

Chapter 8

Measures of Maximal Entropy

In this chapter we will apply some of the machinery of entropy theory developed in earlier chapters to characterize Haar measures using dynamical properties. Here and in Chapter 9 we exhibit various results that force an unknown measure possessing some dynamical property to be algebraic — instances of the phenomena of *measure rigidity*. The results in these two chapters are in each case extremely special instances of far more general results, chosen either to illustrate something that will be expanded upon later or simply for simplicity.

We start by a discussion of maximal entropy measures for general systems and shift spaces in Section 8.1. After this we again turn to algebraic systems.

8.1 Maximal Measures

As mentioned before, a measure $\mu \in \mathcal{M}^T(X)$ with $h_\mu(T) = h_{\text{top}}(T)$ is called a *maximal measure*. It is clear from Theorem 2.33 that the set of maximal measures is a (possibly empty) convex subset of \mathcal{M}^T , and indeed it shares other properties with \mathcal{M}^T .

The next result records one general situation in which at least one maximal measure exists;⁽³⁷⁾ it is an immediate corollary of upper semi-continuity of entropy (Theorem 5.6).

Corollary 8.1. *If $T: (X, d) \rightarrow (X, d)$ is an expansive homeomorphism, then T has a maximal measure.*

PROOF. By Theorem 5.6, the entropy map $\mu \mapsto h_\mu(T)$ is upper semi-continuous, and an upper semi-continuous real-valued map on a compact space attains its bounds. \square

In turn, the following is an immediate corollary to Theorem 2.33.

Corollary 8.2. *Let $T: X \rightarrow X$ be a continuous map on a compact metric space satisfying $h_{\text{top}}(T) < \infty$. If T has an invariant probability measure of maximal entropy, then it has an ergodic probability measure of maximal entropy. Moreover, if there is only one ergodic invariant probability measure of maximal entropy then there is only one invariant probability measure of maximal entropy.*

PROOF. Let μ be a maximal measure; using Theorem 2.7 ([46, Th. 6.2]) write

$$\mu = \int_Y \mu_y \, d\nu(y)$$

for the ergodic decomposition of μ . By Theorem 2.33,

$$h_{\text{top}}(T) = h_\mu(T) = \int_Y h_{\mu_y}(T) \, d\nu(y).$$

Since $h_{\mu_y}(T) \leq h_{\text{top}}(T)$ for each $y \in Y$, we must have

$$h_{\mu_y}(T) = h_\mu(T) = h_{\text{top}}(T)$$

for ν -almost every y , and in particular there must be an ergodic measure of maximal entropy.

If there is only one ergodic measure μ_{erg} of maximal entropy, the same argument shows that ν -almost every y satisfies $\mu_y = \mu_{\text{erg}}$, which proves the corollary. \square

The simplest example of a unique measure of maximal entropy arises for full shifts.

Lemma 8.3. *Let $X = \prod_{-\infty}^{\infty} \{0, 1, \dots, s-1\}$ with the metric from (A.2). The shift map $\sigma: X \rightarrow X$ defined by $(\sigma(x))_k = x_{k+1}$ for all $k \in \mathbb{Z}$ is a homeomorphism of (X, d) . The map σ has a unique measure of maximal entropy, and this unique measure is the Bernoulli measure corresponding to the uniform measure $(\frac{1}{s}, \dots, \frac{1}{s})$ on the alphabet $\{0, 1, \dots, s-1\}$.*

The proof of the lemma is identical to the last part of Example 1.28.

We now give more interesting examples of unique maximal measures.

8.1.1 Uniqueness for a Shift of Finite Type

In this section we will use material from Appendix B on the Perron–Frobenius theorem to show that there is a unique maximal measure for a Markov shift. This will be a particular choice⁽³⁸⁾ of a Markov measure which is naturally associated to the adjacency matrix A_G (see Section A.4.2). Due to the maximality of entropy and the uniqueness one can argue that this choice of a

shift-invariant probability measure is dynamically canonical for the Markov shift.

Assume that $A = A_G$ is irreducible, so that by the Perron–Frobenius theorem there is a unique positive real eigenvalue $\lambda = \lambda_A$, with geometric and algebraic multiplicity one, and with an associated strictly positive left eigenvector \mathbf{u} and an associated strictly positive right eigenvector \mathbf{v} . We normalize these eigenvectors so that

$$\sum_{i=1}^{|V_G|} u_i v_i = 1. \quad (8.1)$$

Definition 8.4. The *Parry measure* associated to the shift dynamical system $(X_G^{(v)}, \sigma)$ is the Markov measure $\mu_{\mathbf{p}, P}$ as defined in Section A.4.2, where the initial probability vector \mathbf{p} is defined by

$$p_i = u_i v_i \quad (8.2)$$

for $1 \leq i \leq |V_G|$ and the transition probabilities are given by

$$P_{i,j} = \frac{a_{i,j} v_j}{\lambda v_i} \quad (8.3)$$

for $1 \leq i, j \leq |V_G|$.

It is easy to check that $\mathbf{p}P = \mathbf{p}$ so that Section A.4.2 indeed gives the construction of the measure $\mu_{\mathbf{p}, P}$. We note that by Theorem A.8 the resulting invariant probability measure is also ergodic; in fact condition (5) of that theorem is the most conveniently checked property since we know that A has λ as an eigenvalue with multiplicity one and P is conjugate to $\frac{1}{\lambda}A$.

Theorem 8.5. *The Parry measure $\mu_{\mathbf{p}, P}$ associated to the shift dynamical system $(X_G^{(v)}, \sigma)$ with $A = A_G$ an irreducible matrix is the unique measure of maximal entropy.*

PROOF. Recall from Corollary 5.15 that

$$h_{\text{top}}(\sigma) = \log \lambda_A,$$

where λ_A is the largest positive eigenvalue of the adjacency matrix $A = A_G$. We first claim that $\mu = \mu_{\mathbf{p}, P}$ is a maximal measure by computing its entropy. As seen in Example 1.29,

$$\begin{aligned}
h_\mu(\sigma) &= - \sum_{i,j=1}^{|V_G|} p_i p_{i,j} \log p_{i,j} \\
&= - \sum_{i,j=1}^{|V_G|} u_i v_i \left(\frac{a_{i,j} v_j}{\lambda v_i} \right) \log \left(\frac{a_{i,j} v_j}{\lambda v_i} \right) \quad (\text{by (8.2) and (8.3)}) \\
&= - \sum_{i,j=1}^{|V_G|} \left(\frac{a_{i,j} u_i v_j}{\lambda} \right) \left(\underbrace{\log a_{i,j}}_A + \underbrace{\log v_j}_B - \underbrace{\log \lambda}_C - \underbrace{\log v_i}_D \right).
\end{aligned}$$

First notice that the sum corresponding to A vanishes since $a_{i,j} \in \{0, 1\}$ for all i and j . Second the sums corresponding to B and D cancel since

$$\sum_{i,j=1}^{|V_G|} \frac{a_{i,j} u_i v_j}{\lambda} \log v_j = \frac{\lambda}{\lambda} \sum_j u_j v_j \log v_j$$

and a similar calculation shows that the sum for D leads to the same expression. Finally, the sum corresponding to C reduces to $\log \lambda$ since

$$\begin{aligned}
\sum_{i=1}^{|V_G|} u_i \sum_{j=1}^{|V_G|} a_{i,j} v_j &= \lambda \sum_{i=1}^{|V_G|} u_i v_i \quad (\text{since } A\mathbf{v} = \lambda\mathbf{v}) \\
&= \lambda \quad (\text{by the normalization (8.1)})
\end{aligned}$$

so $h_\mu(\sigma) = h_{\text{top}}(\sigma)$, and μ is a maximal measure.

As noted above, the Parry measure μ is ergodic by Theorem A.8. Hence, if μ is not the only invariant probability measure with maximal entropy then we may find, using Corollary 8.2, another invariant and ergodic probability measure $\nu \perp \mu$ with

$$h_\nu(\sigma) = h_\mu(\sigma) = \log \lambda.$$

It follows that there is a Borel set $B \subseteq X = \mathbf{X}_G^{(v)}$ with $\mu(B) = 0$ and $\nu(B) = 1$. In the notation from Section A.4.3, recall that the state partition $\xi = \{[i]_0\}$ is a generator for the shift map σ . Using (for example) martingale convergence we see that we may choose a sequence $(A_n)_{n \geq 1}$ of Borel sets so that A_n is a union of atoms in ξ_{-n}^n with $(\mu + \nu)(A_n \Delta B) \rightarrow 0$ as $n \rightarrow \infty$. In particular, we have $\mu(A_n) \rightarrow 0$ and $\nu(A_n) \rightarrow 1$ as $n \rightarrow \infty$.

Write $t_n(A)$ for the number of elements of ξ_{-n}^n that intersect a set $A \subseteq X$. For each $n \geq 1$, let η_n be the partition into A_n and its complement. By the estimate for the entropy of a finite partition in Propositions 1.5 (and the definition of conditional entropy), it follows that

$$H_\nu \left(\xi_{-n}^n \mid \eta_n \right) \leq \nu(A_n) \log t_n(A_n) + (1 - \nu(A_n)) \log t_n(X \setminus A_n). \quad (8.4)$$

Now for any $n \geq 1$ we have

$$\begin{aligned}
(2n+1)h_\nu(\sigma) &\leq H_\nu\left(\xi_{-n}^n\right) && \text{(by Definition 1.14)} \\
&\leq H_\nu(\eta_n) + H_\nu\left(\xi_{-n}^n \mid \eta_n\right) && \text{(by additivity in Prop. 1.7(2))} \\
&\leq \log 2 + \nu(A_n) \log t_n(A_n) + (1 - \nu(A_n)) \log t_n(X \setminus A_n). && \text{(by (8.4))}
\end{aligned}$$

By the maximality property of ν it follows that

$$(2n+1) \log \lambda \leq \log 2 + \nu(A_n) \log t_n(A_n) + (1 - \nu(A_n)) \log t_n(X \setminus A_n). \quad (8.5)$$

Now if

$$C = {}_{-n}[a_{-n} \dots a_n]_n$$

is a non-empty cylinder set (so that $a_{-n} \dots a_n$ is an allowed word in the shift space X), then by definition

$$\mu(C) = u_{a_{-n}} v_{a_{-n}} \prod_{k=-n}^{n-1} \frac{a_{a_k, a_{k+1}}}{\lambda v_{a_k}} v_{a_{k+1}} = \frac{u_{a_{-n}} v_{a_n}}{\lambda^{2n}}$$

since $a_{a_k, a_{k+1}} = 1$ as C is a non-empty cylinder set. It follows that the measure $\mu(C)$ is bounded above and below as

$$\frac{\alpha}{\lambda^{2n}} \leq \mu(C) \leq \frac{\beta}{\lambda^{2n}} \quad (8.6)$$

where $\alpha = \min_{1 \leq i, j \leq |V_G|} u_i v_j > 0$ and $\beta = \max_{1 \leq i, j \leq |V_G|} u_i v_j > 0$.

By taking unions of disjoint sets, we deduce that for any set A that is a union of atoms in ξ_{-n}^n we have

$$\frac{\alpha}{\lambda^{2n}} t_n(A) \leq \mu(A) \leq \frac{\beta}{\lambda^{2n}} t_n(A).$$

These bounds hold in particular for the set A_n and its complement $X \setminus A_n$, which can be substituted into (8.5) to conclude that

$$\begin{aligned}
(2n+1) \log \lambda &\leq \log 2 + \nu(A_n) \log \left(\frac{\mu(A_n) \lambda^{2n}}{\alpha} \right) \\
&\quad + (1 - \nu(A_n)) \log \left(\frac{(1 - \mu(A_n)) \lambda^{2n}}{\alpha} \right)
\end{aligned}$$

or equivalently that

$$\log \lambda \leq \log 2 - \log \alpha + \nu(A_n) \log \mu(A_n) + (1 - \nu(A_n)) \log(1 - \mu(A_n)).$$

Now by construction $\nu(A_n) \rightarrow 1$ and $\mu(A_n) \rightarrow 0$ as $n \rightarrow \infty$, so the right-hand side goes to $-\infty$ as $n \rightarrow \infty$. This contradiction proves the theorem. \square

Let us highlight two important features that make the above proof work: On $X = X_G^{(v)}$ we have a generator ξ independent of the measure, and an ergodic measure μ with the property that the measure of the cylinder sets decay precisely as predicted by entropy (where we allow only a deviation by multiplicative constants as in (8.6)). Generalizing this approach to other expansive maps is not too difficult (we will outline this for hyperbolic automorphisms in Section 8.2). The method also works for certain examples of maps on homogeneous spaces $X = \Gamma \backslash G$ as considered in Section 6.4, but as those are in many cases not expansive this requires more work (see Section 8.5).

8.1.2 An Example Without Maximal Measures

Next we record a simple example to show that there may be no maximal measures at all. The first such examples were found by Gurevič [79].

Example 8.6. For each $N \geq 2$, define a closed σ -invariant subset of the full shift $X = \{0, 1\}^{\mathbb{Z}}$ by

$$X_{(N)} = \{x \in X \mid \text{the block } 0^N \text{ does not appear in } x\}$$

and write $\sigma_{(N)}$ for the shift restricted to $X_{(N)}$. For any $k \geq 1$ there are no more than $(2^N - 1)^k$ blocks of length Nk , because there are $(2^N - 1)$ allowed blocks of length N (but not all of them can be concatenated to make an allowed block). On the other hand, if w_1, \dots, w_k are arbitrary blocks of length $(N-1)$ then $w_1 1 w_2 1 \cdots 1 w_k 1$ is an allowed block of length Nk , so there are at least $(2^{N-1})^k$ blocks of length Nk . By Corollary 5.14 and Lemma 5.21 we deduce that

$$\left(1 - \frac{1}{N}\right) \log 2 \leq h_{\text{top}}(\sigma_{(N)}) \leq \frac{1}{N} \log(2^N - 1),$$

and, in particular, $h_{\text{top}}(\sigma_{(N)}) < \log 2$ while $h_{\text{top}}(\sigma_{(N)}) \rightarrow \log 2$ as $N \rightarrow \infty$.

Choose a metric d_N on each $X_{(N)}$ compatible with its compact topology and with $\text{diam}_{d_N}(X_{(N)}) = 1$. Define a new space X_* to be the disjoint union

$$X_* = \bigsqcup_{N \geq 2} X_{(N)} \sqcup \{\infty\}$$

of all the sets $X_{(N)}$ with an additional point at infinity. We can make X_* into a compact metric space by, for example, defining the metric d on X_* so that it satisfies

$$d(x, y) = \begin{cases} \frac{1}{N^2} d_N(x, y) & \text{if } x, y \in X_{(N)}; \\ \sum_{j=M}^N \frac{1}{j^2} & \text{if } x \in X_{(M)}, y \in X_{(N)} \text{ and } M < N; \\ \sum_{j=M}^{\infty} \frac{1}{j^2} & \text{if } x \in X_{(M)} \text{ and } y = \infty. \end{cases}$$

Note that this implies that every $X_{(N)}$ is a clopen subset of X_* , and moreover that

$$\sigma_*(x) = \begin{cases} \sigma_{(N)}(x) & \text{if } x \in X_{(N)}; \\ \infty & \text{if } x = \infty \end{cases}$$

defines a homeomorphism σ_* of X_* . Now let μ be any σ_* -invariant probability measure on X_* and allow N to denote a member of $\mathbb{N} \cup \{\infty\}$.

Since X_* is a disjoint union of countably many invariant subsets, we may write

$$\mu = \sum_{N \leq \infty} p_N \mu_N,$$

where $p_N \in [0, 1]$, $\sum_{N \leq \infty} p_N = 1$, $\mu_N \in \mathcal{M}^{\sigma_{(N)}}(X_{(N)})$ is the normalized restriction of μ to $X_{(N)}$ for $N \geq 2$, and $\mu_\infty = \delta_\infty$.

By Theorem 2.33 we have

$$h_\mu(\sigma_*) = \sum_{N \leq \infty} p_N h_{\mu_N}(\sigma_*) \quad (8.7)$$

since $h_{\mu_\infty}(\sigma_*) = 0$. Moreover

$$h_{\mu_n}(\sigma_*) \leq h_{\text{top}}(\sigma_{(N)}) < \log 2 \quad (8.8)$$

for any $N \geq 1$. As this holds for any $\mu \in \mathcal{M}^{\sigma_*}(X_*)$ and X_* contains $X_{(N)}$ for all $N \geq 1$ we see that

$$h_{\text{top}}(\sigma_*) = \log 2$$

by Theorem 5.24. However, this, Equations (8.7) and (8.8) together show that $h_\mu(\sigma_*) < h_{\text{top}}(\sigma_*)$ for any $\mu \in \mathcal{M}^{\sigma_*}(X_*)$. Thus (X_*, σ_*) has no measure of maximal entropy.

