A BRIEF INTRODUCTION TO ERGODIC THEORY

ALEX FURMAN

Abstract. These are expanded notes from four introductory lectures on Ergodic Theory, given at the Minerva summer school *Flows on homogeneous spaces* at the Technion, Haifa, Israel, in September 2012.

1. Dynamics on a compact metrizable space

Given a compact metrizable space X, denote by $C(X)$ the space of continuous functions $f : X \to \mathbb{C}$ with the uniform norm

$$\|f\|_u = \max_{x\in X} |f(x)|.$$

This is a *separable* Banach space (see Exs 1.2.(a)). Denote by $\mathrm{Prob}(X)$ the space of all regular probability measures on the Borel σ-algebra of X. By Riesz representation theorem, the dual $C(X)^*$ is the space $\mathrm{Meas}(X)$ of all finite signed regular Borel measures on X with the total variation norm, and $\mathrm{Prob}(X) \subset \mathrm{Meas}(X)$ is the subset of $\Lambda \in C(X)^*$ that are *positive* ($\Lambda(f) \geq 0$ whenever $f \geq 0$) and *normalized* ($\Lambda(\mathbf{1}) = 1$). Thus $\mathrm{Prob}(X)$ is a closed convex subset of the unit ball of $C(X)^*$, it is compact and metrizable with respect to the *weak-* topology* defined by

$$\mu_n \overset{weak*}{\longrightarrow} \mu \qquad \text{iff} \qquad \int_X f\, d\mu_n \longrightarrow \int_X f\, d\mu \quad (\forall f \in C(X)).$$

Definition 1.1. Let X be a compact metrizable space and $\mu \in \mathrm{Prob}(X)$.
A sequence $\{x_n\}_{n=0}^{\infty}$ is <u>μ-equidistributed</u> if $\frac{1}{n}(\delta_{x_0}+\delta_{x_1}+\cdots+\delta_{x_{n-1}})$ weak-* converge to μ, that is if

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(x_n) \;=\; \int_X f\, d\mu \qquad (f \in C(X)).$$

Let $T : X \to X$ be a continuous map and $\mu \in \mathrm{Prob}(X)$. We say that a point $x \in X$ is <u>μ-generic</u> if the sequence $\{T^n x\}_{n=0}^{\infty}$ is μ-equidistributed.

Exercise 1.2. Prove that

(a) If X is compact metrizable then $C(X)$ is separable.
(b) In the definitions of weak-* convergence and in the definition of μ-equidistribution, one can reduce the verification of the convergence "for all $f \in C(X)$" to "for all f from a subset with a dense linear span in $C(X)$".
(c) If $x \in X$ is μ-generic, then μ is a T-invariant measure.
(d) If $A \subset X$ is Borel set, and $\mu(\partial A) = 0$ then for any μ-generic point $x \in X$

$$\frac{\#\{0 \leq n < N \mid T^n x \in A\}}{N} \;\to\; \mu(A).$$

1.1. **Irrational rotation of the circle.**

We consider the example of the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (or the one-dimensional torus).

Lemma 1.3 (Weyl). *A sequence $\{x_n\}_{n=0}^{\infty}$ of points in $\mathbb{T}$ is equidistributed with respect to the Lebesgue measure m on $\mathbb{T}$ if and only if for every $k \in \mathbb{Z} - \{0\}$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i k x_n} = 0.$$

Proof. We only need to prove the "if" direction, as the "only if" is obvious. By Stone-Weierstrass theorem, trigonometric polynomials (same as that is finite linear combinations of $e_k(x) = e^{2\pi i k x}$, $k \in \mathbb{Z}$), are dense in $C(\mathbb{T})$. So the proof is completed by Exercise 1.2.(b). □

Next consider the transformation on the circle $T(x) = x + \alpha$, where $\alpha \in \mathbb{T}$ is irrational, that is $k\alpha \neq 0$ for all non-zero integers k.

Theorem 1.4. *Let $T(x) = x + \alpha$ be an irrational rotation on $X = \mathbb{R}/\mathbb{Z}$. Then every $x \in \mathbb{T}$ is equidistributed for the Lebesgue measure m on $\mathbb{T}$.*

Proof. By Lemma 1.3 it suffices to check that

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i k(x+n\alpha)} \longrightarrow 0 \qquad (k \in \mathbb{Z} \setminus \{0\})$$

Denoting $w = e^{2\pi i k x}$ and $z = e^{2\pi i k \alpha}$ the LHS above is just

$$\frac{1}{N} \sum_{n=0}^{N-1} w z^n = \frac{w}{N} \cdot \frac{1 - z^N}{1 - z} \ \rightarrow \ 0$$

because $|z| = 1$ and $z \neq 1$ due to the irrationality of α. □

1.2. **More on invariant measures.** Recall (Exercise 1.2.(c)) that that one can talk about μ-generic points only of T-invariant measures. The set of all T-invariant measures

$$\mathrm{P}_{\mathrm{inv}}(X) = \{\mu \in \mathrm{Prob}(X) \mid T_*\mu = \mu\}$$

is a closed convex subset. Hence it is compact in the weak-* topology. Crucially, this set is never empty as the following classical result shows:

Theorem 1.5 (Krylov-Bogoloubov, Markov-Kakutani?).
The set $\mathrm{P}_{\mathrm{inv}}(X)$ is non-empty for any continuous map $T : X \to X$ of a compact metrizable space X. In fact, for any sequence $x_n \in X$ and $N_n \to \infty$ every weak- limit point of the sequence*

$$\mu_n = \frac{1}{N_n} \sum_{n=0}^{N_n} \delta_{T^n x_n}$$

of atomic measures is in $\mathrm{P}_{\mathrm{inv}}(X)$.

Proof. Suppose $\mu_n \to \mu$ in weak-* topology. Then for every $f \in C(X)$

$$\begin{aligned}\int f\,d\mu - \int f\,dT_*\mu &= \lim_{n\to\infty} \int (f - f\circ T)\,d\mu_n \\ &= \lim_{n\to\infty} \frac{1}{N_n} \sum_{n=0}^{N_n-1} (f(T^n x_n) - f(T^{n+1}x_n)) \\ &= \lim_{n\to\infty} \frac{1}{N_n}(f(x_n) - f(T^{N_n}x_n)) = 0.\end{aligned}$$

□

Theorem 1.6 (Uniquely ergodicity).
Let $T : X \to X$ be a continuous map of a compact metrizable space. TFAE:

(a) *There is $\mu \in \mathrm{Prob}(X)$ so that every $x \in X$ is μ-generic.*
(b) *There is only one T-invariant measure:* $\mathrm{P_{inv}}(X) = \{\mu\}$.
(c) *For every $f \in C(X)$ the averages*

$$A_n f(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$$

converge uniformly to a constant, which is $\int f\,d\mu$.

Such systems (X, T) are called uniquely ergodic. To clarify: such μ is unique, it is T-invariant and ergodic (see Theorem 1.6 and Definition 1.8 below), and the setting can also be characterized by saying that there is only one T-*ergodic* probability measure on X, useing Proposition 1.9 and Krein-Milman's theorem below.

Proof. (a) $\Longrightarrow$ (b). Take any $\nu \in \mathrm{P_{inv}}(X)$ and arbitrary $f \in C(X)$. The sequence $A_n f(x)$ of averages is uniformly bounded by $\|f\|_u$, and for every $x \in X$ converges to the constant $\int f\,d\mu$. Using T-invariance of ν and Lebesgue's dominated convergence theorem, gives:

$$\int_X f\,d\nu = \int_X A_n f(x)\,d\nu(x) \quad\longrightarrow\quad \int_X (\int_X f\,d\mu)\,d\nu = \int_X f\,d\mu.$$

(b) $\Longrightarrow$ (c). Consider the function $f_0 = f - \int f\,d\mu$ and the linear subspace

$$B = \{g - g\circ T \mid g \in C(X)\}.$$

Let $\Lambda \in C(X)^*$ be an arbitrary functional vanishing on B. By Riesz' representation theorem, Λ is given by integration of a signed measure, which has a unique representation as

$$\lambda = (a_1\mu_1 - a_2\mu_2) + i(b_1\nu_1 - b_2\nu_2)$$

with $a_1, a_2, b_1, b_2 \geq 0$, $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathrm{Prob}(X)$ and $\mu_1 \perp \mu_2$, $\nu_1 \perp \nu_2$. As Λ vanishes on B, we have $\int g\,d\lambda = \int g\circ T\,d\lambda$ for all $g \in C(X)$, meaning $\lambda = T_*\lambda$ and

$$(a_1\mu_1 - a_2\mu_2) + i(b_1\nu_1 - b_2\nu_2) = (a_1T_*\mu_1 - a_2T_*\mu_2) + i(b_1T_*\nu_1 - b_2T_*\nu_2).$$

Hence $\mu_1, \mu_2, \nu_1, \nu_2$ are T-invariant, and therefore equal to μ. Thus $\lambda = c\mu$ for some $c \in \mathbb{C}$ and

$$\Lambda(f_0) = c\cdot\int f_0\,d\mu = 0.$$

We just showed that every functional vanishing on B, vanishes on f_0, hence $f_0 \in \overline{B}$ by Hahn-Banach theorem. Given any $\epsilon > 0$ there is $g \in C(X)$ so that $\|f_0 - h\|_u < \epsilon$ for $h = (g - g \circ T)$. Hence for $n > 2\|g\|_u/\epsilon$ one has

$$\|A_n f - \int f\, d\mu\|_u = \|A_n(f_0)\|_u \leq \|A_n h\|_u + \epsilon \leq \frac{2\|g\|_u}{n} + \epsilon < 2\epsilon.$$

(c) $\Longrightarrow$ (a). Note that if the averages $A_n f$ converges to a constant $a(f)$, then for any T-invariant probability measure ν one has

$$\int_X f\, d\nu = \int_X A_n f\, d\nu \quad \longrightarrow \quad \int_X a(f)\, d\nu = a(f).$$

It follows that $\mathrm{P}_{\mathrm{inv}}(X)$ is a singleton $\{\mu\}$, and $a(f) = \int f\, d\mu$. □

Exercise 1.7. Prove the implication (b) $\Longrightarrow$ (c) using Theorem 1.5.

Definition 1.8. Let μ be an invariant probability measure for a transformation $T : X \to X$. We say that μ is $\underline{T\text{-ergodic}}$ if for every Borel set $E \subset X$ one has

$$\mu(E \triangle T^{-1}E) = 0 \qquad \Longrightarrow \qquad \mu(E) = 0 \quad \text{or} \quad \mu(E) = 1.$$

Recall that if $C \subset V$ is a convex set in some (real) vector space V, a point $c \in C$ is called *extremal* if it is not an interior point of a segment contained in C, that is if

$$c = tc_1 + (1-t)c_2, \quad \text{with} \quad 0 < t < 1,\ c_1, c_2 \in C \qquad \Longrightarrow \qquad c_1 = c_2 = c.$$

The set of extremal points of C is denoted $\mathrm{ext}(C)$. By Krein-Milman theorem, any convex compact subset C in a locally convex topological vector space V, is the closure of the convex hull of the extremal points

$$C = \overline{\mathrm{conv}}(\mathrm{ext}(C)).$$

In particular, any non-empty convex compact set has extremal points.

Proposition 1.9.
Let $T : X \to X$ be a continuous map of a compact metrizable space X.
Then the set $\mathrm{ext}(\mathrm{P}_{\mathrm{inv}}(X))$ is precisely the set $\mathrm{P}_{\mathrm{erg}}(X)$ of T-ergodic measures.

Proof. If $\mu \in \mathrm{P}_{\mathrm{inv}}(X) \setminus \mathrm{P}_{\mathrm{erg}}(X)$, then there is $E \in \mathcal{B}$ with $\mu(E \triangle T^{-1}E) = 0$ and $0 < \mu(E) < 1$. The probability measures

$$\mu_E = \mu(E)^{-1} \cdot \mu|_E, \qquad \mu_{X\setminus E} = (1 - \mu(E))^{-1} \cdot \mu|_{X \setminus E}$$

are T-invariant. Since $\mu = \mu(E) \cdot \mu_E + (1 - \mu(E)) \cdot \mu_{X\setminus E}$, it follows that μ is not extremal: $\mu \notin \mathrm{ext}(\mathrm{P}_{\mathrm{inv}}(X))$.

