{n} \log(\operatorname{cov}(n, \delta(\varepsilon_k), T_2))$$

$$\leq \lim_k \lim_n \frac{1}{n} \log(\text{cov}(n, \varepsilon_{k+1}, T_1)) = h(T_1).$$

This finishes the proof of the proposition. □

Next, we plan to compute the topological entropy of the (forward) Bernoulli shift. Before doing so, we introduce the following terminology. Let $T : X \to X$ be a topological dynamical system. We say that T is (positively) *expansive* if there exists $\kappa > 0$ such that whenever $x, y \in X$ and $x \neq y$, there exists $n \in \mathbb{N}$ for which $d(T^n(x), T^n(y)) \geq \kappa$. Assume moreover that $T : X \to X$ is a homeomorphism. We say that T is *expansive* if there exists $\kappa > 0$ such that whenever $x, y \in X$ and $x \neq y$, there exists $n \in \mathbb{Z}$ for which $d(T^n(x), T^n(y)) \geq \kappa$. We call $\kappa > 0$ a constant of *expansiveness*.

Observe that the notion of expansiveness does not depend on the choice of the compatible metric d on X. Indeed, let d and ρ be compatible metrics on X and assume that $T : X \to X$ is expansive with respect to the metric d with constant of expansiveness $\kappa > 0$. We claim that $T : X \to X$ is also expansive with respect to the metric ρ. Indeed, for every $\varepsilon > 0$, set

$$\delta(\varepsilon) = \sup \{d(x, y) \mid (x, y) \in X \times X,\ \rho(x, y) \leq \varepsilon\}.$$

By compactness and since ρ and d are equivalent compatible metrics on X, we have $\delta(\varepsilon) \to 0$ when $\varepsilon \to 0^+$. Choose $\varepsilon > 0$ small enough so that $\delta(\varepsilon) < \kappa$. For all $x, y \in X$ such that $x \neq y$, there exists $n \in \mathbb{N}$ such that $d(T^n(x), T^n(y)) \geq \kappa > \delta(\varepsilon)$ and so $\rho(T^n(x), T^n(y)) > \varepsilon$. This shows that $T : X \to X$ is expansive with respect to the metric ρ with constant of expansiveness $\varepsilon > 0$.

Example 1.23. Let $r \geq 2$. Set $Y = \{1, \ldots, r\}$ and consider the forward Bernoulli shift $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$. Define the compatible metric $d : Y^{\mathbb{N}} \times Y^{\mathbb{N}} \to \mathbb{R}_+$ by the formula $d((y_n)_n, (z_n)_n) = 2^{-k}$ where $k = \min\{n \in \mathbb{N} \mid y_n \neq z_n\}$. Then S is expansive with constant of expansiveness $\kappa = 1$. Likewise, the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$ is expansive with constant of expansiveness $\kappa = 1$.

The expansiveness property turns out to be useful when computing the topological entropy.

Proposition 1.24. *Let $T : X \to X$ be an expansive topological dynamical system with constant of expansiveness $\kappa > 0$. Then for every $0 < \varepsilon < \kappa$, we have $h_\varepsilon(T) = h(T)$.*

The same statement holds true for homeomorphisms.

Proof. Fix $0 < \gamma < \varepsilon < \kappa$. We show that $h_{2\gamma}(T) = h_\varepsilon(T)$. By monotonicity, it suffices to prove that $h_{2\gamma}(T) \leq h_\varepsilon(T)$. This clearly implies the statement of the proposition.

By expansiveness, for all $x, y \in X$ such that $x \neq y$, there exists $n \in \mathbb{N}$ for which $d(T^n(x), T^n(y)) \geq \kappa > \varepsilon$. Since the set $\{(x, y) \in X \times X \mid d(x, y) \geq \gamma\}$ is compact, there exists $k = k(\gamma, \varepsilon) \in \mathbb{N}$ such that if $d(x, y) \geq \gamma$, then

$d(T^j(x), T^j(y)) > \varepsilon$ for some $0 \le j \le k$. It follows that if $Y \subset X$ is an (n, γ)-separating set, then Y is an $(n+k, \varepsilon)$-separating set. Using lemma 1.18, this implies that

$$\mathrm{cov}(n, 2\gamma, T) \le \mathrm{sep}(n, \gamma, T) \le \mathrm{sep}(n+k, \varepsilon, T) \le \mathrm{cov}(n+k, \varepsilon, T)$$

and so $h_{2\gamma}(T) \le h_\varepsilon(T)$. This finishes the proof of the proposition. □

We can now compute the topological entropy of Bernoulli shifts.

Proposition 1.25. *Let $r \ge 2$. Set $Y = \{1, \dots, r\}$ and consider the forward Bernoulli shift $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$ and the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$. Then $h(S) = h(T) = \log(r)$.*

Proof. We prove that $h(S) = \log(r)$. The proof that $h(T) = \log(r)$ is completely analogous. Set $X = Y^{\mathbb{N}}$. Define the compatible metric $d : X \times X \to \mathbb{R}_+$ by the formula $d((y_n)_n, (z_n)_n) = 2^{-k}$ where $k = \min\{n \in \mathbb{N} \mid y_n \neq z_n\}$. By Example 1.23, S is expansive with constant of expansiveness $\kappa = 1$. Choose $\frac{1}{2} < \varepsilon < 1$. For every $j \in \{1, \dots, r\}$, set $U_j = \{(x_n)_n \in X \mid x_0 = j\}$ and note that $U_j \subset X$ is both open and closed and has d-diameter equal to $\frac{1}{2}$. Moreover, $\mathscr{U} = (U_j)_{1 \le j \le r}$ is a partition of X and for any $j \neq k$ and any $x \in U_j$ and $y \in U_k$, we have $d(x, y) = 1$. This implies that $\mathrm{cov}(1, \varepsilon, S) = r$. More generally, for every $m \in \mathbb{N}$, we may consider the partition $\bigvee_{k=0}^m S^{-k}(\mathscr{U}) = (U_{(i_0, \dots, i_m)})_{(i_0, \dots, i_m) \in \{1, \dots, r\}^{m+1}}$ of X defined by

$$\begin{aligned} U_{(i_0, \dots, i_m)} &= \{(x_n)_n \in X \mid x_0 = i_0, \dots, x_m = i_m\} \\ &= U_{i_0} \cap S^{-1}(U_{i_1}) \cap \dots \cap S^{-m}(U_{i_m}) \end{aligned}$$

for all $(i_0, \dots, i_m) \in \{1, \dots, r\}^{m+1}$. Then $U_{(i_0, \dots, i_m)} \subset X$ is both open and closed and has d_m-diameter equal to $\frac{1}{2}$. Moreover, for any $(i_0, \dots, i_m) \neq (j_0, \dots, j_m)$ and any $x \in U_{(i_0, \dots, i_m)}$ and $y \in U_{(j_0, \dots, j_m)}$, we have $d_m(x, y) = 1$. This implies that $\mathrm{cov}(m, \varepsilon, S) = r^m$. Therefore, using Proposition 1.24, we have

$$h(S) = h_\varepsilon(S) = \lim_m \frac{1}{m} \log(\mathrm{cov}(m, \varepsilon, S)) = \lim_m \frac{1}{m} \log(r^m) = \log(r).$$

This finishes the proof of the proposition. □

CHAPTER 2

Ergodic theory

Throughtout this chapter, a probability space $(X, \mathscr{X}, \nu)$ is a triple where X is a nonempty set, $\mathscr{X}$ is a σ-algebra of subsets of X and ν is a probability measure defined on $\mathscr{X}$. We refer to elements in $\mathscr{X}$ as measurable subsets of X. A *Borel space* Z is a space endowed with a σ-algebra $\mathscr{Z}$ of Borel subsets. A topological space X is naturally a Borel space endowed with the σ-algebra $\mathscr{X}$ generated by open sets. A Borel space Z is *standard* if Z is Borel isomorphic to a Borel subset of a Polish space. A standard Borel space is either finite, countable or Borel isomorphic to the segment $[0, 1]$. A *standard* probability space $(X, \mathscr{X}, \nu)$ is a standard Borel space $(X, \mathscr{X})$ endowed with a Borel probability measure $\nu \in \mathrm{Prob}(X)$. Any standard probability space $(X, \mathscr{X}, \nu)$ such that ν is atom-free is measurably isomorphic to $([0, 1], \mathscr{B}([0, 1]), \mathrm{Leb})$ where Leb denotes the Lebesgue measure on $[0, 1]$. This means that there exists a measurable map $\pi : (X, \mathscr{X}, \nu) \to ([0, 1], \mathscr{B}([0, 1]), \mathrm{Leb})$, conull measurable subsets $Y \subset X$, $Z \subset [0, 1]$ such that $\pi|_Y : Y \to Z$ is bijective, $(\pi|_Z)^{-1} : Z \to Y$ is measurable and $\pi_*\nu = \mathrm{Leb}$.

Recall that a collection of subsets $\mathscr{S} \subset \mathscr{P}(X)$ is said to be a *semi-algebra* if the following properties hold:

- $\emptyset \in \mathscr{S}$.
- For all $A, B \in \mathscr{S}$, $A \cap B \in \mathscr{S}$.
- For all $A \in \mathscr{S}$, the complement $X \setminus A$ is finite union of pairwise disjoint elements in $\mathscr{S}$.

A collection of subsets $\mathscr{A} \subset \mathscr{P}(X)$ is said to be an *algebra* if $\mathscr{A}$ is a semi-algebra and if moreover for all $A \in \mathscr{A}$, we have $X \setminus A \in \mathscr{A}$.

Example 2.1. Let $(X, \mathscr{X}, \nu) = (X_1 \times X_2, \mathscr{X}_1 \otimes \mathscr{X}_2, \nu_1 \otimes \nu_2)$ be a product probability space. The collection

$$\mathscr{S} = \{U_1 \times U_2 \mid U_1 \in \mathscr{X}_1, U_2 \in \mathscr{X}_2\} \subset \mathscr{X}_1 \otimes \mathscr{X}_2$$

that consists of all measurable rectangles is a semi-algebra. The collection $\mathscr{A} \subset \mathscr{X}_1 \otimes \mathscr{X}_2$ that consists of all finite unions of rectangles forms an algebra. Moreover, we have $\sigma(\mathscr{S}) = \sigma(\mathscr{A}) = \mathscr{X}_1 \otimes \mathscr{X}_2$.

We will use the following result without comment (see [**EW11**, Theorem A.1.10]).

Theorem 2.2. *Let $(X, \mathscr{X}, \nu)$ be a probability space and $\mathscr{A} \subset \mathscr{X}$ an algebra for which $\sigma(\mathscr{A}) = \mathscr{X}$. Then for every $\varepsilon > 0$ and every $U \in \mathscr{X}$, there exists $A \in \mathscr{A}$ such that $\nu(U \triangle A) < \varepsilon$.*

For more information, we refer the reader to [**EW11**, Appendix A] and [**KL16**, Appendix A].

A measurable dynamical system is a map $T : X \to X$ for which $T^{-1}(U) \in \mathscr{X}$ for every $U \in \mathscr{X}$. We denote by $T_*\nu$ the *pushforward* probability measure of ν by T defined by $(T_*\nu)(U) = \nu(T^{-1}(U))$ for every $U \in \mathscr{X}$. We say that T is *nonsingular* if the probability measures $T_*\nu$ and ν are equivalent on X. We then say that the quadruple $(X, \mathscr{X}, \nu, T)$ is a nonsingular dynamical system. We say that T is *probability measure preserving* (pmp) if $T_*\nu = \nu$. We then say that the quadruple $(X, \mathscr{X}, \nu, T)$ is a probability measure preserving (pmp) dynamical system. In these lectures notes, we will be mostly interested in pmp dynamical systems.

For every $i \in \{1, 2\}$, let $(X_i, \mathscr{X}_i, \nu_i, T_i)$ be a pmp dynamical system. We say that $(X_2, \mathscr{X}_2, \nu_2, T_2)$ is a pmp *factor* of $(X_1, \mathscr{X}_1, \nu_1, T_1)$ or that $(X_1, \mathscr{X}_1, \nu_1, T_1)$ is a pmp *extension* of $(X_2, \mathscr{X}_2, \nu_2, T_2)$ if there exists a measurable map $\pi : (X_1, \mathscr{X}_1, \nu_1) \to (X_2, \mathscr{X}_2, \nu_2)$ such that $\pi_*\nu_1 = \nu_2$ and $T_2 \circ \pi = \pi \circ T_1$ ν_1-almost everywhere. We say that $(X_2, \mathscr{X}_2, \nu_2, T_2)$ and $(X_1, \mathscr{X}_1, \nu_1, T_1)$ are *measurably conjugate* if there exists a measurable map $\pi : (X_1, \mathscr{X}_1, \nu_1) \to (X_2, \mathscr{X}_2, \nu_2)$, conull measurable subsets $Y_1 \subset X_1$, $Y_2 \subset X_2$ such that $\pi|_{Y_1} : Y_1 \to Y_2$ is bijective and $(\pi|_{Y_1})^{-1} : Y_2 \to Y_1$ is measurable, $\pi_*\nu_1 = \nu_2$ and $T_2 \circ \pi = \pi \circ T_1$ ν_1-almost everywhere.

In this chapter, we follow the presentation given in [**BS02, EW11, KL16**].

1. Ergodicity and recurrence

Firstly, we prove a useful result in order to check that a measurable dynamical system $T : X \to X$ is pmp. We will use this result without reference in what follows.

LEMMA 2.3. *Let $T : X \to X$ be a measurable dynamical system. Let $\mathscr{C} \subset \mathscr{X}$ be a collection of measurable subsets that is stable under finite intersection and such that $\sigma(\mathscr{C}) = \mathscr{X}$. Then T preserves the probability measure ν if and only if $\nu(T^{-1}(U)) = \nu(U)$ for every $U \in \mathscr{C}$.*

PROOF. Assume that $\nu(T^{-1}(U)) = \nu(U)$ for every $U \in \mathscr{C}$. Set

$$\mathscr{Y} = \left\{U \in \mathscr{X} \mid \nu(T^{-1}(U)) = \nu(U)\right\}.$$

We have that $\mathscr{C} \subset \mathscr{Y}$ and that $\mathscr{Y}$ is a monotone class. By the monotone class lemma, we obtain $\mathscr{X} = \sigma(\mathscr{C}) = \mathscr{M}(\mathscr{C}) \subset \mathscr{Y}$ and hence $\mathscr{X} = \mathscr{Y}$. □

Secondly, we show that pmp dynamical systems behave well with respect to L^p-spaces.

LEMMA 2.4. *Let $T : X \to X$ be a measurable dynamical system. Then T preserves the probability measure ν if and only if for every measurable function $f : X \to \mathbb{R}_+$, we have*

$$\int_X f \,\mathrm{d}\nu = \int_X f \circ T \,\mathrm{d}\nu.$$

Moreover, if T preserves the probability measure ν, then for every $p \in [1,+\infty)$ and every $f \in \mathrm{L}^p(X,\mathscr{X},\nu)$, we have $f \circ T \in \mathrm{L}^p(X,\mathscr{X},\nu)$ and $\|f \circ T\|_p = \|f\|_p$.

Proof. Assume that $T : X \to X$ preserves the probability measure ν. Then for every $U \in \mathscr{X}$, we have

$$\int_X \mathbf{1}_U \,\mathrm{d}\nu = \nu(U) = \nu(T^{-1}(U)) = \int_X \mathbf{1}_U \circ T \,\mathrm{d}\nu.$$

By linearity, the above equality also holds for every simple measurable function $f : X \to \mathbb{C}$. Let now $f : X \to \mathbb{R}_+$ be a measurable fonction. Then there exists an increasing sequence $(f_n)_n$ of finite linear combinations of simple nonnegative measurable functions on X such that $f_n \to f$. By monotone convergence theorem, we have

$$\int_X f \,\mathrm{d}\nu = \lim_n \int_X f_n \,\mathrm{d}\nu = \lim_n \int_X f_n \circ T \,\mathrm{d}\nu = \int_X f \circ T \,\mathrm{d}\nu.$$

Let now $f \in \mathrm{L}^p(X,\mathscr{X},\nu)$. Using the first part of the proof, we obtain

$$\|f\|_p^p = \int_X |f|^p \,\mathrm{d}\nu = \int_X |f|^p \circ T \,\mathrm{d}\nu = \int_X |f \circ T|^p \,\mathrm{d}\nu = \|f \circ T\|_p^p.$$

This shows that $f \circ T \in \mathrm{L}^p(X,\mathscr{X},\nu)$. Note that $f \circ T \in \mathrm{L}^p(X,\mathscr{X},\nu)$ only depends on (the class of) $f \in \mathrm{L}^p(X,\mathscr{X},\nu)$. □

For any topological space X endowed with its σ-algebra $\mathscr{X} = \mathscr{B}(X)$ of Borel sets, any topological dynamical system $T : X \to X$ and any Borel probability measure $\nu \in \mathrm{Prob}_T(X)$, we may consider the pmp dynamical system $(X,\mathscr{X},\nu,T)$.

1.1. Ergodicity.

Definition 2.5. Let $(X,\mathscr{X},\nu,T)$ be a pmp dynamical system. We say that T is *ergodic* if for every measurable subset $U \in \mathscr{X}$ such that $T^{-1}(U) = U$, we have $\nu(U) \in \{0,1\}$.

For every pmp dynamical system $(X,\mathscr{X},\nu,T)$, define the *Koopman operator* $\kappa_T : \mathrm{L}^2(X,\mathscr{X},\nu) \to \mathrm{L}^2(X,\mathscr{X},\nu)$ by the formula $\kappa_T(\xi) = \xi \circ T$. For all $\xi,\eta \in \mathrm{L}^2(X,\mathscr{X},\nu)$, we have

$$\begin{aligned}
\langle \kappa_T(\xi), \kappa_T(\eta)\rangle &= \langle \xi \circ T, \eta \circ T\rangle \\
&= \int_X \xi \circ T \cdot \overline{\eta \circ T} \,\mathrm{d}\nu \\
&= \int_X \xi \cdot \overline{\eta} \,\mathrm{d}\nu \\
&= \langle \xi, \eta\rangle.
\end{aligned}$$

It follows that $\kappa_T : \mathrm{L}^2(X,\mathscr{X},\nu) \to \mathrm{L}^2(X,\mathscr{X},\nu)$ is an *isometry*, that is, $\kappa_T^*\kappa_T = 1$. If $(X,\mathscr{X},\nu,T)$ is an invertible pmp dynamical system, then $\kappa_T : \mathrm{L}^2(X,\mathscr{X},\nu) \to \mathrm{L}^2(X,\mathscr{X},\nu)$ is a *unitary*, that is, $\kappa_T^*\kappa_T = 1$ and

$\kappa_T\kappa_T^* = 1$. Observe that $\kappa_T(\mathbf{1}_X) = \mathbf{1}_X$ and so 1 is always an eigenvalue for κ_T with eigenvector given by $\mathbf{1}_X$. The next proposition provides useful characterizations of ergodicity.

PROPOSITION 2.6. *Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system. The following assertions are equivalent:*

(i) *T is ergodic.*
(ii) *For all $U \in \mathscr{X}$ such that $\nu(T^{-1}(U)\triangle U) = 0$, we have $\nu(U) \in \{0, 1\}$.*
(iii) *For all $U \in \mathscr{X}$ such that $\nu(U) > 0$, we have $\nu(\bigcup_{n\in\mathbb{N}} T^{-n}(U)) = 1$.*
(iv) *For all $U, V \in \mathscr{X}$ such that $\nu(U)\nu(V) > 0$, there exists $n \geq 1$ such that $\nu(T^{-n}(U) \cap V) > 0$.*
(v) *Every measurable function $f : X \to \mathbb{C}$ satisfying $f = f \circ T$ ν-almost everywhere is constant ν-almost everywhere.*
(vi) *The eigenvalue 1 is simple for κ_T.*

PROOF. (i) $\Rightarrow$ (ii) Let $U \in \mathscr{X}$ be such that $\nu(T^{-1}(U)\triangle U) = 0$. Define $V = \bigcap_{n\in\mathbb{N}}(\bigcup_{k\geq n} T^{-k}(U)) \in \mathscr{X}$ and observe that $T^{-1}(V) = V$. By ergodicity, we know that $\nu(V) \in \{0, 1\}$. For every $n \in \mathbb{N}$, set $V_n = \bigcup_{k\geq n} T^{-k}(U) \in \mathscr{X}$. Observe that the sequence $(V_n)_n$ is decreasing and $\bigcap_{n\in\mathbb{N}} V_n = V$. For every $n \in \mathbb{N}$, we have $U\triangle V_n \subset \bigcup_{k\geq n} U\triangle T^{-k}(U)$ and $\nu(U\triangle T^{-n}(U)) = 0$ since

- $U\triangle T^{-n}(U) \subset \bigcup_{j=0}^{n-1} T^{-j}(U)\triangle T^{-(j+1)}(U)$ and
- $\nu(T^{-j}(U)\triangle T^{-(j+1)}(U)) = 0$ for all $j \geq 0$.

It follows that $\nu(U\triangle V_n) = 0$ for every $n \in \mathbb{N}$ and hence $\nu(U\triangle V) = 0$. Since $\nu(V) \in \{0, 1\}$, we obtain $\nu(U) \in \{0, 1\}$.

(ii) $\Rightarrow$ (iii) Set $V = \bigcup_{n\in\mathbb{N}} T^{-n}(U)$. Then we have $T^{-1}(V) \subset V$ and $\nu(T^{-1}(V)) = \nu(V)$ since $T_*\nu = \nu$. Then $\nu(T^{-1}(V)\triangle V) = 0$ and hence $\nu(V) \in \{0, 1\}$. Since $U \subset V$ and $\nu(U) > 0$, we obtain $\nu(V) = 1$.

(iii) $\Rightarrow$ (iv) Since $\nu(U) > 0$, we have $\nu(\bigcup_{n\in\mathbb{N}} T^{-n}(U)) = 1$. Since $\nu(V) = \nu(\bigcup_{n\in\mathbb{N}}(T^{-n}(U) \cap V))$ and $\nu(V) > 0$, there exists $n \in \mathbb{N}$ such that $\nu(T^{-n}(U) \cap V) > 0$.

(iv) $\Rightarrow$ (i). Let $U \in \mathscr{X}$ be such that $T^{-1}(U) = U$. Then for every $n \in \mathbb{N}$,

$$0 = \nu(U \cap X \setminus U) = \nu(T^{-n}(U) \cap X \setminus U).$$

It follows that $\nu(U) = 0$ or $\nu(X \setminus U) = 0$, that is, $\nu(U) \in \{0, 1\}$.

(ii) $\Rightarrow$ (v). Upon taking the real and imaginary parts of f, we may assume without loss of generality that the measurable function f is real-valued. For every $t \in \mathbb{R}$, define $U_t = \{x \in X \mid f(x) \geq t\} \in \mathscr{X}$. Since $f \circ T = f$ ν-almost everywhere, we have $\nu(T^{-1}(U_t)\triangle U_t) = 0$. Therefore, $\nu(U_t) \in \{0, 1\}$. Since $t \mapsto \nu(U_t)$ is decreasing and f is real-valued, there exists $t \in \mathbb{R}$ such that $\nu(U_s) = 0$ for all $s > t$ and $\nu(U_s) = 1$ for all $s \leq t$. This implies that $f = t$ ν-almost everywhere.

(v) $\Rightarrow$ (vi) This is trivial.

(vi) $\Rightarrow$ (i) Let $U \in \mathscr{X}$ be such that $T^{-1}(U) = U$. Set $f = \mathbf{1}_U \in \mathrm{L}^2(X, \mathscr{X}, \nu)$. Since $f \circ T = f$, f is constant ν-almost everywhere. Therefore, $\nu(U) \in \{0, 1\}$. $\square$

We use Proposition 2.6 to give examples of ergodic pmp dynamical systems. Denote by $(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}})$ the probability space that consists of the torus endowed with its σ-algebra of Borel subsets and its Haar (Lebesgue) measure. For every $\alpha \in \mathbb{R}$, consider the rotation $T_\alpha : \mathbb{T} \to \mathbb{T} : z \mapsto \exp(\mathrm{i}2\pi\alpha)z$. Since $T_{\alpha *}\lambda_{\mathbb{T}} = \lambda_{\mathbb{T}}$, the dynamical system $(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}}, T_\alpha)$ is pmp.

Proposition 2.7. *The rotation* $(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}}, T_\alpha)$ *is ergodic if and only if* $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Proof. Firstly, assume that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. For every $k \in \mathbb{Z}$, define $e_k : \mathbb{T} \to \mathbb{T} : x \mapsto x^k$. It easy to check that $(e_k)_{k \in \mathbb{Z}}$ forms an orthonormal basis of the Hilbert space $\mathrm{L}^2(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}})$. Let $f \in \mathrm{L}^2(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}})$ be a function satisfying $f \circ T_\alpha = f$. Write $f = \sum_{k \in \mathbb{Z}} a_k e_k$ for the Fourier expansion of f where $a_k \in \mathbb{C}$ for all $k \in \mathbb{Z}$. By uniqueness of the Fourier expansion, we have $c_k = c_k \exp(\mathrm{i}2\pi k\alpha)$ for all $k \in \mathbb{Z}$. Since $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, we have $c_k = 0$ for all $k \in \mathbb{Z} \setminus \{0\}$. Then f is a constant function. This shows that $(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}}, T_\alpha)$ is ergodic.

Secondly, assume that $\alpha \in \mathbb{Q}$. Write $\alpha = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{N} \setminus 0$. The function $x \mapsto x^q$ is T_α-invariant and is not $\lambda_{\mathbb{T}}$-almost everywhere constant. This shows that $(\mathbb{T}, \mathscr{B}, \lambda_{\mathbb{T}}, T_\alpha)$ is not ergodic $\square$

Let $(Y, \mathscr{Y}, \eta)$ be a probability space. Consider the product probability space $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}})$ together with the forward Bernoulli shift $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$. It is plain to see that $S_* \eta^{\otimes \mathbb{N}} = \eta^{\otimes \mathbb{N}}$ and so $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}}, S)$ is a pmp dynamical system. Likewise, consider the product probability space $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}})$ together with the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$. It is plain to see that $T_* \eta^{\otimes \mathbb{Z}} = \eta^{\otimes \mathbb{Z}}$ and so $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}}, T)$ is a pmp dynamical system.

Proposition 2.8. *The forward Bernoulli shift* $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}}, S)$ *is ergodic. Likewise, the Bernoulli shift* $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}}, T)$ *is ergodic.*

Proof. We only give the proof of ergodicity of the forward Bernoulli shift. The proof of ergodicity of the Bernoulli shift is completely analogous. Set $(X, \mathscr{X}, \nu) = (Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}})$. Note that the σ-algebra $\mathscr{X}$ is generated by cylinder sets of the form $\mathscr{C}(U_0, \dots, U_{n_0}) = \prod_n Z_n$ where $Z_n = U_n \in \mathscr{Y}$ for $n \leq n_0$ and $Z_n = Y$ for $n > n_0$.

Let $U \in \mathscr{X}$ satisfying $S^{-1}(U) = U$. For every $\varepsilon > 0$, there exists a finite union of cylinder sets V such that $\nu(U \triangle V) \leq \varepsilon$. Then there exists $m \in \mathbb{N}$ large enough such that

$$\nu(S^{-m}(V) \setminus V) = \nu(S^{-m}(V) \cap X \setminus V) = \nu(S^{-m}(V))\nu(X \setminus V) = \nu(V)\nu(X \setminus V).$$

We have $\nu(S^{-m}(V)\triangle U) = \nu(S^{-m}(V)\triangle S^{-m}(U)) = \nu(V\triangle U) \leq \varepsilon$ and so $\nu(S^{-m}(V)\triangle V) \leq 2\varepsilon$. Therefore, we have

$$\begin{aligned}\nu(U)\nu(X\setminus U) &\leq (\nu(V)+\varepsilon)(\nu(X\setminus V)+\varepsilon)\\ &\leq \nu(V)\nu(X\setminus V) + 2\varepsilon + \varepsilon^2\\ &\leq \nu(S^{-m}(V)\triangle V) + 2\varepsilon + \varepsilon^2\\ &\leq 4\varepsilon + \varepsilon^2.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have $\nu(U)\nu(X\setminus U) = 0$ and so $\nu(U) \in \{0, 1\}$. $\square$

Let G be a compact abelian metrizable group. Denote by $\mathscr{B}(G)$ its σ-algebra of Borel subsets and by m_G its unique Haar Borel probability measure (see Chapter 3). Let $T \in \mathrm{Aut}(G)$ be a continuous automorphism group. Then we have $T_* m_G = m_G$ and so $(G, \mathscr{B}(G), m_G, T)$ is a pmp dynamical system. We denote by $\widehat{G}$ the Pontryagin dual of G that consists of all continuous group homomorphisms (characters) $\chi : G \to \mathbb{T}$.

Theorem 2.9. *The pmp dynamical system $(G, \mathscr{B}(G), m_G, T)$ is ergodic if and only if the identity $\chi \circ T^n = \chi$ for some $n \geq 1$ and character $\chi \in \widehat{G}$ implies that $\chi = \mathbf{1}_G$ is the trivial character.*

Proof. Firstly, assume that there is a nontrivial character $\chi \in \widehat{G}$ such that $\chi \circ T^n = \chi$ for some $n \geq 1$. We may choose $n \geq 1$ to be minimal with this property. Then the continuous function $f = \chi + \chi\circ T + \cdots + \chi \circ T^{n-1}$ is T-invariant. We claim that f is not constant. Indeed, for every $i \in \{0, \dots, n\}$, set $\chi_0 = \mathbf{1}_G$ and $\chi_i = \chi \circ T^{i-1} \in \widehat{G}$ for every $1 \leq i \leq n$. Then the characters $(\chi_i)_{0\leq i\leq n}$ are pairwise distinct and so the family $(\chi_i)_{0\leq i\leq n}$ is linearly independent. This implies that f is not constant. This shows that the pmp dynamical system $(G, \mathscr{B}(G), m_G, T)$ is not ergodic.

Secondly, assume that there is no nontrivial character $\chi \in \widehat{G}$ such that $\chi \circ T^n = \chi$ for some $n \geq 1$. Let $f \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$ be a function that is invariant under T. Write $f = \sum_{\chi\in\widehat{G}} c_\chi \chi$ for the Fourier expansion of $f \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$. We have $\sum_{\chi\in\widehat{G}} |c_\chi|^2 = \|f\|_2^2$. Since $f = f \circ T$, we have $c_{\chi\circ T^k} = c_\chi$ for every $\chi \in \widehat{G}$ and every $k \in \mathbb{Z}$. Let $\chi \in \widehat{G}$. Then either $c_\chi = 0$ or there are finitely many distinct characters among $(\chi \circ T^k)_{k\in\mathbb{Z}}$. In the latter case, there are $p > q$ such that $\chi \circ T^p = \chi \circ T^q$ and so $\chi \circ T^{p-q} = \chi$. By assumption, this implies that $\chi = \mathbf{1}_G$. Therefore, $f = c_{\mathbf{1}_G} \mathbf{1}_G \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$ is constant m_G-almost everywhere. This shows that the pmp dynamical system $(G, \mathscr{B}(G), m_G, T)$ is ergodic. $\square$

As a corollary to Theorem 2.9, we obtain a characterization of ergodicity for toral automorphisms. Let $d \geq 1$ and $A \in \mathrm{GL}_d(\mathbb{Z})$. Regard $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ and denote by $\lambda_{\mathbb{T}^d}$ the Haar (Lebesgue) probability measure on $\mathbb{T}^d$. Consider the continuous group automorphism $T_A : \mathbb{T}^d \to \mathbb{T}^d : x + \mathbb{Z}^d \mapsto Ax + \mathbb{Z}^d$ which satisfies $T_{A*}\lambda_{\mathbb{T}^d} = \lambda_{\mathbb{T}^d}$. Then $(\mathbb{T}^d, \mathscr{B}, \lambda_{\mathbb{T}^d}, T_A)$ is a pmp dynamical system.

COROLLARY 2.10. *The pmp dynamical system* $(\mathbb{T}^d, \mathscr{B}, \lambda_{\mathbb{T}^d}, T_A)$ *is ergodic if and only if no eigenvalue of* A *is a root of unity.*

PROOF. For every $(x_1, \dots, x_d) \in \mathbb{R}^d$, we denote by $[x_i]_i$ the corresponding column vector. We simply denote by $\langle \cdot , \cdot \rangle$ the canonical inner product on $\mathbb{R}^d$. The homomorphism

$$\mathbb{Z}^d \mapsto \widehat{\mathbb{T}^d} : (n_1, \dots, n_d) \mapsto \Big((x_1, \dots, x_d) + \mathbb{Z}^d \mapsto \exp(\mathrm{i}2\pi \langle [n_i]_i, [x_i] \rangle) \Big)$$

is an isomorphism that allows to identify $\mathbb{Z}^d$ with the Pontryagin dual $\widehat{\mathbb{T}^d}$.

Firstly, assume that there exists an eigenvalue λ of A that is a root of unity. Then there exists $n \geq 1$ such that $1 = \lambda^n$ is an eigenvalue of A^n. Then 1 is also an eigenvalue of $(A^n)^*$. Regarding $(A^n)^* \in \mathrm{GL}_d(\mathbb{Q})$, we infer that there exists a nonzero vector $\xi \in \mathbb{Q}^d$ such that $(A^n)^*\xi = \xi$. Upon multiplying by a nonzero integer, we may assume that $\xi \in \mathbb{Z}^d$ and $(A^n)^*\xi = \xi$. Write $\xi = [n_i]_i$ and consider the corresponding character $\chi = (n_1, \dots, n_d) \in \mathbb{Z}^d = \widehat{\mathbb{T}^d}$. Then for every $g = (x_1, \dots, x_d) + \mathbb{Z}^d \in \mathbb{T}^d$, we have

$$\begin{aligned} \chi(T_A^n(g)) &= \exp(\mathrm{i}2\pi \langle \xi, A^n [x_i]_i \rangle) \\ &= \exp(\mathrm{i}2\pi \langle (A^n)^* \xi, [x_i]_i \rangle) \\ &= \exp(\mathrm{i}2\pi \langle \xi, [x_i]_i \rangle) \\ &= \chi(g). \end{aligned}$$

By Theorem 2.9, the pmp dynamical system $(\mathbb{T}^d, \mathscr{B}, \lambda_{\mathbb{T}^d}, T_A)$ is not ergodic.

Secondly, assume the pmp dynamical system $(\mathbb{T}^d, \mathscr{B}, \lambda_{\mathbb{T}^d}, T_A)$ is not ergodic. By Theorem 2.9, there is a nontrivial character $\chi = (n_1, \dots, n_d) \in \mathbb{Z}^d = \widehat{\mathbb{T}^d}$ such that $\chi \circ T_A^n = \chi$ for some $n \geq 1$. Write $\xi = [n_i]_i$. For every $g = (x_1, \dots, x_d) + \mathbb{Z}^d \in \mathbb{T}^d$, we have

$$\begin{aligned} \exp(\mathrm{i}2\pi \langle (A^n)^* \xi, [x_i]_i \rangle) &= \exp(\mathrm{i}2\pi \langle \xi, A^n [x_i]_i \rangle) \\ &= \chi(T_A^n(g)) \\ &= \chi(g) \\ &= \exp(\mathrm{i}2\pi \langle \xi, [x_i]_i \rangle). \end{aligned}$$

This further implies that $(A^n)^*\xi = \xi$ and so 1 is an eigenvalue of $(A^n)^*$ as well as of A^n. This shows that A has an eigenvalue that is a root of unity. $\square$

1.2. Recurrence and applications. We prove Poincaré's recurrence theorem and we investigate some applications. Poincaré's recurrence theorem is a measurable analogue of Birkhoff's recurrence theorem in topological dynamics (see Corollary 1.6).

THEOREM 2.11 (Poincaré). *Let* $(X, \mathscr{X}, \nu, T)$ *be a pmp dynamical system and* $U \in \mathscr{X}$. *Then* ν*-almost every point of* U *returns to* U *infinitely many times. That is, there exists a conull measurable subset* $V \subset U$ *such that* *for every $x \in V$, there exists an increasing sequence $(n_k)_{k\geq 1}$ in $\mathbb{N}$ for which $T^{n_k}(x) \in U$ for every $k \geq 1$.*

Proof. For every $n \in \mathbb{N}$, set $W_n = \bigcup_{k\geq n} T^{-k}(U) \in \mathscr{X}$ and observe that $W_n = T^{-n}(W_0)$. Moreover, the sequence $(W_n)_n$ is decreasing. Since T is pmp, we have $\nu(W_n) = \nu(W_0)$ for every $n \in \mathbb{N}$. Since $\nu(W_0) < +\infty$, it follows that $\nu(\bigcap_{n\in\mathbb{N}} W_n) = \lim_n \nu(W_n) = \nu(W_0)$. Letting $V = U \cap \bigcap_{n\in\mathbb{N}} W_n$, we are done. $\square$

Let $(X, \mathscr{X}, \nu, T)$ be an invertible pmp dynamical system and $U \in \mathscr{X}$ with $\nu(U) > 0$. By Poincaré's recurrence theorem, the *first return time* defined by

$$r_U(x) = \inf\{n \geq 1 \mid T^n(x) \in U\}$$

is finite ν-almost everywhere on U.

Definition 2.12 (Induced transformation). The map $T_U : U \to U$ defined ν-almost everywhere by

$$T_U(x) = T^{r_U(x)}(x)$$

is called the transformation *induced* by T on the measurable subset U.

Observe that $r_U : U \to \mathbb{N}$ and $T_U : U \to U$ are measurable. Indeed, for every $n \geq 1$, set $X_n = \{x \in X \mid r_U(x) = n\}$. We have $X_1 = T^{-1}(U) \in \mathscr{X}$ and

$$\forall n \geq 2, \quad X_n = T^{-n}(U) \setminus \bigcup_{1\leq i<n} X_i \in \mathscr{X}.$$

This implies that $r_U : U \to \mathbb{N}$ is measurable. For every $n \geq 1$, set $U_n = U \cap X_n$. Since T is invertible, we have $T^n(U_n) \in \mathscr{X}$ for every $n \geq 1$. Therefore, the map

$$T_U : U \to U = \bigsqcup_{n\geq 1} (T^n : U_n \to T^n(U_n))$$

is measurable. The measurable subset $U_n \sqcup T(U_n) \sqcup \cdots \sqcup T^{n-1}(U_n)$ is called the nth *Kakutani tower* and $\bigsqcup_{n\geq 1} \bigsqcup_{0\leq j\leq n-1} T^j(U_n)$ is called the *Kakutani skyscraper.*

Set $\mathscr{U} = \{V \cap U \mid V \in \mathscr{X}\}$ and $\nu_U(V) = \frac{1}{\nu(U)}\nu(V)$ for every $V \in \mathscr{U}$. Observe that $\mathscr{U}$ is a σ-algebra on U and ν_U is a probability measure defined on $(U, \mathscr{U})$. We have the following result.

Proposition 2.13. *The induced dynamical system $(U, \mathscr{U}, \nu_U, T_U)$ is pmp. Moreover, if $(X, \mathscr{X}, \nu, T)$ is ergodic, then so is $(U, \mathscr{U}, \nu_U, T_U)$.*

Proof. Since $(X, \mathscr{X}, \nu, T)$ is pmp and invertible and

$$T_U : U \to U = \bigsqcup_{n\geq 1} (T^n : U_n \to T^n(U_n)),$$

it follows that $(U, \mathscr{U}, \nu_U, T_U)$ is pmp.

If $(U, \mathscr{U}, \nu_U, T_U)$ is not ergodic, then there exists a T_U-invariant subset $V \in \mathscr{U}$ such that $0 < \nu(V) < \nu(U)$. Then $W = \bigsqcup_{n \geq 1} \bigsqcup_{0 \leq j \leq n-1} T^j(V \cap U_n)$ is T-invariant. Moreover for every $0 \leq j \leq n-1$, we have $\nu((U \setminus V) \cap T^j(V \cap U_n)) = 0$. This implies that W is nontrivial and so $(X, \mathscr{X}, \nu, T)$ is not ergodic. $\square$

Proposition 2.14 (Kac). *Let $(X, \mathscr{X}, \nu, T)$ be an invertible ergodic pmp dynamical system and $U \in \mathscr{X}$ with $\nu(U) > 0$. Then*

$$\int_U r_U \, \mathrm{d}\nu = 1.$$

Proof. Since $(X, \mathscr{X}, \nu, T)$ is ergodic and since $\bigsqcup_{n \geq 1} \bigsqcup_{0 \leq j \leq n-1} T^j(U_n)$ is T-invariant, we have

$$\bigsqcup_{n \geq 1} \bigsqcup_{0 \leq j \leq n-1} T^j(U_n) = X$$

up to a ν-null measurable subset. By the monotone convergence theorem, it follows that

$$1 = \nu(X) = \sum_{n \geq 1} \sum_{j=0}^{n-1} \nu(T^j(U_n)) = \sum_{n \geq 1} n \nu(U_n) = \int_U r_U \, \mathrm{d}\nu.$$

This finishes the proof. $\square$

We prove now Kakutani–Rokhlin's tower theorem.