Exercises for Section 8.1

Exercise 8.1.1. Let $T : X \rightarrow X$ be a continuous map on a compact metric space. Show that if T has a unique maximal measure, then it is necessarily ergodic.

8.2 Uniqueness for Hyperbolic Automorphisms

In this section we generalize the argument from Theorem 8.5 for certain toral automorphisms.⁽³⁹⁾ We make strong simplifying assumptions about the map to make clear the relationship between the geometry of the map and the uniqueness property for the measure of maximal entropy and skip some of the details as we will strengthen the theorem in two ways later; in Section 8.4 we discuss more general torus automorphisms (dropping the diagonalizability and hyperbolicity assumption) and in Section 8.5 we relax the conditions of hyperbolicity and generalize the argument to homogeneous spaces.

Throughout this section, we let $X = \mathbb{T}^r$ and assume that $T_A: X \rightarrow X$ is the automorphism defined by a linear map $A: \mathbb{R}^r \rightarrow \mathbb{R}^r$ which is hyperbolic and diagonalizable over \mathbb{C} . Recall that A is said to be hyperbolic if $|\lambda| \neq 1$ for each eigenvalue λ of A ; in this case we also call the map T_A hyperbolic. It is easy to see that in this case T_A is an expansive homeomorphism of X .

Theorem 8.7. *Let $T_A: \mathbb{T}^r \rightarrow \mathbb{T}^r$ be a toral automorphism defined by a hyperbolic diagonalizable linear map $A: \mathbb{R}^r \rightarrow \mathbb{R}^r$. Then the Lebesgue measure $m = m_{\mathbb{T}^r}$ on \mathbb{T}^r is the unique measure of maximal entropy.*

Recall from [46, Cor. 2.20] (which uses Fourier analysis) or show geometrically using Hopf's argument (see [46, Ch. 9]) that m is ergodic with respect to T_A .

Also recall that

$$h := h_m(T_A) = h_{\text{top}}(T_A)$$

by Theorem 6.9, so m is a maximal measure. By choosing a small enough positive ε we can ensure that the two-sided Bowen ball

$$D^\pm(\pi(x), n, 2\varepsilon) = \bigcap_{k=-n}^n T^{-k}(B_{2\varepsilon}(T_A^k x))$$

is the image of the corresponding Bowen ball in \mathbb{R}^r . This follows from Theorem 6.5 (applied to T_A and T_A^{-1}). By the same calculation as indicated on page 172, we know by the assumption that A is diagonalizable[†] that

$$\delta^r c_1 e^{-2hn} \leq m(D^\pm(x, n, 2\delta)) \leq \delta^r c_2 e^{-2hn} \quad (8.9)$$

for all $n \geq 1$, for any $\delta \in (0, 2\varepsilon)$ and some constants $c_1, c_2 > 0$ (depending on the choice of the norm used to define d and the map T_A).

These are versions of two of the properties highlighted at the end of Section 8.1.1 on page 204 that help to prove that there cannot be another measure of maximal entropy. We note, however, that on the torus we do not have a generator with properties as good as those of the state partition for the

[†] In the general case, the presence of non-trivial Jordan blocks distorts the estimate in (8.9) by a polynomial factor that destroys the contradiction step later in the argument.

shift spaces. The following will be a geometric replacement of the martingale convergence theorem in the proof of Theorem 8.5; we leave the proof as an exercise (and refer to the more general Lemma 8.21).

Lemma 8.8. *Let (X, d) be a compact metric space, and let ν be a Borel probability measure on X . If $B \subseteq X$ is measurable and (ξ_n) is a sequence of finite partitions for which $\max_{P \in \xi_n} \text{diam}(P) \rightarrow 0$ as $n \rightarrow \infty$, then there exists a sequence $A_n \in \sigma(\xi_n)$ with $\nu(A_n \triangle B) \rightarrow 0$ as $n \rightarrow \infty$.*

PROOF OF THEOREM 8.7. For the construction of a replacement of the state partition (resp. its refinements) we let $F_n \subseteq X$ be a maximal set for which the sets $D^\pm(x, n, \varepsilon)$ are pairwise disjoint for $x \in F_n$, and write $s_n = |F_n|$. Then

$$\bigcup_{x \in F_n} D^\pm(x, n, 2\varepsilon) = X,$$

for if not then F_n would not be maximal. Choosing some ordering

$$F_n = \{x_1, \dots, x_{s_n}\}$$

of F_n , we can inductively construct a partition $\xi_n = \{P_1, \dots, P_{s_n}\}$ by defining

$$\begin{aligned} P_1 &= D^\pm(x_1, n, 2\varepsilon) \setminus \left(\bigcup_{j=2}^{s_n} D^\pm(x_j, n, \varepsilon) \right), \\ P_2 &= D^\pm(x_2, n, 2\varepsilon) \setminus \left(P_1 \cup \bigcup_{j=3}^{s_n} D^\pm(x_j, n, \varepsilon) \right), \\ P_3 &= D^\pm(x_3, n, 2\varepsilon) \setminus \left(\bigcup_{j=1}^2 P_j \cup \bigcup_{j=4}^{s_n} D^\pm(x_j, n, \varepsilon) \right), \\ &\vdots \\ P_{s_n} &= D^\pm(x_{s_n}, n, 2\varepsilon) \setminus \left(\bigcup_{j=1}^{s_n-1} P_j \right). \end{aligned}$$

By construction, $D^\pm(x_i, n, \varepsilon) \subseteq P_i \subseteq D^\pm(x_i, n, 2\varepsilon)$ for $i = 1, \dots, s_n$. By the inequalities (8.9) and disjointness of the sets $D^\pm(x_i, n, \varepsilon)$ for $i = 1, \dots, s_n$, we know that

$$s_n \leq c_3^{-1} e^{2hn} \quad (8.10)$$

for some $c_3 > 0$ and

$$m(P_i) \geq c_4 e^{-2hn} \quad (8.11)$$

for some $c_4 > 0$.

Since A is hyperbolic, we have $\text{diam}(D_A^\pm(x, n, 2\varepsilon)) \rightarrow 0$ as $n \rightarrow \infty$ and so we may apply Lemma 8.8 to the sequence of partitions ξ_n later on. Next

notice that every element of ξ_n is contained in some Bowen ball

$$D_{T_A}^{\pm}(x, n, 2\varepsilon),$$

and that we may assume that 2ε is an expansiveness constant for T_A . This implies that ξ_n is a generator for T_A^{2n} , and so we know that

$$h_{\mu}(T_A) = \frac{1}{2n} h_{\mu}(T_A^{2n}, \xi_n) \leq \frac{1}{2n} H_{\mu}(\xi_n)$$

for any invariant measure μ .

Assume now that $\mu \in \mathcal{M}^{T_A}(\mathbb{T}^r)$ also satisfies $h_{\mu}(T_A) = h_{\text{top}}(T_A)$ but is different from m . Using ergodicity of m and Corollary 8.2 we may assume that $\mu \perp m$. Hence there exists a measurable T_A -invariant set $B \subseteq X$ with $\mu(B) = 1$ and $m(B) = 0$. Applying Lemma 8.8 to the set B and the measure $\nu = \frac{1}{2}(m + \mu)$, we find a sequence of sets $A_n \in \sigma(\xi_n)$ with $\nu(A_n \Delta B) \rightarrow 0$, which implies that

$$m(A_n) \rightarrow 0 \tag{8.12}$$

and

$$\mu(A_n) \rightarrow 1 \tag{8.13}$$

as $n \rightarrow \infty$. Let M_n be the number of elements of ξ_n that are needed to obtain the set A_n . Then by the inequality (8.11) we can rewrite (8.12) as

$$M_n e^{-2nh} \rightarrow 0 \tag{8.14}$$

as $n \rightarrow \infty$.

Now let $\eta_n = \{A_n, X \setminus A_n\} \subseteq \sigma(\xi_n)$, which gives

$$\begin{aligned} H_{\mu}(\xi_n) &= H_{\mu}(\eta_n) + H_{\mu}(\xi_n | \eta_n) \\ &\leq \log 2 + \mu(A_n) H_{\mu_1}(\xi_n) + (1 - \mu(A_n)) H_{\mu_2}(\xi_n) \end{aligned}$$

by (1.1), where we write μ_1 for $\frac{1}{\mu(A_n)}\mu|_{A_n}$ and μ_2 for $\frac{1}{\mu(X \setminus A_n)}\mu|_{X \setminus A_n}$. As the partition ξ_n can be considered as having M_n elements modulo μ_1 and ξ_n has s_n elements, we get

$$2nh \leq \log 2 + \mu(A_n) \log M_n + \mu(X \setminus A_n) \log s_n$$

or by (8.13), (8.14) and (8.10),

$$0 \leq \log 2 + \underbrace{\mu(A_n)}_{\rightarrow 1} \underbrace{\log M_n e^{-2nh}}_{\rightarrow -\infty} + \underbrace{\mu(X \setminus A_n)}_{\rightarrow 0} \underbrace{\log s_n e^{-2nh}}_{\leq -\log c_3}.$$

This contradiction proves the theorem. \square

Exercises for Section 8.2

Exercise 8.2.1. Extend Theorem 8.7 to toral endomorphisms (but still with the assumption of diagonalizability and a condition on the eigenvalues) as follows.

- (a) Prove the appropriate analogue of Theorem 8.7 for the endomorphism of \mathbb{T}^r defined by the matrix $pI_r \in \text{Mat}_{r,r}(\mathbb{Z})$ for some $p \geq 2$.
 (b) Prove Theorem 8.7 for the endomorphism of \mathbb{T}^3 defined by the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 2 \end{pmatrix}.$$

- (c) Generalize Theorem 8.7 to any diagonalizable matrix $A \in \text{Mat}_{r,r}(\mathbb{Z})$ with $\det(A) \neq 0$ and with $|\lambda| \neq 1$ for every eigenvalue λ of A .

8.3 Completely Positive Entropy for Toral Automorphisms

As an application of the results of Chapter 2 and preparation for the next section we prove the following special case of a result due to Rokhlin [180].

Theorem 8.9. *Let $A \in \text{GL}_r(\mathbb{Z})$ be a matrix with the property that the induced map $T_A : \mathbb{T}^r \rightarrow \mathbb{T}^r$ is ergodic with respect to the Lebesgue measure $m = m_{\mathbb{T}^r}$. Then T_A acts with completely positive entropy with respect to the Lebesgue measure m .*

That is, by the characterization mentioned after Definition 2.25, we will show that the Pinsker algebra $\mathcal{P}(T_A)$ of T_A with respect to m is trivial.

For the proof we will need the following elementary observation.

Lemma 8.10. *The set*

$$\{x \in \mathbb{T}^r \mid \text{there exists some } n \geq 1 \text{ with } T_A^n x = x\}$$

of periodic points for a toral automorphism T_A is dense in \mathbb{T}^r .

PROOF. Clearly every point $x \in \mathbb{Q}^r/\mathbb{Z}^r$ has a finite orbit. Since T_A is an automorphism, it follows that any such point is a periodic point. \square

PROOF OF THEOREM 8.9. By Exercise 2.4.2, we know that $\mathcal{P}(T_A) = \mathcal{P}(T_A^m)$ for any $m \geq 1$. Let $x \in \mathbb{T}^r$ be a periodic point, fixed by T_A^m for some $m \geq 1$, and let B be a set in $\mathcal{P}(T_A)$. Then $B \in \mathcal{P}(T_A^m)$, and the partition

$$\bigvee_{i=0}^{m-1} T_A^{-im} \{(B+x), X \setminus (B+x)\}$$

coincides with the partition

$$\bigvee_{i=0}^{n-1} T_A^{-im} \{B, X \setminus B\}$$

translated by x . As the shift preserves the measure, we obtain from the definition of the Pinsker σ -algebra that

$$B + x \in \mathcal{P}(T_A^m) = \mathcal{P}(T_A).$$

Thus, $B \in \mathcal{P}(T_A)$ and x periodic implies that $B + x \in \mathcal{P}(T_A)$. As the set of periodic points is dense by Lemma 8.10, we may choose for any $y \in \mathbb{T}^r$ some sequence of periodic points $x_n \rightarrow y$. Then $B + x_n$ converges to $B + y$ (in L_m^1 after identifying a set with its characteristic function), which implies that $B + y \in \mathcal{P}(T_A)$. (Notice that this step only uses the fact that $\mathcal{P}(T_A)$ is a σ -algebra, so that its L^1 space is complete and hence closed.)

We will need the following lemma.

Lemma 8.11. *Assume that $\mathcal{C} \subseteq \mathcal{B}(\mathbb{T}^r)$ is a σ -algebra for which $C \in \mathcal{C}$ and $x \in \mathbb{T}^r$ implies that $C + x \in \mathcal{C}$. Then there is a closed subgroup $H \leq \mathbb{T}^r$ for which*

$$\mathcal{C} = \{B \in \mathcal{B}(\mathbb{T}^r) \mid B + x = B \text{ for all } x \in H\};$$

that is, \mathcal{C} is (modulo m) the pull-back of the Borel σ -algebra of \mathbb{T}^r/H .

PROOF. Write $L^1(\mathbb{T}^r, \mathcal{C})$ for the space of functions in $L^1(\mathbb{T}^r)$ that are measurable with respect to \mathcal{C} . By assumption, $f \in L^1(\mathbb{T}^r, \mathcal{C})$ implies that the map $t \mapsto f(t - x)$ lies in $L^1(\mathbb{T}^r, \mathcal{C})$ for any $x \in \mathbb{T}^r$. Let $\chi \in C(\mathbb{T}^r)$. We claim that

$$\chi * f(t) = \int \chi(x) f(t - x) dx$$

is the L^1 -limit of the corresponding Riemann sums

$$\frac{1}{K^r} \sum_{Kx \equiv 0} \chi(x) f(t - x).$$

This can be seen from the fact that the map sending x to the map in $L^1(\mathbb{T}^r)$ defined by

$$t \mapsto \chi(x) f(t - x)$$

is continuous. As the Riemann sums belong to $L^1(\mathbb{T}^r, \mathcal{C})$, so will the limit $\chi * f$ for any $\chi \in C(\mathbb{T}^r)$. As $C(\mathbb{T}^r)$ contains approximate identities with respect to convolution and $\chi * f$ is continuous, we also deduce that

$$C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$$

is dense with respect to the L^1 norm in $L^1(\mathbb{T}^r, \mathcal{C})$.

We define

$$H = \{x \in \mathbb{T}^r \mid f(x) = f(0) \text{ for all } f \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})\}.$$

By definition, H is a closed subset of \mathbb{T}^r . Suppose that $x, y \in H$ and let $f \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$. Then $f_y(z) = f(z - y)$ for $z \in \mathbb{T}^r$ defines another function $f_y \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ and so

$$f(x - y) = f_y(x) = f_y(0) = f_y(y) = f(0).$$

As this holds for any $f \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ we have $x - y \in H$, and so H is a closed subgroup of \mathbb{T}^r .

Suppose now that $A \in \mathcal{C}$. Then by density of $C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ in $L^1(\mathbb{T}^r, \mathcal{C})$ we can find some $f_n \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ with $\mathbb{1}_A = \lim_{n \rightarrow \infty} f_n$ in L^1 . This shows that

$$(\mathbb{1}_A)_x = \lim_{n \rightarrow \infty} (f_n)_x = \lim_{n \rightarrow \infty} f_n = \mathbb{1}_A$$

in L^1 for all $x \in H$. Integrating over $x \in H$ with respect to Haar measure m_H we see that the everywhere-defined measurable function

$$f = \int (\mathbb{1}_A)_x \, dm_H(x) \in \mathcal{L}^\infty$$

coincides in L^1 with $\mathbb{1}_A$. Define

$$B = \{y \in \mathbb{T}^r \mid f(y) = 1\}$$

and notice that $m(A \Delta B) = 0$. By the definition of f , we have $B + x = B$ for all $x \in H$. This shows that

$$\mathcal{C} \subseteq \underset{m}{\{B \in \mathcal{B}(\mathbb{T}^r) \mid B + x = B \text{ for all } x \in H\}}.$$

To see the opposite inclusion, we consider the quotient group $Y = \mathbb{T}^r/H$ and let $\pi : \mathbb{T}^r \rightarrow Y$ be the canonical quotient map. We want to show that the composition $F \circ \pi$ lies in $L^1(\mathcal{C})$ for any $F \in \mathcal{L}^1(Y)$. By approximation, it is sufficient to show this in the case $F \in C(Y)$.