Conversely, if μ is not extremal in $\mathrm{P}_{\mathrm{inv}}(X)$, write $\mu = t\mu_1 + (1-t)\mu_2$ with $0 < t < 1$ and $\mu_1 \neq \mu_2 \in \mathrm{P}_{\mathrm{inv}}(X)$. Then $\mu_1(A) \leq 1/t \cdot \mu(A)$ for every Borel set A. Hence μ_1 is absolutely continuous with respect to μ, $\mu_1 \ll \mu$. Let

$$\phi = \frac{d\mu_1}{d\mu}$$

be the Radon-Nikodym derivative, which is a positive unit vector in $L^1(X, \mu)$, uniquely determined by the relation

$$\mu_1(A) = \int_A \phi\, d\mu \qquad\qquad (A \in \mathcal{B}_X).$$

Since both μ_1 and μ are T-invariant, it follows that μ-a.e. $\phi(Tx) = \phi(x)$. Thus for all $a > 0$ the set $E_a = \{x \in X \mid \phi(x) < a\}$ satisfies $\mu(E_a \triangle T^{-1}E_a) = 0$. As a function of $a \in [0, \infty]$ the measure $\mu(E_a)$ is monotonically non-decreasing with $\mu(E_0) = 0$ and $\mu(E_\infty) = 1$. Let

$$a_0 = \sup\{a \geq 0 \mid \mu(E_a) = 0\}, \qquad a_1 = \inf\{a \mid \mu(E_a) = 1\}.$$

The situation $a_0 = a_1$ leads to a contradiction, because it forces ϕ to be μ-a.e. equal to this constant. Consequently $a_0 = a_1 = 1$ and $\mu_1 = \mu = \mu_2$, contrary to the assumption.

Hence $a_0 < a_1$. Let $E = E_a$ for some fixed intermediate value $a_0 < a < a_1$. Then $0 < \mu(E) < 1$ and $\mu(E \triangle T^{-1}E) = 0$, establishing that μ is not T-ergodic. □

Remark 1.10. It follows from Birkhoff's *individual ergodic theorem* (see Theorem 2.3) that if $\mu \in \mathrm{P}_{\mathrm{erg}}(X)$ is T-ergodic, then μ-a.e. point is μ-generic. This implies that distinct T-ergodic measures are mutually singular.

1.3. **An interesting example of a uniquely ergodic system.**

Suppose $T : X \to X$ be a continuous map of a compact metrizable space X, and $\phi : X \to \mathbb{T}$ is a continuous function. Consider the system $(\hat{X}, \hat{T})$ where

$$\hat{X} = X \times \mathbb{T}, \qquad \hat{T}(x, s) = (Tx, s + \phi(x)). \tag{1}$$

Let m denote the normalized Lebesgue measure on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Given a T-invariant measure $\mu \in \mathrm{P}_{\mathrm{inv}}(X)$ the product measure

$$\hat{\mu} = \mu \times m$$

is $\hat{T}$-invariant.

Theorem 1.11 (Furstenberg [2])**.**
Suppose μ is the unique T-invariant measure on X, and $\hat{\mu} = \mu \times m$ is $\hat{T}$-ergodic. Then $\hat{\mu}$ is the unique $\hat{T}$-invariant measure.

Proof. Let ν_0 be some $\hat{T}$-invariant probability measure. For $t \in \mathbb{T}$ consider the translation ν_t of ν_0 by the homeomorphism $\tau_t : (x, s) \mapsto (x, s + t)$ of $\hat{X}$ that commutes with $\hat{T}$. In other words

$$\int_{\hat{X}} f \, d\nu_t = \int_{\hat{X}} f(x, s + t) \, d\nu_0(x, s) \qquad (f \in C(\hat{X})).$$

Then $\nu_t \in \mathrm{P}_{\mathrm{inv}}(\hat{X})$. Consider the "average" $\bar{\nu} = \int \nu_t \, dt$, defined by

$$\begin{aligned}
\int_{\hat{X}} f \, d\bar{\nu} &= \int_{\mathbb{T}} \int_{\hat{X}} f \, d\nu_t \, dt = \int_{\mathbb{T}} \int_{\hat{X}} f(x, s + t) \, d\nu_0(x, s) \, dt \\
&= \int_{\hat{X}} \int_{\mathbb{T}} f(x, s + t) \, dt \, d\nu_0(x, s) = \int_{\hat{X}} \left(\int_{\mathbb{T}} f(x, t) \, dt \right) d\nu_0(x, s) \\
&= \int_{\hat{X}} f \, d\eta \times m
\end{aligned}$$

where $\eta \in \mathrm{Prob}(X)$ is the push-forward of ν_0 under the projection $\hat{X} \to X$. Since such η is necessarily T-invariant, it follows that $\eta = \mu$ and

$$\int_0^1 \nu_t \, dt = \bar{\nu} = \mu \times m.$$

Since $\mu \times m$ is ergodic, it is an extremal point of $\mathrm{P}_{\mathrm{inv}}(\hat{X})$. Hence $\nu_t = \mu \times m$ for (almost) all t. Thus $\nu_0 = \mu \times m$, by applying τ_{-t} to some $\nu_t = \mu \times m$. As ν_0 was an arbitrary $\hat{T}$-invariant probability measure, unique ergodicity of $(\hat{X}, \hat{T}, \hat{\mu} = \mu \times m)$ is proven. □

Corollary 1.12.
Let α be irrational. The transformation

$$T : (x, y) \mapsto (x + \alpha, y + x)$$

of the 2-torus $\mathbb{T}^2$ is uniquely ergodic, the invariant measure is Lebesgue.

Proof. Applying Theorem 1.11 with irrational rotation $x \mapsto x + \alpha$ on $\mathbb{T}$ as the base, and $\phi(x) = x$, we only need to verify that T is ergodic on $\mathbb{T}^2$ with respect to the Lebesgue measure.

Let $E \subset \mathbb{T}^2$ be some Borel set with $m^2(E \triangle T^{-1}E) = 0$. Consider the Fourier series decomposition of the characteristic function $1_E \in L^2(\mathbb{T}^2, m^2)$

$$1_E = \sum_{(k,\ell)\in\mathbb{Z}^2} c_{k,\ell} \cdot u_{k,\ell} \qquad \text{where} \qquad u_{k,\ell}(x, y) = e^{2\pi i(kx+\ell y)}.$$

We have

$$\sum_{(k,\ell)\in\mathbb{Z}^2} c_{k,\ell} \cdot u_{k,\ell} = \sum_{(q,r)\in\mathbb{Z}^2} c_{q,r} \cdot u_{q,r} \circ T = \sum_{(q,r)\in\mathbb{Z}^2} c_{q,r} e^{2\pi i q\alpha} \cdot u_{q+r,r}.$$

Comparing the coefficients we get the identities

$$c_{k,\ell} = c_{k-\ell,\ell} \cdot e^{2\pi i(k-\ell)\alpha}.$$

In particular, $|c_{k,\ell}| = |c_{k-\ell,\ell}| = |c_{k-2\ell,\ell}| = \dots$ Since $c_{k,\ell}$ should be square summable on $\mathbb{Z}^2$, it follows that $c_{k,\ell} = 0$ whenever $\ell \neq 0$. For $\ell = 0$ we have

$$c_{k,0} = e^{2\pi i k\alpha} \cdot c_{k,0}.$$

As $e^{2\pi i k\alpha} \neq 1$ for $k \neq 0$, we get vanishing $c_{k,0} = 0$ for all $k \neq 0$, and conclude that $c_{k,\ell} = 0$ for all $(k, \ell) \neq (0, 0)$. Thus 1_E is an essentially constant function on $\mathbb{T}^2$, which means that E is either null or co-null. This proves ergodicity of the Lebesgue measure, and unique ergodicity of the transformation on the torus. □

Corollary 1.13.
For an irrational α the sequence $\{n^2\alpha\}_{n=0}^{\infty}$ is equidistributed on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

Proof. Given an irrational α take $\beta = 2\alpha$ (still irrational) and consider the iterates of the uniquely ergodic (Corollary 1.12 !) transformation $T(x, y) = (x + \beta, y + x)$ of the torus:

$$T^n(x, y) = (x + n\beta,\ y + nx + \frac{n^2 - n}{2}\beta) = (x + 2n\alpha,\ y + n(x - \alpha) + n^2\alpha).$$

Given $f \in C(\mathbb{T})$ consider $F(x, y) = f(y)$ and apply the equidistribution of $\{T^n(\alpha, 0)\}$ on $\mathbb{T}^2$ to deduce

$$\frac{1}{N}\sum_{n=0}^{N-1} f(n^2\alpha) = \frac{1}{N}\sum_{n=0}^{N-1} F(T^n(\alpha, 0)) \quad \longrightarrow \quad \int_{\mathbb{T}^2} F(x, y)\, dx\, dy = \int_0^1 f(y)\, dy.$$

□

Similar equidistribution on the circle can be shown for such sequences as $\{n^3\alpha\}$ or $\{n^3+\sqrt{5}n^2-\pi n\}$ by showing unique ergodicity of certain transformations of $\mathbb{T}^3$. This approach allowed Furstenberg [2] to give a dynamical proof for the following theorem

Theorem 1.14 (Weyl)**.**
Let $p(x)=a_dx^d+\cdots+a_1x+a_0$ be a polynomial with at least one irrational non-constant coefficient. Then $\{p(n)\}_{n=0}^{\infty}$ is equidistributed on $\mathbb{T}$.

1.4. **Additional remarks and exercises.**
The central notion of topological dynamics is *minimality*.

Definition 1.15. Let $T:X\to X$ be a continuous map of compact metrizable space. The system (X,T) is minimal if X contains no proper closed T-invariant subsets. A system (X,T) is strictly ergodic if it is both uniquely ergodic and minimal.

The two examples that we discussed so far – irrational rotation $x\mapsto\alpha$ on $\mathbb{T}$, and the skew-product map $(x,y)\mapsto(x+\alpha,y+x)$ on $\mathbb{T}^2$ – are minimal systems (hence strictly ergodic), because the unique invariant measure has full support (see Exercise 1.16.(a)). Yet there exist minimal systems that are not uniquely ergodic. In the above mentioned beautiful paper [2] Furstenberg studied skew-products (1) and characterized (unique) ergodicity and minimality for such systems in terms of $\phi:X\to\mathbb{T}$ and (X,T). He then constructed an example an irrational (Liouville) α and a C^∞-smooth $\phi:\mathbb{T}\to\mathbb{T}$ so that

$$T:(x,y)\mapsto(x+\alpha,y+\phi(x))$$

is minimal (C^∞-diffeomorphism) that is not uniquely ergodic; in fact not ergodic for the Lebesgue measure.

Exercise 1.16. Prove that:

(a) A uniquely ergodic (X,μ,T) system is minimal iff $\operatorname{supp}(\mu)=X$.
(b) Check that the transformation $T:x\mapsto x+1$ of the one point compactification $X=\mathbb{Z}\cup\{\infty\}$ of $\mathbb{Z}$ is uniquely ergodic but not minimal.
(c) Prove that the following are equivalent for a dynamical system (X,T):
   (1) (X,T) is minimal,
   (2) Every orbit $\{(T^nx)\}$ is dense in X,
   (3) For every non-empty open $V\subset X$ there finite cover

$$X=V\cup T^{-1}V\cup\cdots\cup T^{-n}V.$$

Let A be a finite set. The infinite product $A^{\mathbb{Z}}$ is a Cantor set (a perfect totally disconnected compact metrizable space). The shift

$$T:(\dots,x_{-1},x_0,x_1,\dots)\mapsto(\dots,x_0,x_1,x_2,\dots)$$

is a homeomorphism of $A^{\mathbb{Z}}$. Given a sequence $x\in A^{\mathbb{Z}}$ the closure of its orbit

$$X=\overline{\{T^nx\in A^{\mathbb{Z}}\mid n\in\mathbb{Z}\}}$$

is a closed T-invariant subset. Consider the resulting dynamical system (X,T).

Exercise 1.17. Let (X,T) be constructed from some point $x\in A^{\mathbb{Z}}$ as above.

(a) Prove that X consists of those $y \in A^{\mathbb{Z}}$ for which every *word* in y appears in x; more precisely, for every $n \in \mathbb{Z}$ and $\ell \in \mathbb{N}$ there is $k \in \mathbb{Z}$ so that

$$(y_{n+1}, y_{n+2}, \ldots, y_{n+\ell}) = (x_{k+1}, x_{k+2}, \ldots, x_{k+\ell}).$$

(b) Prove that (X, T) is minimal iff every finite word $w = (x_k, \ldots, x_m)$ in x reappears in x with *bounded gaps*, namely there exists $n = n(w)$ so that for any $i \in \mathbb{Z}$ there is $j \in \{i, \ldots, i + n(w)\}$ with $(x_j, x_{j+1}, \ldots, x_{j+m-k}) = w$.