Theorem 2.15 (Kakutani–Rokhlin). *Let $(X, \mathscr{X}, \nu, T)$ be an invertible ergodic pmp dynamical system and assume that ν is atom-free. For every $\varepsilon > 0$ and every $n \geq 1$, there exists $V \in \mathscr{X}$ such that*

- *the measurable subsets $V, T(V), \dots, T^{n-1}(V)$ are pairwise disjoint*
- *and $\nu(X \setminus \bigsqcup_{0 \leq k \leq n-1} T^k(V)) < \varepsilon$.*

Proof. Let $\varepsilon > 0$ and $n \geq 1$. Since ν is atom-free, we may choose a measurable subset $U \in \mathscr{X}$ such that $0 < \nu(U) < \frac{\varepsilon}{n}$. Consider the Kakutani skyscraper over U. By ergodicity, we know that

$$\bigsqcup_{k \geq 1} \bigsqcup_{0 \leq j \leq k-1} T^j(U_k) = X$$

up to ν-null measurable subset. Define the measurable subset

$$V = \bigsqcup_{k \geq n} \bigsqcup_{j=0}^{\lfloor k/n \rfloor - 1} T^{jn}(U_k).$$

Observe that $k = \lfloor k/n \rfloor \cdot n + r$ with $0 \leq r \leq n-1$. We obtain that $V, T(V), \dots, T^{n-1}(V)$ are pairwise disjoint. Then we obtain

$$\nu\left(X \setminus \bigsqcup_{0 \leq k \leq n-1} T^k(V)\right) \leq n \sum_{k \geq 1} \nu(U_k) \leq n \, \nu(U) \leq \varepsilon.$$

This finishes the proof. $\square$

We use Theorem 2.15 to infer that invertible ergodic pmp dynamical systems are not *strongly ergodic*.

COROLLARY 2.16. *Let $(X, \mathscr{X}, \nu, T)$ be an invertible ergodic pmp dynamical system and assume that ν is atom-free.*

Then there exists a sequence $(U_n)_n$ in $\mathscr{X}$ such that $\nu(U_n) = \frac{1}{2}$ for every $n \in \mathbb{N}$ and $\lim_n \nu(T(U_n) \triangle U_n) = 0$.

PROOF. For every $n \geq 1$, we apply Theorem 2.15 to $\varepsilon = \frac{1}{n}$ and $2n$. Then there exists a measurable subset $V_n \in \mathscr{X}$ such that $V_n, T(V_n), \dots, T^{2n-1}(V_n)$ are pairwise disjoint and $\nu(X \setminus \bigsqcup_{0 \leq k \leq 2n-1} T^k(V_n)) < \frac{1}{n}$. Set $W_n = V_n \sqcup T(V_n) \sqcup \cdots \sqcup T^{n-1}(V_n)$. Then we have

$$\frac{1}{2} - \frac{1}{2n} \leq \nu(W_n) \leq \frac{1}{2} \quad \text{and} \quad \nu(T(W_n) \triangle W_n) \leq 2\nu(V_n) \leq \frac{1}{n}.$$

For every $n \geq 1$, define $U_n \in \mathscr{X}$ so that $W_n \subset U_n$ and $\nu(U_n) = \frac{1}{2}$. Since for every $n \geq 1$, $\nu(U_n \setminus W_n) \leq \frac{1}{2n}$, we have $\lim_n \nu(T(U_n) \triangle U_n) = 0$. $\square$

2. Invariant measures and unique ergodicity

In this section, we assume that X is a compact metrizable space. We fix a compatible metric d on X. Denote by $(\mathscr{M}(X), \|\cdot\|)$ the Banach space of all complex Borel measures on X where the norm of $\nu \in \mathscr{M}(X)$ is given by $\|\mu\|_{\mathscr{M}(X)} = |\mu|(X)$. Here $|\mu|$ denotes the *modulus* or *absolute value* of the Borel measure $\nu \in \mathscr{M}(X)$. Hence $|\nu|$ is a finite positive Borel measure on X. By Riesz representation theorem, the mapping

$$\mathscr{M}(X) \to \mathrm{C}(X)^* : \nu \mapsto \left(f \mapsto \int_X f \, \mathrm{d}\nu\right)$$

is isometric and surjective. We can define the *weak-$*$* topology on $\mathscr{M}(X)$: a net $(\nu_i)_{i \in I}$ in $\mathscr{M}(X)$ converges to $\nu \in \mathscr{M}(X)$ with respect to the weak-$*$ topology if for every $f \in \mathrm{C}(X)$, we have

$$\lim_i \int_X f \, \mathrm{d}\nu_i = \int_X f \, \mathrm{d}\nu.$$

Observe that the unit ball $\mathrm{Ball}(\mathscr{M}(X))$ is a metrizable compact space hence separable. Indeed, since (X, d) is a compact metric space, the unital Banach algebra $\mathrm{C}(X)$ is separable by Stone–Weierstrass theorem. Let $(f_n)_{n \geq 1}$ be a uniformly dense sequence in $\mathrm{C}(X)$. For all $\nu, \eta \in \mathrm{Ball}(\mathscr{M}(X))$, define

$$\mathsf{d}(\nu, \eta) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|\int_X f_n \, \mathrm{d}\nu - \int_X f_n \, \mathrm{d}\eta|}{1 + |\int_X f_n \, \mathrm{d}\nu - \int_X f_n \, \mathrm{d}\eta|}.$$

Then d is a metric on $\mathrm{Ball}(\mathscr{M}(X))$ that induces the weak-$*$ topology.

Denote by $\mathscr{X} = \mathscr{B}(X)$ the σ-algebra of Borel subsets of X and by $\mathrm{Prob}(X) \subset \mathscr{M}(X)$ the convex subset of all the Borel (positive) probability measures on X. Since $\mathrm{Prob}(X)$ is contained in the unit ball of $\mathscr{M}(X)$ and since $\mathrm{Prob}(X)$ is weak-$*$ closed, it follows that $\mathrm{Prob}(X)$ is weak-$*$ compact. For any topological dynamical system $T : X \to X$, define the map $T_* : \mathrm{Prob}(X) \to \mathrm{Prob}(X) : \nu \mapsto T_*\nu$ by the formula

$$\int_X f \, \mathrm{d}(T_*\nu) = \int_X f \circ T \, \mathrm{d}\nu.$$

here $T_*\nu$ is the pushforward Borel probability measure of ν by T. It is easy to check that $T_* : \mathrm{Prob}(X) \to \mathrm{Prob}(X)$ is affine and weak-$*$ continuous.

2.1. Invariant measures. Let $T : X \to X$ be a topological dynamical system. Denote by

$$\mathrm{Prob}_T(X) = \{\nu \in \mathrm{Prob}(X) \mid T_*\nu = \nu\}.$$

the weak-$*$ convex subset of T-invariant Borel probability measures on X.

LEMMA 2.17. *We have* $\mathrm{Prob}_T(X) \neq \emptyset$.

PROOF. Choose $\nu \in \mathrm{Prob}(X)$ any Borel probability measure on X. For every $n \geq 1$, define

$$\nu_n = \frac{1}{n}\left(\nu + T_*\nu + \cdots + (T^{n-1})_*\nu\right) \in \mathrm{Prob}(X).$$

For every $n \geq 1$, we have

$$\|T_*\nu_n - \nu_n\|_{\mathscr{M}(X)} = \frac{1}{n}\|(T^n)_*\nu - \nu\|_{\mathscr{M}(X)} \leq \frac{2}{n}$$

and so $\lim_n \|T_*\nu_n - \nu_n\|_{\mathscr{M}(X)} = 0$. Since $\mathrm{Prob}(X)$ is weak-$*$ compact, there exists an increasing sequence $(n_k)_k$ in $\mathbb{N}$ and $\eta \in \mathrm{Prob}(X)$ such that $\nu_{n_k} \to \eta$ weak-$*$ as $k \to \infty$. Since $\lim_k \|T_*\nu_{n_k} - \nu_{n_k}\|_{\mathscr{M}(X)} = 0$, we also have $T_*\nu_{n_k} - \nu_{n_k} \to 0$ weak-$*$ as $k \to \infty$ and so $T_*\eta = \eta$. Therefore, $\eta \in \mathrm{Prob}_T(X)$. $\square$

Let $T : X \to X$ be a topological dynamical system and $\nu \in \mathrm{Prob}_T(X)$. We say that ν is T-*ergodic* if the pmp dynamical system $(X, \mathscr{X}, \nu, T)$ is ergodic. We give a characterization of ergodic measures.

PROPOSITION 2.18. *Let $T : X \to X$ be a topological dynamical system and $\nu \in \mathrm{Prob}_T(X)$. Then $(X, \mathscr{X}, \nu, T)$ is ergodic if and only if ν is an extreme point of the convex set* $\mathrm{Prob}_T(X)$.

PROOF. Let $\nu \in \mathrm{Prob}_T(X)$ be a measure that is not T-ergodic. Let $U \in \mathscr{X}$ be such that $\nu(U) \in (0,1)$. Define $\nu_1 = \frac{1}{\nu(U)}\nu|_U$ and $\nu_2 = \frac{1}{\mu(X \setminus U)}\nu|_{X\setminus U}$. Then we have $\nu_1, \nu_2 \in \mathrm{Prob}_T(X)$, $\nu_1 \neq \nu \neq \nu_2$ and $\nu = \nu(U)\,\nu_1 + (1 - \nu(U))\,\nu_2$. Thus, ν is not an extreme point of the convex set $\mathrm{Prob}_T(X)$.

Conversely, let $\nu \in \mathrm{Prob}_T(X)$ be a T-ergodic measure and let $\nu = t\,\nu_1 + (1-t)\,\nu_2$ be a convex combination with $t \in (0,1)$ and $\nu_1, \nu_2 \in \mathrm{Prob}_T(X)$. Since $t > 0$, ν_1 is absolutely continuous with respect to ν and we may consider the Radon–Nikodym derivative $f = \frac{\mathrm{d}\nu_1}{\mathrm{d}\nu} \in \mathrm{L}^1(X, \mathscr{X}, \nu)$.

Define $U = \{x \in X \mid f(x) < 1\}$. We have

$$\nu_1(U) = \nu_1(U \cap T^{-1}(U)) + \nu_1(U \setminus T^{-1}(U))$$

$$
\begin{aligned}
&= \int_{U\cap T^{-1}(U)} f \,\mathrm{d}\nu + \int_{U\setminus T^{-1}(U)} f \,\mathrm{d}\nu \\
\nu_1(T^{-1}(U)) &= \nu_1(T^{-1}(U)\cap U) + \nu_1(T^{-1}(U)\setminus U) \\
&= \int_{T^{-1}(U)\cap U} f \,\mathrm{d}\nu + \int_{T^{-1}(U)\setminus U} f \,\mathrm{d}\nu.
\end{aligned}
$$

Since $\nu_1(U) = \nu_1(T^{-1}(U))$, we have

$$
\int_{U\setminus T^{-1}(U)} f \,\mathrm{d}\nu = \int_{T^{-1}(U)\setminus U} f \,\mathrm{d}\nu
$$

Moreover, we have

$$
\begin{aligned}
\nu(T^{-1}(U)\setminus U) &= \nu(T^{-1}(U)) - \nu(T^{-1}(U)\cap U) \\
&= \nu(U) - \nu(T^{-1}(U)\cap U) \\
&= \nu(U\setminus T^{-1}(U)).
\end{aligned}
$$

Finally, observe that $f(x) < 1$ for all $x \in U \setminus T^{-1}(U)$ while $f(x) \geq 1$ for all $x \in T^{-1}(U)\setminus U$. Therefore $\nu(T^{-1}(U)\setminus U) = \nu(U\setminus T^{-1}(U)) = 0$ and so $\nu(T^{-1}(U)\triangle U) = 0$. Since ν is ergodic, we have $\nu(U) \in \{0, 1\}$. If $\nu(U) = 1$, then $1 = \nu_1(U) = \int_U f \,\mathrm{d}\nu$ and so $f(x) = 1$ for ν-almost every $x \in U$. This is a contradiction. Thus, $\nu(U) = 0$ and so we have $f(x) \geq 1$ for ν-almost every $x \in X$. Likewise, we have $f(x) \leq 1$ for ν-almost every $x \in X$. Therefore, $f = 1$ in $\mathrm{L}^1(X, \mathscr{X}, \nu)$ and so $\nu_1 = \nu$. Thus, ν is an extreme point of the convex set $\mathrm{Prob}_T(X)$. $\square$

Recall that two positive measures ν and η on a measurable space $(E, \mathscr{A})$ are *mutually singular* if there exists a measurable subset $U \in \mathscr{A}$ such that $\nu(U) = 0$ and $\eta(U) = 1$. The above characterization of ergodic measures allows to obtain an interesting dichotomy result for such ergodic measures.

Proposition 2.19. *Let $T : X \to X$ be a topological dynamical system. Let $\nu_1, \nu_2 \in \mathrm{Prob}_T(X)$ be two T-ergodic measures. Then either $\nu_1 = \nu_2$ or ν_1 and ν_2 are mutually singular.*

Proof. Using Radon–Nikodym's theorem, there exists a unique pair (ζ_1, ζ_2) of finite positive Borel measures on X such that $\nu_1 = \zeta_1 + \zeta_2$, where ζ_1 is absolutely continuous with respect to ν_2 and ζ_2 and ν_2 are mutually singular.

If $\zeta_1 = 0$, then $\nu_2 = \zeta_2$ and so ν_1 and ν_2 are mutually singular.

If $\zeta_2 = 0$, then $\nu_1 = \zeta_1$ and so ν_1 is absolutely continuous with respect to ν_2. The same reasoning as in Proposition 2.18 shows that $\nu_1 = \nu_2$.

Finally, by contradiction, we show that $\zeta_1 = 0$ or $\zeta_2 = 0$. If not, then we may write $\nu_1 = t\eta_1 + (1-t)\eta_2$ with $\eta_1, \eta_2 \in \mathrm{Prob}(X)$, $t \in (0,1)$, η_1 is absolutely continuous with respect to ν_2 and η_2 and ν_2 are mutually singular. Observe that $\eta_1 \neq \eta_2$ and $\nu_1 = T_*\nu_1 = t\,T_*\eta_1 + (1-t)\,T_*\eta_2$. It is clear that $T_*\eta_1$ is absolutely continuous with respect to $T_*\nu_2 = \nu_2$. We claim that $T_*\eta_2$ and $T_*\nu_2 = \nu_2$ are mutually singular. Indeed, let $U \in \mathscr{X}$

be a measurable subset such that $\eta_2(U) = 1$ and $\nu_2(U) = 0$. Then we have $\eta_1(U) = 0$ and so $\nu_1(U) = (1-t)\eta_2(U) = (1-t)$. Since $\nu_2(T^{-1}(U)) = 0$, we also have $\eta_1(T^{-1}(U)) = 0$ and so $\nu_1(T^{-1}(U)) = (1-t)\eta_2(T^{-1}(U))$. Since $\nu_1(T^{-1}(U)) = \nu_1(U)$, we have $\eta_2(T^{-1}(U)) = \eta_2(U) = 1$. Therefore, $T_*\eta_2$ and ν_2 are mutually singular. By uniqueness of the decomposition in Radon–Nikodym's theorem, we have $T_*\eta_1 = \eta_1$ and $T_*\eta_2 = \eta_2$, that is, $\eta_1, \eta_2 \in \mathrm{Prob}_T(X)$. This however contradicts the fact that $\nu_1 \in \mathrm{Prob}_T(X)$ is an extreme point by Proposition 2.18. Therefore, we have $\zeta_1 = 0$ or $\zeta_2 = 0$ and the proof is complete. $\square$

2.2. Unique ergodicity. We say that a topological dynamical system $T : X \to X$ is *uniquely ergodic* if $\mathrm{Prob}_T(X)$ is a singleton. We start by proving a general result on weak-$*$ compact convex subsets.

LEMMA 2.20 (Krein–Milman). *Let $(E, \|\cdot\|)$ be a normed complex vector space. Let $K \subset (E^*)_1$ be a nonempty weak-$*$ closed convex subset of the unit ball of E^*. Then K has an extreme point.*

PROOF. Observe that K is weak-$*$ compact. We say that a nonempty subset $A \subset K$ is *extreme* if whenever $x, y \in K$ and $t \in (0,1)$ are such that $t\,x + (1-t)\,y \in A$, we have $x, y \in A$. Define the nonempty set

$$\mathscr{C} = \{A \subset K \mid A \text{ is weak-}*\text{ closed, extreme and nonempty}\}$$

with order relation $<$ given by

$$A_1 < A_2 \quad \text{if and only if} \quad A_2 \subset A_1.$$

It is easy to check that $(\mathscr{C}, <)$ is an inductive set. Indeed, let $\{A_i \mid i \in I\} \subset \mathscr{C}$ be a totally ordered subset. Set $A = \bigcap_{i \in I} A_i$. Then $A \subset K$ is weak-$*$ closed, extreme and nonempty by compactness. Therefore, we have $A \in \mathscr{C}$.

By Zorn's lemma, $\mathscr{C}$ has a maximal element $B \in \mathscr{C}$. We show that B is a singleton. If not, let $f_1, f_2 \in B$ with $f_1 \neq f_2$. Let $v \in E$ be such that $f_1(v) \neq f_2(v)$. We may assume that $\Re f_1(v) < \Re f_2(v)$. Since B is weak-$*$ closed and hence weak-$*$ compact, there exists $f_0 \in B$ such that $\Re f_0(v) = \sup\{\Re f(v) \mid f \in B\}$. Let $B_0 = \{f \in B \mid \Re f(v) = \Re f_0(v)\}$. Then B_0 is a weak-$*$ closed extreme subset of K. Indeed, let $g_1, g_2 \in K$ and $t \in (0,1)$ be such that $t\,g_1 + (1-t)g_2 \in B_0$. Since B is extreme, we have $g_1, g_2 \in B$. Next, by definition of B_0, we moreover have $g_1, g_2 \in B_0$. Then B_0 is a weak-$*$ closed extreme subset of K such that $B < B_0$ and $B_0 \neq B$. This contradicts the maximality of B. Therefore $B = \{f\}$ is a singleton and so $f \in K$ is an extreme point. $\square$

THEOREM 2.21. *Let $T : X \to X$ be a topological dynamical system. The following assertions are equivalent:*

(i) *T is uniquely ergodic.*
(ii) *There is only one ergodic T-invariant Borel probability measure in $\mathrm{Prob}_T(X)$.*

(iii) *For every $f \in \mathrm{C}(X)$, there exists a constant λ_f such that uniformly for all $x \in X$, we have*

$$\lim_n \frac{1}{n}\left(f(x) + \cdots + f(T^{n-1}(x))\right) = \lambda_f.$$

(iv) *There exists a uniformly dense subspace $\mathscr{A} \subset \mathrm{C}(X)$ such that for every $f \in \mathscr{A}$, there exists a constant λ_f such that for all $x \in X$, we have*

$$\lim_n \frac{1}{n}\left(f(x) + \cdots + f(T^{n-1}(x))\right) = \lambda_f.$$

If any of the above assertions holds, then we have $\lambda_f = \int_X f \,\mathrm{d}\nu$, where $\mathrm{Prob}_T(X) = \{\nu\}$.

PROOF. (i) $\Rightarrow$ (ii). If $\mathrm{Prob}_T(X) = \{\nu\}$, then ν is an extreme point of the convex set $\mathrm{Prob}_T(X)$. Thus, ν is ergodic by Theorem 2.18.

(ii) $\Rightarrow$ (i). We show that if $\mathrm{Prob}_T(X)$ is not a singleton, then it has at least two extreme points. This will prove the implication by Proposition 2.18. Let $\nu_1, \nu_2 \in \mathrm{Prob}_T(X)$ be such that ν_1 is an extreme point (see Lemma 2.20) and $\nu_2 \neq \nu_1$. Let $f \in \mathrm{C}(X)$ be such that $\int_X f \,\mathrm{d}\nu_2 \neq \int_X f \,\mathrm{d}\nu_1$. We may assume that f is real-valued and that $\int_X f \,\mathrm{d}\nu_1 < \int_X f \,\mathrm{d}\nu_2$. Since $\mathrm{Prob}_T(X)$ is weak-$*$ compact, there exists $\nu \in \mathrm{Prob}_T(X)$ such that $\int_X f \,\mathrm{d}\nu = \sup\left\{\int_X f \,\mathrm{d}\eta \mid \eta \in \mathrm{Prob}_T(X)\right\}$. Set

$$K = \left\{\eta \in \mathrm{Prob}_T(X) \mid \int_X f \,\mathrm{d}\eta = \int_X f \,\mathrm{d}\nu\right\}.$$

Then $K \subset \mathrm{Prob}_T(X)$ is a nonempty weak-$*$ closed convex subset. By Lemma 2.20, K has an extreme point η. Since $K \subset \mathrm{Prob}_T(X)$ is moreover an extreme subset, η is an extreme point in $\mathrm{Prob}_T(X)$. Since $\int_X f \,\mathrm{d}\nu_1 < \int_X f \,\mathrm{d}\eta$, we have $\nu_1 \neq \eta$.

(i) $\Rightarrow$ (iii). Using the proof of Theorem 2.17, we have that for every $x \in X$, the sequence $\left(\frac{1}{n}\left(\delta_x + T_*\delta_x + \cdots + (T^{n-1})_*\delta_x\right)\right)_{n \geq 1}$ converges with respect to the weak-$*$ topology to the unique invariant Borel probability ν. Then for all $f \in \mathrm{C}(X)$ and all $x \in X$, since $(T^k)_*\delta_x = \delta_{T^k(x)}$, we have

$$\lim_n \frac{1}{n}\left(x + f(x) + \cdots + f(T^{n-1}(x))\right) = \int_X f \,\mathrm{d}\nu.$$

If the above convergence is not uniform on X for some $f \in \mathrm{C}(X)$, then there exist $\varepsilon > 0$, an increasing sequence $(n_k)_k$ in $\mathbb{N}^*$ and $x_{n_k} \in X$ such that for all $k \in \mathbb{N}$,

$$\left|\frac{1}{n_k}\left(x_{n_k} + f(x_{n_k}) + \cdots + f(T^{n_k - 1}(x_{n_k}))\right) - \int_X f \,\mathrm{d}\nu\right| \geq \varepsilon.$$

By weak-$*$ compactness of $\mathrm{Prob}(X)$, upon choosing a further subsequence $(n_k)_k$ in $\mathbb{N}$, we may assume that the sequence $\frac{1}{n_k}(\delta_{x_{n_k}} + \delta_{T(x_{n_k})} + \cdots + \delta_{T^{n_k-1}(x_{n_k})})$ converges to $\eta \in \mathrm{Prob}(X)$ with respect to the weak-$*$ topology.

The same reasoning as in the proof of Theorem 2.17 shows that η is invariant and hence $\eta = \nu$. We obtain

$$\lim_k \frac{1}{n_k}\left(x_{n_k} + f(x_{n_k}) + \cdots + f(T^{n_k-1}(x_{n_k}))\right) = \int_X f \, \mathrm{d}\nu$$

which is a contradiction.

(iii) $\Rightarrow$ (iv). It is trivial.

(iv) $\Rightarrow$ (i). By applying Lebesgue's dominated convergence theorem, for every $\nu \in \mathrm{Prob}_T(X)$, we have $\lambda_f = \int_X f \, \mathrm{d}\nu$ for every $f \in \mathscr{A}$. Therefore, for all $\nu_1, \nu_2 \in \mathrm{Prob}_T(X)$ and all $f \in \mathscr{A}$,

$$\int_X f \, \mathrm{d}\nu_1 = \lambda_f = \int_X f \, \mathrm{d}\nu_2.$$

By uniform density of $\mathscr{A}$ in $\mathrm{C}(X)$, we have $\int_X f \, \mathrm{d}\nu_1 = \int_X f \, \mathrm{d}\nu_2$ for all $f \in \mathrm{C}(X)$ and so $\nu_1 = \nu_2$. Therefore, $\mathrm{Prob}_T(X)$ is a singleton. □

The sum $\mathscr{S}_{f,T,n}(x) = \frac{1}{n}\left(f(x) + \cdots + f(T^{n-1}(x))\right)$ is sometimes called the nth *Birkhoff sum* of the function $f \in \mathrm{C}(X)$ at the point $x \in X$. Theorem 2.21 shows that the Birkhoff sums of $f \in \mathrm{C}(X)$ converge everywhere and uniformly in X to the *space average* $\int_X f \, \mathrm{d}\nu$.

Recall that for $\alpha \in \mathbb{R}$, the circle rotation $T_\alpha : \mathbb{T} \to \mathbb{T}$ is the topological dynamical system defined by $T_\alpha(x) = \exp(\mathrm{i}2\pi\alpha)x$ for all $x \in \mathbb{T}$.

PROPOSITION 2.22. *The circle rotation $T_\alpha : \mathbb{T} \to \mathbb{T}$ is uniquely ergodic if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.*

PROOF. First, assume that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Let $k \in \mathbb{Z}$ and put $f(x) = x^k$ for all $x \in \mathbb{T}$. Then with $x = \exp(\mathrm{i}2\pi t)$, for all $n \in \mathbb{N}$, we have

$$\begin{aligned}
\mathscr{S}_{f,T_\alpha,n}(x) &= \frac{1}{n}\left(f(x) + f(T_\alpha(x)) + \cdots + f(T_\alpha^{n-1}(x))\right) \\
&= \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(\mathrm{i}2\pi k(t + j\alpha)\right) \\
&= \begin{cases} 1 & \text{if } k = 0 \\ \frac{1}{n}\exp(\mathrm{i}2\pi kt)\frac{1-\exp(\mathrm{i}2\pi kn\alpha)}{1-\exp(\mathrm{i}2\pi k\alpha)} & \text{if } k \neq 0 \end{cases} \\
&\to \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \quad \text{as} \quad n \to \infty \\
&= \int_{\mathbb{T}} f \, \mathrm{d}\lambda_{\mathbb{T}}.
\end{aligned}$$

Since the linear span of $\{\mathbb{T} \to \mathbb{T} : x \mapsto x^k : k \in \mathbb{Z}\}$ is uniformly dense in $\mathrm{C}(\mathbb{T})$ by Stone–Weierstrass theorem, T_α follows uniquely ergodic by Theorem 2.21.

Next, assume that $\alpha \in \mathbb{Q}$. Then the Lebesgue measure $\lambda_{\mathbb{T}}$ is invariant but not ergodic for T_α. Since there must exist ergodic measures in $\mathrm{Prob}_{T_\alpha}(\mathbb{T})$ by Lemma 2.20 and Theorem 2.18, T_α is not uniquely ergodic. □

DEFINITION 2.23. Let $\nu \in \mathrm{Prob}(X)$ and $(x_n)_n$ a sequence in X. We say that the sequence $(x_n)_n$ is *equidistributed* with respect to ν if the sequence of empirical measures $\left(\frac{1}{n}(\delta_{x_0} + \cdots + \delta_{x_{n-1}})\right)_n$ converges to ν with respect to the weak-$*$ topology, that is,

$$\forall f \in \mathrm{C}(X), \quad \lim_n \frac{1}{n}(f(x_0) + \cdots + f(x_{n-1})) = \int_X f \, \mathrm{d}\nu.$$

For every $x \in \mathbb{R}$, write $\{x\} = x - \lfloor x \rfloor \in [0, 1)$ for the fractional part of x. As a consequence of Theorem 2.21 and Proposition 2.22, we deduce Weyl's equidistribution theorem.

COROLLARY 2.24 (Weyl's equidistribution theorem). *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ be irrational. Then the sequence $(\{\alpha n\})_n$ is equidistributed with respect to the Lebesgue measure* Leb *on* $[0, 1]$.

PROOF. Using the continuous mapping $[0, 1] \to \mathbb{T} : x \mapsto \exp(\mathrm{i}2\pi x)$, we may identify $[0, 1]/\sim$ with $\mathbb{T}$ as compact spaces, where $0 \sim 1$ in $[0, 1]$. Moreover, we may identify the Lebesgue measure on $[0, 1]/\sim$ with the Haar measure $\lambda_{\mathbb{T}}$ on $\mathbb{T}$.

Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ be irrational. We can write the rotation $T_\alpha : \mathbb{T} \to \mathbb{T} : x \mapsto \{x + \alpha\}$. Then for every $n \in \mathbb{N}$, we have $\{n\alpha\} = T_\alpha^n(0)$. Using a combination of Theorem 2.21 and Proposition 2.22, we infer that

$$\forall f \in \mathrm{C}(\mathbb{T}), \quad \lim_n \frac{1}{n}(f(0) + \cdots + f(\{\alpha(n-1)\})) = \int_{\mathbb{T}} f(x) \, \mathrm{d}\lambda_{\mathbb{T}}(x).$$

Therefore, $(\{\alpha n\})_n$ is equidistributed with respect to the Lebesgue measure Leb on $[0, 1]$. □

Finally, we obtain the following general result about unique ergodicity for rotations on compact metrizable groups.

THEOREM 2.25. *Let G be a compact metrizable group. Denote by $\mathscr{B}(G)$ its σ-algebra of Borel subsets and by m_G its unique Haar Borel probability measure. Let $g \in G$ and consider the rotation $T_g : G \to G : x \mapsto gx$. Then the following assertions are equivalent:*

(i) *The rotation T_g is uniquely ergodic and* $\mathrm{Prob}_{T_g}(G) = \{m_G\}$.
(ii) *The Haar measure m_G is T_g-ergodic.*
(iii) *The subgroup $g^{\mathbb{Z}}$ is dense in G.*
(iv) *The group G is abelian and $\chi(g) \neq 1$ for every $\chi \in \widehat{G} \setminus \{\mathbf{1}_G\}$.*

PROOF. Fix a compatible metric $d : G \times G \to \mathbb{R}_+$. Define the new compatible metric $d_G : G \times G \to \mathbb{R}_+$ by the formula

$$\forall x, y \in G, \quad d_G(x, y) = \int_G d(hx, hy) \, \mathrm{d}m_G(h).$$

Then $d_G : G \times G \to \mathbb{R}_+$ is left invariant.

(i) ⇒ (ii) This is obvious.

(ii) $\Rightarrow$ (iii) Denote by H the closure of the subgroup $g^{\mathbb{Z}}$ in G. Then $H < G$ is a closed abelian subgroup. Consider the continuous function $f_H : G \to \mathbb{R}_+$ defined by the formula

$$\forall x \in G, \quad f_H(x) = \inf \{d_G(x, y) \mid y \in H\}.$$

Observe that $f_H : G \to \mathbb{R}_+$ is indeed continuous since $|f_H(x) - f_H(y)| \leq d_G(x, y)$ for all $x, y \in G$. Since d_G is left invariant, it follows that $f_H \circ T_g = f_H$. By contraposition, if the subgroup $g^{\mathbb{Z}}$ is not dense in G, then $H \neq G$ and so f_H is not constant. This implies that the pmp dynamical system $(G, \mathscr{B}(G), m_G, T_g)$ is not ergodic.

(iii) $\Rightarrow$ (i) Let $\nu \in \mathrm{Prob}_{T_g}(G)$ be a T_g-invariant Borel probability measure. Then ν is T_{g^n}-invariant for every $n \in \mathbb{Z}$. We show that ν is T_x-invariant for every $x \in G$. Indeed, let $f \in \mathrm{C}(G)$ be a continuous function and $x \in G$. Choose a sequence $(n_k)_k$ in $\mathbb{Z}$ such that $\lim_k d_G(g^{n_k}, x) = 0$. Then by Lebesgue's dominated convergence theorem, we have

$$\int_G f(xy) \, \mathrm{d}\nu(y) = \lim_k \int_G f(g^{n_k} y) \, \mathrm{d}\nu(y) = \int_G f(y) \, \mathrm{d}\nu(y).$$

Then ν is T_x-invariant for every $x \in G$. By uniqueness of the Haar probability measure on G, it follows that $\nu = m_G$. Therefore, T_g is uniquely ergodic and $\mathrm{Prob}_{T_g}(G) = \{m_G\}$.

(iii) $\Rightarrow$ (iv) Since $g^{\mathbb{Z}}$ is dense in G, it follows that G is abelian. Let $\chi \in \widehat{G}$ such that $\chi(g) = 1$. Then for every $n \in \mathbb{Z}$, we have $\chi(g^n) = 1$. By continuity and density, we have $\chi = \mathbf{1}_G$.

(iv) $\Rightarrow$ (ii) Let $f \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$ be a T_g-invariant function. Write $f = \sum_{\chi \in \widehat{G}} c_\chi \chi$ for the Fourier expansion of $f \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$. Since $f \circ T_g = f$, we have $c_\chi \chi(g) = c_\chi$ for every $\chi \in \widehat{G}$. Using the assumption, we obtain $c_\chi = 0$ for every $\chi \in \widehat{G} \setminus \{\mathbf{1}_G\}$. It follows that $f = c_{\mathbf{1}_G} \mathbf{1}_G$ is constant m_G-almost everywhere. Therefore, $(G, \mathscr{B}(G), m_G, T_g)$ is ergodic. $\square$

Keep the same notation as in Theorem 2.25. Simply denote by $\kappa_g = \kappa_{T_g} : \mathrm{L}^2(G, \mathscr{B}(G), m_G) \to \mathrm{L}^2(G, \mathscr{B}(G), m_G) : f \mapsto f \circ T_g$ the Koopman unitary operator. For every $\chi \in \widehat{G}$, we have $\kappa_g(\chi) = \chi \circ T_g = \chi(g)\, \chi$. It follows that $(\chi)_{\chi \in \widehat{G}}$ forms an orthonormal basis of eigenvectors of κ_g on $\mathrm{L}^2(G, \mathscr{B}(G), m_G)$. In that case, we have that the pmp dynamical system $(G, \mathscr{B}(G), m_G, T_g)$ has *discrete spectrum*.

3. Ergodic theorems

3.1. von Neumann's mean ergodic theorem. In this subsection, we prove von Neumann's mean ergodic theorem.

THEOREM 2.26 (von Neumann). *Let $(X, \mathscr{X}, \nu, T)$ be an ergodic pmp dynamical system. Then we have*

$$\forall f \in \mathrm{L}^2(X, \mathscr{X}, \nu), \quad \lim_n \left\| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k - \int_X f \, \mathrm{d}\nu \right\|_2 = 0. \tag{2.1}$$

PROOF. Denote by $\kappa_T : \mathrm{L}^2(X, \mathscr{X}, \nu) \to \mathrm{L}^2(X, \mathscr{X}, \nu)$ the Koopman operator associated with $(X, \mathscr{X}, \nu, T)$. Denote by $\mathscr{K}$ the closed subspace of $\mathrm{L}^2(X, \mathscr{X}, \nu)$ generated by $\left\{\kappa_T(g) - g \mid g \in \mathrm{L}^2(X, \mathscr{X}, \nu)\right\}$. Since T is ergodic, we have $\mathscr{K}^\perp = \mathbb{C}\mathbf{1}_X$. Indeed, for all $f \in \mathscr{K}^\perp$ and all $g \in \mathrm{L}^2(X, \mathscr{X}, \nu)$, we have

$$0 = \langle f, \kappa_T(g) - g \rangle = \langle \kappa_T^*(f) - f, g \rangle$$

and so $\kappa_T^*(f) = f$. Then

$$\begin{aligned} \|f - \kappa_T(f)\|_2^2 &= \|f\|_2^2 + \|f\|_2^2 - 2\Re\langle f, \kappa_T(f)\rangle \\ &= \|f\|_2^2 + \|\kappa_T^*(f)\|_2^2 - 2\Re\langle \kappa_T^*(f), f\rangle \\ &= \|f - \kappa_T^*(f)\|_2^2 = 0. \end{aligned}$$

Since T is ergodic and since $\kappa_T(f) = f$, we have $f \in \mathbb{C}\mathbf{1}_X$. Therefore, $\mathscr{K} = \mathrm{L}^2(X, \mathscr{X}, \nu) \ominus \mathbb{C}\mathbf{1}_X$.

If $f = \lambda \mathbf{1}_X \in \mathbb{C}\mathbf{1}_X$, we have $\frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k(f) = \lambda \mathbf{1}_X = (\int_X f \, \mathrm{d}\nu)\mathbf{1}_X$ for all $n \geq 1$ and so $\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k(f) = (\int_X f \, \mathrm{d}\nu)\mathbf{1}_X = \int_X f \, \mathrm{d}\nu$.

If $f = \kappa_T(g) - g$, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k(f) = \frac{1}{n}(\kappa_T^n(g) - g)$$

and so $\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k(f) = 0$ in $\mathrm{L}^2(X, \mathscr{X}, \nu)$. By density of the linear span of $\left\{\kappa_T(g) - g \mid g \in \mathrm{L}^2(X, \mathscr{X}, \nu)\right\}$ in $\mathscr{K}$ and since the operator $\frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k$ is a contraction for every $n \geq 1$, we have $\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \kappa_T^k(f) = 0$ for every $f \in \mathscr{K}$. This finishes the proof of the theorem. $\square$

Using Theorem 2.26, we obtain a new characterization of ergodic pmp dynamical systems.

COROLLARY 2.27. *Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system. Then $(X, \mathscr{X}, \nu, T)$ is ergodic if and only if*

$$\forall U, V \in \mathscr{X}, \quad \lim \frac{1}{n}\sum_{k=0}^{n-1} \nu(U \cap T^{-k}(V)) = \nu(U)\nu(V). \tag{2.2}$$

PROOF. Assume that $(X, \mathscr{X}, \nu, T)$ is not ergodic. Then there exists $U \in \mathscr{X}$ such that $T^{-1}(U) = U$ and $\nu(U) \in (0,1)$. Then we have

$$\forall n \in \mathbb{N}, \quad \frac{1}{n}\sum_{k=0}^{n-1} \nu(U \cap T^{-k}(U)) = \nu(U) \neq \nu(U)^2.$$

Therefore, (2.2) does not hold.

Assume that $(X, \mathscr{X}, \nu, T)$ is ergodic. Let $U, V \in \mathscr{X}$. Using Theorem 2.26, we obtain

$$\begin{aligned}\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \nu(U \cap T^{-k}(V)) &= \lim_n \frac{1}{n}\sum_{k=0}^{n-1} \left\langle \mathbf{1}_U, \mathbf{1}_V \circ T^k \right\rangle \\ &= \left\langle \mathbf{1}_U, \lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_V \circ T^k \right\rangle \\ &= \langle \mathbf{1}_U, \nu(V)\mathbf{1}_X\rangle \\ &= \nu(U)\nu(V).\end{aligned}$$

Therefore, (2.2) holds. □

3.2. Birkhoff's pointwise ergodic theorem. In this subsection, we prove Birkhoff's pointwise ergodic theorem.

THEOREM 2.28 (Birkhoff). *Let $(X, \mathscr{X}, \nu, T)$ be an ergodic pmp dynamical system. Then for every $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, the sequence $(\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k)_n$ converges to $\int_X f \,\mathrm{d}\nu$ ν-almost everywhere and in $\mathrm{L}^1(X, \mathscr{X}, \nu)$.*

Before proving Theorem 2.28, we prove a useful inequality known as the *maximal inequality*.

LEMMA 2.29 (Maximal inequality). *Let $(Y, \mathscr{Y}, \eta, S)$ be a pmp dynamical system. Let $(\varphi_n)_{n\geq 1}$ be a sequence of real-valued functions in $\mathrm{L}^1(X, \mathscr{X}, \nu)$ that satisfy the* subadditivity relation *$\varphi_{m+n} \leq \varphi_m \circ S^n + \varphi_n$ for all $m, n \geq 1$. Set $\varphi = \varphi_1$ and $\varphi^* = \sup_{n\geq 1} \varphi_n$. Then*

$$\int_{\{\varphi^* > 0\}} \varphi \,\mathrm{d}\eta \geq 0.$$

PROOF OF LEMMA 2.29. Set $\psi_0 = \varphi_0 = 0$. For every $n \geq 1$, set $\psi_n = \max\{0, \varphi_1, \ldots, \varphi_n\}$ and $Y_n = \{y \in Y \mid \psi_n(y) > 0\}$. For every $y \in Y_n$, we have $\psi_n(y) = \varphi_{k(y)}(y)$ for some $k(y) \in \{1, \ldots, n\}$ and so

$$\psi_n(y) = \varphi_{k(y)}(y) \leq (\varphi_{k(y)-1} \circ S)(y) + \varphi(y) \leq (\psi_{n-1} \circ S)(y) + \varphi(y).$$

For every $y \in Y \setminus Y_n$, we have $\psi_n(y) = 0$ and $(\psi_{n-1} \circ S)(y) \geq 0$. Therefore,

$$\begin{aligned}\int_{Y_n} \varphi(y) \,\mathrm{d}\eta(y) &\geq \int_{Y_n} \psi_n(y) \,\mathrm{d}\eta(y) - \int_{Y_n} (\psi_{n-1} \circ S)(y) \,\mathrm{d}\eta(y) \\ &\geq \int_Y \psi_n(y) \,\mathrm{d}\eta(y) - \int_Y (\psi_{n-1} \circ S)(y) \,\mathrm{d}\eta(y) \\ &= \int_Y (\psi_n - \psi_{n-1})(y) \,\mathrm{d}\eta(y) \geq 0.\end{aligned}$$

Since $(Y_n)_n$ is increasing and since $\{\varphi^* > 0\} = \bigcup_{n\geq 1} Y_n$, Lebesgue's dominated convergence theorem implies that

$$\int_{\{\varphi^*>0\}} \varphi \, \mathrm{d}\eta = \int_Y \mathbf{1}_{\{\varphi^*>0\}} \varphi \, \mathrm{d}\eta = \lim_n \int_Y \mathbf{1}_{Y_n} \varphi \, \mathrm{d}\eta = \lim_n \int_{Y_n} \varphi \, \mathrm{d}\eta \geq 0.$$

This finishes the proof. □

Proof of Theorem 2.28. Upon taking real and imaginary parts, we may assume that f is real-valued. For every $n \geq 1$, set $f_n = \sum_{k=0}^{n-1} f \circ T^k$. Then for every $n \geq 1$, we have $\int_X \frac{1}{n} f_n \, \mathrm{d}\nu = \int_X f \mathrm{d}\nu \in \mathbb{R}$. In order to prove that $\lim_n \frac{1}{n} f_n(x)$ exists ν-almost everywhere, it suffices to show that for all rational numbers $\alpha < \beta$, the measurable subset

$$X_{\alpha,\beta} = \left\{ x \in X \mid \liminf_n \frac{1}{n} f_n(x) < \alpha < \beta < \limsup_n \frac{1}{n} f_n(x) \right\} \in \mathscr{X}$$

is ν-null. Observe that for every $x \in X$, we have $\lim_n \frac{1}{n}(f_n - f_{n-1} \circ T)(x) = \lim_n \frac{1}{n} f(x) = 0$. This implies that $T^{-1}(X_{\alpha,\beta}) = X_{\alpha,\beta}$. Since $(X, \mathscr{X}, \nu, T)$ is ergodic, we have $\nu(X_{\alpha,\beta}) \in \{0, 1\}$. Assume by contradiction that $\nu(X_{\alpha,\beta}) = 1$. If we apply Lemma 2.29 to $Y = X_{\alpha,\beta}$, $S = T|_{X_{\alpha,\beta}}$ and $(\varphi_n)_n = (f_n - \beta n)_n$ (resp. $Y = X_{\alpha,\beta}$, $S = T|_{X_{\alpha,\beta}}$ and $(\varphi_n)_n = (\alpha n - f_n)_n$), we obtain

$$\int_{X_{\alpha,\beta}} (f - \beta) \, \mathrm{d}\nu \geq 0 \quad \text{and} \quad \int_{X_{\alpha,\beta}} (\alpha - f) \, \mathrm{d}\nu \geq 0.$$

It follows that $\int_{X_{\alpha,\beta}} (\alpha - \beta) \, \mathrm{d}\nu \geq 0$ and so $\nu(X_{\alpha,\beta}) = 0$, which is a contradiction.

