Chapter 4

Entropy and mixing of affine automorphisms

In this chapeter, we shall compute the topological entropy of affine automorphisms of a compact nilmanifold $M = G/\Gamma$. We will also provide a criterion for such automorphisms to be mixing. In particular, it will be established that for a linear automorphism $T \in \mathrm{Aut}(M)$, T is K-mixing if and only if it is weakly mixing. The arguments in this chapter will follow a mixture of [ELW18, Ch. 2 & Ch. 6] and [Par69a].

4.1 Basics of entropy theory

We start by briefly recalling the definitions of topological and measure-theoretic entropies.

Let (X, T) be a topological dynamical system, where X is a compact metric space with metric d. Then for every $n \in \mathbb{N}$, there is a natural embedding $X \to X^n$ by $\iota_N(x) = (x, Tx, \cdots, T^{N-1}x)$, which sends a point to the segment of length n at the beginning of its T-orbit. Denote by $d_N = \iota_N^* d^N$ the pull back of the l^∞-metric $d^N(\mathbf{x}, \mathbf{y}) = \max_{n=0}^{N-1} d(x_n, y_n)$ on X^n, then

$$d_N(x, y) = \max_{n=0}^{N-1} d(T^n x, T^n y), \forall x, y \in X. \tag{4.1}$$

For all compact subsets $Y \subseteq X$, denote

$$S_{N,\epsilon}(Y) = \text{smallest number of } \epsilon\text{-balls in } d_N \text{ needed to cover } Y. \tag{4.2}$$

In the sequel, we will denote respectively by $B_\epsilon(x)$ and $B_{N,\epsilon}(x)$ the open balls centered at x of radius ϵ according to d and d_N. The set $B_{N,\epsilon}(x)$ is called a **Bowen ball**.

Definition 4.1.1. *The* **topological entropy** *of a topological dynamical system* (X, T) *on a compact metric space is*

$$h_{\text{top}}(T) = \lim_{\epsilon > 0} \lim_{N \to \infty} \frac{1}{N} \log S_{N,\epsilon}(X).$$

This is well-defined, though possibly has an infinite value, because of the facts listed below. First, for a given ϵ, $\log S_{N,\epsilon}(X)$ is a subadditive sequence:

$$\log S_{N+M,\epsilon}(X) \leq \log S_{N,\epsilon}(X) + \log S_{M,\epsilon}(X).$$

Moreover, $S_{N,\epsilon}(X)$ is clearly increasing as $\epsilon \to 0$. For more details, see [ELW18, Ch. 6].

One may generalize this notion to non-compact spaces.

Definition 4.1.2. *Suppose* (X, d) *is a locally compact* σ*-compact metric space, and* $T : X \to X$ *be a map that is uniformly continuous with respect to* d*, then the* **topological entropy** *of* (X, T, d) *is*

$$h_{\text{top}}(T) = \sup_{\text{compact } Y \subseteq X} \lim_{\epsilon > 0} \lim_{N \to \infty} \frac{1}{N} \log S_{N,\epsilon}(Y).$$

Here we keep the notations (4.1) and (4.2). The definition again makes sense for the same reasons as before. Furthermore, because $S_{N,\epsilon}(Y)$ is increasing as Y enlarges, instead of $\sup_{\text{compact } Y \subseteq X}$ one may write $\lim_{Y_k \to X}$ where $\{Y_k\}$ is an increasing sequence of compact subsets such that $\bigcup_{k=1}^{\infty} Y_k = X$.

Definitions 4.1.1 and 4.1.2 are related by:

Lemma 4.1.3. *Suppose a discrete group* Γ *acts freely and properly* [1] *on a metric space* (X, d) *from the right by isometries. Let* (X', d') *be the quotient space* X/Γ *equipped with the quotient metric* d'*, so that the projection* $\pi : X \to X'$ *is locally an isometry*[2].

Assume in addition that X' *is compact, and* $T' \circ \pi = \pi \circ T$ *for uniformly continuous maps* $T : X \to X$*,* $T' : X' \to X'$*. Then* $h_{\text{top}}(T) = h_{\text{top}}(T')$*.*

Proof. For every $x \in X$, there is a radius $\delta_x > 0$ such that π is an isometry between $B_{\delta_x}(x)$ and its image $\pi(B_{\delta_x}(x))$. Because Γ acts isometrically, $\delta_x = \delta_{x\gamma}$ for all $x \in X$ and $\gamma \in \Gamma$.

By compactness of X', it is covered by $\bigcup_{i=1}^{k} \pi(B_{\frac{1}{3}\delta_{x_i}}(x_i))$ for finitely many x_i's. For $\delta = \frac{1}{3} \min_{i=1}^{k} \delta_{x_i}$, we claim every ball of radius δ in X is

[1] An action $X \curvearrowleft \Gamma$ is proper if the map $(x, \gamma) \to (x, x\gamma)$ is proper.

[2] The construction of X' relies on the properness of the action

projected injectively to X'. Indeed, if $\pi(z_1) = \pi(z_2)$ and $d(z_1, z_2) < \delta$, choose $x \in \{x_1, \cdots, x_k\}$ such that $\pi(z_1) \in \pi(B_{\frac{1}{2}\delta_x}(x))$. Then there are $\gamma_1, \gamma_2 \in \Gamma$ such that $z_j \in B_{\frac{1}{3}\delta_x}(x\gamma_j)$. Since $d(z_1, z_2) < \delta$, it follows $d(x\gamma_1, x\gamma_2) < \delta + \frac{2}{3}\delta_x \le \delta_x$, or equivalently $d(x, x\gamma_2\gamma_1^{-1}) < \delta_x$. This contradicts the choice of δ_x, by which π is injective on $B_{\delta_x}(x)$.