(c) Prove that (X, T) is uniquely ergodic iff every finite word $w = (x_k, \ldots, x_m)$ in x reappears in x with fixed *frequency*, namely the following limits exist

$$a(w) = \lim_{N\to\infty} \frac{1}{2N} \#\{-N \le j < N \mid (x_j, x_{j+1}, \ldots, x_{j+m-k}) = w\}.$$

(d) Construct minimal non-uniquely ergodic system as above.

Theorem 1.6 shows that in uniquely ergodic systems averages $A_n f$ of any continuous functions uniformly converge to a constant. A closer examination of the proof of that theorem shows that such a conclusion can be obtained for a specific function (or functions) on systems that are not necessarily uniquely ergodic.

Exercise 1.18. Let $T : X \to X$ be a continuous map of a compact metrizable space X. For real valued continuous function $f : X \to \mathbb{R}$ define

$$I_*(f) = \inf_{\mu \in \mathrm{P}_{\mathrm{inv}}(X)} \int_X f\, d\mu, \qquad I^*(f) = \sup_{\mu \in \mathrm{P}_{\mathrm{inv}}(X)} \int_X f\, d\mu.$$

and $A_n f(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ as before.

(a) Show that in the above definition inf and sup can be replaced by min and max, and $\mathrm{P}_{\mathrm{inv}}(X)$ by $\mathrm{P}_{\mathrm{erg}}(X)$.

(b) Prove that for any $f \in C(X, \mathbb{R})$ the following inequalities hold for *every* $x \in X$ and *uniformly* on X:

$$I_*(f) \le \liminf_{n\to\infty} A_n f(x), \qquad \limsup_{n\to\infty} A_n f(x) \le I^*(f).$$

(c) Assuming (X, T) is minimal, for any $f \in C(X, \mathbb{R})$, $\mu \in \mathrm{P}_{\mathrm{erg}}(X)$ prove that

$$\left\{ x \in X \mid \int f\, d\mu \quad \text{is a limit point of} \quad A_n f(x) \right\}$$

is a dense G_δ-set in X. (Hint: this result relies on Birhoff's theorem 2.3).

(d) Deduce, that if (X, T) is minimal then for $f \in C(X, \mathbb{R})$ the set

$$\left\{ x \in X \mid I_*(f) = \liminf_{n\to\infty} A_n f(x), \quad \limsup_{n\to\infty} A_n f(x) = I^*(f) \right\}$$

is a dense G_δ-set in X.

Given a topological dynamical system (X, T) let $U(X, T) \subset C(X)$ be the collection of *uniformly averaged* functions $f \in C(X)$, namely those for which $A_n f$ converges uniformly to a constant $c(f)$. Let $V(X, T) = \{f \in C(X) \mid I_*(f) = I^*(f)\}$. These are closed linear subspaces of $C(X)$. By part (a) we have

$$V(X, T) \subseteq U(X, T)$$

and, if (X, T) is minimal, part (d) implies the equality

$$V(X, T) = U(X, T).$$

2. Ergodic theorems in the measurable context

Dynamics can be studied in a purely measure-theoretical context, where $(X, \mathcal{B}, \mu)$ is a *standard probability space*[1], $T : X \to X$ is a measurable map, preserving the measure μ. The latter means that

$$T_*\mu(E) = \mu(T^{-1}E) = \mu(E) \qquad (E \in \mathcal{B}).$$

Equivalently

$$\int_X f\, dT_*\mu = \int_X f(Tx)\, d\mu(x) = \int_X f(x)\, d\mu(x) \qquad (f \in L^1(X, \mu)).$$

Definition 2.1.
A measure-preserving map T of $(X, \mathcal{B}, \mu)$ is ergodic if every *essentially invariant set* (that is a set $E \in \mathcal{B}$ with $\mu(E \triangle T^{-1}E) = 0$) is *$\mu$-null* ($\mu(E) = 0$) or *$\mu$-co-null* ($\mu(X \setminus E) = 0$).

Here are some other characterizations of ergodicity

Exercise 2.2. Let T be a measure-preserving transformation of $(X, \mathcal{B}, \mu)$. Show that the following conditions are equivalent:

(a) T is ergodic.
(b) Any measurable a.e. *T-invariant function* $f : X \to \mathbb{C}$ (that is one for which μ-a.e. $f(Tx) = f(x)$), is μ-a.e. equal to a constant.
(c) The only *T-invariant vectors* f in the Hilbert space $L^2(X, \mu)$ are constants[2].

2.1. **Birkhoff's Individual Ergodic Theorem.**

The following central result is known as the *individual ergodic theorem* or *pointwise ergodic theorem*; it was proved after the *mean ergodic theorem* (Theorem 2.9).

Theorem 2.3 (Birkhoff)**.**
Let $(X, \mathcal{B}, \mu, T)$ be an ergodic probability measure preserving system and $f \in L^1(X, \mu)$. Then for μ-a.e. $x \in X$ and in L^1-norm the ergodic averages converge to the integral:

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \quad \longrightarrow \quad \int_X f\, d\mu$$

If $(X, \mathcal{B}, \mu, T)$ is a probability measure preserving that is not necessarily ergodic, the ergodic averages converge μ-a.e. and in L^1 to the conditional expectation

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \quad \longrightarrow \quad \mathbb{E}(f \mid \mathcal{B}_{\text{inv}})$$

to the sub-σ-algebra of T-invariant sets $\mathcal{B}_{\text{inv}} = \{E \in \mathcal{B} \mid \mu(E \triangle T^{-1}E) = 0\}$.

We bring here the short elegant proof by Katznelson and Weiss [5].

---

[1]Hereafter we shall assume taht $\mathcal{B}$ is complete with respect to μ.

[2]Recall that vectors in L^2 are equivalence classes of measurable functions

Proof. We think of $f : X \to \mathbb{R}$ as a fixed Borel function. For $n \geq 1$ denote the ergodic sums and ergodic averages by

$$S_n(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x), \qquad A_n(x) = \frac{1}{n} S_n(x).$$

Note the cocycle relation

$$S_{n+m}(x) = S_n(x) + S_m(T^n x).$$

Define measurable functions $f_* : X \to \mathbb{R} \cup \{-\infty\}$ and $f^* : X \to \mathbb{R} \cup \{+\infty\}$ by

$$f_*(x) = \liminf_{n\to\infty} A_n(x), \qquad f^*(x) = \limsup_{n\to\infty} A_n(x).$$

Since $S_{n+1}(x) = f(x) + S_n(Tx)$ we have

$$A_{n+1}(x) = \frac{1}{n+1} f(x) + \frac{n}{n+1} A_n(Tx),$$

and it follows that $f_*(x)$ and $f^*(x)$ are μ-a.e. T-invariant functions.

For the sake of clarity of exposition, we focus on the ergodic case. Then $f_*(x)$ and $f^*(x)$ equal μ-a.e. to some constants $-\infty \leq a_* < \infty$, $-\infty < a^* \leq +\infty$, respectively. We want to show that

$$a_* = \int f \, d\mu = a^*$$

and due to symmetry, it suffices to show $a_* = \int f \, d\mu$. Fix arbitrary

$$\epsilon > 0 \qquad \text{and} \qquad a_* < a$$

(if $-\infty < a_*$ then take $a = a_* + \epsilon$, otherwise take $a = -R$ with $R \gg 1$ large). Define

$$n(x) = \min \{n \in \mathbb{N} \cup \{\infty\} \mid A_n(x) < a\}.$$

Then $n(x)$ is a Borel function with $n(x) < \infty$ for μ-a.e. $x \in X$. For $k \in \mathbb{N}$ let

$$E_k = \{x \in X \mid n(x) > k\}.$$

Then $E_1 \supset E_2 \supset \dots$ with $\mu(E_k) \to 0$ as $k \to \infty$. Choose k large enough so that

$$\int_{E_k} (|f(x) - a|) \, d\mu(x) < \epsilon \tag{2}$$

and consider n large enough so that

$$\frac{k}{n} \cdot \int_X (|f(x) - a|) \, d\mu(x) < \epsilon. \tag{3}$$

Define a measurable function $\tilde{n} : X \to \{1, 2, \dots, k\}$ by

$$\tilde{n}(x) = \begin{cases} n(x) & \text{if} \quad x \in X \setminus E_k \\ 1 & \text{if} \quad x \in E_k. \end{cases}$$

Given $x \in X$ define a sequence of points $x_n \in X$ by $x_0 = x$ and inductively $x_{i+1} = \tilde{n}(x_n)$. Given $n \gg k$ as in (3), let $r \in \mathbb{N}$ be such that

$$\tilde{n}(x_0) + \tilde{n}(x_1) + \cdots + \tilde{n}(x_{r-1}) \leq n < \tilde{n}(x_0) + \tilde{n}(x_1) + \cdots + \tilde{n}(x_r)$$

and set $m = n - \tilde{n}(x_0) + \cdots + \tilde{n}(x_{r-1})$. So $0 \leq m < \tilde{n}(x_r) \leq k$. We are interested in a good upper estimate for (the integral of) the average

$$A_n(x) = \frac{1}{n} S_n(x) = \frac{1}{n}(S_{\tilde{n}(x_0)}(x_0) + \cdots + S_{\tilde{n}(x_{r-1})}(x_{r-1}) + S_m(x_r)).$$

Let $I = \{0 \le i < r \mid x_n \notin E_k\}$ and $J = \{0, \dots, r-1\} \setminus I$. Then the terms $S_{\tilde{n}(x_n)}(x_n)$, $0 \le i < r$, can be estimated as follows: for $i \in I$

$$S_{\tilde{n}(x_n)}(x_n) = S_{n(x_n)}(x_n) < n(x_n) \cdot a = \tilde{n}(x_n) \cdot a,$$

while for $i \in J$

$$S_{\tilde{n}(x_n)}(x_n) = f(x_n) \le |f(x_n) - a| + a = (1_{E_k} \cdot |f - a|)(x_n) + \tilde{n}(x_n) \cdot a.$$

For the last summand we have

$$S_m(x_r) = \sum_{i=0}^{m-1} f(T^\ell x_r) \le \sum_{\ell=n-k}^{n-1} |f(T^\ell x) - a| + ma.$$

Therefore

$$\begin{aligned} A_n(x) &= \frac{1}{n}\left(S_{\tilde{n}(x_0)}(x_0) + \cdots + S_{\tilde{n}(x_{r-1})}(x_{r-1}) + S_m(x_r)\right) \\ &\le \frac{1}{n}\left(\sum_{i=0}^{r-1} \tilde{n}(x_n) \cdot a + ma + \sum_{j \in J}(1_{E_k} \cdot |f - a|)(x_j) + \sum_{\ell=n-k}^{n-1} |f(T^\ell x) - a|\right) \\ &\le a + \frac{1}{n} \cdot \sum_{j=0}^{n-1}(1_{E_k} \cdot |f - a|)(T^j x) + \frac{1}{n} \cdot \sum_{\ell=n-k}^{n-1} |f(T^\ell x) - a|. \end{aligned}$$

Integrating this estimates with respect to x we obtain, using (2) and (3),

$$\int_X f \, d\mu = \int_X A_n(x) \, d\mu \le a + \int_{E_k} |f - a| + \frac{k}{n} \cdot \int_X |f - a| \, d\mu < a + 2\epsilon.$$

As $a > a_*$ and $\epsilon > 0$ were arbitrary, it follows that $a_* \ge \int f \, d\mu$. Similarly (by replacing $f(x)$ by $-f(x)$) we deduce that $a^* \le \int f \, d\mu$, and $a_* = a^* = \int f \, d\mu$ follows, using the obvious $a_* \le a^*$. This proves the μ-a.e. convergence

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

The above estimate also gives $\int \max(A_n(x) - a, 0) \, d\mu < 2\epsilon$, that gives the L^1-convergence when combined with the similar estimate for $-f(x)$. □

Exercise 2.4. Let $T : X \to X$ be a continuous map of a compact metrizable space.

(a) Let $\mu \in \mathrm{P}_{\mathrm{erg}}(X)$ be a T-ergodic measure. Show that μ-a.e. $x \in X$ is μ-generic.

(b) Deduce that any two distinct T-ergodic measures $\mu \ne \nu \in \mathrm{P}_{\mathrm{erg}}(X)$ are mutually singular: $\mu \perp \nu$.

Definition 2.5. A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is mixing if

$$\forall A, B \in \mathcal{B}: \qquad \lim_{n \to \infty} \mu(A \cap T^{-n} B) = \mu(A) \cdot \mu(B).$$

Exercise 2.6. (a) Prove that any mixing system is ergodic.

(b) Show that an irrational rotation is not mixing (hence the converse to (a) is false).

(c) Say that a collection $\mathcal{C} \subset \mathcal{B}$ is *dense* if for every $\epsilon > 0$ and $A \in \mathcal{B}$ there is $C \in \mathcal{C}$ with $\mu(A \triangle C) < \epsilon$.
Prove that if $\mu(C \cap T^{-n} D) \to \mu(C)\mu(D)$ for all sets C, D from a dense collection $\mathcal{C}$, then the transformation is mixing.