Define the measurable function $\lambda_f : X \to \overline{\mathbb{R}}$ by the formula $\lambda_f(x) = \lim_n \frac{1}{n} f_n(x)$ for ν-almost every $x \in X$. Since for every $x \in X$, we have $\lim_n \frac{1}{n}(f_n - (f_{n-1} \circ T))(x) = 0$, we obtain $\lambda_f \circ T = \lambda_f$ ν-almost everywhere. Since T is ergodic, λ_f is a constant function ν-almost everywhere. We moreover have $|\lambda_f| \leq \|f\|_1$. Indeed, upon taking positive and negative parts, we may assume that $f = f^+ \geq 0$ (resp. $f = f^- \geq 0$). By Fatou's lemma, we have

$$0 \leq \lambda_f = \int_X \liminf_n \frac{1}{n} f_n \, \mathrm{d}\nu \leq \liminf_n \frac{1}{n} \int_X f_n \, \mathrm{d}\nu = \int_X f \, \mathrm{d}\nu.$$

It remains to show that $\lambda_f = \int_X f \, \mathrm{d}\nu$ and that $\lim_n \|f_n - \lambda_f\|_1 = 0$. To do this, we firstly assume that $f \in \mathrm{L}^\infty(X, \mathscr{X}, \nu)$ (which is indeed contained in $\mathrm{L}^1(X, \mathscr{X}, \nu)$ since ν is a probability measure). Then by Lebesgue's dominated convergence theorem, we obtain $\lim_n \|\lambda_f - \frac{1}{n} f_n\|_1 = 0$ and so $\lambda_f = \int_X f \, \mathrm{d}\nu$. Secondly, assume that $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$ and fix $\varepsilon > 0$. By L^1-density of $\mathrm{L}^\infty(X, \mathscr{X}, \nu)$ in $\mathrm{L}^1(X, \mathscr{X}, \mu)$, choose $g \in \mathrm{L}^\infty(X, \mathscr{X}, \nu)$ such that $\|f - g\|_1 \leq \varepsilon/3$. We have $\frac{1}{n}\|f_n - g_n\|_1 \leq \varepsilon/3$ for every $n \geq 1$ by triangle inequality and $|\lambda_f - \lambda_g| \leq \|f - g\|_1 \leq \varepsilon/3$ by the observation above. Using again the triangle inequality, we obtain $\limsup_n \|\lambda_f - \frac{1}{n} f_n\|_1 \leq \varepsilon$. Since

$\varepsilon > 0$ is arbitrary, we have $\lim_n \|\lambda_f - \frac{1}{n} f_n\|_1 = 0$ and so $\lambda_f = \int_X f \, \mathrm{d}\nu$. This finishes the proof. $\square$

As a consequence of Birkhoff's pointwise ergodic theorem, we can deduce the *strong law of large numbers*.

Corollary 2.30 (Strong law of large numbers). *Let $(X_n)_{n \geq 1}$ be an infinite sequence of iid integrable real valued random variables. Then almost surely, we have*

$$\frac{1}{n}(X_1 + \cdots + X_n) \to \mathbb{E}(X_1).$$

Proof. Denote by $(\Omega, \mathscr{A}, \mathbb{P})$ the underlying probability space and consider the measurable map $\pi : \Omega \to \mathbb{R}^{\mathbb{N}^*} : \omega \mapsto (X_n(\omega))_n$. Since $(X_n)_n$ is iid, there is a unique Borel probability measure $\eta \in \mathrm{Prob}(\mathbb{R})$ such that $\pi_* \mathbb{P} = \eta^{\otimes \mathbb{N}^*}$. Consider the forward Bernoulli shift $S : \mathbb{R}^{\mathbb{N}^*} \to \mathbb{R}^{\mathbb{N}^*} : (y_n)_n \mapsto (y_n)_{n+1}$. Then the pmp dynamical system $(\mathbb{R}^{\mathbb{N}^*}, \mathscr{B}(\mathbb{R})^{\otimes \mathbb{N}^*}, \eta^{\otimes \mathbb{N}^*}, S)$ is ergodic. Consider the function $f \in \mathrm{L}^1(\mathbb{R}^{\mathbb{N}^*}, \mathscr{B}(\mathbb{R})^{\otimes \mathbb{N}^*}, \eta^{\otimes \mathbb{N}^*})$ defined by $f((y_n)_n) = y_1$ which satisfies $\int_X f \, \mathrm{d}\eta^{\otimes \mathbb{N}^*} = \mathbb{E}(X_1)$. Then for every $n \geq 1$ and every $y = (y_n)_n = \pi(\omega)$, we have $\frac{1}{n} \sum_{k=0}^{n-1} (f \circ S^k)(y) = \frac{1}{n}(X_1(\omega) + \cdots + X_n(\omega))$. Therefore, Theorem 2.28 implies that almost surely, we have $\frac{1}{n}(X_1 + \cdots + X_n) \to \mathbb{E}(X_1)$. $\square$

3.3. Kingman's subadditive ergodic theorem. In this subsection, we prove Kingman's subadditve ergodic theorem.

Theorem 2.31 (Kingman). *Let $(X, \mathscr{X}, \nu, T)$ be an ergodic pmp dynamical system. Let $(f_n)_{n \geq 1}$ be a sequence of real-valued functions in $\mathrm{L}^1(X, \mathscr{X}, \nu)$ that satisfy the* subadditivity relation *$f_{m+n} \leq f_m \circ T^n + f_n$ for all $m, n \geq 1$. Then the sequence $(\frac{1}{n} f_n)_n$ converges to $\inf_{n \geq 1} \frac{1}{n} \int_X f_n \, \mathrm{d}\nu$ ν-almost everywhere.*

Proof. Set $f_0 = 0$. For every $n \geq 1$, set $g_n = f_n - \sum_{k=0}^{n-1} f_1 \circ T^k$. Then $(g_n)_n$ is a sequence of real-valued functions in $\mathrm{L}^1(X, \mathscr{X}, \nu)$ that still satisfy the subadditivity relation $g_{m+n} \leq g_m \circ T^n + g_n$ for all $m, n \geq 1$. Moreover, $g_1 = 0$ and so $g_n \leq 0$ for all $n \geq 1$. By Theorem 2.28, the sequence $(\frac{1}{n} \sum_{k=0}^{n-1} f_1 \circ T^k)_n$ converges to $\int_X f_1 \, \mathrm{d}\nu$ ν-almost everywhere. Therefore, we may assume without loss of generality that $f_n \leq 0$ for every $n \geq 1$. Set $\ell = \inf_{n \geq 1} \frac{1}{n} \int_X f_n \, \mathrm{d}\nu$.

Firstly, since the sequence $(\int_X f_n \, \mathrm{d}\nu)_n$ is subadditive, Lemma 1.17 implies that $\ell = \lim_n \frac{1}{n} \int_X f_n \, \mathrm{d}\nu$.

Secondly, we show that the sequence $(\frac{1}{n} f_n)_n$ converges ν-almost everywhere. For all $\alpha, \beta \in \mathbb{Q}$, define

$$X_\alpha = \left\{ x \in X \mid \liminf_n \frac{1}{n} f_n(x) < \alpha \right\} \in \mathscr{X}$$

$$Y_\beta = \left\{ x \in X \mid \limsup_n \frac{1}{n} f_n(x) > \beta \right\} \in \mathscr{X}.$$

Assume that $\alpha < \beta$. Since $f_{n+1} \leq f_n \circ T + f_1$ for every $n \geq 1$, we have $T^{-1}(X_\alpha) \subset X_\alpha$ and $Y_\beta \subset T^{-1}(Y_\beta)$. Since $(X, \mathscr{X}, \nu, T)$ is ergodic, we have $\nu(X_\alpha), \nu(Y_\beta) \in \{0,1\}$. Assume by contradiction that $\nu(X_\alpha) = \nu(Y_\beta) = 1$. If we apply Lemma 2.29 to $Y = X_\alpha \cap Y_\beta$, $S = T|_{X_\alpha \cap Y_\beta \cap T^{-1}(X_\alpha \cap Y_\beta)}$ and $(\varphi_n)_n = (f_n - \beta n)_n$ (resp. $Y = X_\alpha \cap Y_\beta$, $S = T|_{X_\alpha \cap Y_\beta \cap T^{-1}(X_\alpha \cap Y_\beta)}$ and $(\varphi_n)_n = (\alpha n - f_n)_n$), we obtain

$$\int_{X_\alpha \cap Y_\beta} (f_1 - \beta) \, \mathrm{d}\nu \geq 0 \quad \text{and} \quad \int_{X_\alpha \cap Y_\beta} (\alpha - f_1) \, \mathrm{d}\nu \geq 0.$$

It follows that $\int_{X_\alpha \cap Y_\beta} (\alpha - \beta) \, \mathrm{d}\nu \geq 0$ and so $\nu(X_\alpha \cap Y_\beta) = 0$, which is a contradiction. Therefore, we have $\nu(X_\alpha \cap Y_\beta) = 0$ for all $\alpha, \beta \in \mathbb{Q}$ satisfying $\alpha < \beta$. This implies that the sequence $(f_n)_n$ converges ν-almost everywhere. Define $\psi(x) = \lim_n \frac{1}{n} f_n(x) \leq 0$ for ν-almost every $x \in X$.

For every $n \geq 1$, since $-f_n \geq 0$, Fatou's lemma implies that

$$\int_X -\psi \, \mathrm{d}\nu = \int_X \liminf_n -\frac{1}{n} f_n \, \mathrm{d}\nu \leq \liminf_n \int_X -\frac{1}{n} f_n \, \mathrm{d}\nu = -\ell.$$

Next, we show that $\psi(x) \leq \ell$ for ν-almost every $x \in X$. Recall that $f_n \leq 0$ for every $n \in \mathbb{N}$. Let $q, m \geq 1$ and $0 \leq k \leq m-1$. By iterating the subadditivity relation, we have

$$\begin{aligned}
f_{(q+1)m} &\leq f_{qm+k} + f_{m-k} \circ T^{qm+k} \leq f_{qm+k} \leq f_k + f_{qm} \circ T^k \\
&\leq f_{qm} \circ T^k \\
&\leq f_{(q-1)m} \circ T^k + f_m \circ T^{(q-1)m+k} \\
&\leq f_{(q-2)m} \circ T^k + f_m \circ T^{(q-2)m+k} + f_m \circ T^{(q-1)m+k} \\
&\leq \cdots \\
&\leq \sum_{i=0}^{q-1} f_m \circ T^{im+k}.
\end{aligned}$$

Summing over $k \in \{0, \ldots, m-1\}$ these inequalities and dividing by m, we obtain

$$f_{(q+1)m} \leq \frac{1}{m} \sum_{i=0}^{qm-1} f_m \circ T^i.$$

Dividing both sides by qm and letting $q \to \infty$, by applying Theorem 2.28 to $f_m \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, we obtain for ν-almost every $x \in X$,

$$\psi(x) = \lim_q \frac{1}{qm} f_{qm}(x) \leq \frac{1}{m} \int_X f_m \, \mathrm{d}\nu.$$

Since this holds true for every $m \geq 1$, we obtain $\psi(x) \leq \ell$ for ν-almost every $x \in X$.

If $\ell = -\infty$, then $\psi(x) = \ell = -\infty$ for ν-almost every $x \in X$.

If $-\infty < \ell \leq 0$, then $-\psi(x) + \ell \geq 0$ for ν-almost every $x \in X$ and $\int_X (-\psi + \ell)\, \mathrm{d}\nu \leq 0$. This implies that $\psi(x) = \ell$ for ν-almost every $x \in X$. This finishes the proof. $\square$

4. Strong and weak mixing

Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system. The convergence (2.2) in Corollary 2.27 suggests the following strengthenings of the notion of ergodicity.

DEFINITION 2.32. We say that $(X, \mathscr{X}, \nu, T)$ is

- *strongly mixing* if for all $U, V \in \mathscr{X}$, we have
$$\lim_n \nu(U \cap T^{-n}(V)) = \nu(U)\nu(V).$$
- *weakly mixing* if for all $U, V \in \mathscr{X}$, we have
$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \left| \nu(U \cap T^{-k}(V)) - \nu(U)\nu(V) \right| = 0.$$

Observe that for any sequence $(a_n)_n$ in $\mathbb{R}$, if $\lim_n a_n = 0$, then we have $\lim_n \frac{1}{n} \sum_{k=0}^{n-1} |a_k| = 0$. Therefore, strong mixing implies weak mixing. Moreover, weak mixing implies ergodicity by Corollary 2.27.

Firstly, we give examples of strongly mixing pmp dynamical systems. Let $(Y, \mathscr{Y}, \eta)$ be a probability space. Consider the product probability space $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}})$ together with the forward Bernoulli shift $S : Y^{\mathbb{N}} \to Y^{\mathbb{N}} : (y_n)_n \mapsto (y_{n+1})_n$. Likewise, consider the product probability space $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}})$ together with the Bernoulli shift $T : Y^{\mathbb{Z}} \to Y^{\mathbb{Z}} : (y_n)_n \mapsto (y_{n+1})_n$.

PROPOSITION 2.33. *The forward Bernoulli shift $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}}, S)$ is mixing. Likewise, the Bernoulli shift $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}}, T)$ is mixing.*

PROOF. We only give the proof of strong mixing of the forward Bernoulli shift. The proof of strong mixing of the Bernoulli shift is completely analogous. Set $(X, \mathscr{X}, \nu) = (Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}})$. Note that the σ-algebra $\mathscr{X}$ is generated by cylinder sets of the form $\mathscr{C}(U_0, \dots, U_{n_0}) = \prod_n Z_n$ where $Z_n = U_n \in \mathscr{Y}$ for $n \leq n_0$ and $Z_n = Y$ for $n > n_0$. In order to check the strong mixing condition, we may assume that $U, V \in \mathscr{X}$ are finite unions of cylinder sets. Then there exists $m_0 \in \mathbb{N}$ large enough such that for every $m \geq m_0$, we have
$$\nu(U \cap S^{-m}(V)) = \nu(U)\nu(S^{-m}(V)) = \nu(U)\nu(V).$$
This finishes the proof. $\square$

We make the following observation regarding the connection between topological mixing and strong mixing. Let X be a compact metrizable space and $T : X \to X$ a topological dynamical system. Let $\nu \in \mathrm{Prob}_T(X)$ be a T-invariant Borel probability measure such that $\mathrm{supp}(\nu) = X$. If the pmp dynamical system $(X, \mathscr{X}, \nu, T)$ is strongly mixing, then $T : X \to X$ is topologically mixing.

Secondly, we prove the following characterization of weak mixing.

THEOREM 2.34. *Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system. The following assertions are equivalent:*

(i) *$(X, \mathscr{X}, \nu, T)$ is weakly mixing.*
(ii) *$(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu, T \otimes T)$ is ergodic.*
(iii) *The Koopman operator $\kappa_T : \mathrm{L}^2(X, \mathscr{X}, \nu) \to \mathrm{L}^2(X, \mathscr{X}, \nu)$ has only one eigenvalue, which is 1, and moreover, the eigenvalue 1 is simple for κ_T.*

PROOF. (i) $\Rightarrow$ (ii) Let $U_1, U_2, V_1, V_2 \in \mathscr{X}$. Set $a = \nu(U_1)\nu(V_1)$ and $b = \nu(U_2)\nu(V_2)$. For every $k \in \mathbb{N}$, set $a_k = \nu(U_1 \cap T^{-k}(V_1))$ and $b_k = \nu(U_2 \cap T^{-k}(V_2))$. For every $n \geq 1$, we have

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}&(\nu \otimes \nu)(U_1 \times U_2 \cap (T \otimes T)^{-k}(V_1 \times V_2))\\&= \frac{1}{n}\sum_{k=0}^{n-1} a_k b_k = \frac{1}{n}\sum_{k=0}^{n-1}(a_k - a)b_k + \frac{1}{n}\sum_{k=0}^{n-1} a b_k.\end{aligned}$$

Since $(X, \mathscr{X}, \nu, T)$ is weakly mixing, we have $\lim_n \frac{1}{n}\sum_{k=0}^{n-1}|a_k - a| = 0$ and $\lim_n \frac{1}{n}\sum_{k=0}^{n-1}|b_k - b| = 0$. This further implies that

$$\begin{aligned}\lim_n \frac{1}{n}\sum_{k=0}^{n-1}&(\nu \otimes \nu)(U_1 \times U_2 \cap (T \otimes T)^{-k}(V_1 \times V_2))\\&= \nu(U_1)\nu(V_1)\nu(U_2)\nu(V_2)\\&= (\nu \otimes \nu)(U_1 \times U_2)(\nu \otimes \nu)(V_1 \times V_2).\end{aligned}$$

Since the σ-algebra $\mathscr{X} \otimes \mathscr{X}$ is generated by elements of the form $U \times V$ for $U, V \in \mathscr{X}$, it follows that for all $W, Z \in \mathscr{X} \otimes \mathscr{X}$, we have

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1}(\nu \otimes \nu)(W \cap (T \otimes T)^{-k}(Z)) = (\nu \otimes \nu)(W)(\nu \otimes \nu)(Z).$$

Then $(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu, T \otimes T)$ is ergodic by Corollary 2.27.

(ii) $\Rightarrow$ (i) Let $U, V \in \mathscr{X}$. Since $(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu, T \otimes T)$ is ergodic, Corollary 2.27 implies that

$$\begin{aligned}\lim_n \frac{1}{n}\sum_{k=0}^{n-1}\nu(U \cap T^{-k}(V))^2 &= \lim_n \frac{1}{n}\sum_{k=0}^{n-1}(\nu \otimes \nu)(U \times U \cap (T \otimes T)^{-k}(V \times V))\\&= \nu(U)^2\nu(V)^2 \qquad (2.3)\end{aligned}$$

Observe that by Cauchy–Schwarz inequality, for any sequence $(a_n)_n$ in $\mathbb{R}$, we have

$$\forall n \geq 1, \quad \left(\frac{1}{n}\sum_{k=0}^{n-1} a_k\right)^2 \leq \frac{1}{n}\sum_{k=0}^{n-1} a_k^2. \tag{2.4}$$

Applying (2.4) to the sequence $a_n = |\nu(U \cap T^{-n}(V)) - \nu(U)\nu(V)|$, we obtain

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{k=0}^{n-1} |\nu(U \cap T^{-k}(V)) - \nu(U)\nu(V)|\right)^2 \\
&\quad \leq \frac{1}{n}\sum_{k=0}^{n-1} |\nu(U \cap T^{-k}(V)) - \nu(U)\nu(V)|^2 \\
&\quad = \frac{1}{n}\sum_{k=0}^{n-1} \Big(\nu(U \cap T^{-k}(V))^2 + \nu(U)^2\nu(V)^2 - 2\nu(U \cap T^{-k}(V))\nu(U)\nu(V)\Big)
\end{aligned}$$

Since $(X, \mathscr{X}, \nu, T)$ is ergodic, a combination of (2.3) and Corollary 2.27 implies that

$$\frac{1}{n}\sum_{k=0}^{n-1} |\nu(U \cap T^{-k}(V)) - \nu(U)\nu(V)| = 0.$$

Then $(X, \mathscr{X}, \nu, T)$ is weakly mixing.

(ii) $\Rightarrow$ (iii) Assume that $\lambda \in \mathbb{T}$ is an eigenvalue for κ_T and choose a nonzero λ-eigenvector $\xi \in \mathrm{L}^2(X, \mathscr{X}, \nu)$. We have $\kappa_T(\xi) = \lambda\xi$. Consider $\theta : X \times X \to \mathbb{C} : (x, y) \mapsto \xi(x)\overline{\xi(y)}$. Then we have $\theta \in \mathrm{L}^2(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu)$ and $\kappa_{T\otimes T}(\theta) = \lambda\overline{\lambda}\,\theta = \theta$. Since $(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu, T \otimes T)$ is ergodic, θ is constant $(\nu \otimes \nu)$-almost everywhere. This further implies that ξ is constant ν-almost everywhere and so $\lambda = 1$.

(iii) $\Rightarrow$ (ii) By contraposition, assume that $(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu, T \otimes T)$ is not ergodic. Let $\theta \in \mathrm{L}^2(X \times X, \mathscr{X} \otimes \mathscr{X}, \nu \otimes \nu) \ominus \mathbb{C}\mathbf{1}_{X\times X}$ be a nonzero element such that $\kappa_{T\otimes T}(\theta) = \theta$. Since θ is not constant $(\nu \otimes \nu)$-almost everywhere, it follows that one of the functions

- $(x, y) \mapsto \theta(x, y) + \overline{\theta(y, x)}$
- $(x, y) \mapsto \mathrm{i}(\theta(x, y) - \overline{\theta(y, x)})$

is not constant $(\nu\otimes\nu)$-almost everywhere. Without loss of generality, we may assume that $\theta(x, y) = \overline{\theta(y, x)}$ $(\nu \otimes \nu)$-almost everywhere. Upon subtracting $\int_{X\times X} \theta \,\mathrm{d}(\nu \otimes \nu)$, we may further assume that $\int_{X\times X} \theta \,\mathrm{d}(\nu \otimes \nu) = 0$.

Using Fubini and Cauchy–Schwarz theorems, we may consider the well-defined operator

$$K_\theta : \mathrm{L}^2(X, \mathscr{X}, \nu) \to \mathrm{L}^2(X, \mathscr{X}, \nu) : \xi \mapsto \int_X \theta(\,\cdot\,, y)\xi(y)\,\mathrm{d}\nu(y).$$

Observe that K_θ is nonzero, selfadjoint and satisfies $\|K_\theta\| \leq \|\theta\|_2$. Note that

$$\begin{aligned}\|K_\theta\| &= \sup\left\{|\langle K_\theta(\xi), \eta\rangle| \mid \xi, \eta \in \mathrm{L}^2(X, \mathscr{X}, \nu), \|\xi\|_2, \|\eta\|_2 \leq 1\right\} \\ &= \sup\left\{\left|\int_{X\times X} \theta(x,y)\xi(y)\eta(x)\,\mathrm{d}(\nu\otimes\nu)(x,y)\right| \mid \|\xi\|_2, \|\eta\|_2 \leq 1\right\}.\end{aligned}$$

Since the linear span of $\left\{(x,y) \mapsto \xi(x)\eta(y), \xi, \eta \in \mathrm{L}^2(X, \mathscr{X}, \nu)\right\}$ is $\|\cdot\|_2$-dense in $\mathrm{L}^2(X\times X, \mathscr{X}\otimes\mathscr{X}, \nu\otimes\nu)$, it follows that K_θ is a norm limit of finite rank operators and so K_θ is a compact operator. Then we may choose a nonzero eigenvalue λ of K_θ whose λ-eigenspace $\mathscr{K}_\lambda \subset \mathrm{L}^2(X, \mathscr{X}, \nu)$ is necessarily finite dimensional. We claim that $\mathscr{K}_\lambda$ is κ_T-invariant. Indeed, let $\xi \in \mathscr{K}_\lambda$. Since $K_\theta(\xi) = \lambda\xi$, for ν-almost every $x \in X$, we have

$$\begin{aligned}K_\theta(\kappa_T(\xi))(x) &= \int_X \theta(x,y)\xi(T(y))\,\mathrm{d}\nu(y) \\ &= \int_X \theta(T(x),T(y))\xi(T(y))\,\mathrm{d}\nu(y) \\ &= \int_X \theta(T(x),y)\xi(y)\,\mathrm{d}\nu(y) \\ &= K_\theta(\xi)(T(x)) \\ &= \lambda\xi(T(x)) \\ &= \lambda\kappa_T(\xi)(x).\end{aligned}$$

Thus, we have $K_\theta(\kappa_T(\xi)) = \lambda\kappa_T(\xi)$ and so $\mathscr{K}_\lambda$ is κ_T-invariant. By restriction, we may now regard $\kappa_T : \mathscr{K}_\lambda \to \mathscr{K}_\lambda$ as a linear operator defined on the finite dimensional space $\mathscr{K}_\lambda$. Therefore, κ_T has a nonzero eigenvector $\eta \in \mathscr{K}_\lambda$ with respect to some eigenvalue $\mu \in \mathbb{T}$. Since $\int_{X\times X} \theta\,\mathrm{d}(\nu\otimes\nu) = 0$, $\eta \in \mathrm{L}^2(X, \mathscr{X}, \nu)$ is not constant ν-almost everywhere. Therefore, either $\mu \neq 1$ or $\mu = 1$ and $\dim\ker(\kappa_T - 1) \geq 2$. $\square$

Using Theorem 2.34, we infer that rotations are never weakly mixing. Indeed, let $\alpha \in \mathbb{R}$ and consider the rotation $T_\alpha : \mathbb{T} \to \mathbb{T} : z \mapsto \exp(\mathrm{i}2\pi\alpha)z$. Then the continuous function $\theta : (x,y) \mapsto \exp(\mathrm{i}2\pi(x-y))$ is invariant under $T_\alpha \otimes T_\alpha$ and is not invariant. More generally, rotations on compact metrizable groups are never weakly mixing.

5. Applications to random walks in $\mathrm{SL}_d(\mathbb{R})$

In this section, we give a brief introduction to the topic of matrix random products. For more information, we refer the reader to [**Fu00**].

5.1. Definition of the first Lyapunov exponent. Let $d \geq 2$. Denote by $V = \mathbb{R}^d$ the d-dimensional real vector space endowed with its canonical euclidean structure. Denote by $G = \mathrm{SL}_d(\mathbb{R})$ the special linear group.

We define the norm $\|\cdot\|$ on G by the formula

$$\|g\| = \sup\left\{\frac{\|gv\|_2}{\|v\|_2} \mid v \in V \setminus \{0\}\right\}.$$

Then we have $\|gh\| \leq \|g\|\,\|h\|$ for all $g, h \in G$.

Denote by $K = \mathrm{SO}_d(\mathbb{R}) < G$ the special orthogonal subgroup and observe that $K < G$ is compact. Define the subset $A^+ \subset G$ of diagonal matrices by

$$A^+ = \{\mathrm{diag}(\lambda_1, \dots, \lambda_d) \mid \lambda_1 \geq \dots \geq \lambda_d > 0,\ \lambda_1 \cdots \lambda_d = 1\} \subset G$$

and by $A < G$ the subgroup of diagonal matrices generated by A^+.

Lemma 2.35 (Cartan decomposition). *We have $G = K \cdot A^+ \cdot K$.*

Proof. Let $g \in G$ be a matrix. By polar decomposition, we may write $g = k_0 h$ where $k_0 \in K$ and $h \in G$ is symmetric positive definite. By diagonalization, there exists $k_2 \in K$ such that $k_2 h k_2^{-1} = a \in A^+$. Then $g = k_1 a k_2$ with $k_1 = k_0 k_2^{-1} \in K$. $\square$

As a consequence of Lemma 2.35, we infer that for every $g \in G$, we have $\|g\| \geq 1$ and $\|g^{-1}\| \leq \|g\|^{d-1}$.

Let $\mu \in \mathrm{Prob}(G)$ be a Borel probability measure on G and denote by $G_\mu = \overline{\langle \mathrm{supp}(\mu)\rangle} < G$ the closed subgroup generated by the support of μ. We will assume throughout this section that μ has a *finite first moment* meaning that

$$\int_G \log(\|g\|)\,\mathrm{d}\mu(g) < +\infty.$$

Set $(\Omega, \mathscr{B}, \mathbb{P}) = (G^{\mathbb{N}^*}, \mathscr{B}(G)^{\otimes \mathbb{N}^*}, \mu^{\otimes \mathbb{N}^*})$. Consider the forward Bernoulli shift $S : (\Omega, \mathscr{B}, \mathbb{P}) \to (\Omega, \mathscr{B}, \mathbb{P}) : (y_n)_n \mapsto (y_{n+1})_n$. Recall that the pmp dynamical system $(\Omega, \mathscr{B}, \mathbb{P}, S)$ is ergodic (see Proposition 2.8). For every $n \geq 1$ and every $\omega = (g_k)_k \in \Omega$, define $S_n(\omega) = g_n \cdots g_1$. The sequence of random products $(S_n)_n$ is called the *random walk* on G with law μ.

The following proposition provides a noncommutative analogue of the strong law of large numbers (see Corollary 2.30).

Proposition 2.36. *There exists $\lambda = \lambda_1(\mu) \in \mathbb{R}_+$ such that for $\mathbb{P}$-almost every $\omega = (g_n)_n \in \Omega$, we have*

$$\lambda_1(\mu) = \lim_n \frac{1}{n} \log(\|g_n \cdots g_1\|) = \inf_n \int_\Omega \frac{1}{n} \log(\|g_n \cdots g_1\|)\,\mathrm{d}\mathbb{P}(\omega).$$

Proof. For every $n \geq 1$, set $f_n = \log(\|S_n(\,\cdot\,)\|)$ and observe that $f_n \geq 0$. Since μ has a finite first moment, we have $f_1 \in \mathrm{L}^1(\Omega, \mathscr{B}, \mathbb{P})$. Moreover, for $\mathbb{P}$-almost every $\omega = (g_k)_k \in \Omega$ and all $m, n \geq 1$, we have

$$\begin{aligned} f_{m+n} = \log(\|S_{m+n}(\omega)\|) &= \log(\|S_m(S^n(\omega)) S_n(\omega)\|) \\ &\leq \log(\|S_m(S^n(\omega))\|) + \log(\|S_n(\omega)\|) \\ &= f_m \circ S^n + f_n. \end{aligned}$$

Then the sequence $(f_n)_n$ satisfies the subadditivity relation and in particular $(f_n)_n$ is a sequence in $\mathrm{L}^1(\Omega, \mathscr{B}, \mathbb{P})$. By Theorem 2.31, there exists $\lambda = \lambda_1(\mu) \in \mathbb{R}$ such that for $\mathbb{P}$-almost every $\omega = (g_n)_n \in \Omega$, we have

$$\begin{aligned}\lambda_1(\mu) &= \lim_n \frac{1}{n} \log(\|g_n \cdots g_1\|) \\ &= \lim_n \int_\Omega \frac{1}{n} \log(\|g_n \cdots g_1\|) \, \mathrm{d}\mathbb{P}(\omega) \\ &= \inf_{n \geq 1} \int_\Omega \frac{1}{n} \log(\|g_n \cdots g_1\|) \, \mathrm{d}\mathbb{P}(\omega) \geq 0.\end{aligned}$$

This finishes the proof. □

Definition 2.37. The nonnegative real number $\lambda_1(\mu)$ is called the *first Lyapunov exponent* of the random walk on $\mathrm{SL}_d(\mathbb{R})$ with law μ.

Recall that for every $n \geq 1$, the *convolution product* $\mu^{*n} \in \mathrm{Prob}(G)$ is defined as the pushforward measure $\mu^{*n} = \pi_{n*}\mu^{\otimes n}$, where $\pi_n : G^n \to G : (g_n, \ldots, g_1) \mapsto g_n \cdots g_1$. By definition of the convolution product, we moreover have the formula

$$\lambda_1(\mu) = \lim_n \frac{1}{n} \int_G \log(\|g\|) \, \mathrm{d}\mu^{*n}(g) = \inf_n \frac{1}{n} \int_G \log(\|g\|) \, \mathrm{d}\mu^{*n}(g).$$

5.2. Positivity of the first Lyapunov exponent. In this subsection, we follow the exposition given by Emmanuel Breuillard.

Definition 2.38. We say that a subgroup $H < \mathrm{SL}_d(\mathbb{R})$ is

- *irreducible* if $\{0\}$ and V are the only subspaces invariant under H.
- *strongly irreducible* if $\{\{0\}\}$, $\{V\}$ and $\{\{0\}, V\}$ are the only finite sets of subspaces of V invariant under H.

The main result of this subsection gives a sufficient condition regarding positivity of the first Lyapunov exponent.

Theorem 2.39 (Furstenberg). *Let $\mu \in \mathrm{Prob}(G)$ be a Borel probability measure with a finite first moment. Assume that G_μ is noncompact and strongly irreducible. Then $\lambda_1(\mu) > 0$.*

Theorem 2.39 means that under the assumptions that G_μ is noncompact and strongly irreducible, the norm of the random walk $(S_n)_n$ grows exponentially with exponential rate given by $\lambda_1(\mu) > 0$.

Firstly, we observe that it suffices to prove Theorem 2.39 under the extra assumption that $\mu(\{e\}) > 0$. Indeed, let $\mu \in \mathrm{Prob}(G)$ be a Borel probability measure with a finite first moment. Let $\varepsilon \in (0,1)$ and define $\mu_\varepsilon = \varepsilon\delta_e + (1-\varepsilon)\mu \in \mathrm{Prob}(G)$. Then $\mu_\varepsilon \in \mathrm{Prob}(G)$ still has a finite first moment and $G_{\mu_\varepsilon} = G_\mu$.

Claim 2.40. We $\lambda_1(\mu_\varepsilon) = (1-\varepsilon)\lambda_1(\mu)$.

Indeed, for every $n \geq 1$, we have $\mu_\varepsilon^{*n} = \sum_{k=0}^n C_n^k \, \varepsilon^{n-k}(1-\varepsilon)^k \mu^{*k}$. Using Proposition 2.36, for every $n \geq 1$, we have

$$\begin{aligned}
&\frac{1}{n}\int_G \log(\|g\|)\,\mathrm{d}\mu_\varepsilon^{*n}(g)\\
&\quad = (1-\varepsilon)\sum_{k=1}^n C_{n-1}^{k-1}\varepsilon^{n-k}(1-\varepsilon)^{k-1}\cdot\frac{1}{k}\int_G \log(\|g\|)\,\mathrm{d}\mu^{*k}(g)\\
&\quad \geq (1-\varepsilon)\sum_{k=1}^n C_{n-1}^{k-1}\varepsilon^{n-k}(1-\varepsilon)^{k-1}\cdot\lambda_1(\mu)\\
&\quad = (1-\varepsilon)\lambda_1(\mu).
\end{aligned}$$

Since this holds true for every $n \geq 1$, we infer that $\lambda_1(\mu_\varepsilon) \geq (1-\varepsilon)\lambda_1(\mu)$. Conversely, let $\delta > 0$ and choose $N \in \mathbb{N}$, such that $\frac{1}{k}\int_G \log(\|g\|)\,\mathrm{d}\mu^{*k}(g) \leq \lambda_1(\mu)+\delta$ for every $k \geq N+1$. For every $1 \leq k \leq N$, we have $C_{n-1}^{k-1} \leq n^N$ and so

$$\lim_n \sum_{k=1}^N C_{n-1}^{k-1}\varepsilon^{n-k}(1-\varepsilon)^{k-1}\cdot\frac{1}{k}\int_G \log(\|g\|)\,\mathrm{d}\mu^{*k}(g) = 0.$$

Therefore, we obtain

$$\begin{aligned}
\lambda_1(\mu_\varepsilon) &= \lim_n \frac{1}{n}\int_G \log(\|g\|)\,\mathrm{d}\mu_\varepsilon^{*n}(g)\\
&\leq (1-\varepsilon)\limsup_n \sum_{k=1}^N C_{n-1}^{k-1}\varepsilon^{n-k}(1-\varepsilon)^{k-1}\cdot\frac{1}{k}\int_G \log(\|g\|)\,\mathrm{d}\mu^{*k}(g)\\
&\quad + (1-\varepsilon)\limsup_n \sum_{k=N+1}^n C_{n-1}^{k-1}\varepsilon^{n-k}(1-\varepsilon)^{k-1}\cdot(\lambda_1(\mu)+\delta)\\
&\leq (1-\varepsilon)(\lambda_1(\mu)+\delta).
\end{aligned}$$

Since this holds true for every $\delta > 0$, we infer that $\lambda_1(\mu_\varepsilon) \leq (1-\varepsilon)\lambda_1(\mu)$. Therefore, we have $\lambda_1(\mu_\varepsilon) = (1-\varepsilon)\lambda_1(\mu)$.

We endow the d-dimensional real vector space V with the unique (up to multiplicative constant) infinite Lebesgue measure λ_V. Observe that the linear action $G \curvearrowright V$ preserves the Lebesgue measure λ_V. Set $\mathscr{H} = \mathrm{L}^2(V, \mathscr{B}(V), \lambda_V)$. We may then define the unitary representation $\pi : G \to \mathscr{U}(\mathscr{H})$ by the formula

$$\forall g \in G, \forall \xi \in \mathscr{H}, \quad (\pi(g)\xi)(v) = \xi(g^{-1}v).$$

Moreover, the map $\pi : G \to \mathscr{U}(\mathscr{H})$ is *strongly continuous* in the sense that for every $\xi \in \mathscr{H}$, the map $G \to \mathscr{H} : g \mapsto \pi(g)\xi$ is continuous (see Chapter 3). Then we may define the *Markov operator* $\pi(\mu) : \mathscr{H} \to \mathscr{H}$ by the formula

$$\forall \xi, \eta \in \mathscr{H}, \quad \langle \pi(\mu)\xi, \eta\rangle = \int_G \langle \pi(g)\xi, \eta\rangle\,\mathrm{d}\mu(g).$$

Observe that

$$\forall \xi, \eta \in \mathscr{H}, \quad |\langle \pi(\mu)\xi, \eta\rangle| \leq \int_G |\langle \pi(g)\xi, \eta\rangle| \, \mathrm{d}\mu(g) \leq \|\xi\| \, \|\eta\|.$$

This shows that $\|\pi(\mu)\| \leq 1$. We simply say that $\pi(\mu)$ is a contraction.

The next proposition shows that under the assumptions of Theorem 2.39, the Markov operator $\pi(\mu)$ has a spectral gap.

Proposition 2.41. *Let $\mu \in \mathrm{Prob}(G)$ be a Borel probability measure such that $\mu(\{e\}) > 0$. Assume that G_μ is noncompact and strongly irreducible. Then $\pi(\mu)$ has a* spectral gap*, meaning that $\|\pi(\mu)\| < 1$.*

Let us prove Theorem 2.39 using Proposition 2.41.

Proof of Theorem 2.39. Recall that using Lemma 2.35, for every $g \in G$, we have $\|g\| \geq 1$ and $\|g^{-1}\| \leq \|g\|^{d-1}$. Moreover, using Claim 2.40, we may assume that $\mu(\{e\}) > 0$.

Fix $c > d/2$ and define $\xi \in \mathscr{H}$ by the formula $\xi(v) = \min\{1, \|v\|^{-c}\}$ for every $v \in V$. Then for every $v \in V$ such that $1 \leq \|v\| \leq 2$ and every $g \in G$, we have

$$\xi(g^{-1}v) = \min\{1, \|g^{-1}v\|^{-c}\} \geq 2^{-c}\|g^{-1}\|^{-c} \geq 2^{-c}\|g\|^{-c(d-1)}$$

Then using Fubini's theorem, for every $n \geq 1$, we have $\pi(\mu)^n = \pi(\mu^{*n})$ and so

$$\begin{aligned}
\left\langle \pi(\mu)^n \xi, \mathbf{1}_{\{1 \leq \|v\| \leq 2\}} \right\rangle &= \int_{1 \leq \|v\| \leq 2} (\pi(\mu)^n \xi)(v) \, \mathrm{d}\lambda_V(v) \\
&= \int_G \int_{1 \leq \|v\| \leq 2} \xi(g^{-1}v) \, \mathrm{d}\lambda_V(v) \, \mathrm{d}\mu^{*n}(g) \\
&\geq 2^{-c} \int_G \int_{1 \leq \|v\| \leq 2} \|g\|^{-c(d-1)} \, \mathrm{d}\mu^{*n}(g) \, \mathrm{d}\lambda_V(v) \\
&\geq 2^{-c} \operatorname{vol}_V \left(\{1 \leq \|v\| \leq 2\}\right) \int_G \|g\|^{-c(d-1)} \, \mathrm{d}\mu^{*n}(g).
\end{aligned}$$

Therefore, we obtain

$$\int_G \|g\|^{-c(d-1)} \, \mathrm{d}\mu^{*n}(g) \leq \kappa \|\pi(\mu)\|^n$$

for some constant $\kappa > 0$ independent of $n \geq 1$. Since log is a concave function and since $(g \mapsto \|g\|^{-c(d-1)}) \in \mathrm{L}^1(G, \mathscr{B}(G), \mu^{*n})$, by Jensen's inequality, we have

$$\begin{aligned}
\int_G \log(\|g\|^{-c(d-1)}) \, \mathrm{d}\mu^{*n}(g) &\leq \log\left(\int_G \|g\|^{-c(d-1)} \, \mathrm{d}\mu^{*n}(g)\right) \\
&\leq \log(\kappa) + n \log(\|\pi(\mu)\|).
\end{aligned}$$

Using Proposition 2.41, this finally implies that

$$\lambda_1(\mu) = \lim_n \frac{1}{n} \int_G \log(\|g\|) \, \mathrm{d}\mu^{*n}(g) \geq -\frac{1}{c(d-1)} \log(\|\pi(\mu)\|) > 0.$$

This finishes the proof. $\square$

Recall that the projective space $\mathbb{P}(V)$ is a compact metrizable space. Moreover, for every nonzero vector subspace $W \subset V$, $\mathbb{P}(W) \subset \mathbb{P}(V)$ is a closed subset. We simply denote by $p : V \setminus \{0\} \to \mathbb{P}(V) : v \mapsto \mathbb{R}v$ the canonical map. The linear action $G \curvearrowright V$ naturally induces the projective action $G \curvearrowright \mathbb{P}(V)$. For every nonzero vector subspace $W \subset V$, denote by $P_W : V \to W$ the canonical orthogonal projection. Denote by $\mathrm{Gr}(V)$ the *Grassmannian manifold* that consists of all nonzero vector subspaces $W \subset V$. Define the metric $d : \mathrm{Gr}(V) \times \mathrm{Gr}(V) \to \mathbb{R}_+ : (W_1, W_2) \mapsto \|P_{W_1} - P_{W_2}\|$. Then $(\mathrm{Gr}(V), d)$ is a compact metric space.