Because T is uniformly continuous, there exists $\epsilon_0 \in (0, \delta)$ such that if $d(x, y) < \epsilon_0$, then $d(Tx, Ty) < \delta$. We prove the following claim:

Claim 4.1.4. *For all $\epsilon \in (0, \epsilon_0)$, $N \in \mathbb{N}$, and $x \in X$, π is an isometry (and thus bijective) between $B_{N,\epsilon}(x)$ and $B_{N,\epsilon}(\pi(x))$.*

We now prove the claim above. By the choice above, we know π is injective on $B_{N,\epsilon}(x) \subseteq B_\epsilon(x)$. It suffices to prove the image is $B_{N,\epsilon}(\pi(x))$. For every $z \in B_{N,\epsilon}(x)$ and $0 \le n \le N-1$, $d(T^n x, T^n z) < \epsilon$ and thus

$$d((T')^n\pi(x), (T')^n\pi(z)) = d(\pi(T^n x), \pi(T^n z)) < \epsilon.$$

So $\pi(B_{N,\epsilon}(x)) \subseteq B_{N,\epsilon}(\pi(x))$. On the other hand, if $z' \in B_{N,\epsilon}(\pi(x)) \subseteq B_\epsilon(\pi(x)) = \pi(B_\epsilon(x))$, then $z' = \pi(z)$ for some $z \in B_\epsilon(x)$. It can be shown inductively that $d(T^n x, T^n z) < \epsilon$. Indeed, this is true for $n = 0$ by construction. Suppose $d(T^{n-1}x, T^{n-1}z) < \epsilon$, then $d(T^n x, T^n z) < \delta$ so $T^n z \in B_\delta(T^n x)$ Furthermore, when $n \le N-1$,

$$d(\pi(T^n x), \pi(T^n z)) = d((T')^n\pi(x), (T')^n z') < \epsilon,$$

Since π is an isometry on $B_\delta(T^n x)$, $d(T^n x, T^n z) < \epsilon$. Therefore $z \in B_{N,\epsilon(x)}$, which implies $\pi(B_{N,\epsilon}(x)) = B_{N,\epsilon}(\pi(x))$. The claim is established.

Given the claim, we know that $S_{N,\epsilon}(X') \le S_{N,\epsilon}(Y)$ where $Y \subseteq X$ is a sufficiently large compact set such that $\pi(Y) = X'$ (it suffices to choose, for example, $\bigcup_{i=1}^k \overline{B_{\delta_{x_i}}(x_i)}$.) It follows that $h_{\text{top}}(T') \le h_{\text{top}}(T)$.

On the other hand, any compact subset $Y \subset X$ is covered by a finitely union balls $\bigcup_{i=1}^K B_\delta(y_i)$ of radius δ, where K depends only on π, Y and δ. Each $\pi(B_\delta(y_i))$ can be covered by the union of $S_{N,\epsilon}(X')$ Bowen balls $B_{N,\epsilon}(z'_{ij})$, $1 \le j \le S_{N,\epsilon}(X')$. Here each z'_{ij} is in $\pi(B_\delta(y_i))$ has a lift z_{ij} in $B_\delta(y_i)$. Thus the union of $B_{N,\epsilon}(z_{ij}) = (\pi|_{B_\delta(y_i)})^{-1}(B_{N,\epsilon}(z'_{ij}) \cap B_\delta(y_i))$, over $j = 1, \cdots, S_{N,\epsilon}(X')$, covers $B_\delta(y_i)$. So Y can be covered by $K \cdot S_{N,\epsilon}(X')$ ϵ-balls in distance d_N. Therefore

$$\lim_{n\to\infty} \frac{1}{N} \log S_{N,\epsilon}(Y) \le \lim_{n\to\infty} \frac{1}{N}(\log S_{N,\epsilon}(X') + \log K) = \lim_{n\to\infty} \frac{1}{N} \log S_{N,\epsilon}(X')$$

for all compact subsets Y, and thus $h_{\text{top}}(T) \le h_{\text{top}}(T')$. The proof is completed. $\square$

We now briefly review the theory of measure-theoretic entropy. The notions and theorems below can be found in, for example, [ELW18, Ch. 2].

It is not hard to see that

Lemma 4.1.5. *For $n \in \mathbb{N}$, $h_\mu(T^n|\mathcal{A}) = nh_\mu(T|\mathcal{A})$.*

Hereafter, let $(X, \mathcal{B}, T, \mu)$ be a measure preserving dynamical system, $\mathcal{A} \subseteq \mathcal{B}$ be a countably generated T-invariant σ-subalgebra. From the discussion in §3.2, we have conditional measure $\mu_x^{\mathcal{A}}$, supported on the atom $[x]^{\mathcal{A}}$, for μ-almost every $x \in X$.

Definition 4.1.6. *For a finite measurable partition $\mathcal{Q}$ of X, define the* **conditional information function**

$$I_\mu(\mathcal{Q}|\mathcal{A})(x) = -\log \mu_x^{\mathcal{A}}([x]^{\mathcal{Q}}),$$

where $[x]^{\mathcal{Q}}$ is the atom of $\mathcal{Q}$ containing x. The **conditional entropy** *of $\mathcal{Q}$, with respect to μ and conditional to $\mathcal{A}$, is given by*

$$H_\mu(\mathcal{Q}|\mathcal{A}) = \int_X I_\mu(\mathcal{Q}|\mathcal{A})(x)\mathrm{d}\mu(x).$$

When $\mathcal{A}$ is the trivial σ-algebra modulo μ, i.e. only consists of null and conull sets, one can omit the symbol "$|\mathcal{A}$" in $I_\mu(\mathcal{Q}|\mathcal{A})(x)$ and $H_\mu(\mathcal{Q}|\mathcal{A})$ and the word "conditional" above.

The absolute and conditional entropies are related by

$$H_\mu(\mathcal{Q}|\mathcal{A}) = \int_X H_{\mu_x^{\mathcal{A}}}(\mathcal{Q})\mathrm{d}\mu(x). \tag{4.3}$$

The quantity $H_\mu(\mathcal{Q}|\mathcal{A})$ is increasing in $\mathcal{Q}$ and decreasing in $\mathcal{A}$, where the orderings of partitions and σ-algebras are given by refinements. This can be proved using the Jensen inequality and the fact that $x \to -x \log x$ is a concave function on $[0, 1]$. Moreover, it is subadditive in $\mathcal{Q}$:

$$H_\mu(\mathcal{P} \vee \mathcal{Q}|\mathcal{A}) \le H_\mu(\mathcal{P}|\mathcal{A}) + H_\mu(\mathcal{Q}|\mathcal{A}). \tag{4.4}$$

Here and below, $\mathcal{P} \vee \mathcal{Q}$ denotes the coarsest common refinement $\{P \cap Q : P \in \mathcal{P}, Q \in \mathcal{Q}\}$ of two partitions $\mathcal{Q}$ and $\mathcal{Q}$. Indeed,

$$H_\mu(\mathcal{P} \vee \mathcal{Q}|\mathcal{A}) = H_\mu(\mathcal{P}|\mathcal{Q} \vee \mathcal{A}) + H_\mu(\mathcal{Q}|\mathcal{A}), \tag{4.5}$$

where $\mathcal{Q} \vee \mathcal{A}$ is the σ-algebra generated by $\{Q \cap A : Q \in \mathcal{Q}, A \in \mathcal{A}\}$.