Exercise 2.7 (Bernoulli shift)**.** Let (Z, ζ) be a probability space, and $(X, \mu) = (Z, \zeta)^{\mathbb{N}}$ – the infinite product of probability spaces. Prove that the shift T on X

$$T(z_1, z_2, z_3, \dots) = (z_2, z_3, z_4, \dots)$$

is a measure preserving transformation, which is mixing (hence also ergodic).
Hint: use Exercise 2.6.(c) and the collection $\mathcal{C}$ of so called *cylinder sets*:

$$C = E_1 \times E_2 \times E_k \times Z \times Z \cdots$$

where $k \in \mathbb{N}$ and $E_1, \dots, E_k$ are measurable subsets of Z.

Remark 2.8. Given a probability space (Z, ζ) and some function $\phi \in L^1(Z, \zeta)$, one can apply Birkhoff's ergodic theorem to the Bernoulli shift on $(X, \mu) = (Z, \zeta)^{\mathbb{N}}$ and the function $f : X \to \mathbb{R}$ given by $f(z_1, z_2, \dots) = \phi(z_1)$ to deduce the *Strong Law of Large Numbers* in Probability Theory.

2.2. **Von Neumann's Mean Ergodic Theorem.**

Birkhoff's individual ergodic theorem was predated by von Neumann's *mean ergodic theorem*, asserting that for an ergodic $(X, \mathcal{B}, \mu, T)$ for every $f \in L^2(X, \mu)$ there is an L^2-convergence:

$$\| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) - \int_X f \, d\mu \|_2 \quad \longrightarrow \quad 0.$$

If the system $(X, \mathcal{B}, \mu, T)$ is not ergodic, the constant $\int f \, d\mu$ should be replaced by the conditional expectation $\mathbb{E}(f \mid \mathcal{B}_{\mathrm{inv}})$, which is the orthogonal projection of f to the closed subspace

$$\mathcal{H}_{\mathrm{inv}} = L^2(X, \mathcal{B}_{\mathrm{inv}}, \mu|_{\mathcal{B}_{\mathrm{inv}}}) \subset \mathcal{H} = L^2(X, \mathcal{B}, \mu).$$

In fact, the beauty of von Neumann's proof is that it is a purely Hilbert space argument, that applies to general *isometry* U of a Hilbert space, and not only to ones coming from a transformations of (an) underlying measure space. A linear map $U : \mathcal{H} \to \mathcal{H}$ is an isometry if $\langle Ux, Uy \rangle = \langle x, y \rangle$ for all $x, y \in \mathcal{H}$. Note that in an infinite dimensional Hilbert space, not every isometry is invertible.

Theorem 2.9 (von Neumann)**.**
Let U be a linear isometry of a Hilbert space $\mathcal{H}$. Then

$$\lim_{n \to \infty} \| \frac{1}{n} \sum_{k=0}^{n-1} U^k x \; - \; P_{\mathrm{inv}} x \| = 0 \qquad (x \in \mathcal{H}),$$

where P_{inv} is the orthogonal projection to the subspace $\mathcal{H}_{\mathrm{inv}} = \{ y \in \mathcal{H} \mid Uy = y \}$.

Proof. We claim that $\mathcal{H}_{\mathrm{inv}}$ is the orthogonal complement $\mathcal{H}_0^{\perp}$ of the linear subspace $\mathcal{H}_0 = \{ Uz - z \mid z \in \mathcal{H} \}$ of $\mathcal{H}$. Indeed $y \in \mathcal{H}_0^{\perp}$ iff for every $z \in \mathcal{H}$

$$0 = \langle Uz - z, y \rangle = \langle Uz, y \rangle - \langle z, y \rangle = \langle z, U^* y \rangle - \langle z, y \rangle = \langle z, U^* y - y \rangle$$

that is equivalent to $U^*y = y$ and to

$$\|y\|^2 = \langle U^*y, y\rangle = \langle y, Uy\rangle \le \|y\| \cdot \|Uy\| = \|y\|^2,$$

that is equivalent, by Cauchy-Schwartz, to $Uy = y$.

Hence $\mathcal{H}_0^\perp = \mathcal{H}_{\mathrm{inv}}$, and therefore $\mathcal{H} = \mathcal{H}_{\mathrm{inv}} \oplus \overline{\mathcal{H}_0}$. Given $x \in \mathcal{H}$ and $\epsilon > 0$, there is $z \in \mathcal{H}$ so that

$$\|x - P_{\mathrm{inv}}x - (Uz - z)\| < \epsilon.$$

Applying any power U^k of the isometry U, we have

$$\|U^k x - P_{\mathrm{inv}}x - (U^{k+1}z - U^k z)\| < \epsilon$$

and taking averages we obtain

$$\|\frac{1}{n}\sum_{k=0}^{n-1} U^k x - P_{\mathrm{inv}}x - \frac{1}{n}(U^n z - z)\| < \epsilon$$

and

$$\|\frac{1}{n}\sum_{k=0}^{n-1} U^k x - P_{\mathrm{inv}}x\| < \frac{2}{n}\|z\| + \epsilon < 2\epsilon$$

as soon as $n > 2\|z\|/\epsilon$. □

2.3. **Kingman's Subadditive Ergodic Theorem.**

Ergodic theorems are concerned with ergodic averages of a function along the orbit. There are several important generalizations of these results concerning compositions of non-commuting transformations along an orbit of a measure preserving transformation, such as Oseledets' theorem (see [8], [4], [3]). The following result of Kingman, known as the *subadditive ergodic theorem*, is an essential tool in these considerations.

Definition 2.10.
Let $(X, \mathcal{B}, \mu, T)$ be a probability measure-preserving system. A subadditive cocycle is a sequence $h_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, of measurable functions, satisfying μ-a.e.

$$h_{n+m}(x) \le h_n(x) + h_m(T^n x) \qquad (n, m \in \mathbb{N}).$$

A particular case of subadditive cocycles, are *additive cocycles* – sequences of functions $a_n : X \to \mathbb{R}$, satisfying equalities:

$$a_{n+m}(x) = a_n(x) + a_m(T^n x) \qquad (n, m \in \mathbb{N}).$$

Given any function $f : X \to \mathbb{R}$, the ergodic sums

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

form an additive cocycle. Conversely, any additive cocycle $\{a_n(x)\}$ is given by the ergodic sums $a_n = S_n f$ of the function $f(x) = a_1(x)$.

Subadditive cocycles appear, for example, as

$$h_n(x) = \log \|A(T^{n-1} \cdots A(Tx)A(x)\|$$

where $A : X \to G$ is some measurable function taking values in a group (e.g. $G = \mathrm{GL}_d(\mathbb{R})$), and $\| - \| : G \to \mathbb{R}_+$ is a sub-multiplicative norm.

We formulate the ergodic case of Kingman's subadditive ergodic theorem.

Theorem 2.11 (Kingman)**.**
Let $h_n : X \to \mathbb{R}$*,* $n \in \mathbb{N}$*, be a subadditive cocycle on an ergodic system* $(X, \mathcal{B}, \mu, T)$ *with* $h_1^+(x) = \max(h_1(x), 0) \in L^1(X, \mu)$*.*
Then there is μ*-a.e. and* $L^1(X, \mu)$ *convergence*

$$\lim_{n\to\infty} \frac{1}{n} h_n(x) = L$$

where the constant $L \in [-\infty, \infty)$ *is given by* lim *and* inf *of the integrals*

$$L = \lim_{n\to\infty} \frac{1}{n} \int_X h_n \, d\mu = \inf_{n\geq 1} \frac{1}{n} \int_X h_n \, d\mu.$$

In the general measure preserving case (not necessarily ergodic), the limit is a T-invariant function $L(x)$ given by lim and inf of conditional expectations

$$\mathbb{E}(\frac{1}{n} h_n \mid \mathcal{B}_{\mathrm{inv}}).$$

We refer the reader to Katznelson-Weiss [5] for a short proof of this theorem, along the same lines as in their proof of the Birkhoff's ergodic theorem, presented above.

3. Howe-Moore theorem on vanishing of matrix coefficients

Translations on homogeneous spaces of Lie groups provide very interesting examples of dynamical systems with many applications in different areas. More specifically, consider the following examples:

Example 3.1. Let G be a connected Lie group, $\Gamma < G$ a discrete subgroup so that G/Γ has a G-invariant probability measure $m_{G/\Gamma}$, with the transformation $T_\alpha : g\Gamma \mapsto \alpha g\Gamma$ where $\alpha \in G$ is some fixed element.

Circle rotations correspond to $G = \mathbb{R}$ with $\Gamma = \mathbb{Z}$. Highly non-commutative settings, e.g. $G = \mathrm{SL}_d(\mathbb{R})$ with $\Gamma = \mathrm{SL}_d(\mathbb{Z})$, provide more interesting examples (see next section). The following very general result about unitary representations proves ergodicity (in fact, a stronger property of *mixing*) for many of such examples (Corollary 3.4).

Let G be a semi-simple Lie group with finite center and no non-trivial compact factors. For example $G = \mathrm{SL}_d(\mathbb{R})$, $\mathrm{SO}(n, 1)$, etc. A unitary representation of G is a continuous group homomorphism

$$\pi : G \longrightarrow U(\mathcal{H})$$

into the unitary group of some Hilbert space $\mathcal{H}$, where $U(\mathcal{H})$ is taken with the strong operator topology. In other words, the continuity requirement is that whenever $g_n \to g$ in G for every $x \in \mathcal{H}$ one has $\|\pi(g_n)x - \pi(g)x\| \to 0$.

Theorem 3.2 (Howe-Moore)**.**
Let $G = \prod_{i=1}^{k} G_i$ be a connected, semi-simple Lie group with finite center and no non-trivial compact factors, let $\pi : G \to U(\mathcal{H})$ be a unitary representation, and assume that each factor G_i acts with no non-trivial invariant vectors: $\mathcal{H}^{\pi(G_i)} = \{0\}$. Then for every $u, v \in \mathcal{H}$ one has $\langle \pi(g)u, v\rangle \to 0$ as $g \to \infty$ in G.

For $u, v \in \mathcal{H}$ the function $f_{u,v}(g) = \langle \pi(g)u, v\rangle$ is called matrix coefficient of π. Indeed if $\mathcal{H}$ is finite dimensional and $\{u_1, \dots, u_d\}$ is an orthonormal basis, then $f_{u_n,u_j}(g)$ is the ij-entry of the unitary matrix representing $\pi(g)$. Matrix coefficients are bounded ($|f_{u,v}(g)| \le \|u\|\cdot\|v\|$) continuous functions on G. Howe-Moore theorem states that matrix coefficients *vanish at infinity* on G:

$$f_{u,v} \in C_0(G) \qquad (u, v \in \mathcal{H})$$

provided $\mathcal{H}^{\pi(G_i)} = \{0\}$ for all factors G_i of G. This strengthens and generalizes a prior result of Moore:

Theorem 3.3 (Moore)**.**
Let $G = \prod_{i=1}^{k} G_i$ and $\pi : G \to U(\mathcal{H})$ with $\mathcal{H}^{\pi(G_i)} = \{0\}$ be as above. Then $\mathcal{H}^{\pi(H)} = \{0\}$ for every non-precompact subgroup $H < G$.

Hereafter we shall focus on *simple* Lie groups G, such as $\mathrm{SL}_d(\mathbb{R})$, $\mathrm{SO}(d, 1)$.

Corollary 3.4.
Let G be a connected simple Lie group and $\alpha \in G$ does not lie in a compact subgroup. Then the system $(X, \mu, T) = (G/\Gamma, m_{G/\Gamma}, T_\alpha)$ is ergodic, and furthermore mixing.

Before proving Theorem 3.2, we recall a few facts about connected semi-simple group G. Every such G contains a *maximal compact* subgroup K, a maximal connected Abelian diagonalizable subgroup A (called *Cartan subgroup*); such subgroups are unique up to conjugation. For the key case of $G = \mathrm{SL}_d(\mathbb{R})$ one can take

$$K = \mathrm{SO}(d), \qquad A = \left\{ \mathrm{diag}(e^{t_1}, \dots, e^{t_d}) \mid t_1 + \cdots + t_d = 0 \right\}.$$

Any semi-simple group admits so called *Cartan decomposition*: $G = K \cdot A \cdot K$. In fact, the A component can be assumed to belong to the *positive Weyl chamber* $A^+ \subset A$, so $G = K \cdot A^+ \cdot K$. In the case of $G = \mathrm{SL}_d(\mathbb{R})$

$$A^+ = \left\{ \mathrm{diag}(e^{t_1}, \dots, e^{t_d}) \mid t_1 + \cdots + t_d = 0, \quad t_1 \geq t_2 \geq \cdots \geq t_d \right\},$$

and the Cartan decomposition, stating that every $g \in G$ can be written (not always uniquely) as a product $g = kar$ with $k, r \in \mathrm{SO}(d)$ and $a \in A^+$, is just the polar decomposition of a matrix.