For the proof of Proposition 2.41, we need the following useful result.

LEMMA 2.42 (Furstenberg). *Let $\eta \in \mathrm{Prob}(\mathbb{P}(V))$ be a Borel probability measure on the projective space $\mathbb{P}(V)$. Then at least one of the following assertions holds.*

- *The stabilizer group $\mathrm{Stab}_G(\eta)$ is compact.*
- *The measure η is degenerate in the sense that η is supported on $\mathbb{P}(V_1) \cup \mathbb{P}(V_2)$ where $V_1, V_2 \subset V$ are proper nonzero subspaces.*

PROOF. Assume that $H = \mathrm{Stab}_G(\eta)$ is not compact. Then using Lemma 2.35, there exists a noninvertible matrix $A \in \mathrm{M}_d(\mathbb{R})$ and a sequence $(g_n)_n$ in H such that $\lim_n \frac{1}{\|g_n\|} g_n = A$. Upon passing to a subsequence, we may further assume that $g_n(\ker A) \to V_1$ in $\mathrm{Gr}(V)$ where $V_1 \subset V$ is a nonzero subspace. Set $V_2 = \mathrm{rng}(A)$.

If $p(v) \in \mathbb{P}(\ker A)$, then any cluster point of the sequence $g_n p(v)$ necessarily lies in $\mathbb{P}(V_1)$. If $p(v) \in \mathbb{P}(V) \setminus \mathbb{P}(\ker A)$, then

$$\lim_n g_n p(v) = \lim_n p\left(\frac{1}{\|g_n\|} g_n v\right) = p(Av) \in \mathbb{P}(V_2).$$

Let $\varphi \in \mathrm{C}(\mathbb{P}(V))$ be a continuous function such that $\mathrm{supp}(\varphi) \subset \mathbb{P}(V) \setminus (\mathbb{P}(V_1) \cup \mathbb{P}(V_2))$. Then for every $v \in V \setminus \{0\}$, we have $\lim_n \varphi(g_n p(v)) = 0$. Then Lebesgue's dominated convergence theorem implies that

$$\begin{aligned}\int_{\mathbb{P}(V)} \varphi(p(v))\, \mathrm{d}\eta(p(v)) &= \int_{\mathbb{P}(V)} \varphi(p(v))\, \mathrm{d}({g_n^{-1}}_* \eta)(p(v)) \\ &= \int_{\mathbb{P}(V)} \varphi(g_n p(v))\, \mathrm{d}\eta(p(v)) \to 0.\end{aligned}$$

This shows that η is supported on $\mathbb{P}(V_1) \cup \mathbb{P}(V_2)$. $\square$

We are now ready to prove Proposition 2.41.

PROOF OF PROPOSITION 2.41. Assume that $\|\pi(\mu)\| = 1$. Denote by $\overline{\mu} \in \mathrm{Prob}(G)$ the pushforward measure of μ under the inversion map $G \to G : g \mapsto g^{-1}$. Then denote by $\overline{\mu} * \mu \in \mathrm{Prob}(G)$ the convolution product. We may consider the contractions $\pi(\overline{\mu}) : \mathscr{H} \to \mathscr{H}$ and $\pi(\overline{\mu} * \mu) : \mathscr{H} \to \mathscr{H}$.

A straightforward computation shows that $\pi(\overline{\mu}) = \pi(\mu)^*$ and $\pi(\overline{\mu} * \mu) = \pi(\overline{\mu})\pi(\mu) = \pi(\mu)^*\pi(\mu)$. Then we have

$$\|\pi(\overline{\mu} * \mu)\| = \|\pi(\mu)^*\pi(\mu)\| = \|\pi(\mu)\|^2 = 1.$$

Moreover since $\mu(\{e\}) > 0$, we clearly have $G_{\overline{\mu}*\mu} = G_\mu$. Therefore, upon replacing μ by $\overline{\mu} * \mu$, we may assume that $\mu = \overline{\mu}_0 * \mu_0$ for some Borel probability measure μ_0 and that $\|\pi(\mu)\| = 1$.

Since $\pi(\mu)$ is a selfadjoint positive operator, its spectrum $\sigma(\pi(\mu))$ is contained in the segment $[0, \|\pi(\mu)\|]$. Since $\|\pi(\mu)\| = 1$, we have $1 \in \sigma(\pi(\mu))$. Since $1 = \|\pi(\mu)\| = \sup\{\langle \pi(\mu)\xi, \xi\rangle \mid \xi \in \mathscr{H}, \|\xi\| = 1\}$, there exists a sequence $(\xi_n)_n$ of unit vectors in $\mathscr{H}$ such that $\lim_n \langle \pi(\mu)\xi_n, \xi_n\rangle = 1$.

For every $n \in \mathbb{N}$, we have

$$\int_G \|\pi(g)\xi_n - \xi_n\|^2 \,\mathrm{d}\mu(g) = 2(1 - \langle \pi(\mu)\xi_n, \xi_n\rangle).$$

Then $\lim_n \int_G \|\pi(g)\xi_n - \xi_n\|^2 \,\mathrm{d}\mu(g) = 0$. Upon passing to a subsequence, we may assume that for μ-almost every $g \in G$, we have $\lim_n \|\pi(g)\xi_n - \xi_n\|_2 = 0$. For every $n \in \mathbb{N}$ and every $g \in G$, simply write $g\xi_n = \pi(g)\xi_n \in \mathrm{L}^2(V, \mathscr{B}(V), \lambda_V)$, $g|\xi_n| = \pi(g)|\xi_n| \in \mathrm{L}^2(V, \mathscr{B}(V), \lambda_V)$ and $g|\xi_n|^2 = |g\xi_n|^2 \in \mathrm{L}^1(V, \mathscr{B}(V), \lambda_V)$. Then for μ-almost every $g \in G$, using Cauchy–Schwarz inequality, we have

$$\begin{aligned}\limsup_n \|g|\xi_n|^2 - |\xi_n|^2\|_1 &\leq \limsup_n \|g|\xi_n| + |\xi_n|\|_2 \cdot \|g|\xi_n| - |\xi_n|\|_2 \\ &\leq 2 \limsup_n \|g\xi_n - \xi_n\|_2 = 0.\end{aligned}$$

For every $n \in \mathbb{N}$, denote by $\nu_n \in \operatorname{Prob}(V)$ the Borel probability measure on V whose density with respect to λ_V is given by $|\xi_n|^2 \in \mathrm{L}^1(V, \mathscr{B}(V), \lambda_V)$. Then every $n \in \mathbb{N}$, $g_*\nu_n \in \operatorname{Prob}(V)$ is the Borel probability measure on V whose density with respect to λ_V is given by $g|\xi_n|^2 \in \mathrm{L}^1(V, \mathscr{B}(V), \lambda_V)$. Denote by $(\mathscr{M}(V), \|\cdot\|_{\mathscr{M}(V)})$ the Banach space of all bounded complex Borel measures on V. Then for μ-almost every $g \in G$, we have

$$\begin{aligned}\|g_*\nu_n - \nu_n\|_{\mathscr{M}(V)} &= 2 \sup\{|(g_*\nu_n)(A) - \nu_n(A)| \mid A \in \mathscr{B}(V)\} \\ &\leq 2 \lim_n \|g|\xi_n|^2 - |\xi_n|^2\|_1 = 0.\end{aligned}$$

Recall that $p : V \setminus \{0\} \to \mathbb{P}(V)$ is the canonical map. For every $n \in \mathbb{N}$, denote by $\eta_n = p_*\nu_n \in \operatorname{Prob}(\mathbb{P}(V))$ the pushforward measure of ν_n under p. We have $\lim_n \|g\eta_n - \eta_n\|_{\mathscr{M}(\mathbb{P}(V))} = 0$ for μ-almost every $g \in G$. Since $\mathbb{P}(V)$ is compact, choose $\eta \in \operatorname{Prob}(\mathbb{P}(V))$ a weak-$*$ limit point for the sequence $(\eta_n)_n$. We then have $g\eta = \eta$ for μ-almost every $g \in G$. By continuity of the G-action on $\mathbb{P}(V)$, we have $g\eta = \eta$ for all $g \in G_\mu$. Thus, $G_\mu \subset \operatorname{Stab}_G(\eta)$.

By Lemma 2.42, we obtain that G_μ is compact or that there exist proper nonzero subspaces $V_1, V_2 \subset V$ such that the measure η is supported on $\mathbb{P}(V_1) \cup \mathbb{P}(V_2)$. In the latter case, set

$$r = \min\{\dim W \mid \{0\} \neq W \subset V \text{ and } \eta(\mathbb{P}(W)) \neq 0\}.$$

We know that $1 \leq r < \dim V$. For all subspaces $W_1 \neq W_2$ of dimension r, we have

$$\begin{aligned}\eta(\mathbb{P}(W_1) \cup \mathbb{P}(W_2)) &= \eta(\mathbb{P}(W_1)) + \eta(\mathbb{P}(W_2)) - \eta(\mathbb{P}(W_1 \cap W_2)) \\ &= \eta(\mathbb{P}(W_1)) + \eta(\mathbb{P}(W_2)).\end{aligned}$$

More generally, for every family $(W_j)_{1 \leq j \leq k}$ of pairwise distinct subspaces of dimension r, we have

$$\eta(\mathbb{P}(W_1) \cup \cdots \cup \mathbb{P}(W_k)) = \sum_{j=1}^{k} \eta(\mathbb{P}(W_j)).$$

Thus, for every $\varepsilon > 0$, there are only finitely many subspaces $W \subset V$ of dimension r such that $\eta(\mathbb{P}(W)) \geq \varepsilon$. Set

$$\delta = \max \{\eta(\mathbb{P}(W)) \mid \dim W = r\}$$

and

$$\mathscr{F} = \{W \subset V \mid \dim W = r \text{ and } \eta(\mathbb{P}(W)) = \delta\}.$$

Then $\mathscr{F}$ is a finite set of proper subspaces of V. Since η is G_μ-invariant, for every $W \in \mathscr{F}$ and every $g \in G_\mu$, we have

$$\eta(\mathbb{P}(g^{-1}W)) = g\eta(\mathbb{P}(W)) = \eta(\mathbb{P}(W)) = \delta.$$

Therefore $g^{-1}W \in \mathscr{F}$ for every $g \in G_\mu$ and so the set $\mathscr{F}$ is G_μ-invariant. This implies that G_μ is not strongly irreducible. $\square$

6. Measure entropy

6.1. Information and Shannon entropy. Let $(X, \mathscr{X}, \nu)$ be a probability space. Let $\xi = \{A_1, \dots, A_n\}$ be a finite measurable partition of X (modulo ν-null sets). Define the *information function* of ξ by the formula

$$I_\xi = -\sum_{i=1}^{n} \log(\nu(A_i)) \mathbf{1}_{A_i}.$$

Intuitively, the value $I_\xi(x)$ measures how much information we gain from knowing that $x \in X$ belongs to one of the elements A_i of the partition ξ. Then define the *Shannon entropy* of ξ as the integral of the function I_ξ against the probability measure ν, that is,

$$H(\xi) = \int_X I_\xi \, \mathrm{d}\nu = -\sum_{i=1}^{n} \nu(A_i) \log(\nu(A_i)).$$

Intuitively, the Shannon entropy $H(\xi)$ measures the average information of the elements of the partition ξ. When we want to emphasize that we consider the Shannon entropy of ξ with respect to the probability measure ν, we write $H_\nu(\xi)$ instead of $H(\xi)$. When no confusion is possible, we simply write $H(\xi)$.

We can also define a *conditional* version of the information function and of Shannon entropy. Let $\xi = \{A_1, \dots, A_n\}$ and $\eta = \{B_1, \dots, B_p\}$ be two finite measurable partitions of X (modulo ν-null sets). Define the *conditional information function* of ξ with respect to η by the formula

$$I_{\xi,\eta} = -\sum_{j=1}^{p}\sum_{i=1}^{n} \log\left(\frac{\nu(A_i \cap B_j)}{\nu(B_j)}\right)\mathbf{1}_{A_i\cap B_j}.$$

Intuitively, the value $I_{\xi,\eta}(x)$ measures how much information we gain from knowing that $x \in X$ belongs to one of the elements A_i of the partition ξ given that we already know $x \in X$ belongs to one of the elements B_j of the partition η. Denote by $\sigma(\eta)$ the σ-subalgebra of $\mathscr{X}$ generated by η and denote by $\mathbb{E}_\nu(\,\cdot\,|\sigma(\eta)) : \mathrm{L}^\infty(X, \mathscr{X}, \nu) \to \mathrm{L}^\infty(X, \sigma(\eta), \nu)$ the unique ν-preserving conditional expectation. Then we have

$$\begin{aligned}\sum_{i=1}^{n} -\log\left(\mathbb{E}_\nu(\mathbf{1}_{A_i}|\sigma(\eta))\right)\mathbf{1}_{A_i} &= \sum_{i=1}^{n} -\log\left(\sum_{j=1}^{p}\frac{\nu(A_i\cap B_j)}{\nu(B_j)}\mathbf{1}_{B_j}\right)\mathbf{1}_{A_i}\\ &= -\sum_{i=1}^{n}\sum_{j=1}^{p}\log\left(\frac{\nu(A_i\cap B_j)}{\nu(B_j)}\right)\mathbf{1}_{B_j}\mathbf{1}_{A_i}\\ &= I_{\xi,\eta}.\end{aligned}$$

Then define the *conditional Shannon entropy* of ξ with respect to η as the integral of the function $I_{\xi,\eta}$ against the probability measure ν, that is,

$$H(\xi|\eta) = \int_X I_{\xi,\eta}\,\mathrm{d}\nu = -\sum_{j=1}^{p}\sum_{i=1}^{n}\nu(A_i\cap B_j)\log\left(\frac{\nu(A_i\cap B_j)}{\nu(B_j)}\right).$$

Intuitively, the conditional Shannon entropy $H(\xi|\eta)$ measures the average information of the elements of the partition ξ given the partition η. When we want to emphasize that we consider the conditional Shannon entropy of ξ given η with respect to the probability measure ν, we write $H_\nu(\xi|\eta)$ instead of $H(\xi|\eta)$. When no confusion is possible, we simply write $H(\xi|\eta)$. For every $j \in \{1, \dots, p\}$ such that $\nu(B_j) > 0$, define the probability measure $\nu_j \in \mathrm{Prob}(X)$ by the formula $\nu_j(A) = \frac{\nu(A\cap B_j)}{\nu(B_j)}$. Then we have

$$\sum_{j=1}^{p}\nu(B_j)H_{\nu_j}(\xi) = -\sum_{j=1}^{p}\sum_{i=1}^{n}\nu(B_j)\nu_j(A_i)\log(\nu_j(A_i)) = H(\xi|\eta).$$

If $\tau = \{X\}$ denotes the trivial partition of X, then we have $H(\xi|\tau) = H(\xi)$ for every finite measurable partition ξ of X.

Let $\xi = \{A_1, \dots, A_n\}$ and $\eta = \{B_1, \dots, B_p\}$ be finite measurable partitions of X (modulo ν-null sets). We say that η is a *refinement* of ξ and write $\xi \leq \eta$ if every element B_j of η is contained in some element A_i of ξ ν-almost everywhere. The common *refinement* $\xi \vee \eta$ is the finite measurable partition $\{A_i \cap B_j \mid 1 \leq i \leq n, 1 \leq j \leq p\}$ of X (modulo ν-null sets). We say that ξ and η are *independent* if $\nu(A_i \cap B_j) = \nu(A_i)\nu(B_j)$ for every $i \in \{1, \dots, n\}$ and every $j \in \{1, \dots, p\}$. We say that a sequence $(\xi_n)_n$ of finite measurable partitions of X is *generating* for X if the σ-algebra $\sigma((\xi_n)_n)$ generated by $\bigcup_{n\in\mathbb{N}} \xi_n$ coincides with $\mathscr{X}$ (modulo ν-null sets).

Define the continuous function $\varphi : [0,1] \to \mathbb{R}_+$ by $\varphi(0) = 0$ and $\varphi(x) = -x\log(x)$ for every $x \in (0,1]$. For every $x \in (0,1]$, we have $\varphi''(x) = -\frac{1}{x} < 0$. Then φ is strictly concave, that is, for all $n \geq 1$, all $x_1, \dots, x_n \in [0,1]$ and all $\lambda_1, \dots, \lambda_n > 0$ such that $\sum_{i=1}^n \lambda_i = 1$, we have

$$\varphi\left(\sum_{i=1}^n \lambda_i x_i\right) \geq \sum_{i=1}^n \lambda_i \varphi(x_i)$$

with equality if and only if $x_1 = \cdots = x_n$.

Next, we record the following elementary properties of (conditional) Shannon entropy that we will use without comment.

Proposition 2.43. *Let $\xi = \{A_1, \dots, A_n\}$, $\eta = \{B_1, \dots, B_p\}$ and $\zeta = \{C_1, \dots, C_q\}$ be finite measurable partitions of X. Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system. The following assertions hold:*

(i) $0 \leq H(\xi) \leq \log(n)$ *and* $H(\xi) = \log(n)$ *if and only if* $\nu(A_1) = \cdots = \nu(A_n) = \frac{1}{n}$.
(ii) *If* $\xi \leq \eta$, *then* $H(\xi|\zeta) \leq H(\eta|\zeta)$ *and if* $\zeta \leq \eta$, *then* $H(\xi|\eta) \leq H(\xi|\zeta)$.
(iii) $0 \leq H(\xi|\eta) \leq H(\xi)$ *and* $H(\xi|\eta) = H(\xi)$ *if and only if* ξ *and* η *are independent.*
(iv) $H(\xi|\eta) = 0$ *if and only if* $\xi \leq \eta$.
(v) $H(\xi \vee \eta|\zeta) = H(\xi|\zeta) + H(\eta|\xi \vee \zeta)$.
(vi) $H(\xi \vee \eta) = H(\xi) + H(\eta|\xi) \leq H(\xi) + H(\eta)$.
(vii) $H(T^{-1}(\xi)|T^{-1}(\eta)) = H(\xi|\eta)$ *and* $H(T^{-1}(\xi)) = H(\xi)$.

Proof. (i) By definition, we have $0 \leq H(\xi)$. Applying the strict concavity of φ to $x_i = \nu(A_i)$ and $\lambda_i = \frac{1}{n}$ for every $i \in \{1, \dots, n\}$, we obtain

$$\frac{1}{n}\log(n) = \varphi\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \geq \frac{1}{n}\sum_{i=1}^n \varphi(x_i) = \frac{1}{n}H(\xi).$$

Then we have $H(\xi) \leq \log(n)$ and $H(\xi) = \log(n)$ if and only if $x_1 = \cdots = x_n = \frac{1}{n}$.

(v) For every $i \in \{1, \dots, n\}$, every $j \in \{1, \dots, p\}$, every $k \in \{1, \dots, q\}$ and every $x \in A_i \cap B_j \cap C_k$, we have

$$\begin{aligned}
I_{\xi\vee\eta,\zeta}(x) &= -\log\left(\frac{\nu(A_i \cap B_j \cap C_k)}{\nu(C_k)}\right) \\
&= -\log\left(\frac{\nu(A_i \cap C_k)}{\nu(C_k)} \cdot \frac{\nu(A_i \cap B_j \cap C_k)}{\nu(A_i \cap C_k)}\right) \\
&= -\log\left(\frac{\nu(A_i \cap C_k)}{\nu(C_k)}\right) - \log\left(\frac{\nu(A_i \cap B_j \cap C_k)}{\nu(A_i \cap C_k)}\right) \\
&= I_{\xi,\zeta}(x) + I_{\eta,\xi\vee\zeta}(x).
\end{aligned}$$

This implies that $I_{\xi\vee\eta,\zeta} = I_{\xi,\zeta}+I_{\eta,\xi\vee\zeta}$ and so $H(\xi\vee\eta|\zeta) = H(\xi|\zeta)+H(\eta|\xi\vee\zeta)$ after integrating.

(ii) If $\xi \leq \eta$, then $\xi \vee \eta = \eta$ and since $H(\eta|\xi \vee \zeta) \geq 0$, we have

$$H(\eta|\zeta) = H(\xi \vee \eta|\zeta) = H(\xi|\zeta) + H(\eta|\xi \vee \zeta) \geq H(\xi|\zeta).$$

Next, assume that $\zeta \leq \eta$. For every $i \in \{1, \dots, n\}$, every $j \in \{1, \dots, p\}$, every $k \in \{1, \dots, q\}$, set $x_{i,j} = \frac{\nu(A_i\cap B_j)}{\nu(B_j)}$ and $\lambda_{j,k} = \frac{\nu(B_j\cap C_k)}{\nu(C_k)}$. Since $\zeta \leq \eta$, $\nu(B_j \cap C_k) = \nu(B_j)$ if $B_j \subset C_k$ ν-almost everwhere and $\nu(B_j \cap C_k) = 0$ otherwise. This further implies that for i, k fixed, we have $\sum_{j=1}^{p} \lambda_{j,k}x_{i,j} = \sum_{B_j\subset C_k} \frac{\nu(A_i\cap B_j)}{\nu(C_k)} = \frac{\nu(A_i\cap C_k)}{\nu(C_k)}$. Using concavity of φ, we infer that

$$\begin{aligned}
H(\xi|\eta) &= \sum_{j=1}^{p}\sum_{i=1}^{n} \nu(B_j)\varphi(x_{i,j}) \\
&= \sum_{j=1}^{p}\sum_{i=1}^{n} \left(\sum_{k=1}^{q} \nu(C_k)\lambda_{j,k}\right) \varphi(x_{i,j}) \\
&= \sum_{k=1}^{q}\sum_{i=1}^{n} \nu(C_k) \sum_{j=1}^{p} \lambda_{j,k}\varphi(x_{i,j}) \\
&\leq \sum_{k=1}^{q}\sum_{i=1}^{n} \nu(C_k)\varphi\left(\sum_{j=1}^{p} \lambda_{j,k}x_{i,j}\right) \\
&= \sum_{k=1}^{q}\sum_{i=1}^{n} \nu(C_k)\varphi\left(\frac{\nu(A_i \cap C_k)}{\nu(C_k)}\right) = H(\xi|\zeta).
\end{aligned}$$

(iii) By definition, we have $0 \leq H(\xi|\eta)$. Since $\tau \leq \eta$, we have $H(\xi|\eta) \leq H(\xi|\tau) = H(\xi)$. If ξ and η are independent, then we have

$$\begin{aligned}
H(\xi|\eta) &= -\sum_{j=1}^{p}\sum_{i=1}^{n} \nu(A_i \cap B_j) \log\left(\frac{\nu(A_i \cap B_j)}{\nu(B_j)}\right) \\
&= -\sum_{j=1}^{p}\sum_{i=1}^{n} \nu(A_i \cap B_j) \log(\nu(A_i)) \\
&= -\sum_{i=1}^{n} \nu(A_i) \log(\nu(A_i)) = H(\xi).
\end{aligned}$$

Conversely, assume that $H(\xi|\eta) = H(\xi)$. Then we have

$$\sum_{i=1}^{n} \left(-\sum_{j=1}^{p} \nu(B_j)\varphi\left(\frac{\nu(A_i \cap B_j)}{\nu(B_j)}\right) + \varphi(\nu(A_i))\right) = 0.$$

For every fixed $i \in \{1, \dots, n\}$, set $\lambda_j = \nu(B_j)$ and $x_j = \frac{\nu(A_i\cap B_j)}{\nu(B_j)}$ and apply strict concavity of φ. Then the quantity $\frac{\nu(A_i\cap B_j)}{\nu(B_j)}$ does not depend on j and we set $\lambda_i = \frac{\nu(A_i \cap B_j)}{\nu(B_j)}$. Then we have

$$\lambda_i = \sum_{j=1}^{p} \lambda_i \nu(B_j) = \sum_{j=1}^{p} \nu(A_i \cap B_j) = \nu(A_i).$$

This further implies that ξ and η are independent.

(iv) If $\xi \leq \eta$, then for every $i \in \{1, \ldots, n\}$ and every $j \in \{1, \ldots, p\}$, we have either $\nu(A_i \cap B_j) = 0$ or $\frac{\nu(A_i \cap B_j)}{\nu(B_j)} = 1$. Then we have

$$H(\xi|\eta) = -\sum_{j=1}^{p} \sum_{i=1}^{n} \nu(A_i \cap B_j) \log \left(\frac{\nu(A_i \cap B_j)}{\nu(B_j)} \right) = 0.$$

Conversely, assume that

$$0 = H(\xi|\eta) = \sum_{j=1}^{p} \sum_{i=1}^{n} \nu(B_j) \varphi \left(\frac{\nu(A_i \cap B_j)}{\nu(B_j)} \right).$$

Then for every $j \in \{1, \ldots, p\}$ and every $i \in \{1, \ldots, n\}$ such that $\nu(B_j) > 0$ and $\nu(A_i \cap B_j) > 0$, we have $\varphi(\frac{\nu(A_i \cap B_j)}{\nu(B_j)}) = 0$ which further implies that $\nu(A_i \cap B_j) = \nu(B_j)$ and so $B_j \subset A_i$ ν-almost everywhere. This shows that $\xi \leq \eta$.

(vi) We may apply (v) to ξ, η and $\zeta = \tau$ the trivial partition and we obtain

$$H(\xi \vee \eta) = H(\xi \vee \eta | \tau) = H(\xi|\tau) + H(\eta|\xi \vee \tau) = H(\xi) + H(\eta|\xi) \leq H(\xi) + H(\eta).$$

(vii) Since $T_* \nu = \nu$, this is obvious from the definitions. $\square$

6.2. Measure entropy of a pmp dynamical system. We fix a pmp dynamical system $(X, \mathscr{X}, \nu, T)$. For every finite measurable partition ξ of X and every $n \geq 1$, define the finite measurable partition $\xi_n = \xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-n+1}(\xi)$. For all $m, n \geq 1$, we have

$$\begin{aligned} H_\nu(\xi_{m+n}) &= H_\nu(\xi_m \vee T^{-m}(\xi_n)) \\ &\leq H_\nu(\xi_m) + H_\nu(T^{-m}(\xi_n)) \\ &= H_\nu(\xi_m) + H_\nu(\xi_n). \end{aligned}$$

Since the sequence $(H_\nu(\xi_n))_n$ is subadditive, Lemma 1.17 implies that the sequence $(\frac{1}{n} H_\nu(\xi_n))_n$ is convergent and we set

$$h_\nu(T, \xi) = \lim_n \frac{1}{n} H_\nu(\xi_n) = \inf_n \frac{1}{n} H_\nu(\xi_n).$$

Then $h_\nu(T, \xi)$ is the *measure entropy* of T with respect to the finite measurable partition ξ.

The next proposition shows that $h_\nu(T, \xi)$ is the average information added by the present state on condition that all past states are known.

PROPOSITION 2.44. *Let ξ be a finite measurable partition of X. Then $h_\nu(T, \xi) = \lim_n H_\nu(\xi | T^{-1}(\xi_n))$.*

PROOF. Since the sequence $(T^{-1}(\xi_n))_n$ is increasing, it follows that the sequence $(H_\nu(\xi|T^{-1}(\xi_n))_n$ is decreasing and so it is convergent. For every $n \geq 1$, we have $H_\nu(\xi_{n+1}) = H_\nu(T^{-1}(\xi_n)) + H_\nu(\xi|T^{-1}(\xi_n))$ and so

$$H_\nu(\xi|T^{-1}(\xi_n)) = H_\nu(\xi_{n+1}) - H_\nu(T^{-1}(\xi_n)) = H_\nu(\xi_{n+1}) - H_\nu(\xi_n).$$

By summation, we obtain

$$H_\nu(\xi_n) - H_\nu(\xi) = \sum_{k=1}^{n} H_\nu(\xi|T^{-1}(\xi_k)).$$

Dividing by $n \geq 1$ and passing to the limit, we obtain

$$h_\nu(T, \xi) = \lim_n \frac{1}{n} \sum_{k=1}^{n} H_\nu(\xi|T^{-1}(\xi_k)).$$

By Cesàro average, we necessarily have $h_\nu(T, \xi) = \lim_n H_\nu(\xi|T^{-1}(\xi_n))$. $\square$

We record the following elementary properties that we will use without comment.

PROPOSITION 2.45. *Let ξ and η be finite measurable partitions of X. Then the following assertions hold:*

(i) $h_\nu(T, \xi) = h_\nu(T, T^{-1}(\xi))$. *If $(X, \mathscr{X}, \nu, T)$ is invertible, then we have $h_\nu(T, \xi) = h_\nu(T, T(\xi))$.*
(ii) $h_\nu(T, \xi) = h_\nu(T, \bigvee_{i=0}^{k} T^{-i}(\xi))$ *for every $k \in \mathbb{N}$. If $(X, \mathscr{X}, \nu, T)$ is invertible, then $h_\nu(T, \xi) = h_\nu(T, \bigvee_{i=-k}^{k} T^{-i}(\xi))$ for every $k \in \mathbb{N}$.*
(iii) $h_\nu(T, \xi) \leq h_\nu(T, \eta) + H_\nu(\xi|\eta)$ *and if $\xi \leq \eta$, then $h_\nu(T, \xi) \leq h_\nu(T, \eta)$.*
(iv) $h_\nu(T, \xi \vee \eta) \leq h_\nu(T, \xi) + h_\nu(T, \eta)$.

PROOF. (i) For every $n \geq 1$, we have $T^{-1}(\xi)_n = T^{-1}(\xi) \vee \cdots \vee T^{-n}(\xi) = T^{-1}(\xi_n)$. This implies that

$$h_\nu(T, T^{-1}(\xi)) = \lim_n \frac{1}{n} H_\nu(T^{-1}(\xi_n)) = \lim_n \frac{1}{n} H_\nu(\xi_n) = h_\nu(T, \xi).$$

Assume moreover that $(X, \mathscr{X}, \nu, T)$ is invertible. Then with respect to the transformation T^{-1}, for every $n \geq 1$, we have $T(\xi)_n = T(\xi) \vee \cdots \vee T^n(\xi) = T(\xi_n)$. This implies that

$$h_\nu(T, T(\xi)) = \lim_n \frac{1}{n} H_\nu(T(\xi_n)) = \lim_n \frac{1}{n} H_\nu(\xi_n) = h_\nu(T, \xi).$$

(ii) For every $k \in \mathbb{N}$ and every $n \in \mathbb{N}$, we have $\xi_n \leq (\bigvee_{i=0}^{k} T^{-i}(\xi))_n$ and so $H_\nu(\xi_n) \leq H_\nu((\bigvee_{i=0}^{k} T^{-i}(\xi))_n)$. This implies that

$$h_\nu(T, \xi) = \lim_n \frac{1}{n} H_\nu(\xi_n) \leq \lim_n \frac{1}{n} H_\nu((\bigvee_{i=0}^{k} T^{-i}(\xi))_n) = h_\nu(T, \bigvee_{i=0}^{k} T^{-i}(\xi)).$$

For the reverse inequality, for every $k \in \mathbb{N}$ and every $n \in \mathbb{N}$, we have $(\bigvee_{i=0}^{k} T^{-i}(\xi))_n = \xi_n \vee T^{-n}(\xi \vee \cdots \vee T^{-(k-1)}(\xi))$ and so

$$\begin{aligned} H_\nu(\bigvee_{i=0}^{k} T^{-i}(\xi))_n) &= H_\nu(\xi_n \vee T^{-n}(\xi \vee \cdots \vee T^{-(k-1)}(\xi))) \\ &\le H_\nu(\xi_n) + H_\nu(T^{-n}(\xi \vee \cdots \vee T^{-(k-1)}(\xi))) \\ &\le H_\nu(\xi_n) + kH_\nu(\xi). \end{aligned}$$

This implies that

$$\begin{aligned} h_\nu(T, \bigvee_{i=0}^{k} T^{-i}(\xi)) &= \lim_n \frac{1}{n} H_\nu(\bigvee_{i=0}^{k} T^{-i}(\xi))_n) \\ &\le \lim_n \frac{1}{n} H_\nu(\xi_n) + \lim_n \frac{k}{n} H_\nu(\xi) \\ &= h_\nu(T, \xi). \end{aligned}$$

Assume moreover that $(X, \mathscr{X}, \nu, T)$ is invertible. Then the exact same argument as above shows that $h_\nu(T, \xi) = h_\nu(T, \bigvee_{i=-k}^{k} T^{-i}(\xi))$ for every $k \in \mathbb{N}$.

(iii) For every $n \ge 1$, we have

$$H_\nu(\xi_n|\eta_n) \le \sum_{i=0}^{n-1} H_\nu(T^{-i}(\xi)|\eta_n) \le \sum_{i=0}^{n-1} H_\nu(T^{-i}(\xi)|T^{-i}(\eta)) = n \cdot H_\nu(\xi|\eta)$$

and so

$$H_\nu(\xi_n) \le H_\nu(\eta_n \vee \xi_n) = H_\nu(\eta_n) + H_\nu(\xi_n|\eta_n) = H_\nu(\eta_n) + n \cdot H_\nu(\xi|\eta).$$

This implies that

$$h_\nu(T, \xi) = \lim_n \frac{1}{n} H(\xi_n) \le \lim_n \frac{1}{n} H_\nu(\eta_n) + H_\nu(\xi|\eta) = h_\nu(T, \eta) + H_\nu(\xi|\eta).$$

If $\xi \le \eta$, then $H_\nu(\xi|\eta) = 0$ and so $h_\nu(T, \xi) \le h_\nu(T, \eta)$.

(iv) For every $n \ge 1$, we have $(\xi \vee \eta)_n = \xi_n \vee \eta_n$ and so

$$H_\nu((\xi \vee \eta)_n) = H_\nu(\xi_n \vee \eta_n) \le H_\nu(\xi_n) + H_\nu(\eta_n).$$

This implies that

$$h_\nu(T, \xi \vee \eta) \le \lim_n \frac{1}{n} H_\nu(\xi_n) + \lim_n \frac{1}{n} H_\nu(\eta_n) = h_\nu(T, \xi) + h_\nu(T, \eta).$$

This finishes the proof. □

Definition 2.46. The *measure entropy* of $(X, \mathscr{X}, \nu, T)$ is defined as

$$h_\nu(T) = \sup_\xi h_\nu(T, \xi)$$

where the supremum is taken over all finite measurable partitions ξ of X.

The measure entropy is an invariant of measurable conjugacy meaning that if two pmp dynamical systems $(X_1, \mathscr{X}_1, \nu_1, T_1)$ and $(X_2, \mathscr{X}_2, \nu_2, T_2)$ are measurably conjugate, then they must have the same measure entropy, that is, $h_{\nu_1}(T_1) = h_{\nu_2}(T_2)$.

Let ξ be a finite measurable partition of X. If $(X, \mathscr{X}, \nu, T)$ is noninvertible, then we say that ξ is a *generator* for T if the sequence $\xi_n = \bigvee_{i=0}^{n} T^{-i}(\xi)$ is generating for X. If $(X, \mathscr{X}, \nu, T)$ is invertible, then we say that ξ is a *generator* for T if the sequence $\xi_n = \bigvee_{i=-n}^{n} T^{-i}(\xi)$ is generating for X. The following result due to Kolmogorov and Sinai allows in many situations to calculate $h_\nu(T)$.

THEOREM 2.47. *Let ξ be a finite measurable partition of X. If ξ is a generator for T, then $h_\nu(T) = h_\nu(T, \xi)$.*

Before proving Theorem 2.47, we need the following technical result.

LEMMA 2.48. *Let ξ be a finite measurable partition of X and $\varepsilon > 0$. Then there exists $\delta > 0$ such that for every finite measurable partition η of X with the property that for every element A of ξ, there exists an element B of the σ-algebra $\sigma(\eta)$ generated by η satisfying $\nu(A \triangle B) < \delta$, we have $H_\nu(\xi|\eta) < \varepsilon$.*

PROOF. Write $\xi = \{A_1, \dots, A_n\}$. Let $\delta, \rho > 0$ with δ to be determined in relation to ρ and ρ in relation to ε. Let η be a finite measurable partition of X and denote by $\sigma(\eta)$ the σ-algebra generated by η. Assume that for every $i \in \{1, \dots, n\}$, there exists an element B_i of $\sigma(\eta)$ such that $\nu(A_i \triangle B_i) < \delta$. Define the finite measurable partition $\zeta = \{C_1, \dots, C_n\}$ of X recursively by $C_1 = B_1$, $C_{i+1} = B_{i+1} \setminus (C_1 \cup \cdots \cup C_i)$ for every $i \in \{1, \dots, n-2\}$ and $C_n = X \setminus (C_1 \cup \cdots \cup C_{n-1})$. Recall that

$$H_\nu(\xi|\zeta) = -\sum_{j=1}^{n} \sum_{i=1}^{n} \nu(A_i \cap C_j) \log\left(\frac{\nu(A_i \cap C_j)}{\nu(C_j)}\right).$$

If we choose $\delta > 0$ small enough in relation to ρ, then by construction, we have $\frac{\nu(A_i \cap C_i)}{\nu(C_i)} > 1 - \rho$ for all $i \in \{1, \dots, n\}$ and $\nu(A_i \cap C_j) < \rho$ for all $i, j \in \{1, \dots, n\}$ such that $i \neq j$. If we choose $\rho > 0$ small enough in relation to ε, it is clear that $H_\nu(\xi|\zeta) < \varepsilon$. Since η refines ζ, that is, $\zeta \leq \eta$, we have $H_\nu(\xi|\eta) \leq H_\nu(\xi|\zeta) < \varepsilon$. $\square$

We are now ready to prove Theorem 2.47.

PROOF OF THEOREM 2.47. Since the proofs of the noninvertible case and the invertible case are completely analogous, we only prove the noninvertible case. Let $(X, \mathscr{X}, \nu, T)$ be a noninvertible pmp dynamical system and ξ a generator for T. For every $n \in \mathbb{N}$, set $\xi_n = \bigvee_{i=0}^{n-1} T^{-i}(\xi)$. Let η be a finite measurable partition of X. We show that $h_\nu(T, \eta) \leq h_\nu(T, \xi)$. Let $\varepsilon > 0$ and choose $\delta > 0$ according to Lemma 2.48. Since the σ-algebra generated by $\bigcup_{n\in\mathbb{N}} \xi_n$ coincides with $\mathscr{X}$ (modulo ν-null sets), there exists

$n \in \mathbb{N}$ large enough such that for every element A of η, there exists an element B of the σ-algebra $\sigma(\xi_n)$ generated by ξ_n satisfying $\nu(A \triangle B) < \delta$. Then we have $H_\nu(\eta|\xi_n) \leq \varepsilon$. This further implies that

$$h_\nu(T, \eta) \leq h_\nu(T, \xi_n) + H_\nu(\eta|\xi_n) = h_\nu(T, \xi) + H_\nu(\eta|\xi_n) \leq h_\nu(T, \xi) + \varepsilon.$$

Since this holds true for every $\varepsilon > 0$, it follows that $h_\nu(T, \eta) \leq h_\nu(T, \xi)$. Since this holds true for every finite measurable partition η of X, it follows that $h_\nu(T) = h_\nu(T, \xi)$. □

An increasing sequence $(\zeta_m)_m$ of finite measurable partitions of X is said to be *generating* if the σ-algebra generated by $\bigcup_{m \in \mathbb{N}} \zeta_m$ coincides with $\mathscr{X}$ (modulo ν-null sets). Observe that any standard Borel probability space possesses a generating increasing sequence $(\zeta_m)_m$ of finite measurable partitions of X (see [**KL16**, Appendix A]). By modifying the proof of Theorem 2.47, we can prove the following useful result.