Notice that when $\mathcal{A}$ is trivial modulo μ, $\mu_x^{\mathcal{A}} = \mu$ for μ-a.e. x. It is clear that $0 \le H_\mu(\mathcal{Q}|\mathcal{A}) \le H_\mu(\mathcal{Q})$. From (4.3), it is not hard to see that $H_\mu(\mathcal{Q}|\mathcal{A}) \le \log n$ if every atom of $\mathcal{A}$ intersects no more than n atoms of $\mathcal{Q}$. In particular, $H_\mu(\mathcal{Q}) \le \log(\#\mathcal{Q})$.

Since $\mathcal{A}$ and μ are both T-invariant, $H_\mu(\mathcal{Q}|\mathcal{A}) = H_\mu(T^{-1}\mathcal{Q}|\mathcal{A})$ for the partition $T^{-1}\mathcal{Q} = \{T^{-1}(P) : P \in \mathcal{Q}\}$. Therefore,

$$H_\mu(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}|\mathcal{A}) = \sum_{k=0}^{N-1} H_\mu\big(\mathcal{Q}\big|(\bigvee_{n=1}^{k} T^{-n}\mathcal{Q}) \vee \mathcal{A}\big).$$

By the monotonicty above, $H_\mu\big(\mathcal{Q}\big|(\bigvee_{n=1}^{k} T^{-n}\mathcal{Q}) \vee \mathcal{A}\big)$ is decreasing as k grows and thus the limit

$$h_\mu(T, \mathcal{Q}|\mathcal{A}) = \lim_{N\to\infty} \frac{1}{N} H_\mu(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}|\mathcal{A}) = \lim_{N\to\infty} H_\mu\big(\mathcal{Q}\big|(\bigvee_{n=1}^{N} T^{-n}\mathcal{Q}) \vee \mathcal{A}\big)$$

exists. [3]

Definition 4.1.7. *For a countably generated T-invariant σ-subalebra $\mathcal{A}$ and a T-invariant probability measure μ, the* **conditional measure-theoretic entropy** *of T, with respect to μ, and conditional to $\mathcal{A}$, is $h_\mu(T|\mathcal{A}) = \sup_{\mathcal{Q}} h_\mu(T, \mathcal{Q}|\mathcal{A})$ where the supremum is taken over all finite measurable partitions. This is called the* **measure-theoretic entropy** *or* **Kolmogrov-Sinai entropy** *of T when $\mathcal{A}$ is trivial modulo μ, denoted by $h_\mu(T)$.*

The measure-theoretic analogue to Lemma 4.1.5

Lemma 4.1.8. *For $n \in \mathbb{N}$, $h_\mu(T^n|\mathcal{A}) = n h_\mu(T|\mathcal{A})$.*

Two fundamental theorems about the measure-theoretic entropy are:

Theorem 4.1.9 (Variational Principle)**.** *If $(X, \mathcal{B}, T)$ is a topological dynamical system, then*

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

where the supremum can be taken either over all T-invariant probability measures or all ergodic T-invariant probability measures.

Theorem 4.1.10 (Kolmogrov-Sinai)**.** *If $(\mathcal{Q}_k)$ is an increasing sequence of finite measurable partitions, which together generates the σ-algebra $\mathcal{B}$, then*

$$h_\mu(T|\mathcal{A}) = \lim_{k\to\infty} h_\mu(T, \mathcal{Q}_k|\mathcal{A}) = \sup_{k\to\infty} h_\mu(T, \mathcal{Q}_k|\mathcal{A}).$$

[3] It follows from the Martingale Convergence Theorem that $h_\mu(T, \mathcal{Q}|\mathcal{A}) = H_\mu\big(\mathcal{Q}\big|(\bigvee_{n=1}^{\infty} T^{-n}\mathcal{Q}) \vee \mathcal{A}\big)$, where $\bigvee_{n=1}^{\infty} T^{-n}\mathcal{Q}$ is the σ-algebra generated by all the $T^{-n}\mathcal{Q}$'s.

Another important theorem that we need is the Abramov-Rokhlin formula:

Theorem 4.1.11 (Abramov-Rokhlin formula)**.** *Suppose* $\pi : (X, \mathcal{B}, T, \mu) \to (Y, \mathcal{A}, S, \nu)$ *is a factor map between measure preserving dynamical systems. Then*

$$h_\mu(T) = h_\nu(S) + h_\mu(T|\pi^{-1}\mathcal{A}).$$

Exercises

Exercise 4.1.1. Prove that the translation $Tx = x + b$ on the torus $\mathbb{T}^d$ has topological entropy 0.

Exercise 4.1.2. Prove that if a homeomorphism $T : \mathbb{T}^1 \to \mathbb{T}^1$ is homotopic to identity, then it has topological entropy 0.

Exercise 4.1.3. Let ρ be a joining (see Definition 4.4.3 later) between measure preserving dynamical systems $(X, \mathcal{B}, T, \mu)$ and $(Y, \mathcal{A}, S, \nu)$. Show that $h_\rho(T \times S) \leq h_\mu(T) + h_\mu(S)$.

4.2 Topological entropy of affine automorphisms

We will first study $h_{\text{top}}(T)$ for the affine transform $Tx = Ax + b$ on $X = \mathbb{R}^d$, where $A \in \text{GL}(d, \mathbb{R})$ and $b \in \mathbb{R}^d$. We will use the Euclidean distance on $\mathbb{R}^d$, and the standard volume form $\mathbf{m}_{\mathbb{R}^d}$.

First of all, notice that for all $x, y, v \in \mathbb{R}^d$ and $n \geq 0$, $T^n(x + v) - T^n x = T^n(y + v) - T^n y = A^n v$. Therefore,

$$B_{N,\epsilon}(y) = B_{N,\epsilon}(x) + (y - x), \forall N \in \mathbb{N}, \forall x, y \in \mathbb{R}^d. \tag{4.6}$$

Proposition 4.2.1. *If* $Tx = Ax + b$ *on* $X = \mathbb{R}^d$*, where* $A \in \text{GL}(d, \mathbb{R})$ *and* $b \in \mathbb{R}^d$*, then* $h_{\text{top}}(T)$ *is given by the volume entropy*

$$h_{\text{vol}}(T) = \lim_{\epsilon \to 0} \limsup_{N \to \infty} -\frac{1}{N} \log \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x)),$$

which is independent of x.

Proof. By (4.6), the volume entropy is independent of x.