Proof of Howe-Moore's theorem 3.2.
Let $\pi : G \to U(\mathcal{H})$ be a unitary representation, and assume that for some $u, v \in \mathcal{H}$ the matrix coefficient $f_{u,v}(g)$ does not vanish at infinity, namely

$$|f_{u,v}(g_n)| = |\langle \pi(g_n)u, v\rangle| \geq \epsilon_0 > 0$$

for some fixed $\epsilon_0 > 0$ and a sequence $g_n \in G$ leaving compact subsets of G.

Step 1: non-vanishing along the Cartan.
Using Cartan decomposition $G = K \cdot A^+ \cdot K$ we write

$$g_n = k_n^{-1} a_n r_n$$

where $a_n \in A^+$, $k_n, r_n \in K$. Since K is a compact group, upon passing to a subsequence, we may assume that $k_n \to k$ and $r_n \to r$ in K. Continuity of π implies convergence of the vectors

$$x_n = \pi(r_n)u \longrightarrow x = \pi(r)u, \qquad y_n = \pi(k_n)v \longrightarrow y = \pi(r)v.$$

We have

$$f_{x_n,y_n}(a_n) = \langle \pi(a_n)x_n, y_n\rangle = \langle \pi(a_n r_n)u, \pi(k_n)v\rangle = \left\langle \pi(k_n^{-1} a_n r_n)u, v\right\rangle = f_{u,v}(g_n)$$

and can estimate

$$\begin{aligned} |f_{x_n,y_n}(a_n) - f_{x,y}(a_n)| &\leq |f_{x_n,y_n}(a_n) - f_{x,y_n}(a_n)| + |f_{x,y_n}(a_n) - f_{x,y}(a_n)| \\ &= |\langle \pi(a_n)(x_n - x), y_n\rangle| + |\langle \pi(a_n)x, (y_n - y)\rangle| \\ &\leq \|x_n - x\| \cdot \|y_n\| + \|x\| \cdot \|y_n - y\| \longrightarrow 0 \end{aligned}$$

because $\|y_n\| = \|y\|$ is constant. Thus

$$\lim_{n\to\infty} |f_{x,y}(a_n)| = \lim_{n\to\infty} |\langle \pi(a_n)x_n, y_n\rangle| \geq \epsilon_0 > 0. \tag{4}$$

We observe that since $g_n = k_n^{-1} a_n r_n \to \infty$ one has $a_n \to \infty$.

Step 2: use of Mautner's Lemma.
In a Hilbert space any closed ball is weakly compact (as is any closed bounded convex set). In particular, upon passing to a subsequence we may assume that the sequence $\{\pi(a_n)x\}_{i=1}^{\infty}$ of equal length vectors converges weakly to some vector z: $\pi(a_n)x \xrightarrow{w} z$, where weak convergence means that

$$\forall w \in \mathcal{H} : \qquad \langle \pi(a_n)x, w\rangle \longrightarrow \langle z, w\rangle.$$

A crucial point here is that $z \neq 0$, because by (4)

$$|\langle z, y\rangle| = \lim_{i\to\infty} |\langle \pi(a_n)x, y\rangle| \geq \epsilon_0 > 0.$$

Lemma 3.5 (Generalized Mautner Lemma)**.**
Let $\pi : G \to U(\mathcal{H})$ be a unitary representation of some topological group, and elements $\{a_n\}$ and h in G satisfy $a_n^{-1}ha_n \to e$ in G.
If vectors $y, z \in \mathcal{H}$ are such that $\pi(a_n)x \xrightarrow{w} z$ then $\pi(h)z = z$.
In particular, if $\pi(a_n)z = z$ then $\pi(h)z = z$.

Proof. (Strong) continuity of π gives

$$\|\pi(ha_n)x - \pi(a_n)x\| = \left\|\pi(a_n^{-1}ha_n)x - x\right\| \longrightarrow 0.$$

At the same time $\pi(a_n)x \xrightarrow{w} z$ and $\pi(ha_n)x \xrightarrow{w} \pi(h)z$. Hence $\pi(h)z = z$. □

Step 3: proof for $G = \mathrm{SL}_2(\mathbb{R})$.
We can now prove Theorem 3.2 in the case of $G = \mathrm{SL}_2(\mathbb{R})$. Let $\pi : \mathrm{SL}_2(\mathbb{R}) \to U(\mathcal{H})$ be a unitary representation with some matrix coefficient not vanishing at infinity Applying Step 1 we get a sequence

$$a^{t_n} = \begin{pmatrix} e^{t_n} & 0 \\ 0 & e^{-t_n} \end{pmatrix} \in A^+$$

with $t_n \to \infty$, and *non zero* vectors $x, z \in \mathcal{H}$ with $\pi(a^{t_n})x \xrightarrow{w} z$. Consider the *horocyclic* subgroup

$$H = \left\{ h^s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\}$$

It is normalized by the diagonal subgroup A, and for every $h^s \in H$ one has

$$a^{-t_n} h^s a^{t_n} = h^{e^{-2t_n}\cdot s} \longrightarrow e$$

as $i \to \infty$. Mautner's Lemma 3.5 yields that the non-zero vector z is $\pi(H)$-invariant. The matrix coefficient $f_{z,z}(g) = \langle \pi(g)z, z\rangle$ is a continuous function on G, which is bi-H-invariant:

$$\begin{array}{llcl} \text{(a)} & f_{z,z}(gh) & = & \langle \pi(g)\pi(h)z, z\rangle = \langle \pi(g)z, z\rangle = f_{z,z}(g) \\ \text{(b)} & f_{z,z}(hg) & = & \left\langle \pi(g)z, \pi(h^{-1})z\right\rangle = \langle \pi(g)z, z\rangle = f_{z,z}(g) \end{array}$$

for all $g \in G$ and $h \in H$. By (a) $f_{z,z} : G \to \mathbb{C}$ descends to a continuous function $\phi : G/H \to \mathbb{C}$, that is H-invariant for the left H-action on G/H. Observe that $G/H = \mathbb{R}^2 - \{(0,0)\}$ with the H-action given by

$$h^s(x, y) = (x + sy, y)$$

it follows that ϕ is a constant c_y on each horizontal line $\ell_y = \{(x, y) \mid x \in \mathbb{R}\}$, for $y \neq 0$. By continuity, ϕ is a constant c_0 on the punctured x-axis as well

$$\{(x, 0) \mid x \neq 0\}.$$

Since $\phi(1, 0) = f_{z,z}(e) = \|z\|^2$, we have

$$\|z\|^2 = \phi(e^{2t}, 0) = \left\langle \pi(a^t)z, z\right\rangle \qquad (t \in \mathbb{R})$$

where $a^t = \mathrm{diag}(e^t, e^{-t}) \in A$. By the strict case of the Cauchy-Schwartz inequality it follows that z is $\pi(A)$-invariant.

Consequently z is fixed by the (connected) upper triangular subgroup of G

$$AH = \left\{ \begin{pmatrix} e^t & s \\ 0 & e^{-t} \end{pmatrix} \mid t, s \in \mathbb{R} \right\}$$

and therefore $f_{z,z}$ and ϕ further descend to a continuous function ψ on G/AH which can be identified with the projective line $\mathbf{P}(\mathbb{R}^2) = \mathbb{R} \cup \{\infty\}$ with $G = \mathrm{SL}_2(\mathbb{R})$ acting by fractional linear transformations. In particular, H acts on $\mathbf{P}(\mathbb{R}^2) = \mathbb{R} \cup \{\infty\}$ by translation: $h^s : x \mapsto x + s$. Since ψ is invariant under this action, while the H-orbit of $x = 1$ is dense, it follows that ψ is constant. Hence so are $\phi : G/H \to \mathbb{C}$ and $f_{z,z} : G \to \mathbb{C}$. Therefore

$$\langle \pi(g)z, z \rangle = f_{z,z}(g) = f_{z,z}(e) = \|z\|^2 \qquad (g \in G)$$

implying, by Cauchy-Schwartz, that $0 \neq z \in \mathcal{H}^{\pi(G)}$.

Step 4: proof for $G = \mathrm{SL}_d(\mathbb{R})$.
One of the key facts about semi-simple Lie groups is that they can be built out of copies of $\mathrm{SL}_2(\mathbb{R})$. In the case of $G = \mathrm{SL}_d(\mathbb{R})$ for each $1 \leq k < \ell \leq d$ one considers the image $G_{k,\ell}$ of the isomorphic embedding $\tau_{k,\ell} : \mathrm{SL}_2(\mathbb{R}) \longrightarrow \mathrm{SL}_d(\mathbb{R})$, which maps

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

to the $d \times d$ matrix g which looks like the identity except for the entries $g_{k,k} = \alpha$, $g_{k,\ell} = \beta$, $g_{\ell,k} = \gamma$, $g_{\ell,\ell} = \delta$.

For $i \neq j$ in $\{1, \dots, d\}$ let $H_{i,j}$ denote the matrices $g \in \mathrm{SL}_d(\mathbb{R})$ with 1-s on the diagonal and 0-s in all off diagonal entries, except possibly for the (i,j)-entry. So for $1 \leq k < \ell \leq d$, $H_{k,\ell} < G_{k,\ell}$ is the image of the horocycle ($\alpha = \delta = 1$, $\gamma = 0$, $\beta \in \mathbb{R}$).

Let $\pi : G = \mathrm{SL}_d(\mathbb{R}) \to U(\mathcal{H})$ be a unitary representation, that has some matrix coefficient $f_{u,v}(g)$ that does not vanish at infinity. By Step 1, there is a sequence $a_n \in A^+$, $a_n \to \infty$, and vectors $x, z \neq 0$ so that $\pi(a_n)x \xrightarrow{w} z$. Write

$$a_n = \mathrm{diag}(e^{t_{1,n}}, \dots, e^{t_{d,n}}), \qquad t_{1,n} + \cdots + t_{d,n} = 0, \qquad t_{1,n} \geq \cdots \geq t_{d,n}.$$

As $a_n \to \infty$, one can assume, after passing to a subsequence, that (still denoted by subscript n) so that

$$t_{k,n} - t_{\ell,n} \longrightarrow +\infty \qquad \text{as} \qquad n \longrightarrow +\infty.$$

for some fixed pair of indices $1 \leq k < \ell \leq d$.

Then for every $h \in H_{k,\ell}$ one has $a_n^{-1} h a_n \to e$ and by Lemma 3.5, z is fixed by $\pi(H_{k,\ell})$. Considering the restriction of π to $G_{k,\ell} \cong \mathrm{SL}_2(\mathbb{R})$ we deduce from (the proof of) Step 3, that z is fixed by all of $G_{k,\ell}$. In particular, z is fixed by

$$a_{k,\ell}^n = \tau_{k,\ell}(\mathrm{diag}(e^n, e^{-n})) = \mathrm{diag}(1, \dots, e^n, \dots 1 \dots, e^{-n}, \dots, 1) \in \mathrm{SL}_d(\mathbb{R}).$$

Another application of Lemma 3.5 shows that z is fixed by $\pi(H_{k,j})$ with any $j \neq k$, because for every $h \in H_{k,j}$

$$a_{k,\ell}^{-n}\, h\, a_{k,\ell}^n \longrightarrow e$$

as $n \to \infty$ if $k < j$, and $n \to -\infty$ if $1 \leq j < k$. Applying similar arguments to $H_{k,j}$ we deduce that z is fixed by $\pi(H_{i,j})$ for all $i \neq j$. Since these groups generated $G = \mathrm{SL}_d(\mathbb{R})$, it follows that $\mathcal{H}^{\pi(G)} \neq \{0\}$. □

Proof of Corollary 3.4.
The probability measure space $(X, \mu) = (G/\Gamma, m_{G/\Gamma})$ has a transitive measure-preserving action of the simple Lie group G by translations: $g : h\Gamma \mapsto gh\Gamma$. Consider the orthogonal complement in $L^2(X, \mu)$ to constant functions:

$$\mathcal{H} = L^2_0(X, \mu) = \{f \in L^2(X, \mu) \mid \int f \, d\mu = 0\}.$$

Then $\pi : G \to U(\mathcal{H})$, given by $(\pi(g)f)(h\Gamma) = f(g^{-1}h\Gamma)$, is a unitary representation without non-trivial invariant vectors (exercise!). By Howe-Moore (Theorem 3.2), for any $f_1, f_2 \in \mathcal{H}$

$$\lim_{n\to\infty} \langle \pi(\alpha^n) f_1, f_2 \rangle = 0$$

We claim that this corresponds to the property of mixing for T on (X, μ) that means, by definition, that for any measurable subsets $A, B \subset X$ one has *asymptotic independence* of $T^{-n}B$ from A:

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

Indeed, given a measurable subset $E \subset X$ the projection f_E of the characteristic function 1_E to $\mathcal{H} = L^2_0(X, \mu)$ is

$$f_E(x) = 1_E(x) - \mu(E) = (1 - \mu(E)) \cdot 1_E(x) + (-\mu(E)) \cdot 1_{X \setminus E}(x).$$

One calculates

$$\langle f_A, f_B \rangle = \mu(A \cap B) - \mu(A)\mu(B)$$

and

$$\langle \pi(\alpha^n) f_A, f_B \rangle = \langle f_A, f_{T^{-n}B} \rangle = \mu(A \cap T^{-n}B) - \mu(A)\mu(B).$$

Finally, mixing implies ergodicity, because any set E with $\mu(E \triangle T^{-1}E) = 0$ would have

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

which is possible only if $\mu(E) = 0$ or $\mu(E) = 1$. $\square$

4. A glimpse of entropy

An important class of examples of dynamical systems are homogeneous spaces G/Γ where G is a Lie group (e.g. $\mathrm{SL}_2(\mathbb{R})$) and $\Gamma < G$ is a lattice, that is a discrete subgroup of finite covolume in G. Here we shall focus on uniform lattices – discrete subgroups $\Gamma < G$ with G/Γ being compact.