Notice that any $f \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ satisfies $f(y + x) = f(y)$ for any $x \in H$ and $y \in \mathbb{T}^r$. Hence $f = F \circ \pi$ for some $F \in C(Y)$. The set of all such functions $F \in C(Y)$ is an algebra under pointwise multiplication that contains the constants. It also separates points since $y_1 + H \neq y_2 + H$ implies that there exists some $f \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ with $f(y_2 - y_1) \neq f(0)$ or $f_{y_1}(y_2) \neq f_{y_1}(y_1)$. By the Stone–Weierstrass theorem, it follows that every function $F \in C(Y)$ can be written as a uniform limit of functions $F_n \in C(Y)$ for which

$$f_n = F_n \circ \pi \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C}).$$

This shows that $F \circ \pi \in C(\mathbb{T}^r) \cap L^1(\mathbb{T}^r, \mathcal{C})$ as required. \square

Returning to the proof of Theorem 8.9, we obtain from $\mathcal{P}(T_A)$ some closed subgroup $H \leq \mathbb{T}^r$ with the property that

$$\mathcal{P}(T_A) = \{B \subseteq \mathbb{T}^r \mid B + x = B \text{ for all } x \in H\}.$$

Since $B \in \mathcal{P}(T_A)$ if and only if $T_A B \in \mathcal{P}(T_A)$, we see that $T_A(H) = H$. Write as before $Y = \mathbb{T}^r/H$, so that (as follows from Pontryagin duality or from Lie theory) $Y \cong \mathbb{T}^s$ for some $s \leq r$, and T_A induces an automorphism on Y , which can also be realized as the automorphism $T_{A'}$ for some matrix $A' \in \text{GL}_s(\mathbb{Z})$. In other words, the argument above shows that the Pinsker σ -algebra corresponds to the factor $(Y, T_{A'})$ of the original system (\mathbb{T}^r, T_A) . If $s = 0$, then $\mathcal{P}(T_A) = \{\mathbb{T}^r, \emptyset\}$ and the theorem follows.

So assume that $s > 0$. Then we know by Theorem 6.9 and the definition of the Pinsker factor (Proposition 2.23) that

$$0 = h_{m_{\mathbb{T}^s}}(T_{A'}) = \sum_{|\lambda| > 1} \log |\lambda|,$$

the sum over all eigenvalues (with multiplicities) of A' of norm bigger than one. Since $\det(A') \in \{\pm 1\}$ it follows that all eigenvalues of A' satisfy $|\lambda| = 1$. By Kronecker's lemma (see Exercise 8.3.1) this implies that the eigenvalues of A' are roots of unity. However, this shows that $T_{A'}$ is not ergodic by [46, Cor. 2.20]. On the other hand, we assumed that T_A is ergodic and so this gives a contradiction, since a factor of an ergodic system is ergodic. \square

8.3.1 The Pinsker Algebra for Toral Automorphisms

Theorem 8.9 generalizes to any invariant Borel probability measure in the following way. Let $T_A : \mathbb{T}^r \rightarrow \mathbb{T}^r$ be a toral automorphism associated to the matrix $A \in \text{GL}_r(\mathbb{Z})$, and let $\mu \in \mathcal{M}^{T_A}(\mathbb{T}^r)$ be an invariant Borel probability measure. Then the Pinsker σ -algebra of T_A with respect to μ coincides modulo μ with the measurable hull $\mathcal{B}(V^-)$ of the *stable foliation* of T_A . Here the stable foliation is the equivalence relation whose equivalence classes comprise the stable leaves

$$x + V^-$$

for $x \in \mathbb{T}^r$, where the subspace V^- is the *stable subgroup*, that is the sum of all the generalized eigenspaces with eigenvalues of absolute value less than one, and the measurable hull consists of all sets that are (up to null sets) unions of stable leaves. Thus the claim about the Pinsker algebra is

$$\mathcal{P}_\mu(T_A) = \mathcal{B}_\mu(V^-) = \left\{ B \in \mathcal{B}(\mathbb{T}^r) \mid B + V^- \stackrel{\mu}{=} B \right\}. \quad (8.15)$$

We refer to Exercise 8.3.3-8.3.5 for the details.

Exercises for Section 8.3

Exercise 8.3.1. Prove Kronecker's lemma in the following form. Let

$$f(x) = \prod_{j=1}^d (x - \lambda_j) \in \mathbb{Z}[x]$$

be a monic polynomial with integer coefficients. Show that if $|\lambda_j| \leq 1$ for $1 \leq j \leq d$ then there is some $k \geq 1$ for which $\lambda_j^k = 1$ for $1 \leq j \leq d$.

Exercise 8.3.2. Let μ be a finite measure on a metric space (X, d) . Show that for any point $x \in X$ there exist arbitrarily small $r > 0$ and a constant C such that

$$\mu(\partial_\varepsilon B_r(x)) < C\varepsilon,$$

where for any $B \subseteq X$ we define $\partial_\varepsilon B = \{y \in X \mid B_\varepsilon(x) \not\subseteq B \text{ and } B_\varepsilon(x) \not\subseteq X \setminus B\}$. Show that if X is compact and $r > 0$, then there exists a finite partition ξ and $C > 0$ with

$$\max_{P \in \xi} \text{diam}(P) < r$$

and $\mu(\partial_\varepsilon) < C\varepsilon$ for all $\varepsilon > 0$.

Exercise 8.3.3. Prove (8.15) for a hyperbolic toral automorphism $T_A : \mathbb{T}^r \rightarrow \mathbb{T}^r$ and invariant Borel probability measure μ on \mathbb{T}^r .

Exercise 8.3.4. Prove (8.15) for any toral automorphism.

Exercise 8.3.5. Let $X = \Gamma \backslash G$ be a compact quotient of a Lie group by a lattice. Generalize and prove (8.15) for $T(x) = xa^{-1}$ with $a \in G$ and $x \in X$.

8.4 Uniqueness for Toral Automorphisms

In this section we will use a method of Berg [12] to remove the hypotheses of hyperbolicity and diagonalizability from Theorem 8.7. Further development of this approach gives Berg's theorem, that Haar measure is the unique measure of maximal entropy for any ergodic automorphism of finite entropy. Since we are more interested in the interaction between the geometry of a map and its entropy properties we do not develop these methods beyond the toral case.

Theorem 8.12. *Let $A \in \text{GL}_r(\mathbb{Z})$, and suppose that the induced automorphism $T_A : \mathbb{T}^r \rightarrow \mathbb{T}^r$ is ergodic with respect to $m = m_{\mathbb{T}^r}$. Then m is the unique T_A -invariant Borel probability measure with maximal entropy.*

In the proof we will use convolution of measures, which we recall here. Given measures $\mu, \nu \in \mathcal{M}(\mathbb{T}^r)$ their convolution $\mu * \nu$ is the unique measure characterized by the property that

$$\int f \, d(\mu * \nu) = \iint f(x+y) \, d\mu(x) \, d\nu(y) \quad (8.16)$$

for all $f \in \mathcal{L}^\infty(\mathbb{T}^r)$. In other words, if we define a map

$$\begin{aligned} \pi_c: \mathbb{T}^r \times \mathbb{T}^r &\longrightarrow \mathbb{T}^r \\ (x, y) &\longmapsto x + y \end{aligned}$$

then $\mu * \nu = (\pi_c)_*(\mu \times \nu)$.

Also write

$$\pi_i: \mathbb{T}^r \times \mathbb{T}^r \longmapsto \mathbb{T}^r$$

for $i = 1, 2$ for the coordinate projections.

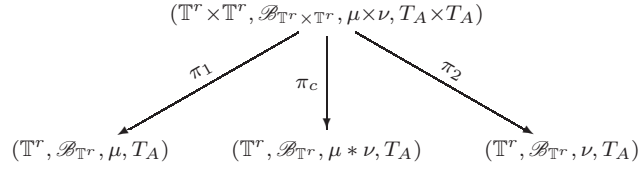


Fig. 8.1: Three factors used in the proof of Theorem 8.12.

Now let μ and ν be T_A -invariant probability measures on \mathbb{T}^r . We consider the measure-preserving system

$$(\mathbb{T}^r \times \mathbb{T}^r, \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}, \mu \times \nu, T_A \times T_A).$$

The maps π_c, π_1, π_2 are all factor maps from this system, as illustrated in Figure 8.1. In particular, $\mu * \nu$ is T_A -invariant if μ and ν are T_A -invariant.

We define the σ -algebras

$$\mathcal{B}_i = \pi_i^{-1} \mathcal{B}_{\mathbb{T}^r} \subseteq \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}$$

for $i = 1, 2$ and

$$\mathcal{B}_c = \pi_c^{-1} \mathcal{B}_{\mathbb{T}^r} \subseteq \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}.$$

It is clear that

$$\mathcal{B}_1 \vee \mathcal{B}_2 = \mathcal{B}_1 \vee \mathcal{B}_c = \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}.$$

Lemma 8.13. *Let $\mu, \nu \in \mathcal{M}^{T_A}(\mathbb{T}^r)$. Then*

$$h_{\mu * \nu}(T_A) \leq h_\mu(T_A) + h_\nu(T_A).$$

PROOF. We know that $h_{\mu \times \nu}(T_A \times T_A) = h_\mu(T_A) + h_\nu(T_A)$ (see Exercise 1.3.2) and $(\mathbb{T}^r, \mu * \nu, T_A)$ is a factor system, so this follows from Theorem 1.20. \square

Proposition 8.14. *Suppose that $\mu, \nu \in \mathcal{M}^{T_A}$. Then*

$$h_\nu(T_A) \leq h_{\mu*\nu}(T_A).$$

PROOF. We have

$$h_\mu(T_A) + h_\nu(T_A) = h_{\mu \times \nu}(T_A \times T_A), \quad (8.17)$$

and

$$h_{\mu \times \nu}(T_A \times T_A) = h_{\mu*\nu}(T_A) + h_{\mu \times \nu}(T_A \times T_A | \mathcal{B}_c)$$

by the Abramov–Rokhlin formula (Corollary 2.21). By definition and the continuity bound of entropy in Proposition 1.17(3) we have

$$h_{\mu \times \nu}(T_A \times T_A | \mathcal{B}_c) = \sup_{\xi, \eta} h_{\mu \times \nu}(T_A \times T_A, \xi \times \eta | \mathcal{B}_c),$$

where the supremum is taken over all finite partitions ξ, η of \mathbb{T}^r . Now

$$\begin{aligned} h_{\mu \times \nu}(T_A \times T_A, \xi \times \eta | \mathcal{B}_c) &= h_{\mu \times \nu}(T_A \times T_A, \xi \times \{\mathbb{T}^r\} | \mathcal{B}_c) \\ &\quad + h_{\mu \times \nu}(T_A \times T_A, \{\mathbb{T}^r\} \times \eta | \xi \times \{\mathbb{T}^r\} \vee \mathcal{B}_c). \end{aligned}$$

Choosing a sequence of partitions (ξ_n) with $\xi_n \nearrow \mathcal{B}_{\mathbb{T}^r}$ as $n \rightarrow \infty$ first, and then a sequence of partitions (η_m) with $\eta_m \nearrow \mathcal{B}_{\mathbb{T}^r}$ as $m \rightarrow \infty$ second, we find (just as in the proof of the Abramov–Rokhlin formula in Corollary 2.21) that

$$\begin{aligned} h_{\mu \times \nu}(T_A \times T_A | \mathcal{B}_c) &= \lim_{n \rightarrow \infty} h_{\mu \times \nu}(T_A \times T_A, \xi_n \times \{\mathbb{T}^r\} | \mathcal{B}_c) \\ &\quad + \lim_{m \rightarrow \infty} h_{\mu \times \nu}(T_A \times T_A, \{\mathbb{T}^r\} \times \eta_m | \mathcal{B}_1 \vee \mathcal{B}_c). \end{aligned}$$

The first limit is bounded above by $h_\mu(T_A)$ and the second vanishes since

$$\mathcal{B}_1 \vee \mathcal{B}_c = \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}.$$

Putting everything together, we deduce that

$$h_\mu(T_A) + h_\nu(T_A) = h_{\mu \times \nu}(T_A \times T_A) \leq h_{\mu*\nu}(T_A) + h_\mu(T_A).$$

Subtracting $h_\mu(T_A) \leq h_{\text{top}}(T_A) < \infty$ gives the proposition. \square

This implies a curious corollary, which we will not need later but record here for interest.

Corollary 8.15. *If $\mu, \nu \in \mathcal{M}^{T_A}(\mathbb{T}^r)$ and $h_\mu(T_A) = 0$, then $h_{\mu*\nu}(T_A) = h_\nu(T_A)$.*

PROOF. By Proposition 8.14,

$$h_\nu(T_A) \leq h_{\mu*\nu}(T_A) \leq h_{\mu \times \nu}(T_A \times T_A) = h_\mu(T_A) + h_\nu(T_A) = h_\nu(T_A).$$

□

Proposition 8.16. *Let (X, \mathcal{B}, μ, T) be an invertible measure-preserving system, and let α, β be countable partitions of X with finite entropy. Assume that $h_\mu(T, \alpha \vee \beta) = h_\mu(T, \alpha) + h_\mu(T, \beta)$. Then*

$$\begin{aligned} H_\mu(T^k \beta \vee T^{k+1} \beta \vee \dots \vee T^{k+n} \beta | \alpha_1^\infty) \\ = H_\mu\left(T^k \beta \vee T^{k+1} \beta \vee \dots \vee T^{k+n} \beta \middle| \bigcap_{j=1}^\infty \alpha_j^\infty\right) \end{aligned}$$

for all $k \in \mathbb{Z}$ and $n \geq 1$. In particular, if the tail σ -algebra

$$\bigcap_{j=1}^\infty \alpha_j^\infty \tag{8.18}$$

is trivial modulo μ , then the σ -algebras $\alpha_{-\infty}^\infty$ and $\beta_{-\infty}^\infty$ are independent.

PROOF. By additivity of dynamical entropy (Proposition 2.19(2)),

$$h_\mu(T, \alpha \vee \beta) = h_\mu(T, \beta) + h_\mu(T, \alpha | \beta_{-\infty}^\infty),$$

so

$$\begin{aligned} H_\mu(\alpha | \alpha_1^\infty) &= H_\mu(\alpha | \alpha_1^\infty \vee \beta_{-\infty}^\infty) \\ &= H_\mu(\alpha | \alpha_1^\infty \vee \gamma) \end{aligned} \tag{8.19}$$

for any

$$\gamma = T^k \beta \vee \dots \vee T^{k+n} \beta. \tag{8.20}$$

Also by additivity of entropy (Proposition 2.13) applied twice we have

$$\begin{aligned} H_\mu(\alpha | \alpha_1^\infty) + H_\mu(\gamma | \alpha \vee \alpha_1^\infty) &= H_\mu(\alpha \vee \gamma | \alpha_1^\infty) \\ &= H_\mu(\gamma | \alpha_1^\infty) + H_\mu(\alpha | \alpha_1^\infty \vee \gamma), \end{aligned}$$

which by (8.19) simplifies to give

$$H_\mu(\gamma | \alpha_0^\infty) = H_\mu(\gamma | \alpha_1^\infty) \tag{8.21}$$

for γ as in (8.20) and for any $k \in \mathbb{Z}$ and $n \geq 1$. Using (8.21) for $T\gamma$ and the fact that T preserves μ , we get

$$H_\mu(\gamma | \alpha_1^\infty) = H_\mu(T\gamma | T\alpha_1^\infty) = H_\mu(T\gamma | \alpha_0^\infty) = H_\mu(T\gamma | \alpha_1^\infty) = H_\mu(\gamma | \alpha_2^\infty).$$

Continuing inductively, we see that

$$H_\mu(\gamma | \alpha_1^\infty) = H_\mu(\gamma | \alpha_\ell^\infty)$$

for all $\ell \geq 1$. By continuity of entropy (see Proposition 2.12, Exercise 2.2.3 and its hint on p. 339) we obtain the first conclusion of the proposition.

Now suppose that the tail in (8.18) is the trivial σ -algebra modulo μ so that $H_\mu(\gamma) = H_\mu(\gamma|\alpha_0^\infty)$ by the first part of the proposition. Using this and (8.21) in the same way as above but for $T^{-1}\gamma$ we obtain

$$H_\mu(\gamma) = H_\mu(\gamma|\alpha_{-\ell}^\infty).$$

By Proposition 2.16 (or Exercise 1.1.3) this shows that γ and $\alpha_{-\ell}^\infty$ are independent. Since this holds for every n, k and ℓ , the proposition follows. \square

As a last preparation before proving Theorem 8.12, we need to prove the existence of a generator. This is actually just a special case of Krieger's theorem (Theorem 4.1), but as the proof here is significantly easier we give the argument.