Let $Y_k \subseteq \mathbb{R}^d$ be the compact ball $\overline{B_k(0)}$ of radius k centered at 0. Then $h_{\text{top}}(T) = \sup_{k\to\infty} \lim_{\epsilon>0} \lim_{n\to\infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k)$. Because $S_{N,\epsilon}(Y) \geq \frac{\mathbf{m}_{\mathbb{R}^d}(Y_k)}{\mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x))}$, we see that

$$\lim_{n\to\infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k) \geq \limsup_{N\to\infty} -\frac{1}{N} \log \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x)), \forall k$$

and $h_{\text{top}}(T) \geq h_{\text{vol}}(T)$ after taking limit in k and ϵ.

On the other hand, assuming $0 < \epsilon < 1$, write $S'_{N,\frac{1}{2}\epsilon}(Y_k)$ for the maximal number of disjoint Bowen balls of the form $B_{N,\frac{1}{2}\epsilon}(z)$ that Y_k can contain. Then $S'_{N,\frac{1}{2}\epsilon}(Y_{k+1}) \geq S_{N,\epsilon}(Y_k)$. In fact, if $B_{N,\epsilon}(z_i)$, $i = 1, \cdots, S'$ are disjoint and contained in Y_{k+1}, then every $z \in Y_k$ must be covered by one of the $B_{N,\epsilon}(z_i)$'s as otherwise $B_{N,\frac{1}{2}\epsilon}(z) \subseteq B_{k+\frac{1}{2}\epsilon}(0) \subseteq Y_{k+1}$ would be disjoint from each $B_{N,\frac{1}{2}\epsilon}(z_i)$ and could be added into the collection of disjoint Bowen balls, contradicting the maximality of this collection.

It follows that $S_{N,\epsilon}(Y_k) \leq S'_{N,\frac{1}{2}\epsilon}(Y_{k+1}) \leq \frac{\mathbf{m}_{\mathbb{R}^d}(Y_{k+1})}{\mathbf{m}_{\mathbb{R}^d}(B_{N,\frac{1}{2}\epsilon}(x))}$, so

$$\lim_{n\to\infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k) \leq \limsup_{N\to\infty} -\frac{1}{N} \log \mathbf{m}_{\mathbb{R}^d}(B_{N,\frac{1}{2}\epsilon}(x)), \forall k$$

and $h_{\text{top}}(T) \leq h_{\text{vol}}(T)$ after taking limit. The proof is complete. □

Therefore, in order to calculate $h_{\text{top}}(T)$, it suffices to estimate the size of $B_{N,\epsilon}(x)$.

Like in §3.3, decompose $\mathbb{C}^d$ as the direct sum $\bigoplus_\lambda V_{\mathbb{C}}^\lambda$ of generalized eigenspaces, where $V_{\mathbb{C}}^{\bar\lambda} = \overline{V_{\mathbb{C}}^\lambda}$. Then $\mathbb{R}^d = \bigoplus_{\operatorname{Im}\lambda\geq 0} V^\lambda$, where

$$V^\lambda = \begin{cases} \ker_{\mathbb{R}^d}(A - \lambda \mathrm{Id})^d, & \text{if } \lambda \in \mathbb{R}; \\ (V_{\mathbb{C}}^\lambda \oplus V_{\mathbb{C}}^{\bar\lambda}) \cap \mathbb{R}^d, & \text{if } \lambda \notin \mathbb{R}. \end{cases}$$

For each eigenvalue λ, write

$$|\lambda|_+ = \max(1, |\lambda|).$$

Proposition 4.2.2. *If $\mathbb{R}^d = V^\lambda$ for some $\lambda \in \mathbb{C}\backslash\{0\}$, then there exists $K > 1$ that depends only on A and d, such that for all $\epsilon > 0$,*

$$B_{K^{-1}N^{-(d-1)}|\lambda|_+^{-N}\epsilon}(x) \subseteq B_{N,\epsilon}(x) \subseteq B_{KN^{(d-1)}|\lambda|_+^{-N}\epsilon}(x).$$

Lemma 4.2.3. *Under the hypothesis of Proposition 4.2.2, there are a decomposition $A = A_0 J$ and a Hilbert norm $|\cdot|_0$ on $\mathbb{R}^d$ such that $|A_0 v|_0 = |\lambda||v|_0$ for all $v \in \mathbb{R}^d$, J is a unipotent matrix, and A_0, J commute.*

Proof. Let

$$A_0 = \begin{cases} \lambda \mathrm{Id}, & \text{if } \lambda \in \mathbb{R}; \\ \lambda \mathrm{Id}|_{V_{\mathbb{C}}^{\lambda}} \oplus \bar{\lambda}\mathrm{Id}|_{V_{\mathbb{C}}^{\bar{\lambda}}}, & \text{if } \lambda \notin \mathbb{R}. \end{cases}$$

Clearly $|A_0 v| = |\lambda| \cdot |v|$ if $\lambda \in \mathbb{R}$. In the imaginary case, let $|\cdot|_0$ be the norm given by a non-degenerate inner product that makes $V_{\mathbb{C}}^{\lambda}$ and $V_{\mathbb{C}}^{\bar{\lambda}}$ orthogonal. Then $|A_0 v|_0 = |\lambda||v|_0$. $J = A_0^{-1} A$ has only eigenvalue 1 and is hence unipotent. Moreover, J preserves both $V_{\mathbb{C}}^{\lambda}$ and $V_{\mathbb{C}}^{\bar{\lambda}}$, on both of which A_0 acts by scalar multiplication, hence A_0 commutes with J.

Finally, in the case when $\lambda \notin \mathbb{R}$, though A_0 is defined as a complex valued matrix, it commutes with complex conjugation and is therefore actually real valued. The lemma is established. □

Since $K_1^{-1}|\cdot| \leq |\cdot|_0 \leq K_1|\cdot|$ for some $K_1 > 1$, to prove Proposition 4.2.2 one may assume without loss of generality that $|\cdot| = |\cdot|_0$, by changing the value of C if necessary.

Proof of Proposition 4.2.2. By (4.6), it is enough to assume $x = 0$. Then $T^n y - T^n 0 = A^n y = J^n A_0^n y$.

Lower bound. If $|y| < K^{-1} N^{-(d-1)} |\lambda|_+^{-N} \epsilon$, then for all $0 \leq n \leq N-1$, $|A_0^n y| = |\lambda|^n |y| < K^{-1} N^{-(d-1)} \epsilon$. Since J is unipotent, all entries of J are polynomials in n of degree less than d, and $|J^n A_0^n y| < \epsilon$ if K is chosen to be sufficiently large. Therefore $|T^n y - T^n 0| < \epsilon$ for all $0 \leq n \leq N-1$, or in other words $y \in B_{N,\epsilon}(0)$.