4.1. **Classification of a class of systems.**

Consider the family of examples of dynamical systems

$$T_\alpha : G/\Gamma \ \to \ G/\Gamma, \qquad T_\alpha : g\Gamma \mapsto \alpha g\Gamma$$

where $G = \mathrm{SL}_2(\mathbb{R})$ and $\Gamma < G$ is a uniform lattice lattice and $g \in G$ is some fixed element. To construct such $\Gamma < G$ take a surface Σ of genus ≥ 2, and choose a Riemannian metric ρ of constant curvature -1 on it. This metric ρ lifts to the universal cover – the hyperbolic plane $\mathbb{H}^2$ – and the fundamental group of Σ acts by isometries on $\mathbb{H}^2$, giving rise to an imbedding

$$j_\rho : \pi_1(\Sigma, *) \longrightarrow \mathrm{Isom}(\mathbb{H}^2) = \mathrm{PSL}_2(\mathbb{R})$$

whose image is a discrete cocompact subgroup $\Gamma_{\rho,*} < \mathrm{PSL}_2(\mathbb{R})$. Changing the base point $* \in \Sigma$ is equivalent to conjugating Γ. But varying ρ on Σ (there are $6 \cdot \mathrm{ginus}(\Sigma) - 6$ many dimensional space of such choices) gives a multi-dimensional family of mutually non-conjugate uniform lattices Γ_ρ. One may also think of them as lifted to the double cover $\mathrm{SL}_2(\mathbb{R}) \to \mathrm{PSL}_2(\mathbb{R})$.

Note that if α is conjugate to β in G, then T_α and T_β are indistinguishable as transformations of $X = G/\Gamma$; indeed, if $\alpha = \gamma\beta\gamma^{-1}$ then $T_\gamma : X \to X$ intertwines T_β with T_α.

Note also, that if $\Lambda = b^{-1}\Gamma b$ then the map $G/\Gamma \ \to \ G/\Lambda$, $g\Gamma \mapsto gb\Lambda$, intertwines the translation by any α on G/Γ and on G/Λ. Therefore, studying translations T_α on G/Γ, it suffices to focus on representatives of conjugacy classes of elements α and on representatives of conjugacy classes of lattices Γ.

Exercise 4.1. Prove that the following three families of matrices are representative classes of all non-trivial conjugacy classes in $G = \mathrm{SL}_2(\mathbb{R})$:

$$r_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \qquad h = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad a^t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

with $\theta \in (0, 2\pi)$, $t \in (0, +\infty)$ (the normalization $t/2$ is a convention).

Elements of the compact group $\mathrm{SO}(2) = \{r_\theta\}$ do not have interesting dynamics on G/Γ. So we can restrict our attention to just two classes of examples:

- $(G/\Gamma, T_h)$ where h is the unipotent matrix,
- $(G/\Gamma, T_{a^t})$ where a^t is a diagonal matrix,

known as the discretized *horocyclic* flow, and discretized *geodesic* flow, according to their geometric meaning when $\mathrm{PSL}_2(\mathbb{R})/\Gamma_\rho$ is identified with the unit tangent bundle to (Σ, ρ).

Let us now list some known facts about this class of dynamical systems.

(P1) $(G/\Gamma, T_h)$ is uniquely ergodic (Furstenberg).

(P2) $(G/\Gamma, T_{a^t})$ for any $t > 0$, has uncountably many different ergodic measures, and minimal sets of arbitrary Hausdorff dimension in $[0, 3]$.
(P3) $(G/\Gamma, T_{a^t})$ has unique measure of *maximal entropy* (see below), which is precisely the G-invariant measure $m_{G/\Gamma}$.

Properties (P2), (P3) follow from properties of *Anosov diffeomorphisms* and *Anosov flows* studied in *hyperbolic dynamics*.

Two topological systems (X_i, T_i) are considered topologically conjugate, here denoted by $(X_1, T_1) \cong_{\text{top}} (X_2, T_2)$, if there exists a homeomorphism

$$\phi : X_1 \xrightarrow{\cong} X_2 \qquad \text{so that} \qquad \phi \circ T_1 = T_2 \circ \phi.$$

In this context one has the following *topological dynamics rigidity* results

(T1) $(G/\Gamma, T_h) \ncong_{\text{top}} (G/\Lambda, T_{a^t})$ for any $t > 0$.
(T2) There exists $\phi : (G/\Gamma, T_h) \cong_{\text{top}} (G/\Lambda, T_h)$ if and only if

$$\exists b \in G : \qquad \Gamma = b\Lambda b^{-1}, \qquad \phi(g\Gamma) = gb\Lambda.$$

(T3) There exists $\phi : (G/\Gamma, T_{a^t}) \cong_{\text{top}} (G/\Lambda, T_{a^s})$ if and only if

$$t = s, \quad \text{and} \quad \exists b \in G : \quad \Gamma = b\Lambda b^{-1}, \quad \phi(g\Gamma) = gb\Lambda.$$

In the context of (ergodic) measure preserving systems the classification is up to *measurable isomorphism*, where two ergodic probability measure-preserving systems (X_i, μ_i, T_i), $(i = 1, 2)$, are measurably isomorphic if there exists measure space isomorphism

$$\theta : (X_1, \mu_1) \xrightarrow{\cong} (X_2, \mu_2) \quad \text{s.t.} \quad \mu_1\{x \mid \theta \circ T_1(x) \neq T_2 \circ \theta(x)\} = 0.$$

Note, that a priori measurable isomorphism is a rather subtle relation as everything is defined up to null sets. Consider the spaces G/Γ with the G-invariant probability measure $m_{G/\Gamma}$, and recall that T_h and every T_{a^t} (any $t > 0$) is ergodic (Corollary 3.4). Then

(E1) All the above systems $(G/\Gamma, m_{G/\Gamma}, T_h)$ and $(G/\Lambda, m_{G/\Lambda}, T_{a^t})$, $t > 0$, are *unitarily equivalent*, meaning that there are isometric isomorphisms between the Hilbert spaces intertwining the unitary operators induced by the underlying measure-space transformations.
(E2) $(G/\Gamma, m_{G/\Gamma}, T_h) \ncong (G/\Lambda, m_{G/\Lambda}, T_{a^t})$ for any $t > 0$.
(E3) There exists $\phi : (G/\Gamma, m_{G/\Gamma}, T_h) \cong (G/\Lambda, m_{G/\Lambda}, T_h)$ iff $\exists b \in G$ so that

$$\Gamma = b\Lambda b^{-1}, \qquad \phi(g\Gamma) = gb\Lambda.$$

(E4) There exists $\phi : (G/\Gamma, m_{G/\Gamma}, T_{a^t}) \cong_{\text{top}} (G/\Lambda, m_{G/\Lambda}, T_{a^s})$ if and only if

$$t = s.$$

Howe-Moore's theorem implies that the unitary operators are mixing; in fact, it can be shown that the operators have countable Lebesgue spectrum, which implies (E1). Fact (E2) and the "only if" direction of (E4) can be shown using the notion of *entropy* (see Corollary 4.15 below), which is a numeric measurable-isomorphism invariant of general systems (X, μ, T). A brief introduction to entropy is given below.

The "if" direction of (E4), is a consequence of the fact that $(G/\Gamma, m_{G/\Gamma}, T_{a^t})$ are Bernoulli shifts (Ornstein-Weiss [7]), and the fact that entropy is the complete isomorphism invariant for Bernoulli shifts (Ornstein [6]). Fact (E3) is a result of

Ratner, that predated her proof of Ragunathan conjecture and related results [9], that in turn, easily implies (E3), (T2) and (P1).

4.2. **Introduction to Entropy.**
Let A be a finite set, $\Sigma = A^{\mathbb{N}}$ the infinite product with the Tychonoff topology (so Σ is a Cantor set, unless A is singleton), and let $\theta : \Sigma \to \Sigma$ denote the shift

$$\theta : (y_1, y_2, y_3, \dots) \mapsto (y_2, y_3, y_4, \dots)$$

Given a word $w = (w_1, w_2 \dots, w_n)$ of length n in the alphabet A (so $w \in A^n$), the *cylinder set* C_w is

$$C_w = \{y \in \Sigma \mid y_1 = w_1, \dots, y_n = w_n\}.$$

For $y \in \Sigma$ let $[y]_n$ denote the finite word $[y]_n = (y_1, y_2, \dots, y_n)$.

There are many θ-invariant probability measures on Σ. Given $\nu \in \mathrm{P}_{\mathrm{inv}}(\Sigma)$ we shall refer to (Σ, ν, θ) as a *shift space*, or a measure-preserving shift transformation.
Given a shift space (Σ, ν, θ) define the functions $I_n : \Sigma \to [0, \infty]$ by

$$I_n(y) = -\log \nu(C_{[y]_n}) = -\log \nu\{z \in \Sigma \mid z_1 = y_1,\ z_2 = y_2, \dots, z_n = y_n\}$$

and denote by H_n their averages

$$H_n = \int_\Sigma I_n(y)\, d\nu(y) = \sum_{w \in A^n} -\nu(C_w) \cdot \log \nu(C_w).$$

Exercise 4.2. Prove that $H_{n+m} \le H_n + H_m$ for any $n, m \in \mathbb{N}$, using the following:

(a) Given words $u \in A^n$, $w \in A^m$, denote by $uw \in A^{n+m}$ their concatenation. Use shift invariance of ν to show

$$\nu(C_u) = \sum_{w \in A^m} \nu(C_{uw}), \qquad \nu(C_w) = \sum_{u \in A^n} \nu(C_{uw})$$

(b) Use convexity of the function $x \mapsto -x \cdot \log x$ to prove the claim.

Subadditivity of the sequence $\{H_n\}$ enables one to define the (Shannon) *entropy of the process* (Σ, ν, θ) to be

$$\mathrm{ent}(\Sigma, \nu, \theta) = \lim_{n \to \infty} \frac{1}{n} H_n = \inf_{n \ge 1} \frac{1}{n} H_n.$$

Definition 4.3 (Kolmogorov-Sinai)**.**
The entropy of a general ergodic system (X, μ, T) is defined to be

$$\mathrm{Ent}(X, \mu, T) = \sup\{\mathrm{ent}(\Sigma, \nu, \theta) \mid \exists\ (X, \mu, T) \xrightarrow{p} (\Sigma, \nu, \theta)\}$$

where the sup is taken over all possible shift systems as quotients.

Remark 4.4. $\mathrm{Ent}(X, \mu, T)$ is a measurable isomorphism invariant.

Remark 4.5. To define a shift space as a quotient $(X, \mu, T) \longrightarrow (\Sigma, \nu, \theta)$ of a measurable system is the same as choosing a measurable partition

$$X = E_1 \sqcup \dots \sqcup E_a.$$

Indeed, given such a partition take $\Sigma = \{1, \dots, a\}^{\mathbb{Z}}$, define $p(x) = y$ with $y_i = k$ if $T^i x \in E_k$, and take $\nu = p_* \mu$. Note that $p \circ T = \theta \circ p$ by construction.
Conversely, given a quotient map as above, construct a partition of X by

$$X = E_1 \sqcup \dots \sqcup E_a, \qquad E_k = p^{-1}(C_k).$$

We state the following key result without proof.