PROPOSITION 2.49. *Let $(\zeta_m)_m$ be a generating increasing sequence of finite measurable partitions of X. Then we have $h_\nu(T) = \lim_m h_\nu(T, \zeta_m)$.*

PROOF. Since the sequence $(\zeta_m)_m$ is increasing, Proposition 2.45 implies that $(h_\nu(T, \zeta_m))_m$ is increasing and so $\lim_m h_\nu(T, \zeta_m) = \sup_m h_\nu(T, \zeta_m) \leq h_\nu(T)$. Let ξ be an arbitrary finite measurable partition of X. We show that $h_\nu(T, \xi) \leq \lim_m h_\nu(T, \zeta_m)$. Let $\varepsilon > 0$ and choose $\delta > 0$ according to Lemma 2.48. Since the σ-algebra generated by $\bigcup_{m \in \mathbb{N}} \zeta_m$ coincides with $\mathscr{X}$ (modulo ν-null sets), there exists $m \in \mathbb{N}$ large enough such that for every element A of ξ, there exists an element B of the σ-algebra $\sigma(\zeta_m)$ generated by ζ_m satisfying $\nu(A \triangle B) < \delta$. Then we have $H_\nu(\xi|\zeta_m) \leq \varepsilon$. This further implies that

$$h_\nu(T, \xi) \leq h_\nu(T, \zeta_m) + H_\nu(\xi|\zeta_m) \leq \lim_n h_\nu(T, \zeta_n) + \varepsilon.$$

Since this holds true for every $\varepsilon > 0$, we have $h_\nu(T, \xi) \leq \lim_m h_\nu(T, \zeta_m)$. By taking the supremum over all finite measurable partitions ξ of X, we obtain $h_\nu(T) \leq \lim_m h_\nu(T, \zeta_m)$. Therefore, we have $h_\nu(T) = \lim_m h_\nu(T, \zeta_m)$. □

We collect some useful properties of measure entropy.

PROPOSITION 2.50. *Let $(X, \mathscr{X}, \nu, T)$ be a pmp dynamical system on a standard probability space.*

(i) *For every $m \in \mathbb{N}$, we have $h_\nu(T^m) = m\, h_\nu(T)$.*
(ii) *If $(X, \mathscr{X}, \nu, T)$ is invertible, then $h_\nu(T^{-1}) = h_\nu(T)$. Thus, for every $m \in \mathbb{Z}$, we have $h_\nu(T^m) = |m|\, h_\nu(T)$.*

For every $i \in \{1, 2\}$, let $(X_i, \mathscr{X}_i, \nu_i, T_i)$ be a pmp dynamical system on a standard probability space.

(iii) *We have $h_{\nu_1 \otimes \nu_2}(T_1 \times T_2) = h_{\nu_1}(T_1) + h_{\nu_2}(T_2)$.*
(iv) *If $(X_2, \mathscr{X}_2, \nu_2, T_2)$ is a pmp factor of $(X_1, \mathscr{X}_1, \nu_1, T_1)$, then we have $h_{\nu_2}(T_2) \leq h_{\nu_1}(T_1)$.*

PROOF. (i) Let $m \geq 1$. Let ξ be a finite measurable partition of X. We have

$$\begin{aligned} h_\nu(T^m, \xi) &= \lim_n \frac{1}{n} H_\nu(\xi \vee T^{-m}(\xi) \vee \cdots \vee T^{-m(n-1)}(\xi)) \\ &\leq m \lim_n \frac{1}{mn} H_\nu(\xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-mn+1}(\xi)) \\ &= m\, h_\nu(T, \xi) \\ &\leq m\, h_\nu(T). \end{aligned}$$

By taking the supremum over all finite measurable partitions ξ of X, we obtain $h_\nu(T^m) \leq m\, h_\nu(T)$.

Conversely, let ξ be a finite measurable partition of X and set $\eta = \xi \vee \cdots \vee T^{-m+1}(\xi)$. We have

$$\begin{aligned} h_\nu(T, \xi) &= \frac{1}{m} \lim_n \frac{1}{n} H_\nu(\xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-mn+1}(\xi)) \\ &= \frac{1}{m} \lim_n \frac{1}{n} H_\nu(\eta \vee T^{-m}(\eta) \vee \cdots \vee T^{-m(n-1)}(\eta)) \\ &= \frac{1}{m} h_\nu(T^m, \eta) \\ &\leq \frac{1}{m} h_\nu(T^m). \end{aligned}$$

By taking the supremum over all finite measurable partitions ξ of X, we obtain $h_\nu(T) \leq \frac{1}{m} h_\nu(T^m)$. Therefore, we have $h_\nu(T^m) = m\, h_\nu(T)$.

(ii) Let ξ be a finite measurable partition of X. We have

$$\begin{aligned} h_\nu(T, \xi) &= \lim_n \frac{1}{n} H_\nu(\xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-n+1}(\xi)) \\ &= \lim_n \frac{1}{n} H_\nu(T^{n-1}(\xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-n+1}(\xi))) \\ &= \lim_n \frac{1}{n} H_\nu(T^{n-1}(\xi) \vee \cdots \vee T(\xi) \vee \xi) \\ &= h_\nu(T^{-1}, \xi). \end{aligned}$$

By taking the supremum over all finite measurable partitions ξ of X, we obtain $h_\nu(T) = h_\nu(T^{-1})$.

(iii) For every $i \in \{1, 2\}$, let $(\zeta^i_m)_m$ be a generating increasing sequence of finite measurable partitions of X_i. For every $m \in \mathbb{N}$, set $\widehat{\zeta}^1_m = \zeta^1_m \times X_2$ and $\widehat{\zeta}^2_m = X_1 \times \zeta^2_m$. Then $(\widehat{\zeta}^1_m)_m$ and $(\widehat{\zeta}^1_m)_m$ are independent and $(\widehat{\zeta}^1_m \vee \widehat{\zeta}^2_m)_m$ is a generating increasing sequence of finite measurable partitions of $X_1 \times X_2$. Using Proposition 2.49, we have

$$\begin{aligned} h_{\nu_1 \otimes \nu_2}(T_1 \times T_2) &= \lim_m h_{\nu_1 \otimes \nu_2}(T_1 \times T_2, \widehat{\zeta}^1_m \vee \widehat{\zeta}^2_m) \\ &= \lim_m \lim_n \frac{1}{n} H_{\nu_1 \otimes \nu_2}((\widehat{\zeta}^1_m \vee \widehat{\zeta}^2_m)_n) \\ &= \lim_m \lim_n \frac{1}{n} \Big(H_{\nu_1}((\widehat{\zeta}^1_m)_n) + H_{\nu_2}((\widehat{\zeta}^2_m)_n) \Big) \end{aligned}$$

$$= \lim_m \left(h_{\nu_1}(T_1, \widehat{\zeta}^1_m) + h_{\nu_2}(T_2, \widehat{\zeta}^2_m) \right)$$
$$= h_{\nu_1}(T_1) + h_{\nu_2}(T_2).$$

(iv) Let η be a finite measurable partition of X_2. Then $\xi = \pi^{-1}(\eta)$ is a finite measurable partition of X_1. For every $n \geq 1$, we have

$$\begin{aligned} H_{\nu_2}(\eta \vee \cdots \vee T_2^{-n+1}(\eta)) &= H_{\nu_1}(\pi^{-1}(\eta \vee \cdots \vee T_2^{-n+1}(\eta))) \\ &= H_{\nu_1}(\xi \vee \cdots \vee T_1^{-n+1}(\xi)). \end{aligned}$$

This implies that

$$\begin{aligned} h_{\nu_2}(T_2, \eta) &= \lim_n \frac{1}{n} H_{\nu_2}(\eta \vee \cdots \vee T_2^{-n+1}(\eta)) \\ &= \lim_n \frac{1}{n} H_{\nu_1}(\xi \vee \cdots \vee T_1^{-n+1}(\xi)) \\ &= h_{\nu_1}(T_1, \xi) \\ &\leq h_{\nu_1}(T_1). \end{aligned}$$

By taking the supremum over all finite measurable partitions η of X_2, we obtain $h_{\nu_2}(T_2) \leq h_{\nu_1}(T_1)$. $\square$

We use Theorem 2.47 to compute the measure entropy of Bernoulli shifts.

Proposition 2.51. *Let $r \geq 2$. Set $Y = \{1, \dots, r\}$ and $\mathscr{Y} = \mathscr{P}(Y)$. Let $\eta \in \operatorname{Prob}(Y)$ be an arbitrary probability measure on Y. Consider the forward Bernoulli shift $(Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}}, S_r)$ as well as the Bernoulli shift $(Y^{\mathbb{Z}}, \mathscr{Y}^{\otimes \mathbb{Z}}, \eta^{\otimes \mathbb{Z}}, T_r)$. Then $h_{\eta^{\otimes \mathbb{N}}}(S_r) = h_{\eta^{\otimes \mathbb{Z}}}(T_r) = \sum_{i=1}^r \varphi(\eta(i))$.*

Proof. We only prove the result for the forward Bernoulli shift. The proof for the Bernoulli shift is completely analogous. Set $(X, \mathscr{X}, \nu) = (Y^{\mathbb{N}}, \mathscr{Y}^{\otimes \mathbb{N}}, \eta^{\otimes \mathbb{N}})$. For every $1 \leq i \leq r$, define the cylinder set

$$A_i = \{(y_n)_n \in X \mid y_0 = i\}.$$

Then $\xi = \{A_1, \dots, A_r\}$ is a finite measurable partition of X that is a generator for S_r. Moreover, we have $\nu(A_i) = \eta(i)$ for every $1 \leq i \leq r$. By Theorem 2.47, we have $h_\nu(S_r) = h_\nu(S_r, \xi) = \lim_n \frac{1}{n} H_\nu(\xi_n)$. By Proposition 2.43, we have $H_\nu(\xi_n) = n \cdot H_\nu(\xi)$ for every $n \geq 1$. Therefore, we have $h_\nu(S_r) = H_\nu(\xi) = \sum_{i=1}^r \varphi(\eta(i))$. $\square$

Observe that when $\eta = \eta_r \in \operatorname{Prob}(Y)$ is the uniform measure, meaning that $\eta_r(i) = \frac{1}{r}$ for every $1 \leq i \leq r$, Proposition 2.51 implies that $h_{\eta_r^{\otimes \mathbb{N}}}(S_r) = h_{\eta_r^{\otimes \mathbb{Z}}}(T_r) = \log(r)$. Since the measure entropy is an invariant of measurable conjugacy, it follows that the Bernoulli shifts

$$(\{1, \dots, r\}^{\otimes \mathbb{Z}}, \mathscr{P}(\{1, \dots, r\})^{\otimes \mathbb{Z}}, \eta_r^{\otimes \mathbb{Z}}, T_r)_{r \geq 1}$$

are pairwise not measurable conjugate. Also, observe that in this case, the measure entropy $h_{\eta_r^{\otimes \mathbb{N}}}(S_r) = h_{\eta_r^{\otimes \mathbb{Z}}}(T_r) = \log(r)$ coincides with the topological entropy $h(S_r) = h(T_r) = \log(r)$ (see Proposition 1.25).

We also compute the measure entropy of rotations on compact metrizable groups. Before doing so, we need the following lemma.

Lemma 2.52. *Let $\xi_1, \dots, \xi_m$ be finite measurable partitions of X. Then we have*

$$H_\nu(\xi_1 \vee \dots \vee \xi_m) \leq H_\nu(\xi_1) + \sum_{j=2}^{m} H_\nu(\xi_j | \xi_1).$$

Proof. Using repeatedly Proposition 2.43, we have

$$\begin{aligned} H_\nu(\xi_1 \vee \dots \vee \xi_m) &= H_\nu(\xi_1 \vee \dots \vee \xi_{n-1}) + H_\nu(\xi_n | \xi_1 \vee \dots \vee \xi_{n-1}) \\ &= \cdots \\ &= H_\nu(\xi_1) + \sum_{j=2}^{m} H_\nu(\xi_j | \xi_1 \vee \dots \vee \xi_{j-1}) \\ &\leq H_\nu(\xi_1) + \sum_{j=2}^{m} H_\nu(\xi_j | \xi_1). \end{aligned}$$

This finishes the proof. □

Proposition 2.53. *Let G be a compact metrizable group and denote by $\mathscr{B}(G)$ its σ-algebra of Borel subsets and by m_G its unique Haar Borel probability measure. Let $g \in G$ and consider the rotation $T_g : G \to G : x \mapsto gx$. Then $h_{m_G}(T_g) = 0$.*

Proof. Simply write $T = T_g$ and $\nu = m_G$. Let $\varepsilon > 0$ and $\xi = \{A_1, \dots, A_m\}$ be a finite measurable partition of G. We start by proving the following key technical result.

Claim 2.54. There exists a finite partition $\mathbb{Z} = C_1 \sqcup \dots \sqcup C_r$ such that for every $1 \leq j \leq r$ and all $p, q \in C_j$, we have $H_\nu(T^p(\xi) | T^q(\xi)) \leq \varepsilon$.

Fix a left invariant compatible metric $d_G : G \times G \to \mathbb{R}_+$. Consider the left regular unitary representation $\lambda_G : G \to \mathscr{U}(\mathrm{L}^2(G, \mathscr{B}(G), m_G))$ (see Chapter 3). Then for every $A \in \mathscr{B}(G)$, the map

$$G \to \mathbb{R}_+ : h \mapsto \nu(hA \triangle A) = \|\lambda_G(h)(\mathbf{1}_A) - \mathbf{1}_A\|_2^2$$

is continuous. Choose $\delta > 0$ according to Lemma 2.48. Then choose $\rho > 0$ such that for every $h \in B(e, \rho)$, we have $\nu(hA_i \triangle A_i) < \delta$ for every $1 \leq i \leq m$. Observe that for every $n \in \mathbb{Z}$, we have $\kappa_{T_{g^n}} = \lambda_G(g^n)$. Since $G = \bigcup_{h \in G} B(h, \rho)$, by compactness, there exist $h_1, \dots, h_r \in G$ such that $G = \bigcup_{j=1}^{r} B(h_j, \rho)$. Define recursively the finite partition $\mathbb{Z} = C_1 \sqcup \dots \sqcup C_r$ by $C_1 = \{n \in \mathbb{Z} \mid g^n \in B(h_1, \rho)\}$ and $C_j = \{n \in \mathbb{Z} \mid g^n \in B(h_j, \rho)\} \setminus (C_1 \cup \dots \cup C_{j-1})$ for every $2 \leq j \leq r$. Then using Lemma 2.48, for every $1 \leq j \leq r$ and all $p, q \in C_j$, we have $H_\nu(T^p(\xi) | T^q(\xi)) \leq \varepsilon$. This finishes the proof of Claim 2.54.

For every $n \in \mathbb{Z}$, set $\xi_n = \bigvee_{i=-n}^{n} T^i(\xi)$. Then for every $1 \leq j \leq r$, set $\xi_n^j = \bigvee_{i \in \{-n, \dots, n\} \cap C_j} T^i(\xi)$ so that $\xi_n = \xi_n^1 \vee \dots \vee \xi_n^r$. Then we have

$H_\nu(\xi_n) \leq \sum_{j=1}^r H_\nu(\xi_n^j)$. If $\{-n, \dots, n\} \cap C_j = \emptyset$, then $H_\nu(\xi_n^j) = 0$. If $\{-n, \dots, n\} \cap C_j \neq \emptyset$, then choosing $n_j \in \{-n, \dots, n\} \cap C_j$ and using Lemma 2.52 and Claim 2.54, we have

$$\begin{aligned} H_\nu(\xi_n^j) &\leq H_\nu(T^{n_j}(\xi)) + \sum_{p \in \{-n, \dots, n\} \cap C_j \setminus \{n_j\}} H_\nu(T^p(\xi) | T^{n_j}(\xi)) \\ &\leq H_\nu(\xi) + |\{-n, \dots, n\} \cap C_j| \cdot \varepsilon. \end{aligned}$$

Then for every $n \geq 1$, we have

$$\begin{aligned} H_\nu(\xi_n) &\leq \sum_{j=1}^r H_\nu(\xi_n^j) \\ &\leq \sum_{j=1}^r (H_\nu(\xi) + |\{-n, \dots, n\} \cap C_j| \cdot \varepsilon) \\ &= r H_\nu(\xi) + (2n+1)\varepsilon. \end{aligned}$$

This implies that

$$h_\nu(T, \xi) = \lim_n \frac{1}{2n+1} H_\nu(T^n(\xi_n)) = \lim_n \frac{1}{2n+1} H_\nu(\xi_n) \leq \varepsilon.$$

Since this holds true for every finite measurable partition ξ of X and every $\varepsilon > 0$, it follows that $h_\nu(T) = 0$. □

6.3. The Shannon–McMillan–Breiman theorem. In this subsection, we assume that $(X, \mathscr{X}, \nu, T)$ is an ergodic pmp dynamical system. Let $\xi = \{A_1, \dots, A_m\}$ be a finite measurable partition of X. For every $n \geq 1$, set $\xi_n = \xi \vee \dots \vee T^{-n+1}(\xi)$. The Shannon–McMillan–Breiman theorem uses entropy to measure how large sets in the nth joint ξ_n are. Typically, they decrease exponentially and the exponential rate is exactly the measure entropy. More precisely, we prove the following theorem.

THEOREM 2.55 (Shannon–McMillan–Breiman). *Keep the same notation as above. Then the sequence $(\frac{1}{n} I_{\xi_n})_n$ converges to $h_\nu(T, \xi)$ ν-almost everywhere and in $\mathrm{L}^1(X, \mathscr{X}, \nu)$.*

For every $n \geq 1$ and every $x \in X$, denote by $\xi_n(x)$ the unique element of ξ_n that contains x. Then Theorem 2.55 implies that for ν-almost every $x \in X$, we have

$$\lim_n -\frac{1}{n} \log \nu(\xi_n(x)) = \lim_n \frac{1}{n} I_{\xi_n}(x) = h_\nu(T, \xi).$$

Before proving Theorem 2.55, we need to introduce some further notation and prove some preliminary results.

Set $\eta_1 = \tau$ and for every $n \geq 2$, set $\eta_n = T^{-1}(\xi) \vee \dots \vee T^{-n+1}(\xi)$. For every $n \geq 1$, denote by $\mathscr{F}_n = \sigma(\eta_n)$ the σ-subalgebra of $\mathscr{X}$ generated by η_n. Denote by $\mathscr{F}_\infty = \bigvee_{n=1}^\infty \mathscr{F}_n$ the σ-subalgebra of $\mathscr{X}$ generated by $\bigcup_{n=1}^\infty \mathscr{F}_n$. For every $n \geq 1$, set $g_n = I_{\xi, \eta_n} = -\sum_{i=1}^m \log(\mathbb{E}_\nu(\mathbf{1}_{A_i} | \mathscr{F}_n)) \mathbf{1}_{A_i}$. Observe that $g_1 = I_{\xi, \tau} = I_\xi$. Set $g_\infty = -\sum_{i=1}^m \log(\mathbb{E}_\nu(\mathbf{1}_{A_i} | \mathscr{F}_\infty)) \mathbf{1}_{A_i}$.

LEMMA 2.56. *The following assertions hold:*

(i) $g^* = \sup_{n\geq 1} g_n \in \mathrm{L}^1(X, \mathscr{X}, \nu)$.

(ii) $g_n \to g_\infty$ *ν-almost everywhere and in* $\mathrm{L}^1(X, \mathscr{X}, \nu)$.

(iii) $\int_X g_\infty \,\mathrm{d}\nu = h_\nu(T, \xi)$.

PROOF. (i) Note that $g_n \geq 0$ for every $n \geq 1$. Let $1 \leq i \leq m$ and $t \in \mathbb{R}_+$. For every $n \geq 1$, define the measurable subset $B_{i,n,t} \in \mathscr{X}$ as the set of all $x \in X$ for which $n \geq 1$ is the smallest integer such that $-\log(\mathbb{E}_\nu(\mathbf{1}_{A_i}|\mathscr{F}_n)) > t$. Then we have $A_i \cap \{g^* > t\} = \bigsqcup_{n\geq 1} A_i \cap B_{i,n,t}$. Moreover, we have

$$\begin{aligned}
\nu(A_i \cap B_{i,n,t}) &= \nu\left(\mathbb{E}_\nu(\mathbf{1}_{A_i}\mathbf{1}_{B_{i,n,t}}|\mathscr{F}_n)\right)\\
&= \nu\left(\mathbf{1}_{B_{i,n,t}}\mathbb{E}_\nu(\mathbf{1}_{A_i}|\mathscr{F}_n)\right)\\
&\leq \nu\left(\mathbf{1}_{B_{i,n,t}}\exp(-t)\right)\\
&= \exp(-t)\nu(B_{i,n,t}).
\end{aligned}$$

This implies that

$$\nu(A_i \cap \{g^* > t\}) = \sum_{n\geq 1}\nu(A_i \cap B_{i,n,t}) \leq \exp(-t)\sum_{n\geq 1}\nu(B_{i,n,t}) \leq \exp(-t).$$

Thus, $\nu(A_i \cap \{g^* > t\}) \leq \min\{\nu(A_i), \exp(-t)\}$. Since $g^* \geq 0$, we have

$$\begin{aligned}
\int_{A_i} g^* \,\mathrm{d}\nu &= \int_0^\infty \nu(A_i \cap \{g^* > t\})\,\mathrm{d}t\\
&\leq \int_0^\infty \min\{\nu(A_i), \exp(-t)\}\,\mathrm{d}t\\
&\leq \int_0^{-\log(\nu(A_i))} \nu(A_i)\,\mathrm{d}t + \int_{-\log(\nu(A_i))}^\infty \exp(-t)\,\mathrm{d}t\\
&= -\nu(A_i)\log(\nu(A_i)) + \nu(A_i).
\end{aligned}$$

This further implies that

$$\int_X g^*\,\mathrm{d}\nu = \sum_{i=1}^m \int_{A_i} g^*\,\mathrm{d}\nu \leq \sum_{i=1}^m(-\nu(A_i)\log(\nu(A_i)) + \nu(A_i)) = H_\nu(\xi) + 1.$$

Therefore, $g^* \in \mathrm{L}^1(X, \mathscr{X}, \nu)$.

(ii) Using the martingale convergence theorem (see Theorem A.1), for every $1 \leq i \leq m$, we have $\mathbb{E}_\nu(\mathbf{1}_{A_i}|\mathscr{F}_n) \to \mathbb{E}_\nu(\mathbf{1}_{A_i}|\mathscr{F}_\infty)$ ν-almost everywhere. This implies that $g_n \to g_\infty$ ν-almost everywhere. Since $g^* = \sup_{n\geq 1} g_n \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, Lebesgue's dominated convergence theorem implies that $g_n \to g_\infty$ in $\mathrm{L}^1(X, \mathscr{X}, \nu)$.

(iii) Combining Proposition 2.44 and (ii), we obtain

$$h_\nu(T, \xi) = \lim_n H_\nu(\xi|\eta_n) = \lim_n \int_X g_n\,\mathrm{d}\nu = \int_X g_\infty\,\mathrm{d}\nu.$$

This finishes the proof. $\square$

We are now ready to prove Theorem 2.55.

Proof of Theorem 2.55. Using repeatedly Proposition 2.43, for every $n \geq 1$, we have

$$\begin{aligned} I_{\xi_n}(x) &= I_{\eta_n}(x) + g_n(x) \\ &= I_{\xi_{n-1}}(T(x)) + g_n(x) \\ &= \cdots \\ &= g_1(T^{n-1}(x)) + \cdots + g_n(x) \\ &= \sum_{j=0}^{n-1} g_{n-j}(T^j(x)) \\ &= \sum_{j=0}^{n-1} g_\infty(T^j(x)) + \sum_{j=0}^{n-1} (g_{n-j} - g_\infty)(T^j(x)). \end{aligned}$$

Using Theorem 2.28, for ν-almost every $x \in X$, we have that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} g_\infty(T^j(x)) = \int_X g_\infty \, \mathrm{d}\nu = h_\nu(T, \xi).$$

It remains to show that $\frac{1}{n} \sum_{j=0}^{n-1} (g_{n-j} - g_\infty)(T^j(x)) \to 0$ for ν-almost every $x \in X$. For every $N \geq 1$, set $G_N = \sup_{n > N} |g_n - g_\infty|$ and $H_N = \sum_{k=1}^{N} (g_k + g^*) \circ T^{N-k} \in \mathrm{L}^1(X, \mathscr{X}, \nu)$. Then we have $\sup_N G_N \leq g_\infty + g^*$ and $G_N \to 0$ ν-almost everywhere. By Lebesgue's dominated convergence theorem, we have

$$\lim_N \int_X G_N \, \mathrm{d}\nu = \int_X \lim_N G_N \, \mathrm{d}\nu = 0.$$

Define $\ell(x) = \limsup_n \frac{1}{n} \sum_{j=0}^{n-1} |g_{n-j} - g_\infty|(T^j(x))$ for ν-almost every $x \in X$. For every $N \geq 1$ and for ν-almost every $x \in X$, using again Theorem 2.28, we have

$$\begin{aligned} \ell(x) &\leq \limsup_n \frac{1}{n} \sum_{j=0}^{n-N-1} |g_{n-j} - g_\infty|(T^j(x)) \\ &\quad + \limsup_n \frac{1}{n} \sum_{n-N}^{n-1} |g_{n-j} - g_\infty|(T^j(x)) \\ &\leq \limsup_n \frac{1}{n} \sum_{j=0}^{n-N-1} G_N(T^j(x)) + \limsup_n \frac{1}{n} H_N(T^{n-N}(x)) \\ &= \int_X G_N \, \mathrm{d}\nu. \end{aligned}$$

Then $\ell(x) \leq \lim_N \int_X G_N \, \mathrm{d}\nu = 0$ for ν-almost every $x \in X$. This finishes the proof of the theorem. □

6.4. The variational principle. In this subsection, we assume that (X, d) is a compact metric space. We prove the *variational principle* for measure entropy due to Dinaburg [**Di71**] and Goodman [**Go71**] which asserts that for a topological dynamical system on X, the topological entropy is the supremum of the measure entropies over all invariant Borel probability measures.

THEOREM 2.57. *Let $T : X \to X$ be a topological dynamical system. Then we have*

$$h(T) = \sup\{h_\nu(T) \mid \nu \in \mathrm{Prob}_T(X)\}.$$

Before proving Theorem 2.57, we need some preparation.

LEMMA 2.58. *Let $\nu, \eta \in \mathrm{Prob}(X)$ and $\xi = \{A_1, \ldots, A_m\}$ a finite measurable partition of X. Then for every $t \in [0,1]$, we have*

$$tH_\nu(\xi) + (1-t)H_\eta(\xi) \leq H_{t\nu+(1-t)\eta}(\xi).$$

PROOF. We use the concavity of the function $\varphi : [0,1] \to \mathbb{R}$ defined by $\varphi(0) = 0$ and $\varphi(x) = -x \log x$ for every $x \in (0,1]$. For every $t \in [0,1]$, we have

$$\begin{aligned} tH_\nu(\xi) + (1-t)H_\eta(\xi) &= \sum_{i=1}^{m} \left(t\varphi(\nu(A_i)) + (1-t)\varphi(\eta(A_i))\right) \\ &\leq \sum_{i=1}^{m} \varphi(t\nu(A_i) + (1-t)\eta(A_i)) \\ &= H_{t\nu+(1-t)\eta}(\xi). \end{aligned}$$

This finishes the proof. □

For every $A \in \mathscr{X}$, denote by $\partial A = \overline{A} \cap \overline{X \setminus A}$ the boundary of A. For every finite measurable partition $\xi = \{A_1, \ldots, A_m\}$ of X, denote by $\partial\xi = \bigcup_{i=1}^m \partial A_i$ the boundary of ξ.

LEMMA 2.59. *Let $\nu \in \mathrm{Prob}(X)$. The following assertions hold:*

(i) *For every $x \in X$ and every $\delta > 0$, there exists $0 < \varepsilon < \delta$ such that $\nu(\partial B(x, \varepsilon)) = 0$.*
(ii) *For every $\delta > 0$, there exists a finite measurable partition ξ of X for which all elements have diameter less than δ and such that $\nu(\partial\xi) = 0$.*
(iii) *Whenever $(\nu_n)_n$ is sequence in $\mathrm{Prob}(X)$ such that $\nu_n \to \nu$ with respect to the weak-$*$ topology and $A \in \mathscr{X}$ is a measurable set such that $\nu(\partial A) = 0$, we have $\nu(A) = \lim_n \nu_n(A)$.*

PROOF. (i) Let $x \in X$ and $\delta > 0$. For every $\varepsilon > 0$, define the sphere $S(x, \varepsilon) = \{y \in X \mid d(x, y) = \varepsilon\}$. Then we have $B(x, \delta) = \bigcup_{0<\varepsilon<\delta} S(x, \varepsilon)$. Since the open interval $(0, \delta)$ is uncountable, there exists $0 < \varepsilon < \delta$ such that $\nu(S(x, \varepsilon)) = 0$. Since $\partial B(x, \varepsilon) \subset S(x, \varepsilon)$, it follows that $\nu(\partial B(x, \varepsilon)) = 0$.

(ii) By compactness and using item (i), we may choose a finite open cover $\mathscr{U} = \{B_1, \ldots, B_m\}$ by open balls of radius less then $\frac{\delta}{2}$ such that $\nu(\partial\mathscr{U}) = 0$. Set $C_1 = \overline{B}_1$ and for every $2 \leq j \leq m$, define recursively $C_j = \overline{B}_j \setminus (\overline{B}_1 \cup \cdots \cup \overline{B}_{j-1})$. Then $\partial\xi = \bigcup_{i=1}^m \partial C_i \subset \bigcup_{i=1}^m \partial B_i$. Therefore, $\xi = \{C_1, \ldots, C_m\}$ is a finite measurable partition of X whose all elements have diameter less than δ and such that $\nu(\partial\xi) = 0$.

(iii) Let $(\nu_n)_n$ be a sequence in $\mathrm{Prob}(X)$ such that $\nu_n \to \nu$ with respect to the weak-$*$ topology. Let $A \in \mathscr{X}$ be a measurable set such that $\nu(\partial A) = 0$. For every $k \in \mathbb{N}$, define $f_k = 1 - \min(kd(\,\cdot\,, \overline{A}), 1) \in \mathrm{C}(X)$. Then $\mathbf{1}_{\overline{A}} \leq f_k$ and $f_k \to \mathbf{1}_{\overline{A}}$ pointwise. Then for every fixed $k \in \mathbb{N}$, we have

$$\limsup_n \nu_n(A) \leq \limsup_n \nu_n(\overline{A}) \leq \limsup_n \nu_n(f_k) = \nu(f_k).$$

Taking the limit as $k \to \infty$, we have

$$\limsup_n \nu_n(A) \leq \lim_k \nu(f_k) = \nu(\overline{A}) = \nu(A).$$

Similarly, we have

$$\limsup_n \nu_n(X \setminus A) \leq \nu(X \setminus A).$$

Thus, we obtain $\lim_n \nu_n(A) = \nu(A)$. □

We are now ready to prove Theorem 2.57.

Proof of Theorem 2.57. We follow the argument due to Misiurewicz [Mi76]. Set $h_{\sup}(T) = \sup\{h_\nu(T) \mid \nu \in \mathrm{Prob}_T(X)\}$. Recall the notation from Chapter 1.

Firstly, we prove the inequality $h(T) \leq h_{\sup}(T)$. Let $\varepsilon > 0$. For every $n \geq 1$, choose an (n, ε)-separating set $\mathscr{F}_{n,\varepsilon} \subset X$ of maximum cardinality, that is, $|\mathscr{F}_{n,\varepsilon}| = \mathrm{sep}(n, \varepsilon, T)$. Then define $\eta_n = \frac{1}{|\mathscr{F}_{n,\varepsilon}|}\sum_{x \in \mathscr{F}_{n,\varepsilon}} \delta_x \in \mathrm{Prob}(X)$ and $\nu_n = \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \eta_n \in \mathrm{Prob}(X)$. Fix $\nu \in \mathrm{Prob}(X)$ and an increasing sequence $(n_k)_k$ in $\mathbb{N}$ such that $\lim_k \frac{1}{n_k}\log(\mathrm{sep}(n_k, \varepsilon, T)) = \limsup_n \frac{1}{n}\log(\mathrm{sep}(n, \varepsilon, T))$ and $\lim_k \nu_{n_k} = \nu$ with respect to the weak-$*$ topology. The proof of Lemma 2.17 shows that $\nu \in \mathrm{Prob}_T(X)$ is T-invariant. Note that ν *a priori* depends on ε.

By Lemma 2.59, we may choose a finite measurable partition ξ of X with elements of diameter less than ε and such that $\nu(\partial\xi) = 0$. For every $A \in \xi_n$, since the d_n-diameter of A is less than ε, either $\eta_n(A) = 0$ or $\eta_n(A) = \frac{1}{|\mathscr{F}_{n,\varepsilon}|}$. This implies that $H_{\eta_n}(\xi_n) = \log(|\mathscr{F}_{n,\varepsilon}|) = \log(\mathrm{sep}(n, \varepsilon, T))$.

Fix $0 \leq k < q < n$ and assume that $n \geq k + q$. Set $a(k) = \lfloor \frac{n-k}{q} \rfloor \geq 1$. Set $C = \{k + rq + i \mid 0 \leq r < a(k) - 1, 0 \leq i \leq q - 1\}$ and $D = \{0, \ldots, n - 1\} \setminus C$. Then we have

$$\xi_n = \bigvee_{j=0}^{n-1} T^{-j}(\xi) = \bigvee_{r=0}^{a(k)-1} T^{-(k+rq)}(\xi_q) \vee \bigvee_{j \in D} T^{-j}(\xi).$$

Since $|D| \le 2q$, this further implies that

$$\begin{aligned}\log(\operatorname{sep}(n,\varepsilon,T)) &= H_{\eta_n}(\xi_n)\\ &\le \sum_{r=0}^{a(k)-1} H_{\eta_n}(T^{-(k+rq)}(\xi_q)) + \sum_{j\in D} H_{\eta_n}(T^{-j}(\xi))\\ &\le \sum_{r=0}^{a(k)-1} H_{T_*^{(k+rq)}\eta_n}(\xi_q) + 2q\log(|\xi|).\end{aligned}$$

Summing over k, dividing by n and using Lemma 2.58, we obtain

$$\begin{aligned}\frac{q}{n}\log(\operatorname{sep}(n,\varepsilon,T)) &= \frac{1}{n}\sum_{k=0}^{q-1} H_{\eta_n}(\xi_n)\\ &\le \sum_{k=0}^{q-1}\left(\sum_{r=0}^{a(k)-1}\frac{1}{n}H_{T_*^{(k+rq)}\eta_n}(\xi_q)\right) + \frac{2q^2}{n}\log(|\xi|)\\ &\le \sum_{j=0}^{n-1}\frac{1}{n}H_{T_*^{j}\eta_n}(\xi_q) + \frac{2q^2}{n}\log(|\xi|)\\ &\le H_{\nu_n}(\xi_q) + \frac{2q^2}{n}\log(|\xi|).\end{aligned}$$

Since for every $n \ge 1$, we have $\nu(\partial\xi_n) = 0$, Lemma 2.59 implies that for every fixed $q \ge 1$, we have

$$\begin{aligned}\limsup_n \frac{1}{n}\log(\operatorname{sep}(n,\varepsilon,T)) &= \lim_k \frac{1}{n_k}\log(\operatorname{sep}(n_k,\varepsilon,T))\\ &\le \frac{1}{q}\lim_k H_{\nu_{n_k}}(\xi_q)\\ &= \frac{1}{q}H_\nu(\xi_q).\end{aligned}$$

Then taking the limit as $q \to +\infty$, we obtain

$$\limsup_n \frac{1}{n}\log(\operatorname{sep}(n,\varepsilon,T)) \le \lim_q \frac{1}{q}H_\nu(\xi_q) = h_\nu(T,\xi) \le h_\nu(T). \tag{2.5}$$

Finally, we obtain

$$h(T) = \lim_{\varepsilon\to 0^+}\limsup_n \frac{1}{n}\log(\operatorname{sep}(n,\varepsilon,T)) \le h_{\sup}(T).$$

Secondly, we prove the inequality $h_{\sup}(T) \le h(T)$. Let $\nu \in \operatorname{Prob}_T(X)$. Let $\xi = \{A_1,\dots,A_m\}$ be a finite measurable partition. Choose $\varepsilon > 0$ so that $m\log(m)\varepsilon \le 1$. By regularity of the Borel probability measure ν, for every $1 \le i \le m$, we may choose a compact subset $B_i \subset A_i$ such that $\nu(A_i \setminus B_i) < \varepsilon$. Set $B_0 = X \setminus \bigcup_{i=1}^m B_i$ and observe that $B_0 \subset X$

is an open subset such that $\nu(B_0) \leq m\varepsilon$. Define the new finite partition $\beta = \{B_0, B_1, \ldots, B_m\}$. Then we have

$$\begin{aligned}
H_\nu(\xi|\beta) &= -\sum_{i=1}^{m}\sum_{j=0}^{m} \nu(A_i \cap B_j)\log\left(\frac{\nu(A_i\cap B_j)}{\nu(B_j)}\right)\\
&= -\sum_{i=1}^{m}\sum_{j=1}^{m} \nu(A_i \cap B_j)\log\left(\frac{\nu(A_i\cap B_j)}{\nu(B_j)}\right)\\
&\qquad -\sum_{i=1}^{m} \nu(A_i \cap B_0)\log\left(\frac{\nu(A_i\cap B_0)}{\nu(B_0)}\right)\\
&= -\nu(B_0)\sum_{i=1}^{m} \frac{\nu(A_i\cap B_0)}{\nu(B_0)}\log\left(\frac{\nu(A_i\cap B_0)}{\nu(B_0)}\right)\\
&\leq \nu(B_0)\log(m) \leq 1.
\end{aligned}$$

Then Proposition 2.45 implies that

$$h_\nu(T,\xi) \leq h_\nu(T,\beta) + H_\nu(\xi|\beta) + \leq h_\nu(T,\beta) + 1.$$

We now consider the open cover $\mathscr{U} = \{B_0 \sqcup B_1, \ldots, B_0 \sqcup B_m\}$. Note that for every $1 \leq i \leq m$,

$$B_0 \sqcup B_j = X \setminus \bigcup_{i\neq j} B_i$$

is indeed open. Let $n \geq 1$. Every element of $\mathscr{U}_n = \mathscr{U} \vee T^{-1}(\mathscr{U}) \vee \cdots \vee T^{-n+1}(\mathscr{U})$ is of the form

$$(B_0 \sqcup B_{i_0}) \cap T^{-1}(B_0 \sqcup B_{i_1}) \cap \cdots \cap T^{-n+1}(B_0 \sqcup B_{i_{n-1}})$$

where $i_0, \ldots, i_{n-1} \in \{1, \ldots, m\}$. Therefore, every element of $\mathscr{U}_n$ can be written as a pairwise disjoint union of 2^n elements of β_n (some of which may be emptyset). This implies that $|\beta_n| \leq 2^n|\mathscr{U}_n|$ and so

$$H_\nu(\beta_n) \leq \log(|\beta_n|) \leq n\log(2) + \log(|\mathscr{U}_n|).$$

Since no subcollection of $\mathscr{U}$ covers X, the same holds true for $\mathscr{U}_n$ and so $|\mathscr{U}_n| = N(\mathscr{U}_n)$. Theorem 1.20 implies that

$$\begin{aligned}
h(T) &\geq \lim_n \frac{1}{n}\log(|\mathscr{U}_n|)\\
&\geq \lim_n \frac{1}{n} H_\nu(\beta_n) - \log(2)\\
&= h_\nu(T,\beta) - \log(2)\\
&\geq h_\nu(T,\xi) - \log(2) - 1.
\end{aligned}$$

Taking the supremum over all finite measurable partitions ξ of X, it follows that $h_\nu(T) \leq h(T) + \log(2) + 1$. Observe that this inequality holds true for every topological dynamical system $S : X \to X$ and every S-invariant Borel probability measure $\eta \in \mathrm{Prob}_S(X)$. In particular, using Propositions 1.22 and 2.50, for every $m \geq 1$, we have

$$h_\nu(T) = \frac{1}{m} h_\nu(T^m) \leq \frac{1}{m} h(T^m) + \frac{\log(2)+1}{m} = h(T) + \frac{\log(2)+1}{m}.$$

Taking the limit as $m \to \infty$, it follows that $h_\nu(T) \leq h(T)$. Since this holds true for every $\nu \in \mathrm{Prob}_T(X)$, we finally obtain $h_{\sup}(T) \leq h(T)$. □

For expansive topological dynamical systems, there always exists an invariant Borel probability measure of maximal entropy.

PROPOSITION 2.60. *Let $T : X \to X$ be an expansive topological dynamical system. Then there exists a T-invariant Borel probability measure $\nu \in \mathrm{Prob}_T(X)$ such that $h(T) = h_\nu(T)$.*

PROOF. Let $\kappa > 0$ be a constant of expansiveness for T. Choose $0 < \varepsilon < \frac{\kappa}{2}$. By Proposition 1.24, we know that $h_{2\varepsilon}(T) = h(T)$. Using Lemma 1.18 and the proof of Theorem 2.57, specifically (2.5), we obtain

$$h(T) = h_{2\varepsilon}(T) \leq \limsup_n \frac{1}{n} \log(\mathrm{sep}(n, \varepsilon, T)) \leq h_\nu(T).$$

Therefore, Theorem 2.57 implies that $h(T) = h_\nu(T)$. □

CHAPTER 3

Topics in homogeneous dynamics

In this chapter, we give an introduction to the theory of locally compact groups and their lattices. We show that $\mathrm{SL}_d(\mathbb{Z})$ is a lattice in $\mathrm{SL}_d(\mathbb{R})$ for every $d \geq 2$. We also prove that $\mathrm{SL}_d(\mathbb{R})$ has the Howe–Moore property for every $d \geq 2$. As an application, we obtain Moore's ergodicity theorem.

1. Locally compact groups

Definition 3.1. Let G be a group endowed with a Hausdorff topology. We say that G is a *topological group* if the map $G \times G \to G : (g, h) \mapsto gh^{-1}$ is continuous. We then say that G is *locally compact* if there exists a compact neighborhood $U \subset G$ of the identity element $e \in G$.