Upper bound. Suppose $y \in B_{N,\epsilon}(x)$, i.e. $|T^n y| < \epsilon$ for all $0 \leq n \leq N-1$. This is equivalent by Lemma 4.2.3 to that

$$|J^n y| < |\lambda|^{-n} \epsilon, \forall n = 0, \cdots, N-1. \tag{4.7}$$

If $|\lambda| \leq 1$, then $|\lambda|_+ = 1$ and we can take $n = 0$ in (4.7). This yields $|y| \leq \epsilon$ and is sufficient for the upper bound we need, with $K = 1$.

We now assume $|\lambda| = |\lambda|_+ > 1$. In this case, let $n = N-1$ in (4.7). Then

$$\begin{aligned} |y| \leq & \|J^{-(N-1)}\| \cdot |J^{N-1} y| \leq K_2 N^{d-1} \cdot |\lambda|^{-(N-1)} \epsilon \\ = & K_2 |\lambda| \cdot N^{d-1} |\lambda|_+^{-N} \epsilon. \end{aligned}$$

Here K_2 is a constant depending only on J, so $K = K_2|\lambda|$ depends only on A. This proves the upper bound. □

We now return to the general case where $\mathbb{R}^d$ is the direct sum of finitely many V^λ's.

Corollary 4.2.4. *In the setting of Proposition 4.2.1, there exists a constant K that depends only on A and d, such that for all $x \in \mathbb{R}^d$,*

$$K^{-1}N^{-d^2}e^{-hN}\epsilon^d \leq \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x)) \leq KN^{d^2}e^{-hN}\epsilon^d,$$

for

$$h = \sum_\lambda \log|\lambda|_+, \tag{4.8}$$

where the sum is taken over all eigenvalues λ of A, with multiplicities counted.

Proof. First, remark that because any two inner products on $\mathbb{R}^d$ bound each other up to a multiplicative constant, the statement of the corollary is not affected by switching to a different inner product on $\mathbb{R}^d$ after taking a different value of K if necessary. By doing so, we can assume without loss of generality that all the subspaces $V^\lambda, \operatorname{Im}\lambda \geq 0$ are orthogonal to each other.

The map T is a direct product of affine transforms T^λ on V^λ, where $T^\lambda x = A^\lambda x + b^\lambda$, where A^λ only has eigenvalues λ and $\bar{\lambda}$, and b^λ is the component of b in V^λ. Similarly, decompose every $x \in \mathbb{R}^d$ as $\sum_\lambda x^\lambda$. Denote by $B^{T^\lambda}_{N,\epsilon}(x^\lambda)$ the Bowen ball of step N and radius parameter ϵ around $x^\lambda \in V^\lambda$. Then

$$\prod_{\operatorname{Im}\lambda\geq 0} B^{T^\lambda}_{N,\frac{\epsilon}{d}}(x^\lambda) \subseteq B_{N,\epsilon}(x) \subseteq \prod_{\operatorname{Im}\lambda\geq 0} B^{T^\lambda}_{N,\epsilon}(x^\lambda). \tag{4.9}$$

By Proposition 4.2.2,

$$K^{-1}(N^{-d^\lambda}|\lambda|_+^{-N}\epsilon)^{d^\lambda} \leq \mathbf{m}_{V^\lambda}(B^{T^\lambda}_{N,\epsilon}(x^\lambda)) \leq K(N^{d^\lambda}|\lambda|_+^{-N}\epsilon)^{d^\lambda}, \tag{4.10}$$

where $d^\lambda = \dim V^\lambda$ and K depends only on A and d^λ.

After taking product, (4.9) and (4.10) together imply that

$$\begin{aligned} K^{-d}N^{-\sum_{\operatorname{Im}\lambda\geq 0}(d^\lambda)^2} \prod_{\operatorname{Im}\lambda\geq 0} |\lambda|_+^{-N}\epsilon^d &\leq \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x)) \\ &\leq K^d N^{\sum_{\operatorname{Im}\lambda\geq 0}(d^\lambda)^2} \prod_{\operatorname{Im}\lambda\geq 0} |\lambda|_+^{-N}\epsilon^d. \end{aligned} \tag{4.11}$$

By switching to a different K and noting that $\sum_{\operatorname{Im}\lambda\geq 0}(d^\lambda)^2 \leq d^2$, (4.12) becomes

$$\begin{aligned} K^{-1}N^{-d^2}\Big(\prod_{\operatorname{Im}\lambda\geq 0} |\lambda|_+^{d_\lambda}\Big)^{-N}\epsilon^d &\leq \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x)) \\ &\leq KN^{d^2}\Big(\prod_{\operatorname{Im}\lambda\geq 0} |\lambda|_+^{d_\lambda}\Big)^{-N}\epsilon^d, \end{aligned} \tag{4.12}$$

where K depends only on A and d, and the product $\prod_{\mathrm{Im}\,\lambda\geq 0}|\lambda|_+^{d_\lambda}$ is taken over all eigevalues λ with non-negative real part, without counting multiplicities. To conclude, it suffices to notice that this product is equal to e^{-h} for the quantity h in (4.8), as imaginary eigevalues appear in conjugate pairs λ, $\bar{\lambda}$. □

We are now ready to state the main theorems of this section.

Proposition 4.2.5. *If $Tx = Ax + b$ on $X = \mathbb{R}^d$, where $A \in \mathrm{GL}(d, \mathbb{R})$ and $b \in \mathbb{R}^d$, then $h_{\mathrm{top}}(T)$ is given by (4.8).*

Proof. This follows from Proposition 4.2.1 and Corollary 4.2.4. □

By the variational principle, we know that $h_{\mathbf{m}_{\mathbb{T}^d}}(T) \leq h_{\mathrm{top}}(T)$. In the case of toral automorphisms, the inequality holds.

Theorem 4.2.6. *If $Tx = Ax + b$ on $X = \mathbb{T}^d$, where $A \in \mathrm{GL}(d, \mathbb{Z})$ and $b \in \mathbb{R}^d$, then*

$$h_{\mathbf{m}_{\mathbb{T}^d}}(T) = h_{\mathrm{top}}(T) = \sum_{\lambda} \log|\lambda|_+,$$

where the sum is taken over all eigenvalues λ of A, with multiplicities counted.

Proof. Denote by $\widetilde{T}$ the affine transform $x \to Ax + b$ on $\mathbb{R}^d$, then by Lemma 4.1.3 and Proposition 4.2.5, $h_{\mathrm{top}}(T) = h_{\mathrm{top}}(\widetilde{T}) = \sum_\lambda \log|\lambda|_+$.