Theorem 4.6 (Shannon-McMillan-Breiman)**.**
Assume $\nu \in \mathrm{P}_{\mathrm{erg}}(\Sigma)$ *is* θ*-ergodic. Then*

$$\frac{1}{n} I_n(y) \longrightarrow \mathrm{ent}(\Sigma, \nu, \theta)$$

where the convergence is in measure, in $L^1(\Sigma, \nu)$*, and* ν*-a.e.*

So $\mathrm{ent}(\Sigma, \nu, \theta)$ is the constant describing the exponential rate of shrinkage on ν-size of points with similar n-step trajectory. The weakest form of convergence above, the convergence in measure, allows to characterize the entropy as the growth rate of the number of n-words needed to capture significant size of the space.

Corollary 4.7.
Let (Σ, ν, θ) *be a shift system and* $h = \mathrm{ent}(\Sigma, \nu, \theta)$*. Then given* $\epsilon > 0$ *there is* N *so that for every* $n \geq N$*:*

- *There is a collection* $\Omega \subset A^n$ *of words of length* n *so that*
$$|\Omega| < e^{(h+\epsilon)n} \qquad \text{and} \qquad \nu(\bigsqcup_{w \in \Omega} C_w) > 1 - \epsilon$$
- *For any collection* $W \subset A^n$ *of words of length* n*:*
$$|W| < e^{(h-\epsilon)n} \qquad \Longrightarrow \qquad \nu(\bigsqcup_{w \in W} C_w) < \epsilon.$$

Proof. Convergence in measure $n^{-1} \cdot I_n(y) \to h$ (that follows from the L^1-convergence) implies that there is N so that for $n \geq N$ the set

$$G = \{y \in \Sigma \mid I_n(y) = -\log \nu(C_{[y]_n}) < (h + \epsilon)n\}$$

has $\nu(G) > 1 - \epsilon$. This set consists of n-cylinders, so the collection

$$\Omega = \{[y]_n \mid y \in G\} \qquad \text{satisfies} \qquad \Sigma_n = \bigsqcup_{w \in \Omega} C_w.$$

For $w \in \Omega$ one has $\nu(C_w) > e^{-(h+\epsilon)n}$. Hence $|\Omega| < e^{(h+\epsilon)n}$. The second statement is left to the reader. □

Exercise 4.8. Prove using Stirling's formula that given $a > 1$ and $\epsilon > 0$ there is $\delta > 0$ and $N \in \mathbb{N}$ so that for all $n \geq N$ one has

$$\binom{n}{\lceil \delta n \rceil} \cdot a^{\delta n} < a^{\epsilon n}. \tag{5}$$

Proposition 4.9.
Given $a \in \mathbb{N}$ *and* $\epsilon > 0$ *there is* $\delta > 0$ *with the following property: for any ergodic system* (X, μ, T) *with quotient maps to the shift on* $\Sigma = \{1, \ldots, a\}^{\mathbb{N}}$

$$p : (X, \mu, T) \longrightarrow (\Sigma, \nu, \theta), \qquad p' : (X, \mu, T) \longrightarrow (\Sigma, \nu', \theta)$$

satisfying $|\nu(C_i) - \nu'(C_i)| < \delta$ *for each* $i \in \{1, \ldots, a\}$*, one has*

$$|\,\mathrm{ent}(\Sigma, \nu, \theta) - \mathrm{ent}(\Sigma, \nu', \theta)| < \epsilon.$$

Proof. Given $a=|A|$ and $\epsilon>0$ use Exercise 4.8 to choose $\delta>0$ so that

$$\binom{n}{\lceil 2\delta n\rceil}\cdot a^{\lceil 2\delta n\rceil}<a^{\frac{1}{2}\epsilon n}.$$

Assume that the measurable quotient maps $p,p':(X,\mu,T)\to(\Sigma,S)$ satisfy

$$E=\{x\in X\mid p(x)_1\neq p'(x)_1\},\qquad \text{has}\qquad \mu(E)<\delta.$$

As a consequence of the ergodic theorem (this follows already from the mean ergodic theorem 2.9) for large n most points $x\in X$ visit E with frequency $<2\delta$. More precisely

$$A_n=\left\{x\in X\mid \#\{1\le k\le n\mid T^k\in E\}<\lceil 2\delta n\rceil\right\}\qquad \text{has}\qquad \mu(A_n)>0.9$$

for all $n>N_1$. From Corollary 4.7, for large enough N_2 for each $n>N_2$ there is a set $\Omega_n\subset A^n$ of words so that

$$|\Omega_n|<e^{(\mathrm{ent}(\Sigma,\nu,\theta)+\frac{1}{2}\epsilon)n}$$

and the set

$$B_n=\{x\in X\mid [p(x)]_n\in\Omega_n\}\qquad \text{has}\qquad \mu(B_n)>0.9.$$

For every $n>\max(N_1,N_2)$ the set $C_n=A_n\cap B_n$ has $\mu(C_n)>0.8$.

Let $W_n\subset A^n$ be the collection of all the words that are obtained as follows: for each $w\in\Omega_n$ look at all subsets $J\subset\{1,\dots,n\}$ of size $|J|=\lceil 2\delta n\rceil$, and include in W_n all n-long words that agree with w for $i\notin J$. Then

$$|W_n|\le\binom{n}{\lceil 2\delta n\rceil}\cdot a^{\lceil 2\delta n\rceil}\cdot|\Omega_n|<e^{\frac{1}{2}\epsilon n}\cdot e^{(\mathrm{ent}(\Sigma,\nu,\theta)+\frac{1}{2}\epsilon)n}=e^{(\mathrm{ent}(\Sigma,\nu,\theta)+\epsilon)n}$$

while for $x\in C_n$ one has $[p'(x)]_n\in W_n$. Thus by Corollary 4.7

$$\mathrm{ent}(\Sigma,\nu',\theta)\le\mathrm{ent}(\Sigma,\nu,\theta)+\epsilon$$

and a symmetric argument completes the proof. □

Exercise 4.10. Prove

(a) For a shift system (Σ,ν,θ) and $k\in\mathbb{N}$ one has

$$\mathrm{ent}(\Sigma,\nu,\theta^k)=k\cdot\mathrm{ent}(\Sigma,\nu,\theta).$$

(b) For a general system (X,μ,T) and $k\in\mathbb{N}$ one has

$$\mathrm{Ent}(X,\mu,T^k)=k\cdot\mathrm{Ent}(X,\mu,T).$$

Let A and B be finite sets, $c:A^k\to B$ be a map (k-to-1 encoding), $\Sigma=A^{\mathbb{N}}$ and $Z=B^{\mathbb{N}}$. Consider the shift equivariant map $p:\Sigma\to Z$ where $p(y)=z$ means that

$$z_i=c(y_i,y_{i+1},y_{i+2},\dots,y_{i+k-1})\qquad(i\in\mathbb{N}).$$

Exercise 4.11. Let $c:A^k\to B$ and $p:\Sigma\to Z$ be as above, $\nu\in\mathrm{P}_{\mathrm{inv}}(\Sigma)$ be shift-invariant measure and $\eta=p_*\nu\in\mathrm{P}_{\mathrm{inv}}(Z)$. Show that

$$\mathrm{ent}(\Sigma,\nu,\theta)\ge\mathrm{ent}(Z,\eta,\theta)$$

and that injectivity of $c:A^k\to B$ is sufficient for equality.

Theorem 4.12 (Kolmogorov-Sinai)**.**
For an ergodic measure ν on a shift space $\Sigma=A^{\mathbb{N}}$ one has

$$\mathrm{Ent}(\Sigma,\nu,\theta)=\mathrm{ent}(\Sigma,\nu,\theta).$$

Proof. The inequality $\mathrm{Ent}(\Sigma,\nu,\theta)\geq\mathrm{ent}(\Sigma,\nu,\theta)$ is clear from the definition of the Kolmogorov-Sinai entropy (4.3). To establish the reverse inequality one needs to show that Shannon entropy of an *arbitrary measurable quotient* process

$$p:(\Sigma,\nu,\theta)\to(\Omega,\eta,\sigma)$$

cannot be larger than that of the original, namely

$$\mathrm{ent}(\Omega,\eta,\sigma)\leq\mathrm{ent}(\Sigma,\nu,\theta).$$

Here $\Omega=B^{\mathbb{N}}$ Note that here $\Sigma_1=A_1^{\mathbb{N}}$ is a shift space on potentially larger alphabet A_1 and the quotient map p is just measurable.

Let $\epsilon>0$ be arbitrary. Using Proposition 4.9 one can find $\delta>0$ so that if $(\Sigma_1,\nu_2,\theta_2)$ is some shift space be the associated to $|A_1|$ and $\epsilon>0$ as in . For every measurable subset $E\subset\Sigma$ there is $k\in\mathbb{N}$ and collection $W\subset A^k$ so that

$$\nu(E\triangle\bigsqcup_{w\in W}C_w)<\delta.$$

Working with individual letters $a\in A_1$ we can approximate $p^{-1}(C_a)$ by □

Exercise 4.13. Let ζ be a probability measure on a finite set $Z=\{1,\dots,k\}$, say $\zeta(\{i\})=p_i$, where $p_1+p_2+\cdots+p_k=1$. Let T be the Bernoulli shift on $(X,\mu)=(Z,\zeta)^{\mathbb{N}}$ as in Exercise 2.7. Prove:

$$\mathrm{Ent}(X,\mu,T)=\sum -p_i\cdot\log p_i.$$

4.3. **Entropy of some homogeneous systems.**

Let us compute the entropy for translations on the homogeneous space of $G=\mathrm{SL}_2(\mathbb{R})$. These special cases illustrate some general phenomena.

Theorem 4.14.
Let $G=\mathrm{SL}_2(\mathbb{R})$ and $\Gamma<G$ a cocompact lattice. Then

(a) $\mathrm{Ent}(G/\Gamma,m_{G/\Gamma},T_{a^t})=t$.
(b) $\mathrm{Ent}(G/\Gamma,m_{G/\Gamma},T_h)=0$.

Corollary 4.15.
Let $\Gamma,\Lambda<G=\mathrm{SL}_2(\mathbb{R})$ be cocompact lattices, and $t,s>0$. Then:

(a) $(G/\Gamma,m_{G/\Gamma},T_h)\not\cong(G/\Lambda,m_{G/\Lambda},T_{a^s})$.
(b) *If* $(G/\Gamma,m_{G/\Gamma},T_{a^t})\cong(G/\Lambda,m_{G/\Lambda},T_{a^s})$ *then* $t=s$.

Proof of Theorem 4.14.
Let $\Gamma<G=\mathrm{SL}_2(\mathbb{R})$ be a uniform lattice. The compact space $X=G/\Gamma$ is equipped with the G-invariant probability measure $\mu=m_{G/\Gamma}$, and transformation T_α for $\alpha=h$ or $\alpha=a^t$ for some fixed $t>0$.

The fact that $\mathrm{Ent}(X,\mu,T_h)=0$ can be shown using ideas similar to the "upper estimate" part of the T_{a^t} case below. However, assuming one knows that this entropy is finite, it is immediate that it must be zero, because T_h is conjugate (by a measure-space isomorphism) to its square

$$(T_h)^2=T_{h^2}=T_{a^tha^{-t}}=(T_{a^t})T_h(T_{a^t})^{-1}\qquad\text{where}\qquad t=\log 2$$

so using Exercise 4.10 one obtains

$$2\cdot \mathrm{Ent}(X,\mu,T_h)=\mathrm{Ent}(X,\mu,T_h^2)=\mathrm{Ent}(X,\mu,T_h).$$

So we focus on the case T_{a^t} for some fixed $t>0$. We shall show the inequalities

$$t\le \mathrm{Ent}(X,\mu,T_{a^t}),\qquad \mathrm{Ent}(X,\mu,T_{a^t})\le t.$$

The Lower estimate.
For each $x\in X$ there is $r_x>0$ so that the map $\phi_x:\mathbb{R}^3\to X$ given by

$$\phi_x(u,v,w)=\begin{pmatrix}1&0\\u&1\end{pmatrix}\begin{pmatrix}e^{v/2}&0\\0&e^{-v/2}\end{pmatrix}\begin{pmatrix}1&w\\0&1\end{pmatrix}.x \tag{6}$$

is a diffeomorphism from a cube $(-r_x,r_x)^3\subset\mathbb{R}^3$ onto an open neighborhood V_x of $x\in X$. Since $\{V_x\}_{x\in X}$ form an open cover of a compact space X, there is an $r>0$ (Lebesgue number) so that for every $x\in X$ we have a diffeomorphism

$$\phi_x:(-r,r)^3\xrightarrow{\cong}W_x\subset X$$

from a fixed cube onto some open neighborhood W_x of x. By continuity of $T=T_{a^t}$ there exists $r>\delta>0$ small enough so that the ϕ_x-image U_x of the smaller cube $(-\delta,\delta)^3$ satisfies

$$T(U_x)\subset W_{T.x}=\phi_{T.x}((-r,r)^3).$$

In the (u,v,w)-coordinates the map $T:X\to X$ is given by

$$\theta_x=\phi_{T.x}^{-1}\circ T\circ\phi_x,\qquad \theta_x(u,v,w)=(e^{-t}u,v,e^tw). \tag{7}$$

In particular, the w-direction is expanded by a factor of e^t. Consider an intersection of consecutive preimages of sets U_x

$$U_{x_1}\cap T^{-1}U_{x_2}\cap\cdot\cap T^{-(n-1)}U_{x_n}$$

for some $x_1,\dots,x_n\in X$. Such intersection is contained in the ϕ_x-image of a very narrow box

$$(-\delta,\delta)\times(-\delta,\delta)\times(-e^{-nt}\delta,e^{-nt}\delta).$$

The measure μ has some (smooth) density with respect to the $du\,dv\,dw$-measure (pushed forward by ϕ_{x_1}), and since the space is compact we get an estimate

$$\mu(U_{x_1}\cap T^{-1}U_{x_2}\cap\cdot\cap T^{-(n-1)}U_{x_n})<Ce^{-nt}$$

with some constant C independent of n and $x_1,\dots,x_n$.