Let G be a locally compact group. We say that G is

- *first countable* if there exists a countable neighborhood basis of $e \in G$.
- *second countable* if there exists a countable basis for the topology on G.
- *σ-compact* if there exists an increasing sequence of compact subsets $Q_n \subset G$ such that $G = \bigcup_{n \in \mathbb{N}} Q_n$.
- *compactly generated* if there exists a compact subset $Q \subset G$ such that $e \in Q$ and $G = \bigcup_{n \geq 1} Q^n$.
- *totally disconnected* if the connected component of $e \in G$ is equal to $\{e\}$.

The identity element $e \in G$ has a neighborhood basis consisting of compact subsets (see [DE14, Corollary A.8.2]). Any open subgroup $H < G$ is also closed since $G \setminus H = \bigcup_{gH \neq H} gH$. Any compactly generated group G is σ-compact. Any locally compact group G has a compactly generated open subgroup $H < G$. Indeed, choose a compact neighborhood $U \subset G$ of $e \in G$. Then $H = \bigcup_{n \geq 1} (U \cup U^{-1})^n$ is a compactly generated open subgroup of G. In particular, any *connected* locally compact group is compactly generated. A locally compact group G is second countable if and and only it is first countable and σ-compact (see [St73]). Moreover, any locally compact second countable group G is metrizable with a proper left invariant metric (see [St73]).

The class of locally compact groups is stable under taking closed subgroups, finite direct products and quotients with respect to closed normal subgroups. More precisely, we record the following facts.

PROPOSITION 3.2. *The following assertions hold:*

(i) *If G is a locally compact group and $H \leq G$ is a closed subgroup, then H endowed with the induced topology is locally compact.*
(ii) *If $d \geq 1$ and $G_1, \dots, G_d$ are locally compact groups, then the product group $G = G_1 \times \cdots \times G_d$ endowed with the product topology is locally compact.*
(iii) *If G is a locally compact group and $N \lhd G$ is a closed normal subgroup, the quotient group G/N endowed with the quotient topology is locally compact.*
(iv) *If G is a locally compact group acting continuously on a locally compact group H by continuous automorphisms, then the semi-direct product group $G \ltimes H$ endowed with the product topology is locally compact.*

The proof of Proposition 3.2 is left to the reader as an exercise.

EXAMPLES 3.3. Here are some examples of locally compact groups. Let $d \geq 1$.

(i) Any group G endowed with the discrete topology is locally compact. In these notes, any countable group will always be endowed with its discrete topology.
(ii) Any compact group K is locally compact. In particular, the following compact groups

$$\mathbb{T}^d = \left\{(z_1, \dots, z_d) \in \mathbb{C}^d \mid \forall 1 \leq i \leq d, |z_i| = 1\right\}$$
$$\mathrm{SO}_d(\mathbb{R}) = \{A \in \mathrm{SL}_d(\mathbb{R}) \mid A^*A = AA^* = 1_d\}$$
$$\mathscr{U}(d) = \{A \in \mathrm{GL}_d(\mathbb{C}) \mid A^*A = AA^* = 1_d\}$$

are locally compact.
(iii) Any (finite dimensional) real Lie group G is locally compact.
   - The abelian group $(\mathbb{R}^d, +)$ endowed with the usual topology is locally compact.
   - The *general linear group* $\mathrm{GL}_d(\mathbb{R})$ can be regarded as the open (dense) subset of invertible matrices in $\mathrm{M}_d(\mathbb{R}) \cong \mathbb{R}^{d^2}$. Endowed with the topology coming from $\mathbb{R}^{d^2}$, the group $\mathrm{GL}_d(\mathbb{R})$ is locally compact.
   - The *special linear group* $\mathrm{SL}_d(\mathbb{R}) = \ker(\det)$ is a closed subgroup of $\mathrm{GL}_d(\mathbb{R})$ and so $\mathrm{SL}_d(\mathbb{R})$ is locally compact.
   - The semi-direct product group $\mathrm{SL}_d(\mathbb{R}) \ltimes \mathbb{R}^d$ is locally compact.
(iv) Any (finite dimensional) p-adic Lie group G is totally disconnected locally compact. In particular, for every prime $p \in \mathscr{P}$, the groups $\mathrm{GL}_d(\mathbb{Q}_p)$ and $\mathrm{SL}_d(\mathbb{Q}_p)$ are totally disconnected locally compact.

(v) Let $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ be a locally finite tree and denote by $\mathrm{Aut}(\mathsf{T})$ the automorphism group of T. Endowed with the topology of pointwise convergence, the group $\mathrm{Aut}(\mathsf{T})$ is totally disconnected locally compact.

Let X be a locally compact space, meaning that every $x \in X$ has a compact neighborhood. We denote by $\mathscr{B}(X)$ the σ-algebra of Borel subsets of X. We say that a Borel measure ν on X, that is, a measure defined on $\mathscr{B}(X)$ is *regular* if the following conditions are satisfied:

(i) For every Borel subset $B \subset X$, we have

$$\nu(B) = \inf \{\nu(V) \mid V \text{ is open and } B \subset V\}.$$

(ii) For every open subset $U \subset X$, we have

$$\nu(U) = \sup \{\nu(K) \mid K \text{ is compact and } K \subset U\}.$$

(iii) For every compact subset $K \subset X$, we have $\nu(K) < +\infty$.

When ν is nonzero, define the *support* of ν by

$$\mathrm{supp}(\nu) = \bigcap \{F \mid F \subset X \text{ is closed and } \nu(X \setminus F) = 0\}.$$

Observe that $\mathrm{supp}(\nu)$ is closed and $\nu(X \setminus \mathrm{supp}(\nu)) = 0$.

If any open subset of X is σ-compact, then any Borel measure on X that satisfies condition (iii) is regular (see [**Ru87**, Theorem 2.18]). In particular, using [**DE14**, Lemma A.8.1(i)], if X is a locally compact second countable space, then any open subset of X is σ-compact and thus any Borel measure on X that satisfies condition (iii) is regular.

Denote by $\mathrm{C}_c(X)$ the space of compactly supported continuous functions on X. We say that a linear functional $\Phi : \mathrm{C}_c(X) \to \mathbb{C}$ is *positive* if $\Phi(f) \geq 0$ for every $f \in \mathrm{C}_c(X)_+$. By Riesz's representation theorem (see [**Ru87**, Theorem 2.14]), for every positive linear functional $\Phi : \mathrm{C}_c(X) \to \mathbb{C}$, there exists a unique regular Borel measure ν on X such that

$$\forall f \in \mathrm{C}_c(X), \quad \Phi(f) = \int_X f(x)\, \mathrm{d}\nu(x).$$

In that case, we will simply write $\Phi = \nu$. Note that for every regular Borel measure ν on X and every $p \in [1, +\infty)$, the space $\mathrm{C}_c(X)$ is $\|\cdot\|_p$-dense in the Banach space $\mathrm{L}^p(X, \mathscr{X}, \nu)$ of all ν-equivalence classes of p-integrable functions on X.

THEOREM 3.4 (Haar). *Let G be a locally compact group. Then there exists a nonzero regular Borel measure m_G on G that is unique up to multiplicative constant and that satisfies one of the following equivalent conditions:*

(i) *For every Borel subset $B \subset G$ and every $g \in G$, $m_G(gB) = m_G(B)$.*

(ii) *For every $f \in \mathrm{C}_c(G)$ and every $g \in G$,*

$$\int_G f(g^{-1}h)\, \mathrm{d}m_G(h) = \int_G f(h)\, \mathrm{d}m_G(h)$$

We say that m_G *is a* left invariant Haar measure *on* G*.*

For a proof of Theorem 3.4, we refer the reader to [**HR79**, Chapter 15]. The locally compact group G is σ-compact if and only if the left invariant Haar measure m_G is σ-finite.

Theorem 3.4 also implies that there exists a nonzero regular Borel measure μ_G on G that is unique up to multiplicative constant and that satisfies one of the following equivalent conditions:

(i) For every Borel subset $B \subset G$ and every $g \in G$, $\mu_G(Bg) = \mu_G(B)$.
(ii) For every $f \in \mathrm{C}_c(G)$ and every $g \in G$,

$$\int_G f(hg)\, \mathrm{d}\mu_G(h) = \int_G f(h)\, \mathrm{d}\mu_G(h)$$

We say that μ_G is a *right invariant Haar measure* on G. Indeed, any left invariant Haar measure m_G on G gives rise to a right invariant Haar measure μ_G on G by the formula

$$\forall B \in \mathscr{B}(G), \quad \mu_G(B) = m_G(B^{-1}).$$

The next proposition shows that any left invariant Haar measure has full support.

Proposition 3.5. *Let* G *be a locally compact group and* m_G *a left invariant Haar measure on* G*. Then* $\operatorname{supp}(m_G) = G$*. Moreover, for every* $f \in \mathrm{C}_c(G)_+$ *such that* $f \neq 0$*, we have* $\int_G f(h)\, \mathrm{d}m_G(h) > 0$*.*

Proof. Since $m_G \neq 0$, Conditions (ii) and (iii) in the definition of regularity imply that there exists a compact subset $K \subset G$ such that $0 < m_G(K) < +\infty$. Let $U \subset G$ be a nonempty open subset. There exist $g_1, \dots, g_n \in G$ such that $K \subset \bigcup_{i=1}^n g_i U$. This implies that

$$0 < m_G(K) \leq m_G(\bigcup_{i=1}^n g_i U) \leq \sum_{i=1}^n m_G(g_i U) = n \cdot m_G(U)$$

and so $m_G(U) > 0$. Thus, $\operatorname{supp}(m_G) = G$.

Moreover, let $f \in \mathrm{C}_c(G)_+$ such that $f \neq 0$. Then there exist $\varepsilon > 0$ and an open subset $U \subset G$ such that $f(h) \geq \varepsilon$ for every $h \in U$. This implies that

$$\int_G f(h)\, \mathrm{d}m_G(h) \geq \int_U \varepsilon\, \mathrm{d}m_G(h) = \varepsilon \cdot m_G(U) > 0.$$

This finishes the proof. □

The next proposition gives a characterization of compact groups in terms of the Haar measure.

Proposition 3.6. *Let* G *be a locally compact group and* m_G *a left invariant Haar measure on* G*.*

Then G *is compact if and only if* $m_G(G) < +\infty$*.*

Proof. Firstly, assume that G is compact. Then by regularity we have $m_G(G) < +\infty$.

Secondly, assume that G is not compact. Take a compact neighborhood $K \subset G$ of $e \in G$ and set $g_0 = e$. We have $m_G(K) > 0$ by Proposition 3.5. Since KK^{-1} is compact, there exists $g_1 \in G$ such that $g_1 \in G \setminus KK^{-1}$. This implies that $g_1K \cap K = \emptyset$. By induction, define $g_n \in G$ so that $g_n \in G \setminus (K \cup g_1K \cup \cdots \cup g_{n-1}K)K^{-1}$. It follows that $(g_nK)_n$ are pairwise disjoint. This implies that

$$m_G(G) \geq m_G(\bigcup_{n\in\mathbb{N}} g_nK) = \sum_{n\in\mathbb{N}} m_G(g_nK) = +\infty \cdot m_G(K) = +\infty.$$

This finishes the proof. □

Let G be a locally compact group and m_G a left invariant Haar measure on G. The measure m_G need not be right invariant. For every $g \in G$, define the nonzero regular Borel measure m_G^g on G by the formula $m_G^g(B) = m_G(Bg)$ for every $B \in \mathscr{B}(G)$. Since m_G^g is a left invariant Haar measure, there exists an element $\Delta_G(g) \in \mathbb{R}_+^*$ such that $m_G^g = \Delta_G(g)\, m_G$. Then $\Delta_G : G \to \mathbb{R}_+^* : g \mapsto \Delta_G(g)$ is a group homomorphism and is called the *modular function* on G. The modular function Δ_G does not depend on the choice of the left invariant Haar measure m_G on G. Moreover, we have

$$\forall f \in \mathrm{C}_c(G), \forall g \in G, \quad \int_G f(hg^{-1})\,\mathrm{d}m_G(h) = \Delta_G(g) \int_G f(h)\,\mathrm{d}m_G(h). \tag{3.1}$$

The left invariant Haar measure m_G is right invariant if and only if $\Delta_G \equiv 1$. In that case, we say that G is *unimodular*. We then simply refer to m_G as a Haar measure on G.

Proposition 3.7. *Let G be a locally compact group and m_G a left invariant Haar measure on G. Then the modular function $\Delta_G : G \to \mathbb{R}_+^*$ is continuous. Moreover, we have*

$$\forall f \in \mathrm{C}_c(G), \quad \int_G f(h^{-1})\,\mathrm{d}m_G(h) = \int_G \Delta_G(h^{-1}) f(h)\,\mathrm{d}m_G(h).$$

Proof. Choose $\varphi \in \mathrm{C}_c(G)$ such that $\kappa = \int_G \varphi(h)\,\mathrm{d}m_G(h) \neq 0$. Set $Q = \mathrm{supp}(\varphi)$. Then we have

$$\forall g \in G, \quad \Delta_G(g) = \frac{\int_G \varphi(hg^{-1})\,\mathrm{d}m_G(h)}{\int_G \varphi(h)\,\mathrm{d}m_G(h)}.$$

Choose a compact neighborhood $K \subset G$ of $e \in G$. Let $\varepsilon > 0$. Since φ is uniformly continuous by Lemma 3.8, there exists a neighborhood U of $e \in G$ such that $U \subset K$, $U^{-1} = U$ and

$$\forall u \in U, \quad \sup\left\{|\varphi(hu^{-1}) - \varphi(h)| \mid h \in G\right\} \leq \frac{\varepsilon\kappa}{m_G(QK)}.$$

Then for every $u \in U$, we have

$$|\Delta_G(u) - 1| \leq \frac{1}{\kappa}\int_G |\varphi(hu^{-1}) - \varphi(h)|\,\mathrm{d}m_G(h)$$

$$\leq \frac{1}{\kappa} m_G(QK) \frac{\varepsilon\kappa}{m_G(QK)} = \varepsilon.$$

This implies that $\Delta_G : G \to \mathbb{R}_+^*$ is continuous at the identity element $e \in G$ and so Δ_G is continuous.

Next, observe that both of the positive linear functionals

$$\mathrm{C}_c(G) \to \mathbb{C} : f \mapsto \int_G f(h^{-1})\, \mathrm{d}m_G(h)$$
$$\mathrm{C}_c(G) \to \mathbb{C} : f \mapsto \int_G \Delta(h^{-1}) f(h)\, \mathrm{d}m_G(h)$$

define a nonzero right invariant regular Borel measure on G. Thus, there exists $c > 0$ such that

$$\forall f \in \mathrm{C}_c(G), \quad \int_G f(h^{-1})\, \mathrm{d}m_G(h) = c \int_G \Delta_G(h^{-1}) f(h)\, \mathrm{d}m_G(h)$$

Define $\widehat{\varphi} \in \mathrm{C}_c(G)$ by the formula $\widehat{\varphi}(h) = \varphi(h^{-1})$ for every $h \in G$. Then we have

$$\begin{aligned} 0 \neq \int_G \varphi(h)\, \mathrm{d}m_G(h) &= \int_G \widehat{\varphi}(h^{-1})\, \mathrm{d}m_G(h) \\ &= c \int_G \Delta_G(h^{-1}) \widehat{\varphi}(h)\, \mathrm{d}m_G(h) \\ &= c \int_G \Delta_G(h^{-1}) \varphi(h^{-1})\, \mathrm{d}m_G(h) \\ &= c^2 \int_G \Delta_G(h^{-1}) \Delta_G(h) \varphi(h)\, \mathrm{d}m_G(h) \\ &= c^2 \int_G \varphi(h)\, \mathrm{d}m_G(h). \end{aligned}$$

This implies that $c = 1$. □

In the proof of Proposition 3.7, we used the following technical result. Denote by $(\mathrm{C}_b(G), \|\cdot\|_\infty)$ the Banach space of all bounded continuous functions on G endowed with the supremum norm. Denote by $\lambda : G \curvearrowright \mathrm{C}_b(G)$ (resp. $\rho : G \curvearrowright \mathrm{C}_b(G)$) the left (resp. right) translation action defined by $(\lambda(g)f)(h) = f(g^{-1}h)$ (resp. $(\rho(g)f)(h) = f(hg)$) for all $g, h \in G$ and all $f \in \mathrm{C}_b(G)$.

Lemma 3.8. *Let G be a locally compact group and $f \in \mathrm{C}_c(G)$ a compactly supported continuous function. Then for every $\varepsilon > 0$, there exists a symmetric neighborhood $U \subset G$ of $e \in G$ such that*

$$\sup \{\|\lambda(u)f - f\|_\infty, \|\rho(u)f - f\|_\infty \mid u \in U\} < \varepsilon.$$

Then we say that $f \in \mathrm{C}_c(G)$ is uniformly continuous.

Proof. Let $f \in \mathrm{C}_c(G)$ and set $Q = \mathrm{supp}(f)$. Let $\varepsilon > 0$ and fix a symmetric compact neighborhood $V \subset G$ of $e \in G$. For every $g \in G$, there exists an open neighborhood $W_g \subset G$ of $g \in G$ such that for all

$w_1, w_2 \in W_g$, we have $|f(w_1) - f(w_2)| < \varepsilon$. For every $g \in G$, choose an open symmetric neighborhood $U_g \subset G$ of $e \in G$ such that $gU_gU_g \cup U_gU_gg \subset W_g$. Then for every $g \in G$, $gU_g \cap U_gg$ is an open neighborhood of $g \in G$. Since VQV is compact, there exist $n \geq 1$ and $g_1, \dots, g_n \in G$ such that $VQV \subset \bigcup_{i=1} g_iU_{g_i} \cap U_{g_i}g_i$. Define $U = V \cap \bigcap_{i=1}^n U_{g_i}$ which is a symmetric neighborhood of the identity $e \in G$. Then for every $u \in U$ and every $g \in G$, we consider the following situations:

- If $g \in VQV$, then there exists $1 \leq i \leq n$ such that $g \in g_iU_{g_i} \cap U_{g_i}g_i$. Since $u \in U \subset U_{g_i}$, we have $gu \in g_iU_{g_i}U_{g_i} \subset W_{g_i}$ and $ug \in U_{g_i}U_{g_i}g_i \subset W_{g_i}$. It follows that $|f(gu) - f(g)| < \varepsilon$ and $|f(gu) - f(g)| < \varepsilon$.
- If $g \notin VQV$, then $gu \notin Q$ and $ug \notin Q$. It follows that $f(g) = f(ug) = f(gu) = 0$.

We have showed that for every $u \in U$ and every $g \in G$, we have $|f(gu) - f(g)| < \varepsilon$ and $|f(gu) - f(g)| < \varepsilon$. $\square$

Let (G, m_G, Δ_G) and (H, m_H, Δ_H) be locally compact groups with their respective left invariant Haar measure and modular function. Let $\sigma : G \curvearrowright H$ be a continuous action by continuous group automorphisms and write $G \ltimes H$ for the locally compact semi-direct product group. Recall that the group law on $G \ltimes H$ is given by

$$\forall g_1, g_2 \in G, \forall h_1, h_2 \in H, \quad (g_1, h_1) \cdot (g_2, h_2) = (g_1g_2, \sigma_{g_2}^{-1}(h_1)h_2).$$

The next proposition provides an explicit calculation of the Haar measure and the modular function on $G \ltimes H$.

Proposition 3.9. *The regular Borel measure $m_{G\ltimes H}$ defined on $G \ltimes H$ by the formulae*

$$\begin{aligned}
\forall f \in \mathrm{C}_c(G \ltimes H),\ &\int_{G\ltimes H} f(g,h)\,\mathrm{d}m_{G\ltimes H}(h) \\
&= \int_H \left(\int_G f(g,h)\,\mathrm{d}m_G(g)\right)\mathrm{d}m_H(h) \\
&= \int_G \left(\int_H f(g,h)\,\mathrm{d}m_H(h)\right)\mathrm{d}m_G(g)
\end{aligned} \tag{3.2}$$

is a left invariant Haar measure on $G \ltimes H$. Moreover, the modular function $\Delta_{G\ltimes H} : G \ltimes H \to \mathbb{R}_+^*$ satisfies

$$\forall (g,h) \in G \ltimes H, \quad \Delta_{G\ltimes H}(g,h) = \rho(g)\,\Delta_G(g)\,\Delta_H(h)$$

where $\rho : G \to \mathbb{R}_+^*$ is the continuous function defined by the formula

$$\forall f \in \mathrm{C}_c(H), \forall g \in G, \quad \int_H f(\sigma_g(h))\,\mathrm{d}m_H(h) = \rho(g)\int_H f(h)\,\mathrm{d}m_H(h).$$

Proof. Fubini's theorem implies that for every $f \in \mathrm{C}_c(G \ltimes H)$, we have

$$\int_H \left(\int_G f(g,h)\,\mathrm{d}m_G(g)\right)\mathrm{d}m_H(h) = \int_G \left(\int_H f(g,h)\,\mathrm{d}m_H(h)\right)\mathrm{d}m_G(g).$$

Denote by $m_{G\ltimes H}$ the unique regular Borel measure on $G \ltimes H$ defined by (3.2). For every $f \in \mathrm{C}_c(G \ltimes H)$ and every $(g_1,h_1) \in G \ltimes H$, we have

$$\begin{aligned}
&\int_{G\ltimes H} f((g_1,h_1)\cdot(g_2,h_2))\,\mathrm{d}m_{G\ltimes H}(g_2,h_2)\\
&\qquad = \int_{G\ltimes H} f(g_1g_2, \sigma_{g_2}^{-1}(h_1)h_2)\,\mathrm{d}m_{G\ltimes H}(g_2,h_2)\\
&\qquad = \int_G \left(\int_H f(g_1g_2,h_2)\,\mathrm{d}m_H(h_2)\right)\mathrm{d}m_G(g_2)\\
&\qquad = \int_H \left(\int_G f(g_2,h_2)\,\mathrm{d}m_G(g_2)\right)\mathrm{d}m_H(h_2)\\
&\qquad = \int_{G\ltimes H} f(g_2,h_2)\,\mathrm{d}m_{G\ltimes H}(g_2,h_2).
\end{aligned}$$

This shows that $m_{G\ltimes H}$ is a left invariant Haar measure on $G \ltimes H$.

Consider the function $\rho : G \to \mathbb{R}_+^*$ as defined above. For every $f \in \mathrm{C}_c(G \ltimes H)$ and every $(g_2,h_2) \in G \ltimes H$, we have

$$\begin{aligned}
&\int_{G\ltimes H} f((g_1,h_1)\cdot(g_2,h_2)^{-1})\,\mathrm{d}m_{G\ltimes H}(g_1,h_1)\\
&\quad = \int_{G\ltimes H} f(g_1g_2^{-1}, \sigma_{g_2}(h_1h_2^{-1}))\,\mathrm{d}m_{G\ltimes H}(g_1,h_1)\\
&\quad = \Delta_H(h_2)\int_G \left(\int_H f(g_1g_2^{-1}, \sigma_{g_2}(h_1))\,\mathrm{d}m_H(h_1)\right)\mathrm{d}m_G(g_1)\\
&\quad = \rho(g_2)\,\Delta_H(h_2)\int_G \left(\int_H f(g_1g_2^{-1}, h_1)\,\mathrm{d}m_H(h_1)\right)\mathrm{d}m_G(g_1)\\
&\quad = \rho(g_2)\,\Delta_G(g_2)\,\Delta_H(h_2)\int_H \left(\int_G f(g_1,h_1)\,\mathrm{d}m_G(g_1)\right)\mathrm{d}m_H(h_1)\\
&\quad = \rho(g_2)\,\Delta_G(g_2)\,\Delta_H(h_2)\int_{G\ltimes H} f(g_1,h_1)\,\mathrm{d}m_{G\ltimes H}(g_1,h_1)
\end{aligned}$$

and hence $\Delta_{G\ltimes H}(g_2,h_2) = \rho(g_2)\,\Delta_G(g_2)\,\Delta_H(h_2)$. □

EXAMPLES 3.10. Here are some examples of unimodular locally compact groups. Let $d \geq 1$.

(i) Any group G endowed with the discrete topology is unimodular. Indeed, in that case the counting measure m_G is a nonzero regular Borel measure on G that is clearly both left and right invariant.

(ii) Any compact group G is unimodular. Indeed, fix a left invariant Haar measure m_G on G. Then $\Delta_G(G) < \mathbb{R}_+^*$ is a compact subgroup and so $\Delta_G(G) = \{1\}$. This shows that $\Delta_G \equiv 1$ and so G is unimodular.

(iii) Any abelian locally compact group G is unimodular. The Lebesgue measure $\mathrm{d}x_1 \cdots \mathrm{d}x_d$ on $\mathbb{R}^d$ is a Haar measure.

(iv) Recall that the general linear group $\mathrm{GL}_d(\mathbb{R})$ can be regarded as the open (dense) subset of invertible matrices in $\mathrm{M}_d(\mathbb{R}) \cong \mathbb{R}^d \times \cdots \times \mathbb{R}^d$. For every $g \in \mathrm{GL}_d(\mathbb{R})$, the Jacobian of the diffeomorphism

$$L_g : \mathrm{M}_d(\mathbb{R}) \to \mathrm{M}_d(\mathbb{R}) : (x_1, \dots, x_d) \mapsto (gx_1, \dots, gx_d)$$

is equal to $|\det(g)|^d$. It follows that a left invariant Haar measure m_G on $G = \mathrm{GL}_d(\mathbb{R})$ is given by

$$\mathrm{d}m_G(g) = \frac{1}{|\det(g)|^d} \prod_{1 \leq i,j \leq d} \mathrm{d}g_{ij}, \quad g = (g_{ij})_{ij}.$$

For every $g \in \mathrm{GL}_d(\mathbb{R})$, since the Jacobian of the diffeomorphism

$$R_g : \mathrm{M}_d(\mathbb{R}) \to \mathrm{M}_d(\mathbb{R}) : x \mapsto xg$$

is also equal to $|\det(g)|^d$, it follows that m_G is right invariant and so $G = \mathrm{GL}_d(\mathbb{R})$ is unimodular.

(v) Recall that the special linear group $\mathrm{SL}_d(\mathbb{R}) < \mathrm{GL}_d(\mathbb{R})$ is defined by $\mathrm{SL}_d(\mathbb{R}) = \ker(\det)$. It is known that the only normal subgroups of $\mathrm{SL}_d(\mathbb{R})$ are $\{1\}$, $\{\pm 1\}$ and $\mathrm{SL}_d(\mathbb{R})$. This implies that $\ker(\Delta_{\mathrm{SL}_d(\mathbb{R})}) = \mathrm{SL}_d(\mathbb{R})$ and so $\mathrm{SL}_d(\mathbb{R})$ is unimodular.

(vi) For every $d \geq 2$, the strict upper triangular subgroup $G = \mathrm{T}_d(\mathbb{R})$ defined as the group of all matrices $g = (g_{ij})_{ij}$ such that $g_{ij} = 0$ for all $1 \leq j < i \leq d$ and $g_{ii} = 1$ for all $1 \leq i \leq d$ is homeomorphic with $\mathbb{R}^{\frac{d(d-1)}{2}}$. Under this identification, the Lebesgue measure on $\mathbb{R}^{\frac{d(d-1)}{2}}$ gives rise to a left and right invariant Haar measure m_G on G defined as

$$\mathrm{d}m_G(n) = \prod_{1 \leq i < j \leq d} \mathrm{d}n_{ij}, \quad n = (n_{ij})_{ij}.$$

Indeed, for all $i < j$ and all $g, n \in \mathrm{T}_d(\mathbb{R})$, we have $(gn)_{ij} = g_{ij} + n_{ij} + \sum_{i<k<j} g_{ik} n_{kj}$. Endow the set $\{(i,j) \mid 1 \leq i < j \leq d\}$ with the lexicographical order. Then for every $g \in \mathrm{T}_d(\mathbb{R})$, the Jacobian matrix of the diffeomorphism $\mathrm{T}_d(\mathbb{R}) \to \mathrm{T}_d(\mathbb{R}) : n \mapsto gn$ is lower triangular with diagonal entries all equal to 1. This implies that the Jacobian of the diffeomorphism $\mathrm{T}_d(\mathbb{R}) \to \mathrm{T}_d(\mathbb{R}) : n \mapsto gn$ is equal to 1. The same argument shows that for every $g \in \mathrm{T}_d(\mathbb{R})$, the Jacobian of the diffeomorphism $\mathrm{T}_d(\mathbb{R}) \to \mathrm{T}_d(\mathbb{R}) : n \mapsto ng$ is equal to 1. Thus, $G = \mathrm{T}_d(\mathbb{R})$ is unimodular.

2. Lattices in locally compact groups

Let G be a locally compact group and $\Gamma < G$ a discrete subgroup. We say that a Borel subset $\mathscr{F} \subset G$ is a *Borel fundamental domain* (for the right translation action $\Gamma \curvearrowright G$) if

$$\forall \gamma_1, \gamma_2 \in \Gamma,\ \gamma_1 \neq \gamma_2 \ \Rightarrow\ \mathscr{F}\gamma_1 \cap \mathscr{F}\gamma_2 = \emptyset \quad \text{and} \quad \bigcup_{\gamma \in \Gamma} \mathscr{F}\gamma = G.$$

Denote by $G/\Gamma = \{g\Gamma \mid g \in G\}$ the quotient space and by $p : G \to G/\Gamma : g \mapsto g\Gamma$ the quotient map. Endow G/Γ with the quotient topology.

PROPOSITION 3.11. *Keep the same notation as above. The following assertions hold:*

(i) *The quotient map $p : G \to G/\Gamma$ is continuous and open and G/Γ is Hausdorff and locally compact. Moreover, the action map $G \times G/\Gamma \to G/\Gamma : (g, x) \mapsto gx$ is continuous.*

(ii) *If G/Γ is compact, then there exists a Borel fundamental domain $\mathscr{F} \subset G$ that is relatively compact in G.*

(iii) *If G is second countable, then G/Γ is second countable. Moreover, there exists a Borel fundamental domain $\mathscr{F} \subset G$ such that for every compact subset $Y \subset G/\Gamma$, the subset $p^{-1}(Y) \cap \mathscr{F} \subset G$ is relatively compact in G.*

PROOF. (i) Endow the quotient space $G/\Gamma = \{g\Gamma \mid g \in G\}$ with the quotient topology. By definition, a subset $V \subset G/\Gamma$ is open if and only if $p^{-1}(V) \subset G$ is open. Then the quotient topology is the finest topology on G/Γ that makes the quotient map $p : G \to G/\Gamma$ continuous. Let now $U \subset G$ be an open set. Then $p^{-1}(p(U)) = p^{-1}(\{h\Gamma \mid h \in U\}) = \bigcup_{\gamma\in\Gamma} U\gamma$ is open and so is $p(U) \subset G/\Gamma$ is open. This shows that $p : G \to G/\Gamma$ is open.

Let $x_1, x_2 \in G/\Gamma$ with $x_1 \neq x_2$. Write $x_1 = g_1\Gamma$ and $x_2 = g_2\Gamma$. Note that $g_2 \notin g_1\Gamma$. Choose a compact neighborhood $U_1 \subset G$ (resp. $U_2 \subset G_2$) of $g_1 \in G$ (resp. $g_2 \in G$). Since $U_2^{-1}U_1 \subset G$ is compact and since $\Gamma < G$ is discrete, the set $\Lambda = \{\gamma \in \Gamma \mid U_1 \cap U_2\gamma \neq \emptyset\}$ is finite. For every $\gamma \in \Lambda$, since $g_1 \neq g_2\gamma$, there exist neighborhoods U_γ of $g_1 \in G$ and V_γ of $g_2\gamma \in G$ such that $U_\gamma \cap V_\gamma = \emptyset$. Set

$$W_1 = U_1 \cap \bigcap_{\gamma\in\Lambda} U_\gamma \quad \text{and} \quad W_2 = U_2 \cap \bigcap_{\gamma\in\Lambda} V_\gamma\gamma^{-1}.$$

Then for every $\gamma \in \Gamma$, we have $W_1 \cap W_2\gamma = \emptyset$. Indeed, if $\gamma \in \Gamma \setminus \Lambda$, then $U_1 \cap U_2\gamma = \emptyset$. If $\gamma \in \Lambda$, then $U_\gamma \cap (V_\gamma\gamma^{-1})\gamma = \emptyset$. Thus, we have $p(W_1) \cap p(W_2) = \emptyset$. This shows that G/Γ is Hausdorff.

Let $x = g\Gamma \in G/\Gamma$. Choose a compact neighborhood $K \subset G$ of $e \in G$. Then gK is a compact neighborhood of $g \in G$ and so $p(gK)$ is a compact neighborhood of $x \in G/\Gamma$. This shows that G/Γ is locally compact.

Define the action map $a : G \times G/\Gamma \to G/\Gamma : (g, x) \mapsto gx$. Recall that the multiplication map $m : G \times G \to G$ is continuous. Since the map $\mathrm{id}_G \times p : G \times G \to G \times G/\Gamma : (g, h) \mapsto (g, h\Gamma)$ is continuous and open, the commutative diagram

$$\begin{tikzcd} G \times G \arrow[r, "m"] \arrow[d, "\mathrm{id} \times p"] & G \arrow[d, "p"] \\ G \times G/\Gamma \arrow[r, "a"] & G/\Gamma \end{tikzcd}$$

shows that the action map $a : G \times G/\Gamma \to G/\Gamma$ is continuous.

(ii) Since $\Gamma < G$ is discrete, there exists an open neighborhood $V \subset G$ of $e \in G$ such that $V \cap \Gamma = \{e\}$. Since the map $G \times G \to G : (g,h) \mapsto g^{-1}h$ is continuous, there exists an open neighborhood $U \subset G$ of $e \in G$ such that $U^{-1}U \subset V$. Replacing U with $U \cap K$ where K is a relatively compact open neighborhood of $e \in G$, we may assume that $U \subset G$ is relatively compact. Since G/Γ is compact and since $(p(gU)_{g\in G})$ is an open covering of G/Γ, there exist $g_1, \dots, g_n \in G$ such that $G/\Gamma = \bigcup_{i=1}^n p(g_i U)$. Define the Borel subset

$$\mathscr{F} = \bigcup_{i=1}^{n} \left(g_i U \setminus \bigcup_{j<i} g_j U \Gamma \right).$$

By construction, $\mathscr{F} \subset G$ is relatively compact. Then we have $\bigcup_{\gamma\in\Gamma} \mathscr{F}\gamma = \bigcup_{i=1}^n g_i U\Gamma = p^{-1}(\bigcup_{i=1}^n p(g_i U)) = p^{-1}(G/\Gamma) = G$. Let $\gamma_1, \gamma_2 \in \Gamma$ be elements such that $\mathscr{F}\gamma_1 \cap \mathscr{F}\gamma_2 \neq \emptyset$. Upon exchanging γ_1 and γ_2, we may assume that there exist $i \geq j$ and $u_1, u_2 \in U$ such that $g_i u_1 \gamma_1 = g_j u_2 \gamma_2$. By construction and since $g_i u_1 = g_j u_2 \gamma_2 \gamma_1^{-1} \in g_i U \cap g_j U\Gamma$, we necessarily have $i = j$. Then $u_1\gamma_1 = u_2\gamma_2$ and so $u_2^{-1}u_1 = \gamma_2\gamma_1^{-1} \in U^{-1}U \cap \Gamma \subset V \cap \Gamma = \{e\}$. This shows that $\gamma_1 = \gamma_2$ and thus $\mathscr{F} \subset G$ is a Borel fundamental domain.

(iii) Choose a countable basis $(U_n)_n$ for the topology on G. Let $V \subset G/\Gamma$ be an open set. Then $p^{-1}(V) = \bigcup_{\gamma\in\Gamma} V\gamma \subset G$ is open and so there exists a subfamily $(U_{n_k})_k$ such that $p^{-1}(V) = \bigcup_k U_{n_k}$. Then we have $V = p(p^{-1}(V)) = \bigcup_k p(U_{n_k})$. This shows that $(p(U_n))_n$ is a countable basis for the quotient topology on G/Γ and so G/Γ is second countable. For every $n \in \mathbb{N}$, choose $g_n \in U_n$.

As before, there exist open neighborhoods $U, V \subset G$ of $e \in G$ such that $U \subset G$ is relatively compact, $U^{-1}U \subset V$ and $V \cap \Gamma = \{e\}$. We claim that $G = \bigcup_{n\in\mathbb{N}} g_n U$. Indeed, for every $g \in G$, $gU^{-1} \subset G$ is an open set and hence there exists $n \in \mathbb{N}$ such that $U_n \subset gU^{-1}$. This implies that there exists $u \in U$ such that $g_n = gu^{-1}$ or equivalently $g = g_n u$ and thus $g \in g_n U$. Define the Borel subset

$$\mathscr{F} = \bigcup_{n\in\mathbb{N}} \left(g_n U \setminus \bigcup_{k<n} g_k U\Gamma \right).$$

Then we have $\bigcup_{\gamma\in\Gamma} \mathscr{F}\gamma = \bigcup_{n\in\mathbb{N}} g_n U\Gamma = G$. Let $\gamma_1, \gamma_2 \in \Gamma$ be elements such that $\mathscr{F}\gamma_1 \cap \mathscr{F}\gamma_2 \neq \emptyset$. Upon exchanging γ_1 and γ_2, we may assume that there exist $m \geq n$ and $u_1, u_2 \in U$ such that $g_m u_1 \gamma_1 = g_n u_2 \gamma_2$. By construction and since $g_m u_1 = g_n u_2 \gamma_2 \gamma_1^{-1} \in g_m U \cap g_n U\Gamma$, we necessarily have $m = n$. Then $u_1\gamma_1 = u_2\gamma_2$ and so $u_2^{-1}u_1 = \gamma_2\gamma_1^{-1} \in U^{-1}U \cap \Gamma \subset V \cap \Gamma = \{e\}$. This shows that $\gamma_1 = \gamma_2$ and thus $\mathscr{F} \subset G$ is a Borel fundamental domain. Let $Y \subset G/\Gamma$ be a compact subset. Since $(p(g_n U))_n$ is an open covering of Y, there exist $n_1 \leq \cdots \leq n_k$ such that $Y \subset \bigcup_{i=1}^k p(g_{n_i} U)$. Then we have $p^{-1}(Y) \cap \mathscr{F} \subset \bigcup_{j=0}^{n_k} (g_j U \setminus \bigcup_{i<j} g_i U\Gamma)$ and so $p^{-1}(Y) \cap \mathscr{F} \subset G$ is relatively compact. $\square$

Observe that when G is a locally compact σ-compact group, any discrete subgroup $\Gamma < G$ is necessarily countable. Indeed, since G is σ-compact, the left invariant Haar measure m_G is σ-finite. We may then choose a Borel probability measure $\mu \in \mathrm{Prob}(G)$ such that $\mu \sim m_G$. We may also choose open neighborhoods $U, V \subset G$ of $e \in G$ such that $UU^{-1} \subset V$ and $V \cap \Gamma = \{e\}$. Then $(\gamma U)_{\gamma \in \Gamma}$ is a family of pairwise disjoint open subsets. Moreover, since $m_G(\gamma U) = m_G(U) > 0$ for every $\gamma \in \Gamma$, it follows that $\mu(\gamma U) > 0$ for every $\gamma \in \Gamma$. This implies that Γ is necessarily countable.

COROLLARY 3.12. *Let G be a locally compact second countable group and $\Gamma < G$ a discrete subgroup. Then there exists a Borel map $\sigma : G/\Gamma \to G$ such that*

- $\sigma(G/\Gamma) = \mathscr{F}$ *is a Borel fundamental domain,*
- $\sigma(\Gamma) = e$*,*
- $x = \sigma(x)\Gamma$ *for every* $x \in G/\Gamma$*,*
- $\sigma(Y) \subset G$ *is relatively compact for every compact subset* $Y \subset G/\Gamma$*.*

We then simply say that $\sigma : G/\Gamma \to G$ is a Borel section*.*

PROOF. Choose a Borel fundamental domain $\mathscr{F} \subset G$ as in Proposition 3.11(iii) such that $e \in \mathscr{F}$. Then $p|_{\mathscr{F}} : \mathscr{F} \to G/\Gamma$ is Borel and bijective. This implies that the map $\sigma = (p|_{\mathscr{F}})^{-1} : G/\Gamma \to G$ is Borel (see [**Zi84**, Theorem A.4]) and satisfies all the required properties. $\square$

DEFINITION 3.13. Let G be a locally compact group and $\Gamma < G$ a discrete subgroup. We say that $\Gamma < G$ is *uniform* or *cocompact* if G/Γ is compact.

We say that $\Gamma < G$ is a *lattice* if there exists a G-invariant regular Borel probability measure $\nu \in \mathrm{Prob}(G/\Gamma)$.

Define the linear mapping $\mathscr{T} : \mathrm{C}_c(G) \to \mathrm{C}_c(G/\Gamma) : f \mapsto \overline{f}$ by the formula

$$\forall g \in G, \quad \overline{f}(g\Gamma) = \sum_{\gamma \in \Gamma} f(g\gamma).$$

We claim that $\mathscr{T} : \mathrm{C}_c(G) \to \mathrm{C}_c(G/\Gamma)$ is surjective. Indeed, let $\varphi \in \mathrm{C}_c(G/\Gamma)$ be a function and denote by $Q = \mathrm{supp}(\varphi) \subset G/\Gamma$ its compact support. Choose a relatively compact open neighborhood $V \subset G$ of $e \in G$. Then there exist $g_1, \dots, g_n \in G$ such that $Q \subset \bigcup_{i=1}^n p(g_i V)$. Set $K = p^{-1}(Q) \cap \bigcup_{i=1}^n g_i \overline{V}$. Then $K \subset G$ is a compact subset such that $p(K) = Q$. By Urysohn's lemma (see e.g. [**DE14**, Lemma A.8.1(ii)]), we may choose $f_K \in \mathrm{C}_c(G)_+$ such that $f|_K \equiv 1_K$.