For $L \in \mathbb{N}$, let $\mathcal{Q} = \mathcal{Q}_L$ be the measurable partition of $\mathbb{T}^d$ into L^d boxes, each of which is a translate of $[0, \frac{1}{L})^d$. Then

$$h_{\mathbf{m}_{\mathbb{T}^d}}(T) \geq h_{\mathbf{m}_{\mathbb{T}^d}}(T, \mathcal{Q}) = \lim_{N\to\infty} \frac{1}{n} H_{\mathbf{m}_{\mathbb{T}^d}}\Big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}\Big).$$

Note that if $\epsilon \geq \frac{\sqrt{d}}{L}$, then for all atoms P of $\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}$ and $x \in P$, $P \subset B_{N,\epsilon}(x)$. When L is sufficiently large, ϵ is bounded by the constant ϵ_0 in Claim 4.1.4, and we know that $B_{N,\epsilon}(x)$ is an isometrically projected copy of the Bowen ball $B_{N,\epsilon}(\tilde{x}) \subset \mathbb{R}^d$ with respect to $\widetilde{T}$. It then follows from Corollary 4.2.4 that

$$\mathbf{m}_{\mathbb{T}^d}(P) \leq \mathbf{m}_{\mathbb{T}^d}(B_{N,\epsilon}(x)) = \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(\tilde{x})) \leq KN^{d^2} e^{-h_{\mathrm{top}}(T)N} \epsilon^d$$

for some constant K that depends only on T. We obtain that

$$I_{\mathbf{m}_{\mathbb{T}^d}}\Big(\bigvee_{n=0}^{N-1} T^{-N}\mathcal{Q}\Big)(x) \geq N h_{\mathrm{top}}(T) - d^2 \log N - \log(K\epsilon^d)$$

for all x and thus

$$\frac{1}{N}H_{\mathbf{m}_{\mathbb{T}^d}}(\bigvee_{n=0}^{N-1}T^{-n}\mathcal{Q}) \geq \frac{1}{N}(Nh_{\text{top}}(T) - d^2\log N - \log(K\epsilon^d))$$
$$= h_{\text{top}}(T) - \frac{1}{N}(d^2\log N + \log(K\epsilon^d)).$$

By taking limit, we conclude that $h_{\mathbf{m}_{\mathbb{T}^d}}(T) \geq h_{\mathbf{m}_{\mathbb{T}^d}}(T,\mathcal{Q}) \geq h_{\text{top}}(T)$. On the other hand, $h_{\mathbf{m}_{\mathbb{T}^d}}(T) \leq h_{\text{top}}(T)$ by variational principle. The theorem is established. □

Exercises

Exercise 4.2.1. Let $X \subset [0,1]$ be the middle-$\frac{1}{3}$ Cantor set, and $T : X \to X$ be the continuous map $Tx = 3x \pmod 1$. Show that $h_{\text{top}}(T) = \log 2$.

Exercise 4.2.2. Let $X \subset (\mathbb{T}^1)^{\mathbb{Z}}$ be the compact set

$$\{(x_n)_{n\in\mathbb{Z}} : x_n \in \mathbb{T}^1, x_{n+1} = 2x_n, \forall n \in \mathbb{Z}\}$$

and let $T : X \to X$ be the shift map defined by $(Tx)_n = x_{n+1}$. Show that $h_{\text{top}}(T) = \log 2$.

4.3 Entropy on principal torus bundles and nilmanifolds

In this section, we adopt the settings from §3.4 with $K = \mathbb{T}^d$. In other words, we have a continuous factor map between two measure preserving dynamical systems $(X,\mathcal{B},T,\mu)$ and $(X_0,\mathcal{B}_0,T_0,\mu_0)$ such that: X a principal $\mathbb{T}^d$-bundle X over a compact base space X_0 where $\mathbb{T}^d$ is a compact abelian group; $\mathcal{B}$ is the product σ-algebra between $\mathcal{B}_0$ and the Borel σ-algebra $\mathcal{B}_{\mathbb{T}^d}$ of $\mathbb{T}^d$, μ_0 is ergodic and μ is invariant under translations by $\mathbb{T}^d$ (i.e. uniform along fibers); and for some automorphism A of $\mathbb{T}^d$ such that $T(zx) = A(z)Tx$ for $z \in \mathbb{T}^d$.

Theorem 4.3.1. *In these settings, $h_\mu(T) = h_{\mu_0}(T_0) + h_{\text{top}}(A)$.*

This theorem appeared in Parry's paper [Par69a], where it was attributed to Yuzvinskii [Juz65]. It was later generalized by Thomas [Tho71] to general compact groups $\mathbb{T}^d$.

Proof. By Abramov-Rokhlin formula (Theorem 4.1.11), it suffices to show:

$$h_\mu(T|\pi^{-1}\mathcal{B}_0) = h_{\text{top}}(A) = h_{\mathbf{m}_{\mathbb{T}^d}}(A).$$

We first set up a coordinate system on X. Fix a piecewise continuous section $\theta : X_0 \to X$, such that $\pi \circ \theta = \text{id}$. It determines a measurable isomorphism ψ between X and $X_0 \times \mathbb{T}^d$ by $\psi(z\theta(x_0)) = (x_0, z)$ for all $x_0 \in X_0$ and $z \in \mathbb{T}^d$.

The image $T\theta(x_0)$ sits in the fiber $\pi^{-1}(T_0x_0)$, so $T\theta(x_0) = w(x_0)\theta(T_0x_0)$ for some piecewise continuous function $w : X_0 \to \mathbb{T}^d$. In other words, $T\psi(x_0, 0)) = \psi(T_0x_0, w(x_0))$, where we view $\mathbb{T}^d$ as an additive group with the identity denoted by 0. It then follows that $T\psi(x_0, z)) = \psi(T_0x_0, w(x_0)+A(z))$. In other words, T is measurably isomorphic to the map $(x_0, z) \to (T_0x_0, w(x_0)+A(z))$ on $X_0 \times \mathbb{T}^d$. By abusing notation, we view X as $X_0 \times \mathbb{T}^d$ and and view T as the map above. For simplicity, we also write $\mathcal{B}_0$ for the σ-subalgebra $\pi^{-1}\mathcal{B}_0$ of $\mathcal{B}$. Define $\iota_{x_0} : \mathbb{T}^d \to X$ by $\iota_{x_0}(z) = (x_0, z)$. Then or $x = (x_0, z)$, $\mu_x^{\mathcal{B}_0}$ is the unique $\mathbb{T}^d$-invariant probability measure along the fiber $\pi^{-1}(x_0) = \mathbb{T}^d x$ and hence coincides with $(\iota_{x_0})_*\mathbf{m}_{\mathbb{T}^d}$.