Lemma 8.17. *Let $\mu \in \mathcal{M}^{T_A}(\mathbb{T}^r)$ be an ergodic measure. Then there exists a countable partition γ with finite entropy which generates with respect to T_A and μ .*

PROOF. If μ has atoms, then by ergodicity $\text{Supp}(\mu)$ is finite and the claim is trivial. So we may assume that μ has no atoms. Let α be a finite partition of \mathbb{T}^r with elements whose diameter is smaller than δ_A so that the last statement in Theorem 6.5 applies for the canonical projection map $\pi: \mathbb{R}^r \rightarrow \mathbb{T}^r$ (with δ_A equal to δ_T as in Theorem 6.5). Two points $x, y \in \mathbb{T}^r$ in the same atom for $\alpha_{-\infty}^\infty$ then must satisfy $y = x + v$ for some $v \in \mathbb{R}^r$ with $\|A^n v\| < \delta_A$ for all $n \in \mathbb{Z}$. This implies that the eigenvalues of A restricted to the span of $A^n v$ for $n \in \mathbb{Z}$ have absolute value one. Therefore, if $v \neq 0$ there exists some $\kappa = \kappa(x, y) \in (0, \delta_A)$ with the property that $\|A^n v\| \geq \kappa$ and hence

$$d(T_A^n x, T_A^n y) \geq \kappa$$

for all $n \in \mathbb{Z}$.

We now use the measure μ to define a countable partition β with finite entropy such that for any $\kappa > 0$ there exists some $Q \in \beta$ of positive measure with $\text{diam}(Q) < \kappa$. Such a partition can be constructed using sets $B_{r_n}(x_n)$ with $r_n < \frac{1}{n}$, points $x_n \in \text{Supp} \mu$, and $\mu(B_{r_n}(x_n)) < \frac{1}{n^2}$.

Now let $\gamma = \alpha \vee \beta$, and suppose that $x \in \mathbb{T}^r$ satisfies the ergodic theorem for the characteristic function of all elements of β . If $y \neq x$ and x are in the same atom with respect to $\gamma_{-\infty}^\infty$, then by the first part of the proof we know that

$$\kappa = \kappa(x, y) \leq d(T_A^n x, T_A^n y) \leq \delta_A$$

for all $n \in \mathbb{Z}$. However, there exists some $Q \in \beta$ of positive μ -measure and diameter less than κ , and some $n \in \mathbb{Z}$ with $T_A^n x \in Q$ by the ergodic theorem (indeed, this holds for a positive asymptotic proportion of $n \in \mathbb{Z}$). This implies that $T_A^n y \notin Q$ and contradicts the assumption that $y \neq x$ and x are

in the same atom with respect to $\gamma_{-\infty}^{\infty}$. Therefore, for μ -almost every $x \in \mathbb{T}^r$ the atom of x with respect to $\gamma_{-\infty}^{\infty}$ coincides with $\{x\}$ and so γ is a generator with the required properties. \square

PROOF OF THEOREM 8.12. Assume that $h_{\mu}(T_A) = h_{\text{top}}(T_A)$, and notice that we may assume that μ is ergodic by Corollary 8.2. By Lemma 8.17, there exists a countable partition γ with $H_m(\gamma) < \infty$ such that γ is a generator with respect to T_A and m . Let $\alpha = \pi_1^{-1}(\gamma)$ and $\beta = \pi_c^{-1}(\gamma)$. Then $\alpha \vee \beta$ generates under $T_A \times T_A$ the σ -algebra $\mathcal{B}_1 \vee \mathcal{B}_c = \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}$, and so

$$\begin{aligned} h_{m \times \mu}(T_A \times T_A, \alpha \vee \beta) &= h_{m \times \mu}(T_A \times T_A) = 2h_{\text{top}}(T_A) \\ &= h_m(T_A) + h_{m * \mu}(T_A) = h_{m \times \mu}(T_A \times T_A, \alpha) + h_{m \times \mu}(T_A \times T_A, \beta). \end{aligned}$$

Since T_A has completely positive entropy with respect to m by Theorem 8.9, the tail σ -algebra in (8.18) is trivial (by Proposition 2.26). Therefore, Proposition 8.16 shows that $\alpha_{-\infty}^{\infty} = \mathcal{B}_1$ and $\beta_{-\infty}^{\infty} = \mathcal{B}_c$ are independent. This shows that $\mu = m$ by Lemma 8.18 below. \square

Lemma 8.18. *Let $\nu \in \mathcal{M}(\mathbb{T}^r)$. If \mathcal{B}_1 and \mathcal{B}_c are independent with respect to the product measure $m \times \nu$, then $\nu = m$.*

PROOF. Let $B_1, B_c \in \mathcal{B}_{\mathbb{T}^r}$. By definition, $m \times \nu(\pi_1^{-1}(B_1)) = m(B_1)$ and, since $\dagger m * \nu = m$, we also have $m \times \nu(\pi_c^{-1}(B_c)) = m * \nu(B_c) = m(B_c)$. Thus the assumed independence gives $m \times \nu(\pi_1^{-1}(B_1) \cap \pi_c^{-1}(B_c)) = m(B_1)m(B_c)$. This also holds for $m \times m$ which implies that $m \times \nu = m \times m$ since we know that $\mathcal{B}_1 \vee \mathcal{B}_c = \mathcal{B}_{\mathbb{T}^r \times \mathbb{T}^r}$, and hence $m = \nu$. \square

Exercises for Section 8.4

Exercise 8.4.1. Suppose T_A is an ergodic automorphism of \mathbb{T}^r with respect to the Lebesgue measure m . Use weak mixing of T_A with respect to m to show that a finite partition α into sufficiently small rectangles is already a generator (so that the second part of the proof of Lemma 8.17 is not needed for m).

Exercise 8.4.2. Generalize Lemma 8.17 to homogeneous flows on compact quotients. That is, let $X = \Gamma \backslash G$ be a compact quotient of a Lie group G by a lattice Γ , fix $a \in G$ and prove the analog of Lemma 8.17 for the transformation $T(x) = xa^{-1}$ for $x \in X$.

Exercise 8.4.3. By assuming the result that an ergodic automorphism of a compact group has completely positive entropy with respect to Haar measure (a result of Yuzvinskii [217]), generalize the arguments of this section to prove that Haar measure is the unique measure of maximal entropy for an ergodic automorphism of a compact group (Berg's theorem).

\dagger This may be seen using the definition in (8.16) and translation-invariance of m , or by considering the Fourier transform of the measures and the relation $\widehat{m * \nu} = \widehat{m} \cdot \widehat{\nu}$.

8.5 Uniqueness of Maximal Measure for Compact Quotients

In this section we will generalize the uniqueness of maximal entropy measures to diagonalizable ergodic flows on compact quotients $\Gamma \backslash G$ for a lattice Γ in a Lie group G . Notice that in the non-commutative setting of $\Gamma \backslash G$ the assumption of diagonalizability is more crucial than it is for toral automorphisms. For example, the horocycle flow on the (non-compact) quotient $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R})$ has many invariant measures, and all the invariant measures have zero entropy. A more advanced remark (which we will not prove here) is that there exists a compact quotient of $\mathrm{SL}_3(\mathbb{R})$ which supports uncountably many ergodic probability measures invariant under right-multiplication by

$$u = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{SL}_3(\mathbb{R}),$$

and for this action the topological entropy vanishes and so once again all of the measures are maximal. Thus some additional assumptions on the element a of G defining the action by

$$T(x) = xa^{-1}$$

for $x \in X = \Gamma \backslash G$ are necessary. Here we will assume that $\mathrm{Ad}_a: \mathfrak{g} \rightarrow \mathfrak{g}$ is diagonalizable and that X is compact. This simplifies the proof considerably, but still permits many interesting examples to be treated. We will see more general results in this direction in a subsequent volume [45].

Theorem 8.19. *Let G be a Lie group, let $\Gamma < G$ be a uniform lattice and let $X = \Gamma \backslash G$ be the compact quotient. Let $a \in G$ be an element with the following properties.*

- $\mathrm{Ad}_a: \mathfrak{g} \rightarrow \mathfrak{g}$ is diagonalizable,[†]
- 1 is the only eigenvalue of Ad_a of absolute value one, and
- the transformation $T(x) = xa^{-1}$ on X is ergodic with respect to the normalized Haar measure m_X .

Then m_X is the unique probability measure of maximal entropy.

The proof will proceed along the same lines as the proof of uniqueness for diagonalizable hyperbolic toral automorphisms in Section 8.2.

We also emphasize again that ergodicity of m and the estimate

$$c^{-1}e^{-2nh}\delta^{\dim G} \leq m(D^\pm(x, n, \delta)) \leq ce^{-2nh}\delta^{\dim G} \quad (8.22)$$

[†] Diagonalizability is only assumed over \mathbb{C} , and over \mathbb{R} we obtain that Ad_a is block-diagonalizable with a 1×1 -block for each real eigenvalue and a 2×2 -block $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ for each complex eigenvalue $a + ib$ with $b \neq 0$.

for the Bowen balls for $x \in X$, $\delta \in (0, \delta_T]$ and $n \geq 1$ from Theorem 6.11 are the crucial ingredients for our proof. However, we note that the more general setting — and the lack of hyperbolicity in particular — makes some more preparation necessary.

Lemma 8.20. *Let $X = \Gamma \backslash G$ and let $T(x) = xa^{-1}$ be the transformation as above. There exists some $\varepsilon > 0$ with the following property for any ergodic T -invariant Borel probability measure μ on X . If ξ is a finite partition for which every element has diameter smaller than ε , then*

$$h_\mu(T) = h_\mu(T, \xi).$$

Moreover, if $n \geq 0$ and ξ_n is a partition for which

$$P \in \xi_n \implies \text{diam}(T^k P) < \varepsilon \tag{8.23}$$

for $|k| \leq n$, then

$$h_\mu(T^{2n+1}) = h_\mu(T^{2n+1}, \xi_n).$$

Notice that if $\varepsilon > 0$ is an expansiveness constant for T , then (8.23) implies that ξ_n is a generator for T^{2n+1} , and the conclusion is simply the Kolmogorov–Sinaï Theorem. However, even in simple situations like the time-one map for the geodesic flow there is no expansiveness constant.

PROOF OF LEMMA 8.20. We will argue just as in the last statement from Theorem 6.5 and the proof of Lemma 8.17.

Suppose that $r > 0$ is an injectivity radius for $G \rightarrow X = \Gamma \backslash G$ (so that the map $g \in B_r^G(I) \mapsto xg \in X$ is an isometry for all $x \in X$), and assume further that r is chosen small enough to ensure that $\log: B_r(I) \rightarrow \mathfrak{g}$ is a diffeomorphism. Let $\varepsilon = \frac{r}{L}$ where $L = \max\{\|\text{Ad}_a\|, \|\text{Ad}_a^{-1}\|\}$, and finally suppose that $\mu \in \mathcal{M}^T(X)$, and that ξ_n is a partition satisfying the implication (8.23).

We define $\mathcal{A} = \bigvee_{k \in \mathbb{Z}} T^{-k(2n+1)} \xi_n$, and assume that $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ for some $x, y \in X$. Then $x, y \in P_0$ for some $P_0 \in \xi_n$ by (8.23) for $k = 0$ and we have $y = xg$ for some $g \in G$ with $d(g, I) < \varepsilon$. If $n \geq 1$ then (8.23) can be applied with $k = 1$ to give

$$d((Tx)aga^{-1}, Tx) = d(Ty, Tx) < \varepsilon.$$

Together with

$$d(aga^{-1}, I) \leq \|\text{Ad}_a\| d(g, I) < r$$

this shows that $d(aga^{-1}, I) < \varepsilon$. Using (8.23) repeatedly in the same way gives

$$d(a^k ga^{-k}, I) < \varepsilon \tag{8.24}$$

for $k = 0, \dots, n$. This also implies for the same reason that

$$d(a^{n+1} ga^{-(n+1)}, I) < r. \tag{8.25}$$

As $T^{2n+1}x, T^{2n+1}y$ belong to the same partition element $P_1 \in \xi_n$ by assumption, we can apply (8.23) for $k = -n$ to see that

$$d(T^{n+1}y, T^{n+1}x) < \varepsilon.$$

Once more, this implies the improvement $d(a^{n+1}ga^{-(n+1)}, I) < \varepsilon$ of the inequality (8.25). Repeating now the first iteration for $k = -n + 1, \dots, n$ we obtain the inequality (8.24) for $k = 0, \dots, 2n$. At this point one can use the fact that $T^{2(2n+1)}x$ and $T^{2(2n+1)}y$ belong to the same element of ξ_n . By induction and the same argument for negative k we obtain the inequality (8.24) for all $k \in \mathbb{Z}$. This implies that $g = \exp v$ for some small $v \in \mathfrak{g}$, which can be written as a sum of eigenvectors with their eigenvalues of absolute value one. Hence there exists some $\kappa = \kappa(x, y) > 0$ for which

$$d(a^n ga^{-n}, I) \geq \kappa$$

and

$$d(T^n y, T^n x) \geq \kappa$$

for all $n \in \mathbb{Z}$.

We claim that this property of the \mathcal{A} -atoms implies the conclusion of the lemma. To see this, we again construct a partition β as in the proof of Lemma 8.17. We can ensure that for every $\rho > 0$ there exists a set $Q \in \beta$ of positive measure with $\text{diam}(Q) < \rho$. Furthermore we can, given some positive δ , construct β such that $H_\mu(\beta) < \delta$ — we only have to ensure that the extra sets of small diameter in β have sufficiently small measure to make the entropy of resulting partition arbitrarily small. The second part of the proof of Lemma 8.17 shows that $\xi_n \vee \beta$ is a generator for T^{2n+1} . In particular, by Proposition 2.19,

$$\begin{aligned} h_\mu(T^{2n+1}) &= h_\mu(T^{2n+1}, \xi_n \vee \beta) \\ &= h_\mu(T^{2n+1}, \xi_n) + h_\mu(T^{2n+1}, \beta | \mathcal{A}) \\ &\leq h_\mu(T^{2n+1}, \xi_n) + \delta \\ &\leq h_\mu(T^{2n+1}) + \delta. \end{aligned}$$

As this holds for any $\delta > 0$, we have shown the lemma (but not necessarily that ξ_n is a generator for T^{2n+1}). \square

Now let H be a locally compact metric group acting continuously on a locally compact metric space X , where we write $h \cdot x \in X$ for the action of $h \in H$ on $x \in X$. Given a compact neighbourhood $U \subseteq H$ of the identity element we define the transverse-to- U diameter of a subset $P \subseteq X$ by

$$\text{diam}_U P = \sup_{x, y \in P} d(U \cdot x, U \cdot y).$$

We now need the following generalization of Lemma 8.8.

Lemma 8.21. *Let $U \subseteq H$ and X be as above, and suppose that $B \subseteq X$ is measurable and H -invariant. If $\nu \in \mathcal{M}(X)$ and ξ_n is a sequence of finite partitions for which $\max_{P \in \xi_n} \text{diam}_U P \rightarrow 0$ as $n \rightarrow \infty$, then there exists a sequence of sets $A_n \in \sigma(\xi_n)$ with $\nu(A_n \triangle B) \rightarrow 0$ as $n \rightarrow \infty$.*

PROOF. Fix $\varepsilon > 0$. By regularity of ν , there exist compact sets $K \subseteq B$ and $L \subseteq X \setminus B$ with

$$\nu(B \setminus K) + \nu(X \setminus (B \cup L)) < \varepsilon.$$

Since B is H -invariant, $U \cdot K$ and $U \cdot L$ are disjoint, so by compactness

$$\delta = d(U \cdot K, U \cdot L) > 0.$$

Assume now that $\max_{P \in \xi_n} \text{diam}_U P < \delta$ and define

$$A_n = \bigcup_{\substack{P \in \xi_n, \\ P \cap K \neq \emptyset}} P.$$

Clearly $A_n \supseteq K$ and so $\nu(B \setminus A_n) < \varepsilon$. We claim that $A_n \cap L = \emptyset$ so that $A_n \setminus B \subseteq X \setminus (B \cup L)$ and so $\nu(A_n \setminus B) < \varepsilon$, giving the lemma.