Define the function $f : G \to \mathbb{C}$ by the formula $f(g) = \frac{\varphi(g\Gamma)}{\mathscr{T}(f_K)(g\Gamma)} f_K(g)$ if $\mathscr{T}(f_K)(g\Gamma) \neq 0$ and $f(g) = 0$ otherwise. Then $\mathrm{supp}(f) \subset \mathrm{supp}(f_K)$ is compact and f is continuous on G since $\mathscr{T}(f_K)(g\Gamma) > 0$ on a neighborhood of Q. Thus, $f \in \mathrm{C}_c(G)$ and we have $\mathscr{T}(f) = \varphi$.

PROPOSITION 3.14. *Let G be a locally compact group and $\Gamma < G$ a uniform discrete subgroup. Then G is unimodular and $\Gamma < G$ is a lattice.*

If G is moreover compactly generated, then $\Gamma < G$ is finitely generated.

PROOF. Fix a right invariant Haar measure μ_G on G. Consider the positive linear functional

$$\Phi : \mathrm{C}_c(G/\Gamma) \to \mathbb{C} : \overline{f} \mapsto \int_G f(g)\, \mathrm{d}\mu_G(g).$$

In order to check that Φ is well-defined, it suffices to show that if $\varphi \in \mathrm{C}_c(G)$ is such that $\overline{\varphi} = 0$, then we have $\int_G \varphi(g)\, \mathrm{d}\mu_G(g) = 0$. Indeed, for every $\psi \in \mathrm{C}_c(G)$, using Fubini's theorem, we have

$$\begin{aligned}\int_G \overline{\varphi}(h\Gamma)\psi(h)\, \mathrm{d}\mu_G(h) &= \sum_{\gamma\in\Gamma} \int_G \varphi(h\gamma)\psi(h)\, \mathrm{d}\mu_G(h) \\ &= \sum_{\gamma\in\Gamma} \int_G \varphi(h)\psi(h\gamma^{-1})\, \mathrm{d}\mu_G(h) \\ &= \int_G \varphi(h)\overline{\psi}(h\Gamma)\, \mathrm{d}\mu_G(h).\end{aligned}$$

Since the map $\mathrm{C}_c(G) \to \mathrm{C}_c(G/\Gamma) : f \mapsto \overline{f}$ is surjective, there exists $\psi \in \mathrm{C}_c(G)$ such that $\overline{\psi} \equiv 1$ on the compact subset $\mathrm{supp}(\varphi)\Gamma \subset G/\Gamma$. Therefore, we obtain

$$\int_G \varphi(h)\, \mathrm{d}\mu_G(h) = \int_G \varphi(h)\overline{\psi}(h\Gamma)\, \mathrm{d}\mu_G(h) = \int_G \overline{\varphi}(h\Gamma)\psi(h)\, \mathrm{d}\mu_G(h) = 0.$$

By Riesz's representation theorem, there exists a unique regular Borel measure ν on G/Γ such that

$$\forall f \in \mathrm{C}_c(G), \quad \int_G f(h)\, \mathrm{d}\mu_G(h) = \int_G \overline{f}(h\Gamma)\, \mathrm{d}\nu(h\Gamma).$$

Note that the above argument does not use the fact that $\Gamma < G$ is uniform.

However, since $\Gamma < G$ is uniform, G/Γ is compact and we have $0 < \nu(G/\Gamma) < +\infty$. Up to normalization, we may assume that $\nu(G/\Gamma) = 1$.

Define the left invariant Haar measure m_G on G by the formula $m_G(B) = \mu_G(B^{-1})$ for every $B \in \mathscr{B}(G)$. Then for every $B \in \mathscr{B}(G)$ and every $g \in G$, we have

$$(g_*\mu_G)(B) = \mu_G(g^{-1}B) = m_G(B^{-1}g) = \Delta_G(g)\, m_G(B^{-1}) = \Delta_G(g)\, \mu_G(B)$$

and so $g_*\mu_G = \Delta_G(g)\, \mu_G$. By uniqueness in the previous construction, we obtain $g_*\nu = \Delta_G(g)\, \nu$ for every $g \in G$. Since $\nu \in \mathrm{Prob}(G/\Gamma)$ is a probability measure, we obtain $\Delta_G(g) = 1$ and $g_*\nu = \nu$ for every $g \in G$. Thus, $\Delta_G \equiv 1$ and so G is unimodular. Moreover, $\nu \in \mathrm{Prob}(G/\Gamma)$ is G-invariant and so $\Gamma < G$ is a lattice.

Assume moreover that G is compactly generated. Choose a compact subset $Q \subset G$ such that $e \in Q$ and $G = \bigcup_{n\geq 1} Q^n$. Since G/Γ is compact, we may choose a compact subset $K \subset G$ such that $p(K) = G/\Gamma$ (see the proof of surjectivity of the map $\mathscr{T} : \mathrm{C}_c(G) \to \mathrm{C}_c(G/\Gamma)$). Upon replacing Q by $Q \cup K$, we may further assume that $Q \cdot \Gamma = G$. Then $S_0 = Q \cap \Gamma$ is finite. Moreover, since Q^2 is compact, there exists a finite subset $S_1 \subset \Gamma$

such that $Q^2 \subset QS_1$. Indeed, otherwise we could find sequences $(g_n)_n$ in Q^2, $(h_n)_n$ in Q and $(\gamma_n)_n$ in Γ such that $g_n = h_n\gamma_n$ for every $n \in \mathbb{N}$ and $(\gamma_n)_n$ are pairwise distinct. This would imply that $\gamma_n = h_n^{-1}g_n \in Q^3 \cap \Gamma$ for every $n \in \mathbb{N}$. Since Q^3 is compact and $\Gamma < G$ is discrete, $Q^3 \cap \Gamma$ must be finite, a contradiction. Set $S = S_0 \cup S_1 \subset \Gamma$. Then $Q \cap \Gamma \subset S$ and for every $n \geq 1$, we have $Q^{n+1} \subset QS^n$. We claim that S is a finite generating set for Γ. Indeed, by construction, we have $Q \cap \Gamma \subset S$. Next, let $n \geq 1$ and $\gamma \in Q^{n+1} \cap \Gamma \subset QS^n \cap \Gamma$. Then $\gamma = g\gamma_n$ where $g \in Q$ and $\gamma_n \in S^n$. This implies that $\gamma\gamma_n^{-1} = g \in Q \cap \Gamma \subset S$. Then $\gamma = g\gamma_n \in SS^n = S^{n+1}$ and hence $Q^{n+1} \cap \Gamma \subset S^{n+1}$. This implies that $\Gamma = \bigcup_{n\geq 1} Q^n \cap \Gamma \subset \bigcup_{n \geq 1} S^n$ and so Γ is finitely generated. $\square$

PROPOSITION 3.15. *Let G be a locally compact group that possesses a lattice $\Gamma < G$. Then G is unimodular. Moreover, there is a unique G-invariant regular Borel probability measure $\nu \in \mathrm{Prob}(G/\Gamma)$.*

PROOF. Let $\nu \in \mathrm{Prob}(G/\Gamma)$ be a G-invariant regular Borel probability measure. We claim that there exists a unique left invariant Haar measure m_G on G such that

$$\forall f \in \mathrm{C}_c(G), \quad \int_G f(h)\,\mathrm{d}m_G(h) = \int_{G/\Gamma} \overline{f}(g\Gamma)\,\mathrm{d}\nu(g\Gamma). \tag{3.3}$$

Indeed, the well-defined positive linear functional

$$\mathrm{C}_c(G) \to \mathbb{C} : f \mapsto \int_{G/\Gamma} \overline{f}(g\Gamma)\,\mathrm{d}\nu(g\Gamma)$$

is left invariant. By Riesz's representation theorem, there exists a unique left invariant Haar measure m_G on G for which (3.3) holds.

Applying (3.1), for every $f \in \mathrm{C}_c(G)$ and every $\gamma \in \Gamma$, letting $f_\gamma = f(\,\cdot\,\gamma^{-1}) \in \mathrm{C}_c(G)$, we have

$$\begin{aligned}
\Delta_G(\gamma)\int_G f(h)\,\mathrm{d}m_G(h) &= \int_G f_\gamma(h)\,\mathrm{d}m_G(h) \\
&= \int_{G/\Gamma} \overline{f_\gamma}(h\Gamma)\,\mathrm{d}\nu(h\Gamma) \\
&= \int_{G/\Gamma} \overline{f}(h\Gamma)\,\mathrm{d}\nu(h\Gamma) \\
&= \int_G f(h)\,\mathrm{d}m_G(h).
\end{aligned}$$

This implies that $\Delta_G(\gamma) = 1$ for every $\gamma \in \Gamma$. Consider the well-defined continuous mapping $\overline{\Delta} : G/\Gamma \to \mathbb{R}_+^* : g\Gamma \mapsto \Delta_G(g)$. Then $\eta = \overline{\Delta}_*\nu \in \mathrm{Prob}(\mathbb{R}_+^*)$ is a Borel probability measure that is invariant under multiplication by $\Delta_G(g)$ for every $g \in G$. This implies that $\Delta_G \equiv 1$ and so G is unimodular.

Observe that (3.3) together with surjectivity of $\mathscr{T} : \mathrm{C}_c(G) \to \mathrm{C}_c(G/\Gamma)$ imply that there is a unique G-invariant regular Borel probability measure $\nu \in \mathrm{Prob}(G/\Gamma)$. ☐

The next proposition provides a group-theoretic characterization of uniform lattices in locally compact groups.

PROPOSITION 3.16. *Let G be a locally compact group and $\Gamma < G$ a lattice. The following assertions are equivalent:*

(i) *$\Gamma < G$ is uniform.*

(ii) *There exists a compact neighborhood $U \subset G$ of $e \in G$ such that for every $g \in G$, we have $g\Gamma g^{-1} \cap U = \{e\}$.*

PROOF. (i) $\Rightarrow$ (ii) Assume that $\Gamma < G$ is uniform. Since $\Gamma < G$ is discrete, we may choose a compact neighborhood $W \subset G$ of $e \in G$ such that $\Gamma \cap W = \{e\}$. Next, we may choose a symmetric compact neighborhood $V \subset W$ of $e \in G$ such that $VVV \subset W$. Observe that for every $h \in V$, we have

$$h\Gamma h^{-1} \cap V \subset h(\Gamma \cap h^{-1}Vh)h^{-1} \subset h(\Gamma \cap W)h^{-1} = \{e\}.$$

By compactness of G/Γ, there exist $n \geq 1$ and $g_1, \dots, g_n \in G$ such that $G/\Gamma = \bigcup_{i=1}^n g_i p(V)$. Set $U = \bigcap_{i=1}^n g_i V g_i^{-1}$. Then for every $g \in G$, there exist $1 \leq i \leq n$ and $h \in V$ such that $g\Gamma = g_i h\Gamma$ and hence

$$g\Gamma g^{-1} \cap U = g_i h\Gamma h^{-1} g_i^{-1} \cap U \subset g_i(h\Gamma h^{-1} \cap V)g_i^{-1} = \{e\}.$$

(ii) $\Rightarrow$ (i) Denote by $\nu \in \mathrm{Prob}(G/\Gamma)$ the unique G-invariant regular Borel probability measure and by m_G the unique Haar measure on G such that (3.3) holds. Assume that there exists such a compact neighborhood $U \subset G$ of $e \in G$. Choose a compact neighborhood $V \subset G$ of $e \in G$ such that $V^{-1}V \subset U$. Choose a nonnegative function $\varphi \in \mathrm{C}_c(G)$ such that $0 \leq \varphi \leq 1$ and $\mathrm{supp}(\varphi) \subset V$. Set $\varepsilon = \int_G \varphi(h) \, \mathrm{d}m_G(h)$.

For every $g \in G$, define $\varphi_g = \varphi(\,\cdot\, g^{-1}) \in \mathrm{C}_c(G)$. Note that $0 \leq \varphi_g \leq 1$ and $\mathrm{supp}(\varphi_g) \subset Vg$. Moreover, we have $\mathrm{supp}(\overline{\varphi_g}) \subset Vg\Gamma$. Since m_G is right invariant, we have

$$\begin{aligned}
\varepsilon &= \int_G \varphi(h) \, \mathrm{d}m_G(h) \\
&= \int_G \varphi_g(h) \, \mathrm{d}m_G(h) \\
&= \int_{G/\Gamma} \overline{\varphi_g}(h\Gamma) \, \mathrm{d}\nu(h\Gamma) \\
&= \int_{Vg\Gamma} \overline{\varphi_g}(h\Gamma) \, \mathrm{d}\nu(h\Gamma) \\
&= \int_{Vg\Gamma} \sum_{\gamma \in \Gamma} \varphi_g(h\gamma) \, \mathrm{d}\nu(h\Gamma).
\end{aligned}$$

We claim that for every $h \in Vg\Gamma$, there is at most one $\gamma \in \Gamma$ such that $h\gamma \in Vg$. Indeed, if $\gamma_1, \gamma_2 \in \Gamma$ are elements such that $h\gamma_1, h\gamma_2 \in Vg$, then $g\gamma_1^{-1}\gamma_2 g^{-1} \in V^{-1}V \subset U$. Since $g\Gamma g^{-1} \cap U = \{e\}$, we have $\gamma_1 = \gamma_2$. Since $0 \leq \varphi_g \leq 1$ and $\mathrm{supp}(\varphi_g) \subset V_g$, it follows that

$$\varepsilon = \int_{Vg\Gamma} \sum_{\gamma \in \Gamma} \varphi_g(h\gamma)\, \mathrm{d}\nu(h\Gamma) \leq \int_{Vg\Gamma} 1\, \mathrm{d}\nu(h\Gamma) = \nu(Vg\Gamma).$$

We have showed that $\nu(Vg\Gamma) \geq \varepsilon$ for every $g \in G$.

Let $F \subset G$ be a finite subset for which for every $g, h \in F$ such that $g \neq h$, we have $Vg\Gamma \cap Vh\Gamma = \emptyset$. Then we have

$$\sharp F \cdot \varepsilon \leq \sum_{g \in F} \nu(Vg\Gamma) = \nu(\bigcup_{g \in F} Vg\Gamma) \leq 1$$

and hence $\sharp F \leq \varepsilon^{-1}$. We may then choose a maximal finite subset $F \subset G$ with the aforementioned property. It follows that for every $g \in G$, we have $Vg\Gamma \cap VF\Gamma \neq \emptyset$ and hence $g\Gamma \in V^{-1}VF\Gamma \subset UF\Gamma$. Since $UF\Gamma \subset G/\Gamma$ is compact, it follows that $G/\Gamma = UF\Gamma$ is compact. $\square$

When G is a locally compact second countable group, we prove a very useful criterion to ensure that a discrete subgroup $\Gamma < G$ is a lattice.

THEOREM 3.17. *Let G be a locally compact second countable group and $\Gamma < G$ a discrete subgroup. The following assertions are equivalent:*

(i) *$\Gamma < G$ is a lattice.*
(ii) *G is unimodular and there is a Borel fundamental domain $\mathscr{F} \subset G$ for the right translation action $\Gamma \curvearrowright G$ such that $0 < m_G(\mathscr{F}) < +\infty$.*
(iii) *G is unimodular and there is a Borel subset $\mathfrak{S} \subset G$ such that $\mathfrak{S} \cdot \Gamma = G$ and $0 < m_G(\mathfrak{S}) < +\infty$.*

PROOF. Recall that since G is a locally compact second countable group, the discrete subgroup $\Gamma < G$ is necessarily countable.

(i) $\Rightarrow$ (ii) We already know that G is unimodular by Proposition 3.15. Denote by $\nu \in \mathrm{Prob}(G/\Gamma)$ the unique G-invariant regular Borel probability measure. Denote by m_G the unique Haar measure on G satisfying (3.3). Since G is locally compact second countable, (3.3) holds for every nonnegative Borel function $f : G \to \mathbb{R}_+$. In particular, for $f = \mathbf{1}_{\mathscr{F}}$, we have $\overline{f} \equiv 1$ and so

$$m_G(\mathscr{F}) = \int_G f(h)\, \mathrm{d}m_G(h) = \int_{G/\Gamma} \overline{f}\, \mathrm{d}\nu(h\Gamma) = 1 < +\infty.$$

Since $m_G(G) > 0$, $G = \bigcup_{\gamma \in \Gamma} \mathscr{F}\gamma$ and $m_G(\mathscr{F}\gamma) = m_G(\mathscr{F})$ for every $\gamma \in \Gamma$, we also have $m_G(\mathscr{F}) > 0$.

(ii) $\Rightarrow$ (iii) It is trivial.

(iii) $\Rightarrow$ (i) Following the proof of Proposition 3.14 and since m_G is right invariant, we may consider the well-defined nonzero left invariant linear functional

$$\Phi : \mathrm{C}_c(G/\Gamma) \to \mathbb{C} : \overline{f} \mapsto \int_G f(g)\, \mathrm{d}m_G(g).$$

By Riesz's representation theorem, there exists a unique nonzero G-invariant regular Borel measure ν on G/Γ such that (3.3) holds. Since G is locally compact second countable, (3.3) holds for every nonnegative Borel function $f : G \to \mathbb{R}_+$. In particular, for $f = \mathbf{1}_{\mathfrak{S}}$, we have $\overline{f} \geq 1$ and so

$$\nu(G/\Gamma) \leq \int_{G/\Gamma} \overline{f}\, \mathrm{d}\nu(h\Gamma) = \int_G f(h)\, \mathrm{d}m_G(h) = m_G(\mathfrak{S}) < +\infty.$$

Then $\frac{1}{\nu(G/\Gamma)}\nu \in \mathrm{Prob}(G/\Gamma)$ is a G-invariant regular Borel probability measure and so $\Gamma < G$ is a lattice. $\square$

Let us point out that when $\Gamma < G$ is a lattice, all Borel fundamental domains for the right translation action $\Gamma \curvearrowright G$ have the same finite Haar measure. Indeed, whenever $\mathscr{F}_1, \mathscr{F}_2 \subset G$ are Borel fundamental domains, since the Haar measure m_G on G is right invariant, we have

$$\begin{aligned} m_G(\mathscr{F}_1) &= \sum_{\gamma \in \Gamma} m_G(\mathscr{F}_1 \cap \mathscr{F}_2 \gamma) \\ &= \sum_{\gamma \in \Gamma} m_G(\mathscr{F}_1 \gamma^{-1} \cap \mathscr{F}_2) \\ &= m_G(\mathscr{F}_2). \end{aligned}$$

Examples 3.18. Here are some examples of lattices in locally compact groups.

(i) For every $d \geq 1$, the discrete subgroup $\mathbb{Z}^d < \mathbb{R}^d$ is a uniform lattice.
(ii) More generally, any lattice $\Gamma < G$ in a locally compact second countable abelian group G is necessarily uniform.
(iii) The discrete Heisenberg group $\mathrm{H}_3(\mathbb{Z}) < \mathrm{H}_3(\mathbb{R})$ is a uniform lattice in the continuous Heisenberg group $\mathrm{H}_3(\mathbb{R})$:

$$\mathrm{H}_3(\mathbb{Z}) = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{Z} \right\}$$
$$\mathrm{H}_3(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}.$$

(iv) More generally, any lattice $\Gamma < G$ in a locally compact second countable nilpotent group G is necessarily uniform.

3. $\mathrm{SL}_d(\mathbb{Z})$ is a lattice in $\mathrm{SL}_d(\mathbb{R})$, $d \geq 2$

In this section, we prove the following theorem due to Minkowski.

THEOREM 3.19 (Minkowski). *For every $d \geq 2$, the discrete subgroup $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$ is a nonuniform lattice.*

Before proving Theorem 3.19, we need to prove some preliminary results that are also of independent interest.

Let $d \geq 1$. Endow $\mathbb{R}^d$ with its canonical euclidean structure. Denote by $K = \mathrm{SO}_d(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R})$ the special orthogonal subgroup and observe that $K < \mathrm{SL}_d(\mathbb{R})$ is compact. Denote by $A < \mathrm{SL}_d(\mathbb{R})$ the subgroup of diagonal matrices with positive entries, that is,

$$A = \{a = \mathrm{diag}(\lambda_1, \dots, \lambda_d) \mid \lambda_1, \dots, \lambda_d > 0,\ \lambda_1 \cdots \lambda_d = 1\} < \mathrm{SL}_d(\mathbb{R}).$$

Denote by $N = \mathrm{T}_d(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R})$ the strict upper triangular subgroup as in Example 3.10(vi).

LEMMA 3.20 (Iwasawa decomposition). *The map $K \times A \times N \to \mathrm{SL}_d(\mathbb{R}) : (k, a, n) \mapsto kan$ is a homeomorphism. We simply write $\mathrm{SL}_d(\mathbb{R}) = K \cdot A \cdot N$.*

PROOF. Denote by $(e_1, \dots, e_d)$ the canonical basis of $\mathbb{R}^d$. The map $\Psi : K \times A \times N \to \mathrm{SL}_d(\mathbb{R}) : (k, a, n) \mapsto kan$ is clearly continuous. Conversely, let $g \in \mathrm{SL}_d(\mathbb{R})$ and write $v_i = ge_i \in \mathbb{R}^d$ for every $1 \leq i \leq d$. By Gram–Schmidt's orthogonalization process, set $w_1 = v_1$ and $w_{i+1} = v_{i+1} - P_{V_i}(v_{i+1})$ where $V_i = \mathrm{Vect}(v_1, \dots, v_i)$ for every $1 \leq i \leq d-1$. Then $(\frac{w_1}{\|w_1\|}, \dots, \frac{w_d}{\|w_d\|})$ is an orthonormal basis for $\mathbb{R}^d$ and we may find $k \in \mathrm{O}_d(\mathbb{R})$ such that $ke_i = \frac{w_i}{\|w_i\|}$ for every $1 \leq i \leq d$. Then the matrix $k^{-1}g$ is upper triangular and $(k^{-1}g)_{ii} = \|w_i\|$ for every $1 \leq i \leq d$. It follows that $\det(k^{-1}) = \det(k^{-1}g) = \|w_1\| \cdots \|w_d\| > 0$ and hence $k \in \mathrm{SO}_d(\mathbb{R})$. Letting $a = \mathrm{diag}(\|w_1\|, \dots, \|w_d\|) \in A$, we have $g = kan$ and the map $\mathrm{SL}_d(\mathbb{R}) \to K \times A \times N : g \mapsto (k, a, n)$ is continuous. Since its inverse is Ψ, we have showed that $\Psi : K \times A \times N \to \mathrm{SL}_d(\mathbb{R}) : (k, a, n) \mapsto kan$ is a homeomorphism. $\square$

LEMMA 3.21. *Endow $(K, \mathrm{d}k)$, $(A, \mathrm{d}a)$, $(N, \mathrm{d}n)$ with their respective Haar measure. Then the pushforward measure of*

$$\prod_{1 \leq i < j \leq d} \frac{\lambda_i}{\lambda_j} \,\mathrm{d}k \,\mathrm{d}a \,\mathrm{d}n$$

under the map $K \times A \times N \to \mathrm{SL}_d(\mathbb{R}) : (k, a, n) \mapsto kan$ is a Haar measure on $\mathrm{SL}_d(\mathbb{R})$.

PROOF. Consider the product map $\Psi : K \times AN \to \mathrm{SL}_d(\mathbb{R}) : (k, p) \mapsto k^{-1}p$. Since $\mathrm{SL}_d(\mathbb{R})$ is unimodular, the regular Borel measure $(\Psi^{-1})_* m_{\mathrm{SL}_d(\mathbb{R})}$ on $K \times AN$ is right invariant. Then $(\Psi^{-1})_* m_{\mathrm{SL}_d(\mathbb{R})}$ is a right invariant Haar measure on the locally compact second countable group $K \times AN$ and hence $(\Psi^{-1})_* m_{\mathrm{SL}_d(\mathbb{R})} = \mu_K \otimes \mu_{AN}$ where μ_K is a right invariant Haar measure on

K and μ_{AN} is a right invariant Haar measure on AN. Since K is compact, μ_K is also left invariant and hence we may assume that $\mathrm{d}\mu_K(k) = \mathrm{d}k$. It remains to prove that $\prod_{1 \leq i < j \leq d} \frac{\lambda_i}{\lambda_j} \,\mathrm{d}a \,\mathrm{d}n$ is a right invariant Haar measure on AN.

As explained in Examples 3.10(vi), we may assume that $\mathrm{d}m_N(n) = \mathrm{d}n = \prod_{1 \leq i < j \leq d} \mathrm{d}n_{ij}$. Observe that $N \lhd AN$ is a normal subgroup and define the conjugation action $\mathrm{Ad} : A \curvearrowright N$ by $\mathrm{Ad}(a)(n) = ana^{-1}$ for $a \in A$, $n \in N$. Then $AN = A \ltimes N$ and $\mathrm{d}a \,\mathrm{d}n$ is a left invariant measure on AN by Proposition 3.9. A simple calculation shows that $\mathrm{Ad}(a)_* m_N = (\prod_{1 \leq i < j \leq d} \frac{\lambda_i}{\lambda_j})^{-1} \cdot m_N$. Then Proposition 3.9 implies that $\prod_{1 \leq i < j \leq d} \frac{\lambda_i}{\lambda_j} \,\mathrm{d}a \,\mathrm{d}n$ is a right invariant Haar measure on AN. $\square$

For all $t, u > 0$, set

$$\begin{aligned} A_t &= \{a = \mathrm{diag}(\lambda_1, \dots, \lambda_d) \in A \mid \forall 1 \leq i \leq d-1, \lambda_i \leq t\lambda_{i+1}\} \\ N_u &= \{n = (n_{ij})_{ij} \in N \mid \forall 1 \leq i < j \leq d, |n_{ij}| \leq u\} \\ \mathfrak{S}_{t,u} &= K \cdot A_t \cdot N_u. \end{aligned}$$

The Borel subset $\mathfrak{S}_{t,u} \subset G$ is called a *Siegel domain*. We now have all the tools to prove Theorem 3.19.

PROOF OF THEOREM 3.19. For every $t \geq \frac{2}{\sqrt{3}}$ and every $u \geq \frac{1}{2}$, we show that $\mathrm{SL}_d(\mathbb{R}) = \mathfrak{S}_{t,u} \cdot \mathrm{SL}_d(\mathbb{Z})$ and that $\mathfrak{S}_{t,u}$ has finite Haar measure. By Theorem 3.17, this implies that $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$ is a lattice. We divide the proof into a series of claims.

CLAIM 3.22. For all $t, u > 0$, the Siegel domain $\mathfrak{S}_{t,u}$ has finite Haar measure.

Indeed, note that since K and N_u are both compact in $\mathrm{SL}_d(\mathbb{R})$, using Lemma 3.21 it suffices to prove that

$$\kappa_t = \int_{A_t} \prod_{1 \leq i < j \leq d} \frac{\lambda_i}{\lambda_j} \,\mathrm{d}a < +\infty.$$

Observe that the map

$$\Theta : A \to \mathbb{R}^{d-1} : \mathrm{diag}(\lambda_1, \dots, \lambda_d) \mapsto \left(\log \frac{\lambda_2}{\lambda_1}, \dots, \log \frac{\lambda_d}{\lambda_{d-1}}\right)$$

is a topological group isomorphism. We may choose the Haar measure $\mathrm{d}a$ on A that is the pushforward of the Lebesgue measure on $\mathbb{R}^{d-1}$ by Θ^{-1}. We then have

$$\begin{aligned} \kappa_t &= \int_{\mathbb{R}^{d-1}} \prod_{1 \leq i < j \leq d} \exp(-(s_i + \cdots + s_{j-1})) \mathbf{1}_{\{s_1, \dots, s_{d-1} \geq -\log t\}} \,\mathrm{d}s_1 \cdots \mathrm{d}s_{d-1} \\ &= \prod_{k=1}^{d-1} \int_{-\log t}^{+\infty} \exp(-k(d-k)s_k) \,\mathrm{d}s_k < +\infty. \end{aligned}$$

CLAIM 3.23. For every $u \geq \frac{1}{2}$, we have $N = N_u \cdot (N \cap \mathrm{SL}_d(\mathbb{Z}))$.

Indeed, it suffices to prove Claim 3.23 for $u = \frac{1}{2}$. We proceed by induction over $d \geq 1$. For $d = 1$, there is nothing to prove. Assume that the result is true for $d - 1 \geq 1$ and let us prove it for d. Let $n \in N = \mathrm{T}_d(\mathbb{R})$ that we write

$$n = \begin{pmatrix} 1 & * \\ 0 & n_0 \end{pmatrix} \quad \text{where} \quad n_0 \in \mathrm{T}_{d-1}(\mathbb{R}).$$

By induction hypothesis, there exists $\gamma_0 \in \mathrm{T}_{d-1}(\mathbb{R}) \cap \mathrm{SL}_{d-1}(\mathbb{Z})$ such that $n_1 = n_0\gamma_0^{-1} \in \mathrm{T}_{d-1}(\mathbb{R})_{1/2}$. Write

$$n \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0^{-1} \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & n_1 \end{pmatrix} \quad \text{where} \quad x \in \mathbb{R}^{d-1}.$$

Choose $y \in \mathbb{Z}^{d-1}$ such that $x - y \in [-1/2, 1/2]^{d-1}$. Then

$$\begin{aligned} n &= \begin{pmatrix} 1 & x \\ 0 & n_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & x - y \\ 0 & n_1 \end{pmatrix} \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0 \end{pmatrix} \end{aligned}$$

where

$$\begin{pmatrix} 1 & x - y \\ 0 & n_1 \end{pmatrix} \in N_{1/2} \quad \text{and} \quad \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0 \end{pmatrix} \in N \cap \mathrm{SL}_d(\mathbb{Z}).$$

This shows the result is true for d and finishes the proof of Claim 3.23.

CLAIM 3.24. For every $t \geq \frac{2}{\sqrt{3}}$, we have $\mathrm{SL}_d(\mathbb{R}) = K \cdot A_t \cdot N \cdot \mathrm{SL}_d(\mathbb{Z})$.

Indeed, it suffices to prove Claim 3.24 for $t = \frac{2}{\sqrt{3}}$. We proceed by induction over $d \geq 1$. For $d = 1$, there is nothing to prove. Assume that the result is true for $d - 1 \geq 1$ and let us prove it for d. Denote by $(e_1, \ldots, e_d)$ the canonical basis of $\mathbb{R}^d$. Let $g \in \mathrm{SL}_d(\mathbb{R})$. Since $\Lambda = g\mathbb{Z}^d$ is a lattice in $\mathbb{R}^d$, there must exist a vector $v_1 \in \Lambda \setminus \{0\}$ such that

$$\|v_1\| = \min \{\|v\| \mid v \in \Lambda \setminus \{0\}\}.$$

By minimality of the norm of $v_1 \in \Lambda \setminus \{0\}$, we may find $v_2, \ldots, v_d \in \Lambda \setminus \{0\}$ such that $(v_1, \ldots, v_d)$ is a basis of Λ (see e.g. [**Ca71**, Corollary I.3]). Upon further replacing v_1 by $-v_1$, there exists $\gamma \in \mathrm{SL}_d(\mathbb{Z})$ such that $\gamma e_i = g^{-1}v_i$ for every $1 \leq i \leq d$. Note that $g\gamma e_1 = v_1$.

Next, consider the Iwasawa decomposition $g\gamma = kan$ and write

$$an = \begin{pmatrix} \lambda^{d-1} & * \\ 0 & \lambda^{-1}g_0 \end{pmatrix} \quad \text{where} \quad \lambda \in \mathbb{R}_+^*,\ g_0 \in \mathrm{SL}_{d-1}(\mathbb{R}).$$

By induction hypothesis, there exist $k_0 \in \mathrm{SO}_{d-1}(\mathbb{R})$ and $\gamma_0 \in \mathrm{SL}_{d-1}(\mathbb{Z})$ such that $k_0^{-1}g_0\gamma_0^{-1} \in (A_{d-1})_{2/\sqrt{3}} \cdot \mathrm{T}_{d-1}(\mathbb{R})$. If we consider

$$h = \begin{pmatrix} 1 & 0 \\ 0 & k_0^{-1} \end{pmatrix} k^{-1}g\gamma \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0^{-1} \end{pmatrix} = \begin{pmatrix} \lambda^{d-1} & * \\ 0 & \lambda^{-1}\,k_0^{-1}g_0\gamma_0^{-1} \end{pmatrix} \in AN$$

we obtain that the diagonal coefficients of h satisfy $h_{i,i} \leq \frac{2}{\sqrt{3}} h_{i+1,i+1}$ for every $2 \leq i \leq d-1$. It remains to prove that $h_{1,1} \leq \frac{2}{\sqrt{3}} h_{2,2}$. Observe that for every $w \in \mathbb{Z}^d \setminus \{0\}$, we have

$$\|he_1\| = \|g\gamma \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0^{-1} \end{pmatrix} e_1\| = \|g\gamma e_1\| = \|v_1\| \leq \|g\gamma \begin{pmatrix} 1 & 0 \\ 0 & \gamma_0^{-1} \end{pmatrix} w\| = \|hw\|.$$

Using Claim 3.23, write $h = \operatorname{diag}(h_{11}, \ldots, h_{dd}) n_1 \gamma_1$ where $n_1 \in N_{1/2}$ and $\gamma_1 \in N \cap \mathrm{SL}_d(\mathbb{Z})$. Then $he_1 = \operatorname{diag}(h_{11}, \ldots, h_{dd}) e_1 = h_{11} e_1$ and with $w = \gamma_1^{-1} e_2 \in \mathbb{Z}^d \setminus \{0\}$, we have $hw = \operatorname{diag}(h_{11}, \ldots, h_{dd}) n_1 e_2 = h_{11} n_{12} e_1 + h_{22} e_2$. Then we obtain

$$h_{11}^2 = \|he_1\|^2 \leq \|hw\|^2 = h_{11}^2 n_{12}^2 + h_{22}^2 \leq \frac{1}{4} h_{11}^2 + h_{22}^2$$

and so $h_{11}^2 \leq \frac{4}{3} h_{22}^2$. This finishes the proof of Claim 3.24.

A combination of Claims 3.22, 3.23, 3.24 and Theorem 3.17 implies that $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$ is a lattice.

It remains to prove that $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$ is nonuniform. Indeed, regard $\mathrm{SL}_2(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R})$ as a subgroup in the top left corner and set

$$\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{Z}).$$

Then a simple calculation shows that

$$g_n \gamma g_n^{-1} = \begin{pmatrix} 1 & n^{-2} \\ 0 & 1 \end{pmatrix} \to e \quad \text{with} \quad g_n = \begin{pmatrix} n^{-1} & 0 \\ 0 & n \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R}).$$

Then Proposition 3.16 implies that $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$ is nonuniform. $\square$

4. Howe–Moore's property and Moore's ergodicity theorem

4.1. Generalities on unitary representations. Let $(\mathscr{H}, \langle \cdot , \cdot \rangle)$ be a (complex) Hilbert space. We always assume that $\langle \cdot , \cdot \rangle$ is conjugate linear in the second variable. We denote by

$$\mathscr{U}(\mathscr{H}) = \{u \in \mathrm{B}(\mathscr{H}) \mid u^* u = u u^* = 1_{\mathscr{H}}\}$$

the group of unitary operators on $\mathscr{H}$. We simply write $1 = 1_{\mathscr{H}}$. We endow $\mathscr{U}(\mathscr{H})$ with the *strong operator topology* defined as the initial topology on $\mathscr{U}(\mathscr{H})$ that makes the maps $\mathscr{U}(\mathscr{H}) \to \mathbb{R} : u \mapsto \|(u-1)\xi\|$ continuous for all $\xi \in \mathscr{H}$. Then $\mathscr{U}(\mathscr{H})$ is a topological group but $\mathscr{U}(\mathscr{H})$ need not be locally compact. When $\mathscr{H}$ is separable, $\mathscr{U}(\mathscr{H})$ is a Polish group.

Definition 3.25. Let G be a locally compact group. We say that the mapping $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ is a *strongly continuous unitary representation* if the following conditions hold:

(i) $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ is a group homomorphism.
(ii) $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ is strongly continuous, meaning that π is a continuous map when $\mathscr{U}(\mathscr{H}_\pi)$ is endowed with the strong operator topology as above.

When $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ only satisfies condition (i), we simply say that π is a *unitary representation*. When G is discrete, condition (ii) is trivially satisfied.

The next result shows that in order to prove that the unitary representation $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ is strongly continuous, it is enough to show that the coefficients of π are measurable functions.

LEMMA 3.26. *Let G be a locally compact group, $\mathscr{H}_\pi$ a separable Hilbert space and $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ a unitary representation. Assume that for all $\xi, \eta \in \mathscr{H}_\pi$, the map $\varphi_{\xi,\eta} : G \to \mathbb{C} : g \mapsto \langle \pi(g)\xi, \eta\rangle$ is measurable. Then π is strongly continuous.*

PROOF. Let $\xi \in \mathscr{H}_\pi$ be a vector. It suffices to show that the map $G \to \mathscr{H}_\pi : g \mapsto \pi(g)\xi$ is continuous at $e \in G$. Let $Q \subset G$ be a symmetric compact neighborhood of $e \in G$. Consider the compactly generated open subgroup $H = \bigcup_{n\geq 1} Q^n < G$. It further suffices to show that the map $H \to \mathscr{H}_\pi : g \mapsto \pi(g)\xi$ is continuous at $e \in H$. Upon replacing G by H, we may as well assume that G is σ-compact.

As usual, we denote by m_G a left invariant Haar measure on G. Let $\varepsilon > 0$ and set $B = \{g \in G \mid \|\pi(g)\xi - \xi\| < \varepsilon/2\}$. Then $B \subset G$ is a measurable subset since $B = \left\{g \in G \mid 2\Re(\langle \pi(g)\xi, \xi\rangle) > 2\|\xi\|^2 - \varepsilon^2/4\right\}$. Moreover, we have $B^{-1} = B$ and $B^2 = BB^{-1} \subset \{g \in G \mid \|\pi(g)\xi - \xi\| < \varepsilon\}$. Since $\pi(G)\xi \subset \mathscr{H}_\pi$ is separable, there exists a sequence $(g_n)_n$ in G such that $(\pi(g_n)\xi)_n$ is dense in $\pi(G)\xi$. This implies that $\bigcup_{n\in\mathbb{N}} g_n B = G$ and so $m_G(B) > 0$. Since G is σ-compact, upon replacing B by $B \cap K$ for a suitable symmetric compact subset, we may further assume that $B = B^{-1}$, $B \subset K$ and $0 < m_G(B) < +\infty$. Then $\mathbf{1}_B \in \mathrm{L}^2(G, \mathscr{B}(G), m_G)$ and $\varphi = \mathbf{1}_B * \mathbf{1}_B \in \mathrm{C}_c(G)$ with $\mathrm{supp}(\varphi) \subset \overline{BB} \subset KK$. Since $\varphi(e) = m_G(B) > 0$, the subset $U = \varphi^{-1}(0, +\infty)$ is open, $e \in U$ and $U \subset BB \subset \{g \in G \mid \|\pi(g)\xi - \xi\| < \varepsilon\}$. $\square$

DEFINITION 3.27. Let G be a locally compact group and $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ a strongly continuous unitary representation. We say that

- π has *invariant vectors* and we write $1_G \subset \pi$ if the subspace of $\pi(G)$-invariant vectors
$$(\mathscr{H}_\pi)^G = \{\xi \in \mathscr{H}_\pi \mid \forall g \in G, \pi(g)\xi = \xi\}$$
is nonzero. Otherwise, we say that π is *ergodic* and we write $1_G \not\subset \pi$.
- π has *almost invariant vectors* and we write $1_G \prec \pi$ if for every $\varepsilon > 0$ and every compact subset $Q \subset G$, there exists a unit vector $\xi \in \mathscr{H}_\pi$ such that
$$\sup_{g\in Q} \|\pi(g)\xi - \xi\| < \varepsilon.$$
Otherwise, we say that π has *spectral gap* and we write $1_G \not\prec \pi$.

It is clear that if $1_G \subset \pi$, then $1_G \prec \pi$.

For every $i \in \{1, 2\}$, let $\pi_i : G \to \mathscr{U}(\mathscr{H}_{\pi_i})$ be a strongly continuous unitary representation. We say that π_1 and π_2 are *unitarily equivalent* if there exists a unitary operator $U : \mathscr{H}_{\pi_1} \to \mathscr{H}_{\pi_2}$ such that for every $g \in G$, we have $\pi_2(g) = U\pi_1(g)U^*$. In this situation, we will identify π_1 with π_2.

4.2. Examples of unitary representations. Let G be a locally compact group.

The left regular representation λ_G. Let m_G be a left invariant Haar measure on G and simply denote by $\mathrm{L}^2(G) = \mathrm{L}^2(G, \mathscr{B}(G), m_G)$ the corresponding Hilbert space of L^2-integrable functions on G. Define the *left regular representation* $\lambda_G : G \to \mathscr{U}(\mathrm{L}^2(G))$ by the formula

$$\forall g \in G, \forall \xi \in \mathrm{L}^2(G), \quad (\lambda_G(g)\xi)(h) = \xi(g^{-1}h).$$

The left regular representation $\lambda_G : G \to \mathscr{U}(\mathrm{L}^2(G))$ is a strongly continuous unitary representation. This follows from the well known facts that the subspace $\mathrm{C}_c(G)$ of compactly supported continuous functions on G is $\|\cdot\|_2$-dense in $\mathrm{L}^2(G)$ and the left translation action $\lambda : G \curvearrowright \mathrm{C}_c(G)$ is $\|\cdot\|_\infty$-continuous (see Lemma 3.8).