Let $(\mathcal{Q}_L^{X_0})$ be an increasing sequences of measurable partitions of X_0, which generate $\mathcal{B}_0$ togther as a sequence. As in the proof of Theorem 4.2.6, let $\mathcal{Q}_L^{\mathbb{T}^d}$ be the measurable partition of $\mathbb{T}^d$ into L^d translated copies of $[0, \frac{1}{L})^d$. Define $\mathcal{Q}_L = \mathcal{Q}_L^{X_0} \times \mathcal{Q}_L^{\mathbb{T}^d}$. Then the $\mathcal{Q}_L$'s generate together the σ-algebra B. We know that $h_\mu(T|\mathcal{B}_0) = \lim_{i\to\infty} h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0)$ and $h_{\mathbf{m}_{\mathbb{T}^d}}(A) = \lim_{i\to\infty} h_{\mathbf{m}_{\mathbb{T}^d}}(A, \mathcal{Q}_L)$. Therefore it suffices to show: given any $\delta > 0$, for sufficiently large L,

$$|h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0) - h_{\mathbf{m}_{\mathbb{T}^d}}(A, \mathcal{Q}_L)| < \delta. \tag{4.13}$$

Because

$$\begin{aligned}
H_\mu\Big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}_L|\mathcal{B}_0\Big) &= \int_X H_{\mu_x^{\mathcal{B}_0}}\Big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}_L\Big)\mathrm{d}\mu(x) \\
&= \int_{X_0} H_{(\iota_{x_0})_*\mathbf{m}_{\mathbb{T}^d}}\Big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}_L\Big)\mathrm{d}\mu_0(x_0) \\
&= H_{\mathbf{m}_{\mathbb{T}^d}}\Big(\iota_{x_0}^{-1}\big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}_L\big)\Big)\mathrm{d}\mu_0(x_0).
\end{aligned}$$

we know

$$\begin{aligned} h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0) &= \lim_{N\to\infty} \frac{1}{N} \int_{X_0} H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\iota_{x_0}^{-1}\big(\bigvee_{n=0}^{N-1} T^{-n}\mathcal{Q}_L\big)\Big) \mathrm{d}\mu_0(x_0) \\ &= \lim_{N\to\infty} \frac{1}{N} \int_{X_0} H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} \iota_{x_0}^{-1} T^{-n}\mathcal{Q}_L\Big) \mathrm{d}\mu_0(x_0) \end{aligned} \tag{4.14}$$

Moreover,

$$h_{\mathbf{m}_{\mathbb{T}^d}}(A, \mathcal{Q}_L) = \lim_{N\to\infty} \frac{1}{N} H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big). \tag{4.15}$$

Because $T^n\iota_{x_0}(z) = T^n(x_0, z) = T^n(x_0, 0) + A^n(z)$, $\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L$ is a translate of $A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}$. This implies that for $L \geq 2$, every atom of $\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L$ intersects at most 2^d atoms of $A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}$; and every atom of $\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L$ intersects at most 2^d atoms of $A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}$. In consequence, for all $n \in \mathbb{N}$,

$$H(\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L|A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}) \leq \log 2^d, \;\; H(A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}|\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L) \leq \log 2^d.$$

By (4.5), this implies that for all x_0,

$$\begin{aligned} & H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} \iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L\Big) - H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big) \\ \leq & H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} \iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L\Big| \bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big) \\ \leq & \sum_{n=0}^{N-1} H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L\Big| \bigvee_{m=0}^{N-1} A^{-m}\mathcal{Q}_L^{\mathbb{T}^d}\Big) \\ \leq & \sum_{n=0}^{N-1} H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L\Big| A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big) \leq Nd\log 2. \end{aligned} \tag{4.16}$$

Similarly,

$$H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big) - H_{\mathbf{m}_{\mathbb{T}^d}} \Big(\bigvee_{n=0}^{N-1} \iota_{x_0}^{-1}T^{-n}\mathcal{Q}_L\Big) \leq Nd\log 2. \tag{4.17}$$

Combining the relations (4.14)-(4.17), we obtain the inequality (4.17) for $\delta = d\log 2$. In order to improve this to all $\delta > 0$, it suffices to note that for all

$n \in \mathbb{N}$, thanks to Lemma 4.1.8, $n|h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0) - h_{\mathbf{m}_{\mathbb{T}^d}}(A, \mathcal{Q}_L)| < d\log 2$ by applying this special case to T^n. The general case of (4.17) then follows by letting $n \to \infty$. We have therefore completed the proof of the theorem. □

Proposition 4.3.2. *If in Theorem 4.3.1, μ is not assumed to be $\mathbb{T}^d$-invariant, but still is a T-invariant extension of μ_0, then $h_\mu(T) \le h_{\mu_0}(T_0) + h_{\text{top}}(A)$.*

Proof. By Abramov-Rokhlin formula, it suffices to prove $h_\mu(T|\mathcal{B}_0) \le h_{\text{top}}(A)$. Without the assumption that μ is uniform along fibers, the same proof of the theorem would yield instead that

$$h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0) \le \lim_{N\to\infty} \frac{1}{N} \sup_\nu H_\nu\Big(\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}\Big), \tag{4.18}$$

where the supremum is taken over all probability measures ν on $\mathbb{T}^d$.

For $\epsilon = \frac{1}{2L}$, cover $\mathbb{T}^d$ with $S_{N,\epsilon}$ Bowen balls $B_{N,\epsilon}(x)$ with respect to A. For all x and n, $B_\epsilon(A^n x)$ intersects at most 2^d atoms of $\mathcal{Q}_L^{\mathbb{T}^d}$. Therefore $B_{N,\epsilon}(x)$ intersects no more than 2^{Nd} atoms of $\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}$. This shows the total number of atoms in $\bigvee_{n=0}^{N-1} A^{-n}\mathcal{Q}_L^{\mathbb{T}^d}$ is bounded by $2^{Nd}S_{N,\epsilon}$, and thus by (4.19),

$$h_\mu(T, \mathcal{Q}_L|\mathcal{B}_0) \le \lim_{N\to\infty} \frac{1}{N}(\log 2^{Nd}S_{N,\epsilon}) = \lim_{N\to\infty} \frac{1}{N}\log S_{N,\epsilon} + d\log 2.$$

By letting L tend to ∞, we obtain that

$$h_\mu(T|\mathcal{B}_0) \le h_{\text{top}}(A) + d\log 2. \tag{4.19}$$

Like in the proof of Theorem 4.3.1, one can get rid of the term $d\log 2$ by applying(4.19) to all powers T^n with n approaching ∞, using Lemmas 4.1.5 and 4.1.8. □

Theorem 4.3.3. *For an affine transform $Tx = bA(x)$ on a compact nilmanifold $M = G/\Gamma$, where $A \in \text{Aut}(M)$ and $b \in G$, then*

$$h_{\mathbf{m}_M}(T) = h_{\text{top}}(T) = \sum_\lambda \log|\lambda|_+,$$

where the sum is taken over all eigenvalues λ of D_eA, with multiplicities counted.