To complete the proof, assume that $A_n \cap L \neq \emptyset$, so that $P \cap L \neq \emptyset$ for some $P \in \xi_n$ with $P \cap K \neq \emptyset$. Let $x \in P \cap K$ and $y \in P \cap L$, so that $d(U \cdot x, U \cdot y) \geq \delta$ by definition of δ . However, by our assumption on ξ_n we should have

$$d(U \cdot x, U \cdot y) \leq \text{diam}_U P < \delta.$$

This shows that A_n and L must be disjoint as required. \square

PROOF OF THEOREM 8.19. This proof is very similar to the proof of Theorem 8.12 (and to Theorem 8.5) but more involved to the lack of expansiveness (or hyperbolicity) of the dynamics of a . Recall that by Theorem 6.11 we have

$$h_m(T) = h_{\text{top}}(T) =: h.$$

Suppose that $\mu \perp m$ is another ergodic T -invariant probability measure on X with $h_\mu(T) = h$. Define

$$B = \{x \in X \mid x \text{ is generic for } T \text{ and } m\}.$$

Then by [46, Cor. 4.20] we have $m(B) = 1$ and $\mu(B) = 0$. We also define the centralizer of a ,

$$H = C_G(a) = \{g \in G \mid ga = ag\}$$

and claim that B is H -invariant. To see this, suppose that $x \in B$ is generic for T and m , and let $f \in C(X)$. Then $f^h(x) = f(xh^{-1})$ defines for each $h \in H$ a function $f^h \in C(X)$. Hence

$$\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(xh^{-1})) &= \frac{1}{n} \sum_{k=0}^{n-1} f(xh^{-1}a^{-k}) && \text{(by definition of } T) \\
&= \frac{1}{n} \sum_{k=0}^{n-1} f(xa^{-k}h^{-1}) && \text{(since } h \in C_G(a)) \\
&= \frac{1}{n} \sum_{k=0}^{n-1} f^h(T^k x) && \text{(by definition of } f_h \text{ and } T) \\
&\longrightarrow \int f_h \, dm = \int f \, dm && \text{(since } x \text{ is generic)}
\end{aligned}$$

as $n \rightarrow \infty$, which shows that xh^{-1} is generic for m , so B is H -invariant.

By the assumption that Ad_a is diagonalizable we have the estimates in (8.22) for any $x \in X$, $\delta \in (0, \varepsilon]$, and $n \geq 1$. Fix some compact neighbourhood $U = U^{-1} \subseteq H$ of the identity. We claim that if δ is sufficiently small, then

$$\text{diam}_U(D^\pm(x, n, 2\delta)) \longrightarrow 0 \quad (8.26)$$

as $n \rightarrow \infty$ uniformly for $x \in X$. To see this, notice first that the two-sided Bowen balls $D^\pm(x, n, 2\delta)$ in X are the images of the Bowen balls $D_{\theta_a}^\pm(I, n, 2\delta)$ for the conjugation map $\theta_a(g) = aga^{-1}$ with respect to the local isometry

$$D_{\theta_a}^\pm(I, n, 2\delta) \ni g \longmapsto xg \in D^\pm(I, n, 2\delta)$$

(see the proof of Theorem 6.11 for the one-sided case). Hence for the proof of (8.26) it suffices to consider $D_{\theta_a}^\pm(I, n, 2\delta)$. Supposing the opposite we find some $\eta > 0$ and along some subsequence of n , two points $g_n, g'_n \in D_{\theta_a}^\pm(I, n, 2\delta)$ for which

$$d(g_n U, g'_n U) \geq \eta.$$

Let $v_n = \log g_n$. Since $g_n \in D_{\theta_a}^\pm(I, n, 2\delta)$, we have that $\|\text{Ad}_a^k(v_n)\|$ is uniformly bounded for $|k| \leq n$. Therefore, choosing a subsequence if necessary, we obtain some limit point $v \in \mathfrak{g}$ of (v_n) which is a sum of eigenvectors for Ad_a with eigenvalues of absolute value one. By assumption, this shows that

$$\text{Ad}_a(v) = v,$$

and so v belongs to the Lie algebra of H . In other words, along some subsequence (n_k) we have $g_{n_k} \rightarrow h \in H$. Applying the same argument to g'_n , we find a common subsequence (n_k) with $g_{n_k} \rightarrow h$ and $g'_{n_k} \rightarrow h' \in H$, and with $h, h' \in \overline{B_\delta(I)}$. By the assumption on η , we have

$$d(hU, h'U) \geq \eta.$$

However, if δ is sufficiently small then

$$\left(\overline{B_\delta(I)} \cap H\right)^2 \subseteq U,$$

and so $u = h^{-1}h' \in U$ which shows that

$$0 = d(hu, h') \geq d(hU, h'U).$$

This contradiction shows that (8.26) holds uniformly for $x \in X$.

We continue much as in the proof of Theorem 8.7. Let $n \geq 1$ and let F_n be a maximal set such that $D^\pm(x, n, \delta)$ for $x \in F_n$ are pairwise disjoint. By the estimates (8.22) we have

$$a_n = |F_n| \ll e^{2nh}, \quad (8.27)$$

where we write $a_n \ll b_n$ to mean that $a_n < Cb_n$ for some constant C independent of n . By maximality of F_n , we also have

$$\bigcup_{x \in F_n} D^\pm(x, n, 2\delta) = X.$$

Now let $\xi_n = \{P_x \mid x \in F_n\}$ be a partition of X with

$$D^\pm(x, n, \delta) \subseteq P_x \subseteq D^\pm(x, n, 2\delta),$$

(see the construction on page 207) so that

$$e^{-2nh} \ll m(P_x) \ll e^{-2nh} \quad (8.28)$$

for all $P_x \in \xi_n$. Now apply Lemma 8.21 to $X \setminus B$ and $H = C_G(a)$ as above, $\nu = \frac{1}{2}(m + \mu)$, and the partition ξ_n that we have just constructed, which in particular used a choice of δ sufficiently small to ensure that (8.26) holds. It follows that there exists some choice of $A_n \in \sigma(\xi_n)$ with

$$m(A_n) \rightarrow 0 \quad (8.29)$$

and with

$$\mu(A_n) \rightarrow 1 \quad (8.30)$$

as $n \rightarrow \infty$. By Lemma 8.20 we see that

$$H_\mu(\xi_n) \geq h_\mu(T^{2n+1}, \xi_n) = h_\mu(T^{2n+1}) = (2n+1)h.$$

Let M_n denote the number of elements of ξ_n needed to cover A_n . Then we know from (8.28), (8.29) and (8.30) that

$$M_n e^{-2nh} \rightarrow 0 \quad (8.31)$$

as $n \rightarrow \infty$. Let $\eta_n = \{A_n, X \setminus A_n\}$. Then

$$\begin{aligned} 2nh &\leq H_\mu(\xi_n) = H_\mu(\eta_n) + H_\mu(\xi_n | \eta_n) \\ &\leq \log 2 + \mu(A_n) \log M_n + (1 - \mu(A_n)) \log a_n. \end{aligned} \quad (\text{by (1.1)})$$

Therefore, by (8.27) there exists some $c > 0$ such that

$$0 \leq \log 2 + \underbrace{\mu(A_n)}_{\rightarrow 1} \underbrace{\log M_n e^{-2nh}}_{\rightarrow -\infty} + (1 - \mu(A_n)) \log c,$$

where we used (8.29) and (8.31). This is a contradiction, and shows that m is the unique measure of maximal entropy. \square

Exercises for Section 8.5

Exercise 8.5.1. Generalize Theorem 8.19 by dropping the assumption that 1 is the only eigenvalue of Ad_a of absolute value one and assuming instead that m_X is mixing (or at least weak mixing) with respect to T .

8.6 Equidistribution of Periodic Orbits for the Geodesic Flow

In this section we wish to combine the proof of the variational principle (Theorem 5.24) and the uniqueness of the measure of maximal entropy (Theorem 8.19) to obtain the equidistribution of periodic orbits on compact hyperbolic spaces.

8.6.1 Anosov Shadowing and Closing Lemmas

Our main tool used to produce many periodic orbits will be the following very special cases of two fundamental results from the area of smooth dynamics.

Proposition 8.22 (Anosov shadowing for $\Gamma \backslash G$). *Let G be a Lie group, let $\Gamma < G$ be a discrete subgroup and let $X = \Gamma \backslash G$. Assume that*

$$A = \{a_t \mid t \in \mathbb{R}\} < G$$

is a one-parameter subgroup with the property that Ad_{a_t} is diagonalizable over \mathbb{R} for all $t \in \mathbb{R}$. Then for every $\varepsilon > 0$ there exists some $\delta > 0$ such that $d(x_-, x_+) < \delta$ for $x_-, x_+ \in X$ implies that there exists some $z \in X$ and some

$$m \in C_G(A) = \{g \in G \mid ga = ag \text{ for all } a \in A\}$$

with $d(m, e) < \varepsilon$ and the following orbit shadowing property: In the past the orbit of z shadows the orbit of x_- , meaning that $d(a_t \cdot x_-, a_t \cdot z) \leq \varepsilon$ for $t \leq 0$ and $d(a_t \cdot x_-, a_t \cdot z) \rightarrow 0$ as $t \rightarrow -\infty$. In the future shadowing holds for $m \cdot z$ and x_+ , meaning $d(a_t \cdot x_+, a_t m \cdot z) \leq \varepsilon$ for $t \geq 0$ and $d(a_t \cdot x_+, a_t m \cdot z) \rightarrow 0$ as $t \rightarrow \infty$.

We note that the conclusion of the proposition is most useful if $C_G(A)^o$ coincides with A , as in this case the element m simply corresponds to a small time shift in the one-parameter subgroup A .

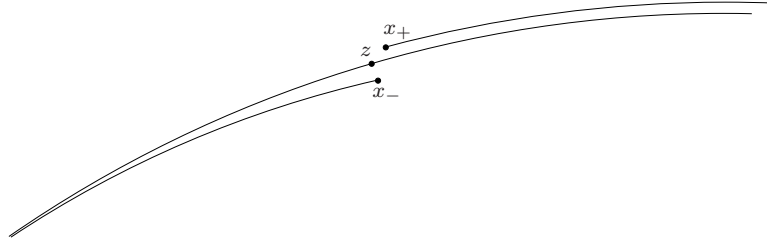


Fig. 8.2: The orbit of z approaches the orbit of x_- in the past and is — because of the shift by $m \in C_G(A)$ — ‘parallel’ to the orbit of x_+ in the future.

PROOF OF PROPOSITION 8.22. Let \mathfrak{g} be the Lie algebra of G , and define the Lie sub-algebras

$$\begin{aligned}\mathfrak{g}^- &= \{v \in \mathfrak{g} \mid \text{Ad}_{a_t}(v) \rightarrow 0 \text{ as } t \rightarrow \infty\}, \\ \mathfrak{g}^+ &= \{v \in \mathfrak{g} \mid \text{Ad}_{a_t}(v) \rightarrow 0 \text{ as } t \rightarrow -\infty\},\end{aligned}$$

and the Lie algebra of $C_G(A)$,

$$\mathfrak{g}^0 = \{v \in \mathfrak{g} \mid \text{Ad}_{a_t}(v) = v \text{ for all } t \in \mathbb{R}\}.$$

By the assumption on A , we have

$$\mathfrak{g} = \mathfrak{g}^- \oplus \mathfrak{g}^0 \oplus \mathfrak{g}^+ \tag{8.32}$$

and the map

$$\begin{aligned}\phi: \mathfrak{g}^- \oplus \mathfrak{g}^0 \oplus \mathfrak{g}^+ &\longrightarrow G \\ (v^-, w, v^+) &\longmapsto \exp(v^-) \exp(w) \exp(v^+)\end{aligned}$$

is a diffeomorphism in a neighbourhood of the origin onto a neighbourhood of the identity in G . In fact this follows since its derivative at $0 \in \mathfrak{g}$ is the map

$$(v^-, w, v^+) \longmapsto v^- + w + v^+,$$

which is a bijection by (8.32).

Choose $\delta > 0$ small enough to ensure that $g \in B_\delta^G(e)$ implies $g = u_- m u_+$, where $u_\pm = \exp(v^\pm)$ with $v^\pm \in \mathfrak{g}^\pm$ and $m = \exp(w)$ for some $w \in \mathfrak{m}$. For a given $\varepsilon > 0$ we may choose $\delta > 0$ with $d(m, e) < \varepsilon$, $d(a_t u_- a_{-t}, e) < \varepsilon$ for $t \geq 0$ and $d(a_t u_+ a_{-t}, e) < \varepsilon$ for $t \leq 0$.

If now $d(x_-, x_+) < \delta$ then $x_+ = g \cdot x_-$ for some $g \in B_\delta^G(e)$ and we obtain $g = u_- m u_+$ as above. We define $z = u_+ \cdot x_-$ and obtain

$$d(a_t \cdot x_-, a_t \cdot z) = d(x_- a_t^{-1}, x_- u_+^{-1} a_t^{-1}) \leq d(a_t^{-1}, u_+^{-1} a_t^{-1}) = d(a_t u_+ a_t^{-1}, e)$$

which we know is less than ε for $t \leq 0$, and converges to 0 as $t \rightarrow -\infty$. By our definitions, we also have $x_+ = u_- m \cdot z$ which gives

$$\begin{aligned} d(a_t \cdot x_+, a_t m \cdot z) &= d(z m^{-1} u_-^{-1} a_t^{-1}, z m^{-1} a_t^{-1}) \\ &\leq d(m^{-1} u_-^{-1} a_t^{-1}, m^{-1} a_t^{-1}) \\ &= d(a_t u_-^{-1} a_t^{-1}, e) \end{aligned}$$

which is less than ε for $t \geq 0$ and converges to 0 as $t \rightarrow \infty$. \square

Using the same notation as in the above proof we note that \mathfrak{g}^\pm are nilpotent Lie algebras. In fact, since

$$[\text{Ad}_{a_1} u, \text{Ad}_{a_1} v] = \text{Ad}_{a_1} [u, v]$$

for all $u, v \in \mathfrak{g}$ we see that $[u, v]$ is another eigenvector if $u, v \in \mathfrak{g}^\pm$ are eigenvectors for Ad_{a_1} and also that the eigenvalue for $[u, v]$ is the product of the eigenvalues for u and v . Hence the claim follows since the finitely many eigenvalues for Ad_{a_1} on \mathfrak{g}^+ are either all bigger than one and on \mathfrak{g}^- are all smaller than one.

A similar argument can be used in the case where $G \subseteq \text{GL}_d(\mathbb{R})$ to see that \mathfrak{g}^\pm consists of nilpotent matrices, that $\exp(\cdot)$ is a polynomial map on \mathfrak{g}^\pm which has a polynomial inverse on its closed image $U^\pm = \exp(\mathfrak{g}^\pm)$.

In the general case we recall that $\exp(\text{Ad}_{a_1}(u)) = a_1 \exp(u) a_1^{-1}$ for all $u \in \mathfrak{g}$. Using this, and since $\exp(\cdot)$ is locally a diffeomorphism, we may also derive that $\exp(\cdot)$ when restricted to \mathfrak{g}^\pm is a diffeomorphism from \mathfrak{g}^\pm onto the image[†] $U^\pm = \exp(\mathfrak{g}^\pm)$.

Finally we note that $C_G(A)$ normalizes the subgroups U^\pm .

Proposition 8.23 (Anosov closing for $\Gamma \backslash G$). *Let G be a Lie group, Γ a discrete subgroup of G and let $X = \Gamma \backslash G$. Let $A = \{a_t \mid t \in \mathbb{R}\} < G$ be a one-parameter subgroup such that $\text{Ad}_{a_t} : \mathfrak{g} \rightarrow \mathfrak{g}$ is diagonalizable over \mathbb{R} for all $t \in \mathbb{R}$. Then, for every $\varepsilon > 0$, there exists some $\delta > 0$ such that for all sufficiently large T and all $x \in X$ with $d(a_T \cdot x, x) < \delta$ there exists some $y \in X$*

[†] Depending on the ambient group G , the image U^\pm is not always a closed subgroup, but it is a subgroup and we may always consider it as an immersed submanifold.

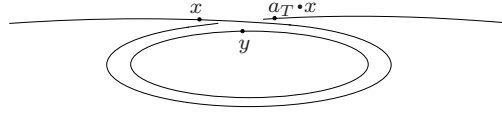


Fig. 8.3: If x and its image under the action of a_T are close, then the orbit between x and its image is close to a periodic orbit of y , but in general this may not be quite true because the orbit of y may only close up to a small shift $m \in C_G(A)$.

with $d(a_t \cdot y, a_t \cdot x) < \varepsilon$ for all $t \in [0, T]$ and $a_T \cdot y = m \cdot y$ for some $m \in C_G(A)$ with $d(m, e) < \varepsilon$.

Once more in the case where $C_G(A)^o = A$ the element m corresponds to a small time shift, and so the conclusion in fact finds a periodic orbit near the original orbit.

PROOF OF PROPOSITION 8.23. We will use the same notation as was used in the proof of Proposition 8.22, but will use \mathfrak{g}^+ and \mathfrak{g}^- in the reversed order. By the same argument, there exists some $\delta_0 > 0$ such that the diffeomorphism

$$\begin{aligned} \psi: B_{\delta_0}^G(e) &\longrightarrow U^+ \times C_G(A) \times U^- \\ g &\longmapsto (\psi_+(g), \psi_0(g), \psi_-(g)) \end{aligned}$$

exists as the local inverse of the map sending $(u_+, m, u_-) \in U^+ \times C_G(A) \times U^-$ to $u_+ m u_- \in G$. We may assume (by shrinking δ_0 if necessary) that there exists a constant $c > 0$ such that the total derivative of the map

$$B_{\delta_0/2}^{U^-} \ni u_- \longmapsto u_- g \in B_{\delta_0}^G \longmapsto \psi_-(u_- g) \in U^-$$

has norm less than c for all $g \in B_{\delta_0/2}^G$. Moreover, for every $\varepsilon \in (0, 1)$ there exists some $\delta \in (0, \frac{\delta_0}{2}]$ such that $u_- \in B_\delta^{U^-}$ and $g \in B_\delta^G$ implies that $\psi_-(u_- g) \in B_\varepsilon^{U^-}$, $\psi_+(u_- g) \in B_\varepsilon^{U^+}$, and $\psi_0(u_- g) \in B_\varepsilon^{C_G(A)}$.