Proposition 3.28. *Keep the same notation as above. Then $1_G \subset \lambda_G$ if and only if G is compact.*

Proof. If G is compact, then the left invariant Haar measure m_G is finite. This implies that the constant function $\mathbf{1}_G$ belongs to $\mathrm{L}^2(G)$ and is $\lambda_G(G)$-invariant. Conversely, assume that there exists a nonzero $\lambda_G(G)$-invariant vector $\xi \in \mathrm{L}^2(G)$.

Claim 3.29. There exists a σ-compact open subgroup $H < G$ such that $\xi = \mathbf{1}_H\xi$.

Indeed, define the measurable subsets $B = \{h \in G \mid \xi(h) \neq 0\}$ and $B_n = \left\{h \in G \mid |\xi(h)| \geq n^{-1}\right\}$ for every $n \geq 1$. Then $B = \bigcup_{n\geq 1} B_n$ and $m_G(B_n) < +\infty$ for every $n \geq 1$. By regularity, for every $n \geq 1$, there exists an open set $U_n \subset G$ such that $B_n \subset U_n$ and $m_G(U_n) < +\infty$. To prove the claim, it suffices to show that every open set $U \subset G$ with finite Haar measure is contained in a σ-compact open subgroup $H < G$.

Let $U \subset G$ be a nonempty open set such that $m_G(U) < +\infty$. Let $L < G$ be a σ-compact open subgroup. Then the set $\Lambda = \{gL \in G/L \mid U \cap gL \neq \emptyset\}$ is at most countable. Letting $H < G$ be the subgroup generated by L and Λ, we have that $U \subset H$ and $H < G$ is σ-compact and open. This finishes the proof of Claim 3.29.

Using Claim 3.29 and the assumption, for every $g \in G$, we have

$$\mathbf{1}_H\xi = \xi = \lambda_G(g)\xi = \lambda_G(g)(\mathbf{1}_H\xi) = \mathbf{1}_{gH}\xi.$$

Since $\xi \neq 0$, we have $gH = H$ for every $g \in G$ and so $H = G$. This shows that G is σ-compact.

We may now apply Fubini's theorem. Indeed, since for every $g \in G$ and m_G-almost every $h \in G$, we have $\xi(g^{-1}h) = \xi(h)$, Fubini's theorem implies that there exists $h \in G$ such that for m_G-almost every $g \in G$, we have $\xi(g^{-1}h) = \xi(h)$. This further implies that ξ is essentially constant. If we denote by $c > 0$ the essential value of $|\xi|^2$, we obtain $c \cdot m_G(G) = \|\xi\|^2 < +\infty$ and so $m_G(G) < +\infty$. Then G is compact by Proposition 3.6. $\square$

The Koopman representation κ. Let G be a locally compact second countable group and $(X, \mathscr{X}, \nu)$ a standard probability space. We endow G with its σ-algebra $\mathscr{B}(G)$ of Borel subsets. Endow the product space $G \times X$ with the product σ-algebra $\mathscr{B}(G) \otimes \mathscr{X}$. Let $G \curvearrowright (X, \mathscr{X}, \nu)$ be a *probability measure preserving* (pmp) action, meaning that the action map $G \times X \to X : (g, x) \mapsto gx$ is measurable and that $g_*\nu = \nu$ for every $g \in G$. Denote by $\mathrm{L}^2(X, \mathscr{X}, \nu)$ the Hilbert space of L^2-integrable functions on X. Since $(X, \mathscr{X}, \nu)$ is a standard probability space, $\mathrm{L}^2(X, \mathscr{X}, \nu)$ is separable. Define the *Koopman representation* $\kappa : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu))$ associated with the pmp action $G \curvearrowright (X, \mathscr{X}, \nu)$ by the formula

$$\forall g \in G, \forall \xi \in \mathrm{L}^2(X, \mathscr{X}, \nu), \quad (\kappa(g)\xi)(x) = \xi(g^{-1}x).$$

The Koopman representation $\kappa : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu))$ is a strongly continuous unitary representation. This follows from Lemma 3.26 after noticing that for all $\xi, \eta \in \mathrm{L}^2(X, \mathscr{X}, \nu)$, the map

$$\varphi_{\xi,\eta} : G \to \mathbb{C} : g \mapsto \langle \kappa(g)\xi, \eta \rangle = \int_X \xi(g^{-1}x)\overline{\eta(x)}\, \mathrm{d}\nu(x)$$

is measurable thanks to Fubini's theorem. The constant function $\mathbf{1}_X$ is $\kappa(G)$-invariant. For this reason, it is natural to consider the restriction of the Koopman representation to the orthogonal complement $\mathrm{L}^2(X, \mathscr{X}, \nu)^0 = \mathrm{L}^2(X, \mathscr{X}, \nu) \ominus \mathbb{C}\mathbf{1}_X$ that we denote by $\kappa^0 : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$.

We say that a measurable subset $Y \subset X$ is

- *ν-almost everywhere G-invariant* if $\nu(gY \triangle Y) = 0$ for every $g \in G$.
- *strictly G-invariant* if $gY = Y$ for every $g \in G$.

The next lemma clarifies the difference between the two notions.

LEMMA 3.30. *For any ν-almost everywhere G-invariant measurable subset $Y \subset X$, there is a strictly G-invariant measurable subset $Z \subset X$ such that $\nu(Y \triangle Z) = 0$.*

PROOF. Fix a left invariant Haar measure m_G on G. By assumption and using Fubini's theorem, the measurable subset

$$X_0 = \left\{x \in X \mid G \to \mathbb{C} : g \mapsto \mathbf{1}_Y(g^{-1}x) \text{ is } m_G\text{-a.e. constant}\right\}$$

is ν-conull in X. For every $x \in X_0$, denote by $f(x)$ the unique essential value of the measurable function $G \to \mathbb{C} : g \mapsto \mathbf{1}_Y(g^{-1}x)$. For every $x \in X \setminus X_0$, set $f(x) = 0$. Note that $f(X) \subset \{0, 1\}$. Fubini's theorem implies that the function $f : X \to \mathbb{C}$ is measurable and $f(x) = \mathbf{1}_Y(x)$ for ν-almost every $x \in X$. For every $x \in X_0$ and every $h \in G$, the measurable function $G \to \mathbb{C} : g \mapsto \mathbf{1}_Y(g^{-1}h^{-1}x)$ is m_G-almost everywhere constant, hence $h^{-1}x \in X_0$ and $f(h^{-1}x) = f(x)$. This further implies that f is strictly G-invariant meaning that $f(g^{-1}x) = f(x)$ for every $g \in G$ and every $x \in X$. Set

$Z = \{x \in X \mid f(x) = 1\}$. Then $Z \subset X$ is a strictly G-invariant measurable subset such that $\nu(Y \triangle Z) = 0$. $\square$

From now on, we simply say that the measurable subset $Y \subset X$ is G-invariant if for every $g \in G$, we have $\nu(gY \triangle Y) = 0$. We say that the pmp action $G \curvearrowright (X, \mathscr{X}, \nu)$ is *ergodic* if every G-invariant measurable subset $Y \subset X$ is null or conull.

Proposition 3.31. *Keep the same notation as above. Then* $1_G \subset \kappa^0$ *if and only if the pmp action* $G \curvearrowright (X, \mathscr{X}, \nu)$ *is not ergodic.*

Proof. If the pmp action $G \curvearrowright (X, \mathscr{X}, \nu)$ is not ergodic, then there exists a G-invariant measurable subset $Y \subset X$ such that $0 < \nu(Y) < 1$. Then the nonzero vector $\xi = \mathbf{1}_Y - \nu(Y)\mathbf{1}_X \in \mathrm{L}^2(X, \mathscr{X}, \nu)^0$ is $\kappa^0(G)$-invariant. Conversely, assume that there exists a nonzero $\kappa^0(G)$-invariant vector $\xi \in \mathrm{L}^2(X, \mathscr{X}, \nu)^0$. Upon taking the real or imaginary part of ξ, we may assume that ξ is real-valued. Next, upon taking $\xi^+ = \max(\xi, 0)$ or $\xi^- = \max(-\xi, 0)$, we may further assume that $\xi \in \mathrm{L}^2(X, \mathscr{X}, \nu)$ is $\kappa(G)$-invariant, nonnegative and $\xi \notin \mathbb{C}\mathbf{1}_X$. For every $t > 0$, define the G-invariant measurable subset $X_t = \left\{x \in X \mid \xi(x)^2 \geq t\right\}$. Then the function $\mathbb{R}_+^* \to \mathbb{R}_+ : t \mapsto \nu(X_t)$ is measurable, decreasing and satisfies $\|\xi\|^2 = \int_0^{+\infty} \nu(X_t)\, \mathrm{d}t$. We claim that there exists $t > 0$ such that $0 < \nu(X_t) < 1$. Indeed otherwise there would exist $s > 0$ such that $\nu(X_t) = 0$ for every $t > s$ and $\nu(X_t) = 1$ for every $t \leq s$. This would imply that ξ is ν-almost everywhere constant equal to $\sqrt{s}$ and thus $\xi \in \mathbb{C}\mathbf{1}_X$, a contradiction. Therefore, there exists $t > 0$ such that $0 < \nu(X_t) < 1$. This shows that the pmp action $G \curvearrowright (X, \mathscr{X}, \nu)$ is not ergodic. $\square$

The quasi-regular representation $\lambda_{G/\Gamma}$. Let G be a locally compact second countable group and $\Gamma < G$ a lattice. We endow the locally compact second countable space $X = G/\Gamma$ with its σ-algebra $\mathscr{X} = \mathscr{B}(G/\Gamma)$ of Borel subsets (see Proposition 3.11(iii)). We denote by $\nu \in \mathrm{Prob}(X)$ the unique G-invariant Borel probability measure (see Proposition 3.15). Then the action $G \curvearrowright (X, \mathscr{X}, \nu)$ is pmp. In that case, we denote by $\lambda_X : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu))$ the Koopman representation and we call it the *quasi-regular representation*. Since $G \curvearrowright X$ is transitive, Lemma 3.30 implies that $G \curvearrowright (X, \mathscr{X}, \nu)$ is ergodic and Proposition 3.31 implies that $\lambda_X^0 : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$ is ergodic. We can strengthen the above result when $\Gamma < G$ is a *uniform* lattice.

Proposition 3.32. *Assume that* $\Gamma < G$ *is a uniform lattice. Then* λ_X^0 *has spectral gap.*

Proof. We may choose a Borel section $\sigma : X \to G$ such that $\sigma(X)$ is relatively compact in G (see Proposition 3.11 and Corollary 3.12). We further choose the Haar measure m_G on G such that $\sigma_*\nu = m_G|_{\sigma(X)}$. Set $Q = \sigma(X)\sigma(X)^{-1} \subset G$. Observe that $Q = Q^{-1}$ is relatively compact in G and so $m_G(Q) < +\infty$. Let $(\xi_n)_n$ be a bounded sequence of vectors in

$\mathrm{L}^2(X, \mathscr{X}, \nu)^0$ such that $\lim_n \sup_{g\in Q} \|\lambda_X^0(g)\xi_n - \xi_n\| = 0$. Using Fubini's theorem, we obtain

$$\begin{aligned}
\int_X |\xi_n(x)|^2 \,\mathrm{d}\nu(x) &= \frac{1}{2}\int_X \left(\int_{\sigma(X)\sigma(x)^{-1}} |\xi_n(gx) - \xi_n(x)|^2 \,\mathrm{d}m_G(g)\right) \mathrm{d}\nu(x) \\
&\leq \frac{1}{2}\int_X \left(\int_Q |\xi_n(gx) - \xi_n(x)|^2 \,\mathrm{d}m_G(g)\right) \mathrm{d}\nu(x) \\
&= \frac{1}{2}\int_Q \left(\int_X |\xi_n(gx) - \xi_n(x)|^2 \,\mathrm{d}\nu(x)\right) \mathrm{d}m_G(g) \\
&= \frac{1}{2}\int_Q \|\lambda_X^0(g^{-1})\xi_n - \xi_n\|^2 \,\mathrm{d}m_G(g) \\
&= \frac{1}{2} m_G(Q) \cdot \sup_{g\in Q} \|\lambda_X^0(g^{-1})\xi_n - \xi_n\| \to 0 \quad \text{as} \quad n \to +\infty.
\end{aligned}$$

This implies that $\lim_n \|\xi_n\| = 0$ and thus λ_X^0 has spectral gap. $\square$

4.3. Howe–Moore's property for $\mathrm{SL}_d(\mathbb{R})$, $d \geq 2$. Let $\mathscr{H}$ be a (complex) Hilbert space and denote by $\mathrm{B}(\mathscr{H})$ the unital Banach $*$-algebra of all bounded linear operators $T : \mathscr{H} \to \mathscr{H}$. Besides the norm topology on $\mathrm{B}(\mathscr{H})$ given by the supremum norm

$$\forall T \in \mathrm{B}(\mathscr{H}), \quad \|T\|_\infty = \sup\left\{\|T\xi\| \mid \xi \in \mathscr{H}, \|\xi\| \leq 1\right\},$$

we can define two weaker locally convex topologies on $\mathrm{B}(\mathscr{H})$ as follows.

- The *strong operator topology* on $\mathrm{B}(\mathscr{H})$ is defined as the initial topology on $\mathrm{B}(\mathscr{H})$ that makes the maps $\mathrm{B}(\mathscr{H}) \to \mathbb{C} : T \mapsto \|T\xi\|$ continuous for all $\xi \in \mathscr{H}$.
- The *weak operator topology* on $\mathrm{B}(\mathscr{H})$ is defined as the initial topology on $\mathrm{B}(\mathscr{H})$ that makes the maps $\mathrm{B}(\mathscr{H}) \to \mathbb{C} : T \mapsto |\langle T\xi, \eta\rangle|$ continuous for all $\xi, \eta \in \mathscr{H}$.

Note that we already defined the strong operator topology on $\mathscr{U}(\mathscr{H})$. As a matter of fact, on $\mathscr{U}(\mathscr{H})$, strong and weak operator topologies coincide. Observe that when $\mathscr{H}$ is separable, both strong and weak operator topologies are metrizable on the unit ball of $\mathrm{B}(\mathscr{H})$ denoted by $\mathrm{Ball}(\mathrm{B}(\mathscr{H}))$. Moreover, $\mathrm{Ball}(\mathrm{B}(\mathscr{H}))$ is weakly compact.

Let G be a locally compact group and $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ a strongly continuous unitary representation. We say that π is *mixing* if $\pi(g) \to 0$ weakly as $g \to \infty$. Note that when G is noncompact, the left regular representation $\lambda_G : G \to \mathscr{U}(\mathrm{L}^2(G))$ is mixing. Let $G \curvearrowright (X, \mathscr{X}, \nu)$ be a pmp action on a standard probability space. We say that $G \curvearrowright (X, \mathscr{X}, \nu)$ is *mixing* if the Koopman representation $\kappa^0 : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$ is mixing. It is easy to check that $G \curvearrowright (X, \mathscr{X}, \nu)$ is mixing if and only if

$$\forall A, B \in \mathscr{X}, \quad \lim_{g\to\infty} \nu(A \cap gB) = \nu(A)\nu(B).$$

Any mixing strongly continuous unitary representation is ergodic. In that respect, we introduce the following terminology.

DEFINITION 3.33. Let G be a noncompact locally compact group. We say that G has the *Howe–Moore property* if any ergodic strongly continuous unitary representation $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$ is mixing.

Observe that when G has the Howe–Moore property, for every nontrivial strongly continuous unitary representation $\pi : G \to \mathscr{U}(\mathscr{H}_\pi)$, the subrepresentation $\pi^0 : G \to \mathscr{U}(\mathscr{H}_\pi \ominus (\mathscr{H}_\pi)^G)$ is ergodic hence mixing. Here are some properties enjoyed by locally compact groups with the Howe–Moore property.

PROPOSITION 3.34. *Let G be a noncompact locally compact group with the Howe–Moore property. The following assertions hold:*

(i) *For every closed normal subgroup $N \lhd G$, either N is compact or $N = G$.*
(ii) *For every open subgroup $H < G$, either H is compact or $H = G$.*
(iii) *For every ergodic pmp action $G \curvearrowright (X, \mathscr{X}, \nu)$ and every noncompact closed subgroup $H < G$, the action $H \curvearrowright (X, \mathscr{X}, \nu)$ is mixing.*

PROOF. (i) Let $N \lhd G$ be a proper closed normal subgroup. Define the quasi-regular representation $\pi : G \to \mathscr{U}(\mathrm{L}^2(G/N))$ and note that $\pi = \lambda_{G/N} \circ p$ where $p : G \to G/N$ is the canonical factor map and $\lambda_{G/N} : G/N \to \mathscr{U}(\mathrm{L}^2(G/N))$ is the left regular representation of the locally compact group G/N. Since $N \neq G$, we have $\mathrm{L}^2(G/N)^G \neq \mathrm{L}^2(G/N)$. By Howe–Moore property, the subrepresentation $\pi^0 : G \to \mathscr{U}(\mathrm{L}^2(G/N) \ominus \mathrm{L}^2(G/N)^G)$ is mixing. Since $\pi|_N \equiv 1$, it follows that $\pi^0|_N \equiv 1$ and thus N is compact.

(ii) Let $H < G$ be a proper open subgroup. Then the homogeneous space G/H is discrete and nontrivial. Define the strongly continuous unitary representation $\pi : G \to \mathscr{U}(\ell^2(G/H))$ by the formula

$$\forall g, h \in G, \quad \pi(g)\delta_{hH} = \delta_{ghH}.$$

Since $H \neq G$, the unit vector $\delta_H \in \ell^2(G/H)$ is not $\pi(G)$-invariant and so $\ell^2(G/H)^G \neq \ell^2(G/H)$. By Howe–Moore property, the subrepresentation $\pi^0 : G \to \mathscr{U}(\ell^2(G/H) \ominus \ell^2(G/H)^G)$ is mixing. Since the nonzero vector $\xi = \delta_H - P_{\ell^2(G/H)^G}(\delta_H) \in \ell^2(G/H) \ominus \ell^2(G/H)^G$ is $\pi(H)$-invariant, it follows that H is compact.

(iii) Let $G \curvearrowright (X, \mathscr{X}, \nu)$ be an ergodic pmp action and $H < G$ a noncompact closed subgroup. By Proposition 3.31, the Koopman representation $\kappa^0 : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$ is ergodic. By Howe–Moore property, $\kappa^0 : G \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$ is mixing and so is $\pi|_H : H \to \mathscr{U}(\mathrm{L}^2(X, \mathscr{X}, \nu)^0)$. Therefore, $H \curvearrowright (X, \mathscr{X}, \nu)$ is mixing. □

The main theorem of this subsection is the following well-known result due to Howe–Moore [**HM77**].

THEOREM 3.35 (Howe–Moore). *For every $d \geq 2$, $\mathrm{SL}_d(\mathbb{R})$ has the Howe–Moore property.*

As a consequence of Theorem 3.35 and Proposition 3.34(iii), we obtain the following ergodicity result due to Moore [**Mo65**].

COROLLARY 3.36 (Moore). *Let $d \geq 2$ and set $G = \mathrm{SL}_d(\mathbb{R})$. Let $\Gamma < G$ be a lattice and denote by $\nu \in \mathrm{Prob}(G/\Gamma)$ the unique G-invariant Borel probability measure. For every noncompact closed subgroup $H < G$, the pmp action $H \curvearrowright (G/\Gamma, \mathscr{B}(G/\Gamma), \nu)$ is ergodic.*

In particular, for every $g \in G$ that is not contained in a compact subgroup, the pmp dynamical system $(G/\Gamma, \mathscr{B}(G/\Gamma), \nu, T_g)$ is ergodic.

Before proving Theorem 3.35, we need to prove some preliminary results that are also of independent interest.

Define the following subgroups of $\mathrm{SL}_2(\mathbb{R})$:

$$U^+ = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \mid x \in \mathbb{R} \right\}$$
$$U^- = \left\{ \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \mid x \in \mathbb{R} \right\}$$
$$A = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \mid \lambda > 0 \right\}.$$

Observe that $\mathrm{SL}_2(\mathbb{R})$ is generated by $U^+ \cup U^-$.

LEMMA 3.37. *Let $\pi : \mathrm{SL}_2(\mathbb{R}) \to \mathscr{U}(\mathscr{H}_\pi)$ be a strongly continuous unitary representation. Every $\pi(U^+)$-invariant vector is $\pi(\mathrm{SL}_2(\mathbb{R}))$-invariant.*

PROOF. Let $\xi \in \mathscr{H}_\pi$ be a $\pi(U^+)$-invariant unit vector. Define the continuous function $\varphi : G \to \mathbb{C} : g \mapsto \langle \pi(g)\xi, \xi\rangle$. By assumption, φ is U^+-bi-invariant. For every $n \geq 1$, set

$$g_n = \begin{pmatrix} 0 & -n \\ \frac{1}{n} & 0 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

A simple calculation shows that for every $\lambda > 0$, we have

$$\begin{pmatrix} 1 & \lambda n \\ 0 & 1 \end{pmatrix} g_n \begin{pmatrix} 1 & \frac{n}{\lambda} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ \frac{1}{n} & \lambda^{-1} \end{pmatrix} \to \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Since φ is continuous and U^+-bi-invariant, it follows that

$$\forall a \in A, \quad \varphi(a) = \lim_n \varphi(g_n) = \varphi(1) = 1.$$

This further implies that $\pi(a)\xi = \xi$ for every $a \in A$. It follows that φ is A-bi-invariant.

Another simple calculation shows that for every $x \in \mathbb{R}$, we have

$$\begin{pmatrix} n & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{x}{n^2} & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since φ is continuous and A-bi-invariant, it follows that for every $u \in U^-$, we have $\varphi(u) = 1$ and so $\pi(u)\xi = \xi$.

We have showed that ξ is both $\pi(U^+)$-invariant and $\pi(U^-)$-invariant. Since $\mathrm{SL}_2(\mathbb{R})$ is generated by $U^+ \cup U^-$, it follows that ξ is $\pi(\mathrm{SL}_2(\mathbb{R}))$-invariant. $\square$

Let $d \geq 2$. For all $1 \leq a \neq b \leq d$ and all $x \in \mathbb{R}$, denote by $E_{ab}(x) \in \mathrm{SL}_d(\mathbb{R})$ the elementary matrix defined by $(E_{ab}(x))_{ij} = 1$ if $i = j$, $(E_{ab}(x))_{ij} = x$ if $i = a$ and $j = b$, $(E_{ab}(x))_{ij} = 0$ otherwise. We leave as an exercise to check that $\mathrm{SL}_d(\mathbb{R})$ is generated by $\{E_{ab}(x) \mid 1 \leq a \neq b \leq d, x \in \mathbb{R}\}$. For every $2 \leq k \leq d$, regard $\mathrm{SL}_k(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R})$ as the following subgroup:

$$\mathrm{SL}_k(\mathbb{R}) \cong \left\{ \begin{pmatrix} A & 0_{d-k,k} \\ 0_{k,d-k} & 1_{d-k,d-k} \end{pmatrix} \mid A \in \mathrm{SL}_k(\mathbb{R}) \right\}.$$

For all $1 \leq \ell_1 < \ell_2 \leq d$, denote by $H_{\ell_1,\ell_2} < \mathrm{SL}_d(\mathbb{R})$ the (ℓ_1, ℓ_2)-*copy* of $\mathrm{SL}_2(\mathbb{R})$ in $\mathrm{SL}_d(\mathbb{R})$ that consists in all matrices $g \in \mathrm{SL}_d(\mathbb{R})$ such that $g_{\ell_1\ell_1} = \alpha$, $g_{\ell_1\ell_2} = \beta$, $g_{\ell_2\ell_1} = \gamma$, $g_{\ell_2\ell_2} = \delta$, $g_{ii} = 1$ for all $i \neq \ell_1, \ell_2$, $g_{ij} = 0$ for all $i \neq j$ and $\{i, j\} \neq \{\ell_1, \ell_2\}$ and such that

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

LEMMA 3.38. *Let $d \geq 2$ and $\pi : \mathrm{SL}_d(\mathbb{R}) \to \mathscr{U}(\mathscr{H}_\pi)$ be a strongly continuous unitary representation. Let $\xi \in \mathscr{H}_\pi$ be a $\pi(H_{\ell_1,\ell_2})$-invariant vector for some $1 \leq \ell_1 < \ell_2 \leq d$. Then ξ is $\pi(\mathrm{SL}_d(\mathbb{R}))$-invariant.*

PROOF. Upon permuting the indices, we may assume that $\ell_1 = 1$ and $\ell_2 = 2$. We proceed by induction over $2 \leq k \leq d$. By assumption, ξ is $\pi(\mathrm{SL}_2(\mathbb{R}))$-invariant. Assume that ξ is $\pi(\mathrm{SL}_k(\mathbb{R}))$-invariant for $2 \leq k \leq d - 1$ and let us show that ξ is $\pi(\mathrm{SL}_{k+1}(\mathbb{R}))$-invariant. Let $1 \leq j \leq k$ and $x \in \mathbb{R}$. For every $n \geq 1$, denote by $g_n \in \mathrm{SL}_k(\mathbb{R}) < \mathrm{SL}_{k+1}(\mathbb{R})$ any diagonal matrix such that $(g_n)_{ii} = \frac{1}{n}$ if $i = j$. Then a simple computation shows that $g_n E_{j(k+1)}(x) g_n^{-1} = E_{j(k+1)}(\frac{x}{n}) \to 1$ as $n \to \infty$ and $g_n^{-1} E_{(k+1)j}(x) g_n = E_{(k+1)j}(\frac{x}{n}) \to 1$ as $n \to \infty$. Since $\pi(g_n)\xi = \xi$, we have

$$\begin{aligned} \|\pi(E_{j(k+1)}(x))\xi - \xi\| &= \lim_n \|\pi(E_{j(k+1)}(x))\pi(g_n)^*\xi - \pi(g_n)^*\xi\| \\ &= \lim_n \|\pi(g_n E_{j(k+1)}(x) g_n^{-1})\xi - \xi\| = 0 \end{aligned}$$

and so $\pi(E_{j(k+1)}(x))\xi = \xi$. Likewise, we have $\pi(E_{(k+1)j}(x))\xi = \xi$. Since $\mathrm{SL}_{k+1}(\mathbb{R})$ is generated by

$$\mathrm{SL}_k(\mathbb{R}) \cup \left\{ E_{j(k+1)}(x), E_{(k+1)j}(x) \mid 1 \leq j \leq k, x \in \mathbb{R} \right\},$$

it follows that ξ is $\pi(\mathrm{SL}_{k+1}(\mathbb{R}))$-invariant. By induction over $2 \leq k \leq d$, we have that ξ is $\pi(\mathrm{SL}_d(\mathbb{R}))$-invariant. $\square$

Let $d \geq 2$. Denote by $K = \mathrm{SO}_d(\mathbb{R}) < \mathrm{SL}_d(\mathbb{R})$ the special orthogonal subgroup and observe that $K < \mathrm{SL}_d(\mathbb{R})$ is compact. Define the subset

$A^+ \subset \mathrm{SL}_d(\mathbb{R})$ of diagonal matrices by

$$A^+ = \{\mathrm{diag}(\lambda_1, \dots, \lambda_d) \mid \lambda_1 \geq \dots \geq \lambda_d > 0,\ \lambda_1 \cdots \lambda_d = 1\} \subset \mathrm{SL}_d(\mathbb{R}).$$

We now have all the tools to prove Theorem 3.35.

Proof of Theorem 3.35. Let $d \geq 2$ and $\pi : \mathrm{SL}_d(\mathbb{R}) \to \mathscr{U}(\mathscr{H}_\pi)$ be a strongly continuous unitary representation. Assuming that π is not mixing, we show that there exists a nonzero $\pi(\mathrm{SL}_d(\mathbb{R}))$-invariant vector. Since $\mathrm{SL}_d(\mathbb{R})$ is second countable, $\pi(G)\xi$ is separable for every $\xi \in \mathscr{H}_\pi$ and so we may assume assume that $\mathscr{H}_\pi$ is separable. Since π is not mixing, there exists a sequence $(g_n)_n$ in G such that $g_n \to \infty$ and $\pi(g_n) \not\to 0$ weakly. Upon taking a subsequence, we may assume that there exists $T \in \mathrm{B}(\mathscr{H})$ such that $T \neq 0$ and $\pi(g_n) \to T$ weakly. Using Lemma 2.35, there exist sequences $(k_{1,n})_n$ and $(k_{2,n})_n$ in K and $(a_n)_n$ in A^+ such that $g_n = k_{1,n} a_n k_{2,n}$ for every $n \in \mathbb{N}$. Upon taking another subsequence, we may assume that $k_{1,n} \to k_1$ in K and $k_{2,n} \to k_2$ in K. This implies that $\pi(k_{1,n}) \to \pi(k_1)$ and $\pi(k_{2,n}) \to \pi(k_2)$ strongly. This further implies that $\pi(a_n) \to \pi(k_1)^* T \pi(k_2)^*$ weakly. Set $S = \pi(k_1)^* T \pi(k_2)^* \in \mathrm{B}(\mathscr{H})$ and observe that $S \neq 0$.

For every $n \in \mathbb{N}$, write $a_n = \mathrm{diag}(\lambda_{1,n}, \dots, \lambda_{d,n})$ with $\lambda_{1,n} \geq \dots \geq \lambda_{d,n}$ and $\lambda_{1,n} \cdots \lambda_{d,n} = 1$. Since $a_n \to \infty$, it follows that $\frac{\lambda_{1,n}}{\lambda_{d,n}} \to +\infty$. A simple computation shows that for every $x \in \mathbb{R}$,

$$a_n^{-1} E_{1d}(x) a_n = E_{1d}(\frac{\lambda_{d,n}}{\lambda_{1,n}} x) \to 1.$$

This implies that for every $x \in \mathbb{R}$, we have $\pi(E_{1d}(x))S = S$ since

$$\begin{aligned}
\forall \eta_1, \eta_2 \in \mathscr{H}_\pi, \quad \langle \pi(E_{1d}(x)) S\eta_1, \eta_2\rangle &= \lim_n \langle \pi(E_{1d}(x))\pi(a_n)\eta_1, \eta_2\rangle \\
&= \lim_n \langle \pi(a_n^{-1} E_{1d}(x) a_n)\eta_1, \pi(a_n^{-1})\eta_2\rangle \\
&= \langle \eta_1, S^* \eta_2\rangle \\
&= \langle S\eta_1, \eta_2\rangle.
\end{aligned}$$

Choose $\eta \in \mathscr{H}_\pi$ so that $\xi = S\eta \neq 0$. Then $\xi \in \mathscr{H}_\pi$ is a nonzero $\pi(E_{1d}(\mathbb{R}))$-invariant vector. Denote by $H_{1d} < \mathrm{SL}_d(\mathbb{R})$ the $(1, d)$-copy of $\mathrm{SL}_2(\mathbb{R})$. By Lemma 3.37, ξ is $\pi(H_{1d})$-invariant and by Lemma 3.38, ξ is $\pi(\mathrm{SL}_d(\mathbb{R}))$-invariant. This finishes the proof of Theorem 3.35. $\square$

APPENDIX A

Appendix

Martingale convergence theorem

Let $(X, \mathscr{X}, \nu)$ be a probability space. Let $\mathscr{Y} \subset \mathscr{X}$ be a σ-subalgebra. Regard $\mathrm{L}^1(X, \mathscr{Y}, \nu) \subset \mathrm{L}^1(X, \mathscr{X}, \nu)$ and denote by $\mathbb{E}_\nu(\,\cdot\,|\mathscr{Y}) : \mathrm{L}^1(X, \mathscr{X}, \nu) \to \mathrm{L}^1(X, \mathscr{Y}, \nu)$ the *conditional expectation* which is the unique ν-preserving linear positive contraction such that $\mathbb{E}_\nu(f|\mathscr{Y}) = f$ for every $f \in \mathrm{L}^1(X, \mathscr{Y}, \nu)$.

In this section, we prove Doob's martingale convergence theorem.

Theorem A.1. *Let $(\mathscr{Y}_n)_n$ be an increasing sequence of σ-subalgebras of $\mathscr{X}$ and denote by $\mathscr{Y} = \sigma((\mathscr{Y}_n)_n)$ the σ-subalgebra of $\mathscr{X}$ generated by $\bigcup_{n\in\mathbb{N}} \mathscr{Y}_n$. Then for every $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, the sequence $(\mathbb{E}_\nu(f|\mathscr{Y}_n))_n$ converges to $\mathbb{E}_\nu(f|\mathscr{Y})$ ν-almost everywhere and in $\mathrm{L}^1(X, \mathscr{X}, \nu)$.*

Proof. Firstly, we prove that for every $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, the associated sequence $(\mathbb{E}_\nu(f|\mathscr{Y}_n))_n$ converges to $\mathbb{E}_\nu(f|\mathscr{Y})$ in $\mathrm{L}^1(X, \mathscr{X}, \nu)$. Let $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$ and $\varepsilon > 0$. Since the subspace $\bigcup_{n\in\mathbb{N}} \mathrm{L}^1(X, \mathscr{Y}_n, \nu)$ is $\|\cdot\|_1$-dense in $\mathrm{L}^1(X, \mathscr{Y}, \nu)$, there exists $n_0 \in \mathbb{N}$ and $g \in \mathrm{L}^1(X, \mathscr{Y}_{n_0}, \nu)$ such that $\|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 \leq \frac{\varepsilon}{2}$. For every $n \geq n_0$, using the triangle inequality and the contraction property of the conditional expectation and since $g \in \mathrm{L}^1(X, \mathscr{Y}_n, \nu)$, we have

$$\begin{aligned}
\|\mathbb{E}_\nu(f|\mathscr{Y}) - \mathbb{E}_\nu(f|\mathscr{Y}_n)\|_1 &\leq \|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 + \|g - \mathbb{E}_\nu(f|\mathscr{Y}_n)\|_1 \\
&= \|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 + \|\mathbb{E}_\nu(g - \mathbb{E}_\nu(f|\mathscr{Y})|\mathscr{Y}_n)\|_1 \\
&\leq 2\|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 \leq \varepsilon.
\end{aligned}$$

Therefore, the sequence $(\mathbb{E}_\nu(f|\mathscr{Y}_n))_n$ converges to $\mathbb{E}_\nu(f|\mathscr{Y})$ in $\mathrm{L}^1(X, \mathscr{X}, \nu)$.

Secondly, we prove that for every $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$, the associated sequence $(\mathbb{E}_\nu(f|\mathscr{Y}_n))_n$ converges to $\mathbb{E}_\nu(f|\mathscr{Y})$ ν-almost everywhere. Recall that for every $g \in \mathrm{L}^1(X, \mathscr{X}, \nu)$ and every $a > 0$, we have $\nu(\{|g| > a\}) \leq \frac{1}{a}\int_X |g| \,\mathrm{d}\nu$ (Chebyshev's inequality). We prove the following key result.

Claim A.2. Let $g \in \mathrm{L}^1(X, \mathscr{X}, \nu)$ be such that $g \geq 0$. Set $G = \sup_n \mathbb{E}_\nu(g|\mathscr{Y}_n) \geq 0$. Then for every $a > 0$, we have

$$\nu(\{G > a\}) \leq \frac{1}{a}\int_X g \,\mathrm{d}\nu.$$

Let $a > 0$. For every $n \in \mathbb{N}$, denote by $Z_n \in \mathscr{X}$ the measurable subset consisting of all elements $x \in X$ for which $\mathbb{E}_\nu(g|\mathscr{Y}_{n+1})(x) > a$ and

$\max\{\mathbb{E}_\nu(g|\mathscr{Y}_k)(x) \mid 0 \le k \le n\} \le a$. Then we have $\{G > a\} = \bigsqcup_n Z_n$. For every $n \in \mathbb{N}$, since $Z_n \in \mathscr{Y}_{n+1}$, we have

$$\begin{aligned}
\nu(Z_n) &\le \frac{1}{a}\int_X \mathbf{1}_{Z_n}\mathbb{E}_\nu(g|\mathscr{Y}_{n+1})\,\mathrm{d}\nu \\
&\le \frac{1}{a}\int_X \mathbb{E}_\nu(\mathbf{1}_{Z_n} g|\mathscr{Y}_{n+1})\,\mathrm{d}\nu \\
&= \frac{1}{a}\int_X \mathbf{1}_{Z_n} g\,\mathrm{d}\nu \\
&= \frac{1}{a}\int_{Z_n} g\,\mathrm{d}\nu.
\end{aligned}$$

Summing over $\mathbb{N}$, we obtain

$$\nu(\{G > a\}) = \sum_{n\in\mathbb{N}} \nu(Z_n) \le \sum_{n\in\mathbb{N}} \frac{1}{a}\int_{Z_n} g\,\mathrm{d}\nu \le \frac{1}{a}\int_X g\,\mathrm{d}\nu.$$

This finishes the proof of the claim.

Let $f \in \mathrm{L}^1(X, \mathscr{X}, \nu)$. Upon taking the real and imaginary parts, we may assume that f is real-valued. Let $\varepsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ and $g \in \mathrm{L}^1(X, \mathscr{Y}_{n_0}, \nu)$ such that $\|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 \le \varepsilon^2$. We may assume that g is also real valued. For every $n \ge n_0$, using the triangle inequality and since $g \in \mathrm{L}^1(X, \mathscr{Y}_n, \nu)$, we have

$$\begin{aligned}
|\mathbb{E}_\nu(f|\mathscr{Y}) - \mathbb{E}_\nu(f|\mathscr{Y}_n)| &\le |\mathbb{E}_\nu(f|\mathscr{Y}) - g| + |g - \mathbb{E}_\nu(f|\mathscr{Y}_n)| \\
&= |\mathbb{E}_\nu(f|\mathscr{Y}) - g| + |\mathbb{E}_\nu(g - \mathbb{E}_\nu(f|\mathscr{Y})|\mathscr{Y}_n)| \\
&\le |\mathbb{E}_\nu(f|\mathscr{Y}) - g| + \mathbb{E}_\nu(|\mathbb{E}_\nu(f|\mathscr{Y}) - g||\mathscr{Y}_n).
\end{aligned}$$

Using Chebyshev's inequality and Claim A.2, this further implies that

$$\begin{aligned}
&\nu\left(\limsup_n |\mathbb{E}_\nu(f|\mathscr{Y}) - \mathbb{E}_\nu(f|\mathscr{Y}_n)| \ge 2\varepsilon\right) \\
&\quad\le \nu\left(|\mathbb{E}_\nu(f|\mathscr{Y}) - g| \ge \varepsilon\right) + \nu\left(\sup_n \mathbb{E}_\nu(|\mathbb{E}_\nu(f|\mathscr{Y}) - g||\mathscr{Y}_n) \ge \varepsilon\right) \\
&\quad\le \frac{1}{\varepsilon}\|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 + \frac{1}{\varepsilon}\|\mathbb{E}_\nu(f|\mathscr{Y}) - g\|_1 \le 2\varepsilon.
\end{aligned}$$

Since this holds true for every $\varepsilon > 0$, it follows that $\limsup_n |\mathbb{E}_\nu(f|\mathscr{Y}) - \mathbb{E}_\nu(f|\mathscr{Y}_n)| = 0$ ν-almost everywhere. This finishes the proof. $\square$

Bibliography

[BS02] M. Brin, G. Stuck, *Introduction to dynamical systems.* Cambridge University Press, Cambridge, 2002. xii+240 pp.

[Ca71] J.W.S. Cassels, *An introduction to the geometry of numbers.* Corrected reprint of the 1971 edition. Classics in Mathematics. Springer-Verlag, Berlin, 1997. viii+344 pp.

[DE14] A. Deitmar, S. Echterhoff, *Principles of harmonic analysis.* Second edition. Universitext. Springer, Cham, 2014. xiv+332 pp.

[Di71] E.I. Dinaburg, *A connection between various entropy characterizations of dynamical systems.* Izv. Akad. Nauk SSSR Ser. Mat. **35** (1971), 324–366.

[EW11] M. Einsiedler, T. Ward, *Ergodic theory with a view towards number theory.* Graduate Texts in Mathematics, **259**. Springer-Verlag London, Ltd., London, 2011. xviii+481 pp.

[Fu00] A. Furman, *Random walks on groups and random transformations.* Handbook of dynamical systems, Vol. 1A, 931–1014, North-Holland, Amsterdam, 2002.

[Go71] T.N.T. Goodman, *Relating topological entropy and measure entropy.* Bull. London Math. Soc. **3** (1971), 176–180.

[HR79] E. Hewitt, K.A. Ross, *Abstract harmonic analysis. Vol.* I. Structure of topological groups, integration theory, group representations. Second edition. Grundlehren der Mathematischen Wissenschaften, **115**. Springer-Verlag, Berlin-New York, 1979. ix+519 pp.

[HM77] R.E. Howe, C.C. Moore, *Asymptotic properties of unitary representations.* J. Functional Analysis **32** (1979), 72–96.

[KL16] D. Kerr, H. Li, *Ergodic theory. Independence and dichotomies.* Springer Monogr. Math. Springer, Cham, 2016. xxxiv+431 pp.

[Mi76] M. Misiurewicz, *A short proof of the variational principle for a $\mathbb{Z}_+^N$ action on a compact space.* International Conference on Dynamical Systems in Mathematical Physics (Rennes, 1975), pp. 147–157. Astérisque, No. 40. Société Mathématique de France, Paris, 1976

[Mo65] C.C. Moore, *Ergodicity of flows on homogeneous spaces.* Amer. J. Math. **88** (1966), 154–178.

[Ru87] W. Rudin, *Real and complex analysis.* Third edition. McGraw-Hill Book Co., New York, 1987. xiv+416 pp.

[St73] R.A. Struble, *Metrics in locally compact groups.* Compositio Math. **28** (1974), 217–222.

[Zi84] R.J. Zimmer, *Ergodic theory and semisimple groups.* Monographs in Mathematics, **81**. Birkhäuser Verlag, Basel, 1984. x+209 pp.