Let $X=E_1\sqcup\cdots\sqcup E_a$ be a measurable partition into sets that are small enough to ensure $E_i\subset U_{z_i}$ for some points $z_1,\dots z_a\in X$. Define the equivariant quotient $p:X\to\Sigma=\{1,\dots,a\}^{\mathbb{N}}$ as in Remark 4.5, and let $\nu=p_*\mu$ to obtain the quotient map:

$$p:(X,\mu,T)\longrightarrow(\Sigma,\nu,S).$$

Then for every $y\in\Sigma$ and n one has (for some $x_1,\dots,x_n\in\{z_1,\dots,z_a\}$)

$$\begin{aligned}I_n(y)&=-\log\nu(C_{[y]_n})=-\log\mu(E_{y_1}\cap T^{-1}E_{y_2}\cap\cdots\cap T^{-(n-1)}E_{y_n})\\&\ge-\log\mu(U_{x_1}\cap T^{-1}U_{x_2}\cap\cdots\cap T^{-(n-1)}U_{x_n})\\&\ge-\log C+nt.\end{aligned}$$

Dividing by n, integrating $d\mu(x)$, and passing to the limit, gives

$$\mathrm{Ent}(X,\mu,T)\ge\mathrm{ent}(\Sigma,\nu,S)=t. \tag{8}$$

The Upper estimate.
We shall need the concept of *topological entropy*, which is a numeric invariant associated to a general topological dynamical system (X,T) where $T : X \to X$ is a continuous map of a compact metrizable space X.

Given an open cover $\mathcal{U} = \{U_i\}$ of X, let $N(\mathcal{U})$ denote the minimal cardinality of a subcover $\{U_{i_1}, \dots, U_{i_N}\}$. Given two open covers $\mathcal{U} = \{U_i\}$, $\mathcal{V} = \{V_j\}$, their *join* is given by

$$\mathcal{U} \vee \mathcal{V} = \{U_i \cap V_j \mid U_i \in \mathcal{U},\ V_j \in \mathcal{V}\}$$

We also denote $T^{-1}\mathcal{U} = \{T^{-1}U \mid U \in \mathcal{U}\}$.

Exercise 4.16. Prove that for covers $\mathcal{U}$, $\mathcal{V}$ of X:
(a) $N(T^{-1}\mathcal{U}) \le N(\mathcal{U})$ and $N(\mathcal{U} \vee \mathcal{V}) \le N(\mathcal{U}) \cdot N(\mathcal{V})$.
(b) Deduce the subadditivity $h_{n+m}(\mathcal{U}) \le h_n(\mathcal{U}) + h_m(\mathcal{U})$ of the numbers

$$h_n(\mathcal{U}) = \log N(\mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-(n-1)}\mathcal{U}).$$

Definition 4.17. The topological entropy of (X,T) is defined to be

$$\mathrm{Ent}^{\mathrm{top}}(X,T) = \sup_{\mathcal{U}} \left(\lim_{n\to\infty} \text{ or } \inf_{n\ge 1} \frac{1}{n} h_n(\mathcal{U}) \right)$$

where the supremum is taken over all open covers $\mathcal{U}$ of X.

Theorem 4.18.
For any topological dynamical system (X,T) for any $\mu \in \mathrm{P}_{\mathrm{inv}}(X)$ one has

$$\mathrm{Ent}(X,\mu,T) \le \mathrm{Ent}^{\mathrm{top}}(X,T).$$

In fact, the supremum of $\mathrm{Ent}(X,\mu,T)$ over all invariant (or ergodic) probability measures is precisely $\mathrm{Ent}^{\mathrm{top}}(X,T)$. This is known as the *Variational Principle.*

Proof of Theorem 4.18. Let $X = X_1 \sqcup \cdots \sqcup X_a$ be some measurable partition corresponding to a quotient shift $p : (X,\mu,T) \to (\Sigma,\nu,S)$ on $\Sigma = \{1,\dots,a\}^{\mathbb{N}}$ as in Remark 4.5. Denote

$$H = \mathrm{Ent}^{\mathrm{top}}(X,T)$$

and let $\epsilon > 0$ be an arbitrary small number. Let $\delta > 0$ be small enough to ensure the estimate (5) as in Exercise 4.8. Choose compact subsets $K_i \subset X_i$ so that

$$E = X \setminus (K_1 \cup \cdots \cup K_a) \qquad \text{has} \qquad \mu(E) < \delta$$

and set

$$U_i = K_i \cup E, \qquad \mathcal{U} = \{U_1, \dots, U_a\}.$$

By the ergodic theorem (using only convergence in measure that follows from the Mean Ergodic Theorem 2.9) one knows that for all large enough n the set $A_n \subset X$ of points $x \in X$ with

$$\#\{1 \le i \le n \mid T^i x \in E\} < \lceil \delta n \rceil \qquad \text{has} \qquad \mu(A_n) > \frac{1}{2}.$$

At the same time, for all large enough n one has

$$\frac{1}{n} \log N(\mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-(n-1)}\mathcal{U}) < H + \epsilon.$$

The latter statement means that there is a set $W_n \subset A^n$ of words of size

$$|W_n| < e^{(H+\epsilon)n}$$

with the property that for every $x \in X$, there is $w \in W_n$ with

$$T^i x \in U_{w_i} = K_{w_i} \sqcup E \subset X_{w_i} \cup E \qquad (i \in \{1, \dots, n\}).$$

For $y = p(x) \in \Sigma = \{1, \dots, a\}^{\mathbb{N}}$, the word $v = [y]_n$ is determined by the condition

$$T^i x \in X_{v_i} \qquad (i \in \{1, \dots, n\}).$$

Therefore, for $x \in A_n$ the words $w \in W_n$ and $v = [p(x)]_n$ may differ at no more than $\lceil \delta n \rceil$ places $i \in \{1, \dots, n\}$.

Let $\Omega_n \subset A_n$ be the set of words, obtained from all $w \in W_n$ by all possible alterations of $\lceil \delta n \rceil$ places. Using (5) we estimate

$$|\Omega_n| \le \binom{n}{\lceil \delta n \rceil} \cdot a^{\lceil \delta n \rceil} \cdot |W_n| < a^{\epsilon n} \cdot e^{(H+\epsilon)n} < e^{(H+\epsilon \cdot (1+\log a))n}$$

while

$$\nu\{y \in \Sigma \mid [y]_n \in \Omega_n\} \ge \mu(A_n) > \frac{1}{2}.$$

By Corollary 4.7 it follows that

$$\mathrm{ent}(\Sigma, \nu, S) \le \mathrm{Ent}^{\mathrm{top}}(X, T) + \epsilon'$$

for arbitrarily small $\epsilon' > 0$. This implies the result. □

Proposition 4.19.
For $\Gamma < G = \mathrm{SL}_2(\mathbb{R})$ *cocompact lattice,* $\quad \mathrm{Ent}^{\mathrm{top}}(G/\Gamma, T_{a^t}) \le t$.

Proof. Let $\mathcal{U}$ be an open cover of $X = G/\Gamma$. Recall the family of diffeomorphisms $\phi_x : (-r, r)^3 \to W_x \subset X$ defined by (6). Using the "Lebesgue number" argument, there exists $\delta \in (0, r)$ so that each ϕ_x-image V_x of $(-\delta, \delta)^3$ is entirely contained in some set U_x from the given open cover $\mathcal{U}$. We have

$$x \in V_x = \phi_x((-\delta, \delta)^3) \subset U_x \in \mathcal{U}.$$

This system of "standard" open sets $\{V_x\}_{x \in X}$ forms an open cover of the compact space X. hence there is a finite set $Z \subset X$ of points so that

$$X = \bigcup_{z \in Z} V_z.$$

For $n \in \mathbb{N}$ let $J_n = \lceil e^{nt} \rceil$ and for $z \in Z$ and $j \in \{-J_n, \dots, J_n\}$ denote

$$z_j^{(n)} = \phi_z(0, 0, j e^{-nt}).$$

We shall use these points to index a certain subcover

$$\mathcal{W} \subset \mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-(n-1)}\mathcal{U}.$$

This will show that

$$h_n(\mathcal{U}) \le |\mathcal{W}| = |Z| \cdot (2J_n + 1)$$

and consequently

$$\lim_{n \to \infty} \frac{1}{n} h_n(\mathcal{U}) \le t$$

as claimed. It remains to construct the subcover

$$\mathcal{W} = \{W_{z_j^{(n)}} \mid z \in Z, \ -J_n \le j \le J_n\}.$$

For $0 \le k < n$ the element $U_{T^i z_j^{(n)}}$ of the open cover $\mathcal{U}$ contains the standard set $V_{T^i z_j^{(n)}} \ni T^i z_j^{(n)}$. Hence (7) implies that $T^{-i} U_{T^i z_j^{(n)}}$ contains the ϕ_z-image of

$$(-\delta, \delta) \times (-\delta, \delta) \times ((je^{-nt} - e^{-it})\delta, (je^{-nt} + e^{-it})\delta)$$

that contains

$$(-\delta, \delta) \times (-\delta, \delta) \times (\frac{j-1}{e^{nt}}\delta, \frac{j+1}{e^{nt}}\delta).$$

For each $z_j^{(n)}$ consider the following element of $\mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-(n-1)}\mathcal{U}$

$$W_{z_j^{(n)}} = \bigcap_{i=0}^{n-1} T^{-i} U_{T^i z_j^{(n)}}.$$

The above computation shows that

$$\bigcup_{z \in Z} \bigcup_{j=-J_n}^{J_n} W_{z_j^{(n)}} = \bigcup_{z \in Z} V_z = X$$

proving that $\mathcal{W}$ is indeed a cover. □

We can complete the proof of Theorem 4.14 by combining the lower estimate (8), Theorem 4.18, and Proposition 4.19, to deduce

$$t \le \mathrm{Ent}(G/\Gamma, m_{G/\Gamma}, T_{a^t}) \le \mathrm{ent}(G/\Gamma, T_{a^t}) \le t.$$

Similar ideas can be used to show

$$0 \le \mathrm{Ent}(G/\Gamma, m_{G/\Gamma}, T_h) \le \mathrm{ent}(G/\Gamma, T_h) \le 0.$$

□

References

[1] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London Ltd., London, 2011.

[2] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

[3] A. Karlsson and F. Ledrappier, *Noncommutative ergodic theorems*, Geometry, rigidity, and group actions, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 2011, pp. 396–418.

[4] A. Karlsson and G. A. Margulis, *A multiplicative ergodic theorem and nonpositively curved spaces*, Comm. Math. Phys. **208** (1999), no. 1, 107–123.

[5] Y. Katznelson and B. Weiss, *A simple proof of some ergodic theorems*, Israel J. Math. **42** (1982), no. 4, 291–296.

[6] D. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math. **4** (1970), 337–352 (1970).

[7] D. S. Ornstein and B. Weiss, *Geodesic flows are Bernoullian*, Israel J. Math. **14** (1973), 184–198.

[8] M. S. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, Israel J. Math. **32** (1979), no. 4, 356–362.

[9] M. Ratner, *Interactions between ergodic theory, Lie groups, and number theory*, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 157–182.