Since Ad_{a_t} contracts U^- for $t > 0$ we can find some T_0 with the property that

$$T \geq T_0 \implies \overline{a_T B_2^{U^-} a_T^{-1}} \subseteq B_\delta^{U^-}.$$

Hence for these T we obtain, for any $g \in B_\delta^G$, a smooth map

$$\phi_T: \overline{B_2^{U^-}} \longrightarrow B_\varepsilon^{U^-}$$

defined by

$$u_- \longmapsto a_T u_- a_T^{-1} \longmapsto \psi_-(a_T u_- a_T^{-1} g) \in B_\varepsilon^{U^-}.$$

As the first step of this map contracts U^- uniformly and the derivative of the second step is bounded by a uniform constant c we may assume that T_0 is chosen large enough to ensure that ϕ_T is a contraction on $\overline{B_2^{U^-}}$. By Banach's

contraction mapping theorem this shows that for every $T \geq T_0$ and $g \in B_\delta^G$ there exists some $\tilde{u}_- \in B_\varepsilon^{U^-}$ with

$$a_T \tilde{u}_- a_T^{-1} g = u_+ m \tilde{u}_- \quad (8.33)$$

for some $u_+ \in B_\varepsilon^{U^+}$ and $m \in B_\varepsilon^{C_G(A)}$. Notice that as $C_G(A)$ normalizes U^- we can deduce that $g \in C_G(A)U^-$ implies that $u_+ = e$ and $g \in mU^-$.

Assume now that $x \in X$ and $d(a_T \cdot x, x) < \delta$ for some $T \geq T_0$. Then there exists $g \in B_\delta^G$ with $a_T \cdot x = g \cdot x$, and applying the argument above we find $\tilde{u}_-, m, u_+ \in B_\varepsilon^G$ satisfying (8.33). We wish to replace x by $z = \tilde{u}_- \cdot x$ and see that

$$\begin{aligned} a_T \cdot z &= (a_T \tilde{u}_- a_T^{-1}) a_T \cdot x && \text{(by definition of } z) \\ &= a_T \tilde{u}_- a_T^{-1} g \cdot x && \text{(by assumption on } g) \\ &= u_+ m \tilde{u}_- \cdot x && \text{(by (8.33))} \\ &= u_+ m \cdot z. && \text{(by definition of } z) \end{aligned}$$

Since $\tilde{u}_- \in U^-$ it is easy to see that $d(a_t \tilde{u}_- a_t^{-1}, e) \ll \varepsilon$ for all $t \geq 0$, and so $d(a_t \cdot z, a_t \cdot x) \ll \varepsilon$ for all $t \geq 0$. In other words, by switching from x to z we have simplified the situation as the new displacement between $a_T \cdot z$ and z is $u_+ m \in U^+ C_G(A)$.

We may now[†] apply the above argument again as follows: We define

$$x' = a_T \cdot z = u_+ m \cdot z$$

and $a'_T = a_T^{-1}$ so that

$$a'_T \cdot x' = a_T^{-1} \cdot x' = z = (u_+ m)^{-1} \cdot x' = m^{-1} u_+^{-1} \cdot x'.$$

This time switch also interchanges the roles of U^+ and U^- . Hence we may apply the above argument to find some $\tilde{u}_+ \in B_\varepsilon^{U^+}$ and the point $y = \tilde{u}_+ \cdot x'$ satisfying $d(a'_t \cdot y, a'_t \cdot x') \ll \varepsilon$ for all $t \geq 0$. Moreover, the remark after (8.33) applies now also and gives $a'_T \cdot y = m^{-1} \cdot y$.

Equivalently y satisfies $a_T \cdot y = m \cdot y$. Using the properties of z , the definition of $a'_t = a_t^{-1}$ and of x' , and finally the properties of y we obtain

$$\begin{aligned} d(a_t \cdot x, a_t \cdot y) &\leq d(a_t \cdot x, a_t \cdot z) + d(a_t \cdot z, a_t \cdot y) \\ &\ll \varepsilon + d(a_{t-T} \cdot x', a_{t-T} m \cdot y) \\ &\ll \varepsilon + d(a_{t-T} \cdot x', a_{t-T} \cdot y) + d(a_{t-T} \cdot y, m a_{t-T} \cdot y) \\ &\ll \varepsilon + \varepsilon + d(I, m) \ll \varepsilon \end{aligned}$$

for all $t \in [0, T]$. Since ε was arbitrary, this proves the proposition. \square

[†] In many instances it is possible to find the point y as in the proposition by setting y to be $a_T^{-1} \tilde{u}_+ a_T \cdot z$ and $a_T \cdot y = m \cdot y$ and then solving the resulting linear equation for \tilde{u}_+ .

As was mentioned above, the case $C_G(A)^o = A$ leads to the strongest conclusions. However, the above is still highly useful if $C_G(A)^o = AM$ for some compact group M as is the case for the orthogonal groups considered below.

We wish to apply these results to quotients of hyperbolic spaces. For this, we discuss next the orthogonal groups of signature $(d, 1)$ for $d \geq 1$, and the relationship between them and hyperbolic spaces of higher dimension for $d \geq 2$.

8.6.2 Special Orthogonal Groups

For $d = 1$ the special orthogonal group

$$\mathrm{SO}_{1,1}(\mathbb{R}) = \{g \in \mathrm{SL}_2(\mathbb{R}) \mid Q_{1,1} \circ g = Q_{1,1}\}$$

is the orthogonal group of the quadratic form $Q_{1,1}$ defined by

$$Q_{1,1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 - x_2^2.$$

Since $x_1^2 - x_2^2 = (x_1 - x_2)(x_1 + x_2)$, we can also use the quadratic form

$$Q'_{1,1} \begin{pmatrix} x \\ z \end{pmatrix} = xz$$

to define a conjugated copy of $\mathrm{SO}_{1,1}(\mathbb{R})$, namely

$$\mathrm{SO}'_{1,1}(\mathbb{R}) = \{g \in \mathrm{SL}_2(\mathbb{R}) \mid Q'_{1,1} \circ g = Q'_{1,1}\}.$$

We note that

$$\mathrm{SO}_{1,1}(\mathbb{R}) = \left\{ \pm \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

and

$$\mathrm{SO}'_{1,1}(\mathbb{R}) = \left\{ \pm \begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix} \mid t \in \mathbb{R} \right\}.$$

We only prove the second statement, starting by noting that the inclusion

$$\mathrm{SO}'_{1,1}(\mathbb{R}) \supseteq \left\{ \pm \begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

is easy to check by a direct calculation. Suppose therefore that

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SO}'_{1,1}(\mathbb{R}).$$

Then

$$0 = Q'_{1,1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = Q'_{1,1} \begin{pmatrix} a \\ b \end{pmatrix} = ab$$

implies that $a = 0$ or $b = 0$. Similarly,

$$0 = Q'_{1,1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = Q'_{1,1} \begin{pmatrix} c \\ d \end{pmatrix} = cd.$$

Since g is invertible, we must have either $a = d = 0$ or $b = c = 0$. In the first case

$$1 = Q'_{1,1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (a + b)(c + d) = bc$$

and

$$g = \begin{pmatrix} 0 & b \\ b^{-1} & 0 \end{pmatrix},$$

respectively in the second

$$1 = Q'_{1,1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (a + b)(c + d) = ad$$

and

$$g = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}.$$

Since we also have $\det g = 1$ it follows that

$$g = \begin{pmatrix} a & \\ & a^{-1} \end{pmatrix}$$

for some $a \in \mathbb{R} \setminus \{0\}$, giving the claim.

The same argument also shows that

$$O'_{1,1}(\mathbb{R}) = \{g \in \mathrm{GL}_2(\mathbb{R}) \mid Q'_{1,1} \circ g = Q'_{1,1}\} = \mathrm{SO}'_{1,1}(\mathbb{R}) \sqcup \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \mathrm{SO}'_{1,1}(\mathbb{R}),$$

showing that the special orthogonal group of signature $(1, 1)$ has two connected components, and the orthogonal group has four connected components.

We next treat the more interesting higher-dimensional case, for which the special case $d = 1$ will be helpful. Setting

$$Q_{d,1} \begin{pmatrix} x_1 \\ \vdots \\ x_{d+1} \end{pmatrix} = x_1^2 + x_2^2 + \cdots + x_d^2 - x_{d+1}^2,$$

we define the associated special orthogonal group

$$\mathrm{SO}_{d,1}(\mathbb{R}) = \{g \in \mathrm{SL}_{d+1}(\mathbb{R}) \mid Q_{d,1} \circ g = Q_{d,1}\}$$

and highlight three important[†] subgroups

$$K = \left\{ \begin{pmatrix} k & \\ & 1 \end{pmatrix} \mid k \in \mathrm{SO}_d(\mathbb{R}) \right\},$$

$$A = \left\{ a_t = \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & I_{d-1} & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix} \mid t \in \mathbb{R} \right\}, \quad (8.34)$$

$$M = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix} \mid k \in \mathrm{SO}_{d-1}(\mathbb{R}) \right\} \quad (8.35)$$

corresponding to orthogonal transformations in $\langle e_1, \dots, e_d \rangle, \langle e_1, e_{d+1} \rangle$, respectively $\langle e_2, \dots, e_d \rangle$.

Using the same argument as in the case $d = 1$ as above, we may also consider the quadratic form

$$Q'_{d,1} \begin{pmatrix} x \\ y_1 \\ \vdots \\ y_{d-1} \\ z \end{pmatrix} = y_1^2 + y_2^2 + \dots + y_{d-1}^2 - 2xz,$$

the corresponding special orthogonal group $\mathrm{SO}'_{d,1}(\mathbb{R})$, the subgroups

$$A' = \left\{ a'_t = \begin{pmatrix} e^t & 0 & 0 \\ 0 & I_{d-1} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \mid t \in \mathbb{R} \right\} \quad (8.36)$$

and

$$M = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix} \mid k \in \mathrm{SO}_{d-1}(\mathbb{R}) \right\}$$

respectively. The quadratic form $Q'_{d,1}$ has the advantage[‡] that it is quite straightforward to define two more important subgroups of $\mathrm{SO}'_{d,1}(\mathbb{R})$, namely

[†] The subgroups K , A , and M are not canonically defined but rather represent convenient choices of subgroups that are uniquely defined up to conjugation: K is a maximal connected compact subgroup, A is a maximal connected \mathbb{R} -diagonalizable subgroup, and $M = C_G(A) \cap K$ is the maximal connected compact subgroup of $C_G(A)$.

[‡] Defining K within $\mathrm{SO}'_{d,1}(\mathbb{R})$ is less convenient, but as $\mathrm{SO}_{d,1}(\mathbb{R})$ and $\mathrm{SO}'_{d,1}(\mathbb{R})$ are conjugated there is also a corresponding subgroup in $\mathrm{SO}'_{d,1}(\mathbb{R})$. Similarly, there exist two subgroups $U^\pm < \mathrm{SO}_{d,1}(\mathbb{R})$ that have the same relation to A as the subgroups here have to A' .

$$U^+ = \left\{ \begin{pmatrix} 1 & v^t & \frac{\|v\|^2}{2} \\ 0 & I_{d-1} & v \\ 0 & 0 & 1 \end{pmatrix} \mid v \in \mathbb{R}^{d-1} \right\}$$

and

$$U^- = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ v & I_{d-1} & 0 \\ \frac{\|v\|^2}{2} & v^t & 1 \end{pmatrix} \mid v \in \mathbb{R}^{d-1} \right\}.$$

Since 1 is the only eigenvalue of all elements of U^\pm , these subgroups and their elements are called unipotent. To see that $U^+ < \mathrm{SO}'_{d,1}(\mathbb{R})$ it is enough to let $v \in \mathbb{R}^{d-1}$ and calculate

$$\begin{pmatrix} 1 & v^t & \frac{\|v\|^2}{2} \\ 0 & I_{d-1} & v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + v_1 y_1 + \cdots + v_{d-1} y_{d-1} + \frac{\|v\|^2}{2} z \\ y_1 + v_1 z \\ \vdots \\ y_{d-1} + v_{d-1} z \\ z \end{pmatrix}$$

giving

$$\begin{aligned} Q'_{d,1} \left(\begin{pmatrix} 1 & v^t & \frac{\|v\|^2}{2} \\ 0 & I_{d-1} & v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right) \\ = y_1^2 + 2y_1 v_1 z + v_1^2 z^2 + \cdots + y_{d-1}^2 + 2y_{d-1} v_{d-1} z + v_{d-1}^2 z^2 \\ - 2 \left(x + v_1 y_1 + \cdots + v_{d-1} y_{d-1} + \frac{\|v\|^2}{2} z \right) z \\ = y_1^2 + \cdots + y_{d-1}^2 - 2xz \end{aligned}$$

as required. The calculation for U^- is similar.

8.6.3 Hyperbolic spaces

These groups are directly related to hyperbolic spaces as follows. Indeed, the hyperbolic d -space may itself be defined using the quadratic form $Q_{d,1}$, by setting

$$\mathbb{H}^d = \{v \in \mathbb{R}^{d+1} \mid Q_{d,1}(v) = -1, v_{d+1} > 0\},$$

as illustrated in Figure 8.4.

One can easily verify that the tangent space of \mathbb{H}^d at $v_0 \in \mathbb{H}^d$ (when viewed as a submanifold of \mathbb{R}^{d+1}) is

$$T_{v_0} \mathbb{H}^d = \{w \in \mathbb{R}^{d+1} \mid \langle w, v_0 \rangle_{Q_{d,1}} = 0\},$$

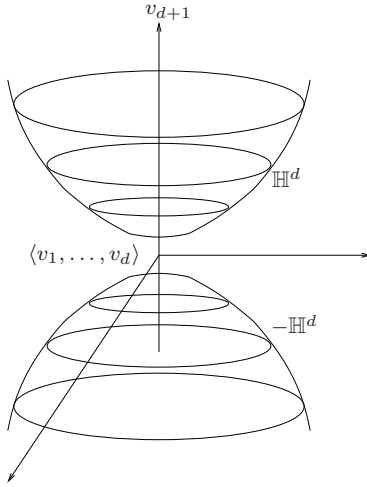


Fig. 8.4: Real hyperbolic d -space.

that is, the orthogonal complement v_0^\perp with respect to the quadratic form $Q_{d,1}$. The Riemannian metric on $T_{v_0}\mathbb{H}^d$ is defined as the restriction of $Q_{d,1}$ to $T_{v_0}\mathbb{H}^d$ (notice that the constraint $Q_{d,1}(v_0) = -1$ ensures that this restriction is positive-definite by Sylvester's law of inertia).

We claim that \mathbb{H}^d is connected, which we will prove by finding a path from $v_0 = e_{d+1}$, the designated origin, to any other specified point $v \in \mathbb{H}^d$. Since

$$-1 = Q_{d,1}(v) = v_1^2 + \cdots + v_d^2 - v_{d+1}^2,$$

there exists some $T \geq 0$ such that $v_{d+1}^2 = \cosh^2 T$ and $v_1^2 + \cdots + v_d^2 = \sinh^2 T$. We also define $w = \frac{(v_1, \dots, v_d, 0)}{\sinh T}$ and choose $k \in K \cong \text{SO}_d(\mathbb{R})$ with $ke_1 = w$. Then

$$\gamma_t = k(e_1 \sinh t + e_{d+1} \cosh t) = ka_t v_0$$

for $t \in [0, T]$ (where a_t is defined in (8.34)) defines a continuous path that connects $\gamma_0 = v_0$ and $\gamma_T = v$, proving the claim. In fact the path $t \mapsto \gamma_t$ has unit speed with respect to the Riemannian metric, as may be easily checked.

We note that

$$\text{SO}_{d,1}(\mathbb{R})^+ = \{g \in \text{SO}_{d,1}(\mathbb{R}) \mid g_{d+1,d+1} > 0\}$$

is the connected component of the identity $\text{SO}_{d,1}(\mathbb{R})^\circ < \text{SO}_{d,1}(\mathbb{R})$. To see this, notice first that $g \in \text{SO}_{d,1}(\mathbb{R})$ implies

$$Q_{d,1}(ge_{d+1}) = Q_{d,1}(e_{d+1}) = -1,$$

which forces $g_{d+1,d+1} \neq 0$ (since $Q_{d,1}|_{\mathbb{R}^d \times \{0\}}$ is positive-definite). This shows that $\mathrm{SO}_{d,1}(\mathbb{R})^+ \supseteq \mathrm{SO}_{d,1}(\mathbb{R})^o$. To see that $\mathrm{SO}_{d,1}(\mathbb{R})^+ = \mathrm{SO}_{d,1}(\mathbb{R})^o$, let g be an arbitrary element of $\mathrm{SO}_{d,1}(\mathbb{R})^+$. Define $v = gv_0 \in \mathbb{H}^d$ and find $T \geq 0$ and an element $k \in K \cong \mathrm{SO}_d(\mathbb{R})$ satisfying $gv_0 = v = ka_Tv_0$ as above. In particular, $k' = a_T^{-1}k^{-1}g$ satisfies $k'v_0 = v_0$ and so $k' \in K \cong \mathrm{SO}_d(\mathbb{R})$. Since $\mathrm{SO}_d(\mathbb{R})$ is connected[†] we see that $g = ka_Tk' \in \mathrm{SO}_{d,1}(\mathbb{R})^+$ belongs to $\mathrm{SO}_{d,1}(\mathbb{R})^o$.

The above argument gives the Cartan decomposition for $G = \mathrm{SO}_{d,1}(\mathbb{R})^+$: Every element $g \in G$ can be written as

$$g = k_1 a_T k_2$$

with k_1, k_2 in K and a_T in A for some $T \geq 0$.

We have also shown that G acts transitively on \mathbb{H}^d with $K = \mathrm{Stab}_G(v_0)$, so that $G/K \cong \mathbb{H}^d$. Clearly the action of G is smooth on \mathbb{H}^d (since it is even linear on \mathbb{R}^{d+1}), and so we may also consider the induced action on the tangent bundle

$$\mathrm{T}\mathbb{H}^d = \bigsqcup_{v \in \mathbb{H}^d} \mathrm{T}_v \mathbb{H}^d.$$

By definition of the Riemannian metric, we see that this action leaves the unit tangent bundle

$$\mathrm{T}^1\mathbb{H}^d = \{(v, w) \mid v \in \mathbb{H}^d, w \in \mathrm{T}_v \mathbb{H}^d, Q_{d,1}(w) = 1\}$$

invariant. Moreover, since the stabilizer subgroup $\mathrm{Stab}_G(v_0) = \mathrm{SO}_d(\mathbb{R})$ of v_0 acts transitively on the sphere $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d$ and $\mathrm{T}_{v_0}^1 \mathbb{H}^d = v_0 \times \mathbb{S}^{d-1}$ we see that G still acts transitively on $\mathrm{T}^1\mathbb{H}^d$.

However there is a structural difference between the case $d = 2$ and the case $d \geq 3$. If $d = 2$ then $\mathrm{SO}_2(\mathbb{R})$ acts simply transitively on $\mathrm{T}_{v_0}^1 \mathbb{H}^2$ and so G acts simply transitively on $\mathrm{T}^1\mathbb{H}^2 \cong G$ (and $M = \{e\}$). However, if $d \geq 3$ then M as defined in (8.35) is the stabilizer of the vector $(v_0, e_1) \in \mathrm{T}_{v_0}^1 \mathbb{H}$ and so $\mathrm{T}^1\mathbb{H}^d \cong G/M$. In any case right multiplication by A is defined on G/M , and corresponds to following a unit speed geodesic path on \mathbb{H}^d with the given initial derivative in $\mathrm{T}^1\mathbb{H}^d$. We note that right multiplication by A on G is the ‘frame bundle flow’ where, in addition to following a given geodesic, an orthogonal frame in the complement in $\mathrm{T}_v \mathbb{H}^d$ of the given direction undergoes ‘parallel transport’ along the path.

[†] This may be shown via a similar argument, using induction on the dimension and the fact that \mathbb{S}^d is connected for all $d \geq 2$.

8.6.4 Dynamics and Equidistribution on Quotients of \mathbb{H}^d

If now Γ is a discrete subgroup of $G = \mathrm{SO}_{d,1}(\mathbb{R})^+$ with no torsion elements then

$$Y = \Gamma \backslash \mathbb{H}^d \cong \Gamma \backslash G / K$$

is a hyperbolic d -dimensional manifold,

$$\mathrm{T}^1 Y = \Gamma \backslash \mathrm{T}^1 \mathbb{H}^d \cong \Gamma \backslash G / M$$

is its unit tangent bundle, right multiplication by A is the geodesic flow, and finally $\Gamma \backslash G$ is the frame bundle consisting of all positively oriented orthogonal frames in $\mathrm{T}_v Y$ for all $v \in Y$, and we will call right multiplication by A the frame bundle flow.

Theorem 8.24 (Equidistribution of periodic points). *Let G be the group $\mathrm{SO}_{d,1}(\mathbb{R})^+$ and let $\Gamma < G$ be a uniform torsion-free lattice so that*

$$Y = \Gamma \backslash G / M$$

is a compact manifold, and let A be the diagonal subgroup defined in (8.35). For any $T > 0$ let

$$\mathcal{P}_T = \{y \in Y \mid ya_t^{-1} = y \text{ for some } t \in (0, T]\}$$

be the set of periodic points for the geodesic flow on Y with period no greater than T . Then \mathcal{P}_T is a finite union of compact orbits

$$\mathcal{P}_T = \bigsqcup_{k=1}^N \{y_k a_t^{-1} \mid t \in [0, \text{length}(y_k A)]\}$$

where $\text{length}(y_k A) = \min\{t > 0 \mid y_k a_t^{-1} = y_k\}$ is the period of y_k . As $T \rightarrow \infty$ the normalized sum

$$\mu_T = \frac{1}{\sum_{k=1}^N \text{length}(y_k A)} \sum_{k=1}^N \int_{y_k A} \delta_{y_k a_t^{-1}} dt$$

of the Lebesgue measure on these compact orbits equidistribute to the uniform measure m_Y (which is obtained from the Haar measure m_X via the canonical projection map $X = \Gamma \backslash G \rightarrow Y = \Gamma \backslash G / M$).

Our proof will rely on entropy, and so we will need to know the following fact.

Lemma 8.25 (Topological entropy). *Let $G = \mathrm{SO}_{d,1}(\mathbb{R})^+$ and suppose that $\Gamma < G$ is a uniform lattice, so that $X = \Gamma \backslash G$ is compact. Then*

$$h_{\text{top}}(a_t) = (d-1)t.$$

This follows from Theorem 6.11 and an analysis of the adjoint map for a_t on the Lie algebra of G . The calculation is somewhat simpler using the quadratic form $Q'_{d,1}$, the subgroups U^\pm , and the diagonal elements a'_t defined in (8.36) which corresponds to $a_t \in G$ under the isomorphism between $\mathrm{SO}_{d,1}(\mathbb{R})$ and $\mathrm{SO}'_{d,1}(\mathbb{R})$, see Exercise 8.6.1 for a more detailed outline.

The following is helpful to understand the biquotients arising in Theorem 8.24.

Lemma 8.26 (Uniform injectivity radius). *Let G be a connected Lie group, $M < G$ a compact subgroup, $\Gamma < G$ a discrete subgroup, and suppose that \mathbf{d}_G is a left G -invariant and right M -invariant metric induced from a left G -invariant and right M -invariant Riemannian metric. Then \mathbf{d}_X defined by*

$$\mathbf{d}_X(\Gamma g_1, \Gamma g_2) = \min_{\gamma_1, \gamma_2 \in \Gamma} \mathbf{d}_G(\gamma_1 g_1, \gamma_2 g_2) = \min_{\gamma \in \Gamma} \mathbf{d}_G(g_1, \gamma g_2)$$

is a right M -invariant metric on $X = \Gamma \backslash G$, and \mathbf{d}_Y defined by

$$\mathbf{d}_Y(\Gamma g_1 M, \Gamma g_2 M) = \min_{\gamma \in \Gamma, m \in M} \mathbf{d}_G(g_1, \gamma g_2 m)$$

is a metric that induces the quotient topology on $Y = \Gamma \backslash G/M$. If Γ is torsion-free (or $M = \{I\}$) and X is compact, then there exists some η_0 such that[†]

$$\mathbf{d}_Y(\Gamma g M, \Gamma gh M) = \mathbf{d}_{G/M}(eM, hM)$$

for any $g \in G$ if $h \in B_{\eta_0}^G$. Moreover, if h belongs to a fixed local transverse to M then $\mathbf{d}_{G/M}(eM, hM) \asymp \mathbf{d}_G(e, h)$.

By a local transverse to a subgroup $H < G$ we mean a set of the form $\exp(B_\rho^{\mathfrak{g}} \cap W)$, where $\rho > 0$, \mathfrak{g} is the Lie algebra of G , and W is a linear complement of the Lie algebra of H within \mathfrak{g} . Notice that we have used the definition $\mathbf{d}_{G/M}(g_1 M, g_2 M) = \min_{m \in M} \mathbf{d}_G(g_1, g_2 m)$ for $g_1, g_2 \in G$ implicitly.

PROOF OF LEMMA 8.26. The first two claims about \mathbf{d}_X and \mathbf{d}_Y are easy to verify, so we will only give the details for the last two claims of the lemma. Also notice that the inequality

$$\mathbf{d}_Y(\Gamma g M, \Gamma gh M) \leq \mathbf{d}_{G/M}(eM, hM)$$

follows directly from the definitions and the assumption that \mathbf{d}_G is left G -invariant.

Suppose now that for every $\eta_n = \frac{1}{n}$ there exists some $x_n = \Gamma g_n \in X$ and $h_n \in B_{1/n}^G(e)$ with $\mathbf{d}_Y(x_n M, x_n h_n M) < \mathbf{d}_{G/M}(eM, h_n M)$. Since X is assumed to be compact there is a compact subset $F \subseteq G$ such that $G = \Gamma F$

[†] If $M = \{I\}$ then η_0 is a uniform injectivity radius on X . Also note that the definition of $\mathbf{d}_{G/M}$ is included in the definition of \mathbf{d}_Y by using $\Gamma = \{I\}$.

and we may assume that $g_n \in F$. By definition of d_Y and of $d_{G/M}$, there exist sequences (γ_n) in $\Gamma \setminus \{e\}$ and (m_n) in M with

$$d_G(g_n, \gamma_n g_n h_n m_n) < d_{G/M}(eM, h_n M) < \frac{1}{n}. \quad (8.37)$$

As g_n, h_n, m_n all belong to a fixed compact set for all $n \geq 1$, we see that γ_n also belongs to a fixed compact subset of G , and hence to a finite subset of Γ . We may therefore choose a subsequence (n_k) such that $g_{n_k} \rightarrow g$, $m_{n_k} \rightarrow m$ in M and $\gamma_{n_k} = \gamma$ in Γ is constant as $k \rightarrow \infty$. Next notice that $\gamma \neq e$, since the first inequality in (8.37) is strict. It follows that $g = \gamma g m$, or equivalently

$$\gamma = g m^{-1} g^{-1} \in g M g^{-1}.$$

Hence $\Gamma \cap g M g^{-1}$ is non-trivial, discrete, and compact. However, this implies that the intersection is finite, which contradicts our assumption that Γ has no torsion elements.

For the final claim, let $V = \exp(B_\rho^g \cap W)$ be a local transverse to M . Notice that we may also assume that the product map $V \times M \ni (v, m) \mapsto v m \in G$ is a diffeomorphism near the identity for which the derivative of the inverse is bounded. This implies that

$$d_G(e, v m) \asymp d_G(e, v) + d_G(e, m)$$

for the metric d_G induced from the Riemmanian metric whenever vm is close to the identity. Choosing V sufficiently small, this implies the final claim of the lemma. \square

To be able to produce periodic orbits, we will use ergodicity as provided by the next lemma.

Lemma 8.27 (Ergodicity). *If $X = \Gamma \backslash G$ and A is as in Theorem 8.24, then every non-trivial element $a = a_t \in A \setminus \{e\}$ acts ergodically with respect to m_X . The same also holds for the action of a on Y with respect to m_Y .*

PROOF. We apply essentially the same argument as is used in [46, Sect. 11.3]. Let $f \in L^2_{m_X}(X)$ be invariant under a , and suppose that $u \in U^+$ so that $a^n u a^{-n} \rightarrow e$ as $n \rightarrow -\infty$. Then using invariance and unitarity of the regular representation π_g for $g \in G$ we obtain

$$\|\pi_u f - f\|_2 = \|\pi_u \pi_{a^{-n}} f - \pi_{a^{-n}} f\|_2 = \|\pi_{a^n u a^{-n}} f - f\|_2.$$

However, by continuity of the representation of G on $L^2_{m_X}(X)$ we see that the last expression converges to 0 as $n \rightarrow -\infty$. Hence $\pi_u f = f$ for any $u \in U^+$ and, by a similar argument using $n \rightarrow \infty$, for all $u \in U^-$. However, by Exercise 8.6.2 we know that $\langle U^+, U^- \rangle = G$, so f is G -invariant and hence equal to a constant in $L^2_{m_X}$.

The final claim follows, since (Y, m_Y) is a factor of (X, m_X) for the action of a . \square

Assuming that we can find periodic points, the connection to entropy should become apparent in the next lemma.

Lemma 8.28 (Separation of periodic points). *Let $Y = \Gamma \backslash G/M$ and A be as above. Then there exists some $\varepsilon > 0$ (depending on Γ) with the following property. If $T \geq 10\varepsilon$ and*

$$y_1 = a_{t_1} \cdot y_1, y_2 = a_{t_2} \cdot y_2$$

are periodic points in Y with periods $t_1, t_2 \in [T, T + \varepsilon]$, then either $t_1 = t_2$ and $y_1 A = y_2 A$ are the same periodic orbit or y_1 and y_2 are (n, ε) -separated for $n = \lceil T \rceil$ with respect to the time-one map $y \mapsto a_1 \cdot y$. In particular, the number of periodic orbits with period in $[T, T + \varepsilon]$ is $\ll e^{(d-1)n+o(n)}$.

PROOF. Suppose initially only that ε is small enough to be used in the proof of Theorem 6.5 (where the uniform injectivity radius from Lemma 8.26 of $\Gamma \backslash G$ plays a crucial role). Assume further that $y_1, y_2 \in Y$ are periodic with periods t_1, t_2 lying in $[T, T + \varepsilon]$ and that y_1, y_2 are not (n, ε) -separated. More precisely, this means that

$$\begin{aligned} y_1 &= \Gamma g_1 M, \\ y_2 &= \Gamma g_2 M, \end{aligned}$$

and the representatives of these points in G satisfy

$$g_i a_{t_i}^{-1} m_i = \gamma_i g_i$$

for $i = 1, 2$, some $\gamma_1, \gamma_2 \in \Gamma$ and $m_1, m_2 \in M$. Moreover, we can choose the representatives such that $g_2 = g_1 h$ for some $h \in B_\varepsilon^G$. We may also assume that $g_1, g_2 \in F$ for some right M -invariant compact subset $F \subseteq G$ with the property that $\Gamma F = G$ that only depends on G and Γ .

Using the coordinate system given by the decomposition $U^- U^+ A M$ near the identity, we may write $h = u_- u_+ a m$ and assume without loss of generality that $m = e$ (since g_2 can be modified on the right by any element of M) and so

$$h = u_- u_+ a.$$

Notice that after the replacement we still have $d(h, e) \leq c\varepsilon$, where the multiplicative constant c comes from the last claim in Lemma 8.26. We now apply the argument used in proving Theorem 6.5. In fact $d(a_1 h a_1^{-1}, e) \leq c \|\text{Ad}_{a_1}\| \varepsilon$ and the points $a_1 \cdot y_1 = \Gamma g_1 a_1^{-1} M$ and $a_1 \cdot y_2 = \Gamma g_1 a_1^{-1} (a_1 h a_1^{-1}) M$ are less than ε apart. We may assume that ε is sufficiently small so that this gives $d(a_1 h a_1^{-1}, e) \leq c\varepsilon$ by Lemma 8.26.

We may repeat the argument up to $n = \lceil T \rceil$ and obtain $d(a_t h a_t^{-1}, e) \leq c_1 \varepsilon$ for $t \in [0, T]$ and some constant c_1 depending on a_1 (but not on T). Rewriting the defining equations for γ_1 and γ_2 we obtain

$$g_1^{-1}\gamma_1g_1 = a_{t_1}^{-1}m_1 \quad (8.38)$$

and

$$g_1^{-1}\gamma_2g_1 = ha_{t_2}^{-1}m_2h^{-1}, \quad (8.39)$$

and so

$$\begin{aligned} g_1^{-1}\gamma_1^{-1}\gamma_2g_1 &= m_1a_{t_1}ha_{t_2}^{-1}m_2h^{-1} \\ &= m_1(a_{t_1}ha_{t_1}^{-1})a_{t_1-t_2}m_2h^{-1} \in MB_{e_1\varepsilon}^GB_\varepsilon^GMB_\varepsilon^G \end{aligned}$$

since $|t_1 - t_2| < \varepsilon$ by assumption. Since M is a compact subgroup (and we may still choose ε as small as we wish) we see that $g_1^{-1}\gamma_1^{-1}\gamma_2g_1$ belongs to a small neighbourhood of M .

As $g_1 \in F$ for some fixed compact subset $F \subseteq G$, we see that $\gamma_1^{-1}\gamma_2$ belongs to a small neighbourhood of the compact set

$$M^F = \bigcup_{g \in F} gMg^{-1} \subseteq G.$$

Notice that $\Gamma \cap M^F = \{e\}$ (since $\Gamma \cap M^g \neq \{e\}$ for some $g \in G$ would force the existence of a torsion element of Γ). If ε is sufficiently small, then this implies that $\gamma_1 = \gamma_2$.

Equation (8.38) shows that the eigenvalues of γ_1 are e^{t_1} , e^{-t_1} and possibly some eigenvalues of absolute value one, while in (8.39) they are e^{t_2} , e^{-t_2} and possibly some eigenvalues of absolute value one. It follows that $t_1 = t_2$, and combining (8.38) and (8.39) we see that h has to map the eigenvector with eigenvalue e^{t_1} to a multiple of itself, and similarly for e^{-t_1} . In particular, h maps $\langle e_1, e_{d+1} \rangle$ to itself and its orthogonal complement $\langle e_2, \dots, e_d \rangle$ to itself. It follows that $h = a_s m$ for some $s \in \mathbb{R}$ and $m \in M$, and the first part of the lemma follows.

The last claim in the lemma follows from the first and Lemma 8.25. \square

We now have, with the entropy calculation in Lemma 8.25, the ergodicity in Lemma 8.27, the Anosov shadowing and closing lemmas in Propositions 8.22 and 8.23, and the upper bound on the number of periodic orbits in Lemma 8.28, all the tools needed for the construction of many periodic orbits to hand.

Proposition 8.29 (Construction of periodic orbits). *Let $Y = \Gamma \backslash G/M$ and A be as in Theorem 8.24. Then there exists some $L > 0$ such that for all $n \geq 1$ there are at least $e^{(d-1)n+o(n)}$ many periodic orbits with periods in $[n, n+L)$.*

PROOF. We set $t = 1$, and will use the time one map of the geodesic flow on X and the frame bundle flow on Y . Let $\varepsilon_1 \in (0, 1]$ be chosen small enough to

ensure that we have at least $e^{(d-1)n+o(n)}$ many (n, ε_1) -separated points in X (see Lemma 8.25 and Theorem 6.11) for all $n \in \mathbb{N}$. We apply the Anosov closing lemma (Proposition 8.23) on X to $\varepsilon = \frac{\varepsilon_1}{5}$ to find $\varepsilon_2 \in (0, \frac{\varepsilon_1}{5})$ so that the closing property holds for any point with an ε_2 -near return to itself. We apply Anosov shadowing (Proposition 8.22) on X to $\varepsilon = \frac{\varepsilon_2}{4}$ to find $\varepsilon_3 \in (0, \frac{\varepsilon_2}{4})$ so that the shadowing property holds for any two points of distance less than ε_3 . We also use compactness to partition X into finitely many sets $\{P_1, \dots, P_r\}$ with positive measure so that $\text{diam } P_j < \varepsilon_3$ for $j = 1, \dots, r$. Finally, we apply ergodicity of $a = a_1$ on X in Lemma 8.27 to find, for every pair j_1, j_2 of indices, a point $x_{j_1, j_2} \in P_{j_1}$ and some $\ell \in \mathbb{N}$ with $a^\ell \cdot x_{j_1, j_2} \in P_{j_2}$. As there are only finitely many indices, there exists some $L > 0$ with the property that ℓ can be chosen in $[1, L] \cap \mathbb{N}$. We will use these patches of orbits linking any part of X to any other part to produce enough periodic orbits.

Now let $n \in \mathbb{N}$. By our choice of ε_1 , Lemma 8.25, and Theorem 6.11 there exists an (ε_1, n) -separated set $F \subseteq X$ with $|F| \geq e^{(d-1)n+o(n)}$. For any $x \in F$ we therefore have $a^n \cdot x \in P_{j_1}$ and $x \in P_{j_2}$ for some $j_1, j_2 \in \{1, \dots, r\}$, and can apply the Anosov shadowing lemma (Proposition 8.22) to $a^n \cdot x$ and x_{j_1, j_2} to change x to some z with $d(a^j \cdot x, a^j \cdot z) \leq \frac{\varepsilon_2}{4}$ for $j = 1, \dots, n$ and also

$$\begin{aligned} d(a^{n+\ell} \cdot z, z) &\leq d(a^{n+\ell} \cdot z, a^{n+\ell} m \cdot z) + d(a^{n+\ell} m \cdot z, a^\ell \cdot x_{j_1, j_2}) + d(a^\ell \cdot x_{j_1, j_2}, z) \\ &< \varepsilon_2 \end{aligned}$$

for some $\ell \in [1, L] \cap \mathbb{N}$ and $m \in M$ with $d(m, e) < \varepsilon_2$. Therefore we can apply the Anosov closing lemma (Proposition 8.23) to find \tilde{x} with

$$d(a^j \cdot x, a^j \cdot \tilde{x}) \leq \frac{\varepsilon_2}{4} + \frac{\varepsilon_1}{5} < \frac{\varepsilon_1}{4} \quad (8.40)$$

for $j = 1, \dots, n$ and some $m \in M$ and $t \in \mathbb{R}$ with $|t| < \frac{\varepsilon_1}{10}$ so that

$$a^{n+\ell+t} \cdot \tilde{x} = m \cdot \tilde{x}.$$

Applying this procedure to any $x \in F$ gives us a new set \tilde{F} of points that is $(n, \frac{\varepsilon_1}{2})$ -separated by (8.40) and has the same cardinality.

We note that any point $\tilde{x} \in \tilde{F}$ gives rise to a periodic point $\tilde{x}M \in Y$. We claim that the number of points $\tilde{x} \in \tilde{F}$ that can give rise to the same periodic orbit $yA \subseteq Y$ is bounded by a quantity $\ll_{M, \varepsilon_1} (n + L + 1)$. For the proof of the claim, we will use the fact that M is compact, the period of these points is bounded by $n + L + 1$, and the points in \tilde{F} are $(n, \frac{\varepsilon_1}{2})$ -separated,

Indeed, if we use the Riemannian metric on the manifold $\tilde{x}MA$ that dominates the Riemannian metric on X restricted to $\tilde{x}MA$ for any $\tilde{x} \in \tilde{F}$, then the A action is isometric. Therefore any two points $\tilde{x}_1, \tilde{x}_2 \in \tilde{F}$ with $\tilde{x}_1MA = \tilde{x}_2MA$ would have to have distance at least $\frac{\varepsilon_1}{2}$, which implies the claim. Since $n + L + 1 = e^{o(n)}$, we may select a subset of \tilde{F} (again

denoted by \tilde{F}) so that $\tilde{x}_1 \neq \tilde{x}_2$ in \tilde{F} implies $\tilde{x}_1MA \neq \tilde{x}_2MA$ and we still have $|\tilde{F}| = e^{(d-1)n+o(n)}$.

Next we note that the actual period of the periodic point \tilde{x}_1M may not be in $[n, n + L + 1]$ but instead in $(0, \frac{n+L+1}{2}]$. However, by Lemma 8.28, the number of such periodic orbits is at most $e^{(d-1)\frac{n}{2}+o(n)}$. Subtracting these points from \tilde{F} does not change the asymptotic growth of $|\tilde{F}|$, which implies the proposition. \square

PROOF OF THEOREM 8.24. Let $L > 0$ be as in Proposition 8.29 and $\varepsilon > 0$ be as in Lemma 8.28. Let $F_n \subseteq X$ be a maximal set of points $x \in X$ with disjoint periodic orbits xMA in Y and periods in $[n, n + L)$. By Proposition 8.29, we have

$$p_n = |F_n| = e^{(d-1)n+o(n)}.$$

By Lemma 8.28 points with ε -close periods are dynamically separated (even when considered in Y). We split the interval $[n, n + L]$ into $L_1 = \lceil \frac{L}{\varepsilon} \rceil$ subintervals of the form $[T, T + \varepsilon)$. Accordingly, we split the collection F_n of points with periodic orbits in Y into L_1 subcollections

$$F_{n,k} = \{x \in F_n \mid \text{the period of } xM \text{ belongs to } [n + k\varepsilon, n + (k + 1)\varepsilon)\}$$

for $k = 0, \dots, L_1 - 1$. We have

$$p_n = |F_n| = \sum_{j=1}^{L_1} |F_{n,k}| = e^{(d-1)n+o(n)}.$$

We say that the index $k \in \{1, \dots, L\}$ is *thick* if

$$|F_{n,k}| \geq \frac{1}{Ln} p_n$$

and *thin* if

$$|F_{n,k}| < \frac{1}{Ln} p_n.$$

Note that for every $n \in \mathbb{N}$ at least one index is thick. Assume first that we are in the thick case, which implies that

$$|F_{n,k}| = e^{(d-1)n+o(n)}.$$

By Lemma 8.28, the set $F_{n,k}$ is $(\lceil n + (k + 1)\varepsilon \rceil, \varepsilon)$ -separated. By the proof of the second half of the variational principle that constructs invariant measures with large entropy (see p. 147) any weak* limit of the sequence $(\mu_{n,k})$ defined by

$$\mu_{n,k} = \frac{1}{\lceil n + (k + 1)\varepsilon \rceil} \frac{1}{|F_{n,k}|} \sum_{j=0}^{\lceil n+(k+1)\varepsilon \rceil} \sum_{x \in F_{n,k}} \delta_{a^j \cdot x}$$

as $n \rightarrow \infty$, under the assumption that the index is thick for all n , will have entropy at least $(d - 1)$. By Theorem 8.19, this implies that the weak* limit is the normalized Haar measure m_X on X . As the limit does not depend on the chosen subsequence, we deduce that $\mu_{n,k} \rightarrow m_X$ as $n \rightarrow \infty$ in the weak* topology assuming that the indices are all thick. Let $f \in C(Y)$ and define

$$F(y) = \int_0^1 f(a_t \cdot x) dt$$

for $y \in Y$. Applying the weak* convergence, we obtain

$$\int_Y F d\mu_{n,k} \rightarrow \int_Y F dm_Y = \int_Y f dm_Y$$

as $n \rightarrow \infty$. We note that the definition of F and $\mu_{n,k}$ implies that $\int_Y F d\mu_{n,k}$ is equal to an integral of f over pieces of orbits in Y . In fact for every $x \in F_{n,k}$ we obtain an integral of f over the time parameter on the periodic orbit xMA . This implies that

$$\frac{1}{\sum_{x \in F_{n,k}} \text{length}(xMA)} \sum_{x \in F_{n,k}} \int_{xMA} f dt \rightarrow \int_Y f dm_Y$$

as $n \rightarrow \infty$, and once more assuming that k is chosen to be thick. Indeed, as $n \rightarrow \infty$, the precise lengths of the orbits matter less and less as they are constrained in an interval of the form $[T, T + \varepsilon)$ and the effect of the continuous integral over f not matching precisely the discrete sum over F near the period also matters less and less.

Taking the sum over all thick indices $k \in \{1, \dots, L\}$ and noticing that the thin indices have asymptotically negligible weight in the average below, we also obtain

$$\frac{1}{\sum_{x \in F_n} \text{length}(xMA)} \sum_{x \in F_n} \int_{xMA} f dt \rightarrow \int_Y f dm_Y$$

as $n \rightarrow \infty$.

The equidistribution in the theorem follows from this by decomposing $[0, T]$ into intervals of the form $[0, \ell), [\ell, 2\ell), \dots, [m\ell, (m+1)\ell)$ and a last interval $[(m+1)\ell, T]$. Adapting the definition of thick and thin to the last interval while using the number of periodic points with period in $[m\ell, (m+1)\ell)$ as the basis, we again see that the periodic orbits coming from the last interval either equidistribute (because there are sufficiently many to apply our argument) or can be ignored (because they are negligible in comparison to the previous collection of periodic orbits). \square

Exercises for Section 8.6

Exercise 8.6.1. Prove Lemma 8.25 by following the steps below.

(a) Show that $\text{Ad}_{a'_t}$ on $\mathfrak{sl}_{d+1}(\mathbb{R})$ has eigenvalues 1, $e^{\pm t}$, and $e^{\pm 2t}$. Describe the eigenspaces.

(b) Show that the Lie algebra of $\text{SO}'_{d,1}$ is given by

$$\mathfrak{so}'_{d,1} = \left\{ w \in \mathfrak{sl}_{d+1}(\mathbb{R}) \mid \langle w(v), v \rangle' = 0 \text{ for all } v \in \mathbb{R}^d \right\},$$

where $\langle \cdot, \cdot \rangle'$ denotes the inner product with respect to the quadratic form $Q'_{d,1}$.

(c) Since $a'_t \in \text{SO}'_{d,1}(\mathbb{R})$ the Lie algebra $\mathfrak{so}'_{d,1}$ is also a sum of eigenspaces for $\text{Ad}_{a'_t}$. Show that there are no eigenvectors for the eigenvalues $e^{\pm 2t}$, and that the Lie algebras of the subgroups U^\pm are the eigenspaces for the eigenvalues $e^{\pm t}$.

(d) Use (c) to calculate the entropy of $\text{Ad}_{a'_t}$ acting on the Lie algebra $\mathfrak{so}'_{d,1}$, and apply Theorem 6.11.

Exercise 8.6.2. Show that $\langle U^+, U^- \rangle = \text{SO}_{d,1}(\mathbb{R})^+$ using either of the following two approaches.

(a) Recall that $\mathfrak{so}_{d,1}(\mathbb{R})$ is a simple real Lie algebra. Using the identity

$$\text{Ad}_g[w_1, w_2] = [\text{Ad}_g w_1, \text{Ad}_g w_2]$$

and the Jacobi identity, show that the Lie algebra generated by the eigenspaces of Ad_{a_t} with non-constant eigenvalues (in our case the Lie algebra generated by the Lie algebras of U^\pm) is a Lie ideal.

(b) Using the quadratic form $Q'_{d,1}$ notice that for any $v \in \mathbb{R}^{d-1}$

$$w_v^+ = \begin{pmatrix} 0 & v^t & 0 \\ 0 & 0 & v \\ 0 & 0 & 0 \end{pmatrix}$$

belongs to the Lie algebra of U^+ and $w_v^- = (w_v^+)^t$ to the Lie algebra of U^- . Show that for any $v \neq 0$ the element $[w_v^+, w_v^-]$ generates the Lie algebra of A . Show that for two orthogonal vectors $u, v \in \mathbb{R}^{d-1}$ we have

$$[w_u^+, w_v^-] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (uv^t - vu^t) & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (8.41)$$

which belongs to the Lie algebra of M . Since $M \cong \text{SO}_{d-1}(\mathbb{R})$, the Lie algebra of M consists of all matrices of the form

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & w & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with $w^t = -w$. Show (or recall) that any such matrix can be block-diagonalized via an orthogonal matrix in M . Use this to conclude that any element in the Lie algebra of M can be written as a sum of $\lfloor \frac{d-1}{2} \rfloor$ commutators as in (8.41).

Exercise 8.6.3. (a) Generalize [46, Th. 11.25] to the group $\text{SO}_{d,1}(\mathbb{R})^+$ considered here.

(b) Fix some $\varepsilon > 0$, and use mixing of A to show that there are $e^{(d-1)t+o(t)}$ many different periodic orbits with period in $[t, t + \varepsilon)$.

(c) Use (b) and the proof of Theorem 8.24 to strengthen Theorem 8.24.

Notes to Chapter 8

⁽³⁷⁾(Page 199) Clearly a uniquely ergodic map has a maximal measure; what is much less clear is whether this arises in an interesting way. The examples we have seen of uniquely ergodic maps have zero entropy. A deep result of Hahn and Katznelson [81] is that there are minimal uniquely ergodic maps with positive entropy. Corollary 8.1 was shown by Dinaburg [42, 43] under the assumption that $\dim(X) < \infty$. Corollary 8.1 as stated is taken from Goodman [75].

⁽³⁸⁾(Page 200) This measure was constructed by Parry [163], who also showed its characterization as the unique maximal measure (which will be shown in Section 8.1.1).

⁽³⁹⁾(Page 206) We follow the approach of Bowen [26] and of Adler and Weiss [5], [6] closely here.