Proof. We first show that $h_{\mathbf{m}_{\mathbb{T}^d}}(T) = \sum_\lambda \log|\lambda|_+$. Recall that M has the structure of a tower of principal torus bundles

$$M = M_s \to M_{s-1} \to \cdots \to M_1 \to M_0 = \{\mathrm{pt}\},$$

where the fiber between M_i and M_{i-1} is a torus quotient of $G_{(i)}/G_{(i+1)} \cong \mathbb{R}^{d_i}$. The affines transform T_i that T induces on M_i is of the form $T_i x = b_i A_i(x)$ where b_i and A_i are respectively projections of b and A to $G/G_{(i+1)}$. In particular, for $x \in M_i$ and $z \in G_{(i)}/G_{(i+1)}$, $T_i(zx) = Tx + A_i(z)$. Since the restriction of A_i to $G_{(i)}/G_{(i+1)}$ is given by the matrix ψ_i in Proposition 3.1.5, which is from $\mathrm{GL}(d_i, \mathbb{Z})$ according to the proof of Corollary 3.1.9, where $\mathbb{Z}^{d_i}$ is identified with $\Gamma_{(i)}/\Gamma_{(i+1)}$.

By Theorem 4.3.1, $h_{\mathbf{m}_M}(T) = h_{\mathbf{m}_{M_{i+1}}}(T_{i+1}) + h_{\mathrm{top}}(\psi_i)$. Thus

$$\begin{aligned} h_{\mathbf{m}_M}(T) =& h_{\mathbf{m}_{M_s}}(T_s) = \sum_{i=1}^{s} h_{\mathrm{top}}(\psi_i) = \sum_{i=1}^{s} \sum_{\substack{\lambda:\ \text{eigenvalue of } \psi_i \\ \text{counting multiplicities}}} \log|\lambda|_+ \\ =& \sum_{\substack{\lambda:\ \text{eigenvalue of } D_eA \\ \text{counting multiplicities}}} \log|\lambda|_+, \end{aligned}$$

here $h_\top(\psi_i)$ is given by Theorem 4.2.6, and the last inequality is based on Proposition 3.1.5.

It remains to show that $h_{\mathrm{top}}(T) = h_{\mathbf{m}_M}(T)$. By the variational principle., $h_{\mathrm{top}}(T) \geq h_{\mathbf{m}_M}(T)$. On the other hand, for all T-invariant probability measures μ, we can show $h_\mu(T) \leq h_{\mathbf{m}_M}(T)$ by the argument above and Proposition 4.3.2. So the variational principle also implies $h_{\mathrm{top}}(T) \leq h_{\mathbf{m}_M}(T)$. □

Exercises

Exercise 4.3.1. Let T be an affine automorphism of a compact nilmanifold $M = G/\Gamma$, where G is a simply connected s-step nilpotent Lie group. Suppose $h : M \to G_{(s)}$ is a continuous function that is constant along $G_{(s)}$-orbits. Prove that the map $\widetilde{T} : x \to h(x).Tx$ on M has the same topological entropy as T.

4.4 Mixing of affine automorphisms of nilmanifolds

This section will first characterize weakly mixing affine automorphisms of a compact nilmanifold $M = G/\Gamma$. It will then be shown that they are all mixing and in fact K-mixing. As usual, let T_1 be the projection of an affine automorphism T to the horizontal torus $M_1 = G/G_{(2)}\Gamma \cong \mathbb{T}^{d_1}$. If $Tx = bA(x)$ then $T_1x = A_1x + b$ where $A_1 \in \mathrm{GL}(d_1, \mathbb{Z})$, identified with D_eA_1, is a projection of the matrix D_eA.

Theorem 4.4.1. *For an affine automorphism $Tx = bA(x)$ of M, the following are equivalent:*
(1) T is weakly mixing;
(2) T_1 is weakly mixing;
(3) A_1 has not roots of unity among its eigenvalues.

Proof. **(1)⇔(2).** Since $T_1 \times T_1$ is the projection of the affine automorphism $T \times T$ of $M \times M$ to its horizonal torus $M_1 \times M_1$. By Theorem 3.2.9 and Theorem 3.5.1, (1) $\Leftrightarrow$ $T \times T$ is ergodic $\Leftrightarrow$ $T_1 \times T_1$ is ergodic $\Leftrightarrow$ (2).

(2)→(3). Suppose A_1 has a non-trivial root of unity among its eigenvalues, then T_1 is not ergodic, and hence not weakly mixing, by Lemma 3.3.3.

If 1 is an eigenvalue of A_1, then by the discussion at the end of §3.3, T_1 has a maximal rotation factor S which is a translation of a torus Y. Because $S \times S$ preserves every translate the diagonal subtorus $\triangle_Y = \{(y, y) : y \in Y\}$ in $Y \times Y$, the uniform probability measure on $Y \times Y$ decomposes into the average of the uniform measures on these translated subtori, each of which is $S \times S$-invariant. So $S \times S$ is not ergodic and therefore S is not weakly mixing. It follows that T_1 is not weakly mixing either.

(3)→(2). If the eigenvalues of A_1 include no roots of unity, then neither do those of $A_1 \times A_1$, which is the linear part of $T_1 \times T_1$. So by Theorem 3.3.5, $T_1 \times T_1$ is ergodic and thus T_1 is weakly mixing. □

The remainder of this section will be used to prove:

Theorem 4.4.2. *If an affine automorphism T of a compact nilmanifold M is weakly mixing, then it is also K-mixing, and thus mixing.*

For this we need the notions of joining and disjointness, introduced by Furstenberg [Fur67]. The facts stated below can be found in [Par81, §4.3-4.4].

Definition 4.4.3. *Let $(X, \mathcal{B}_X, T, \mu)$ and $(Y, \mathcal{B}_Y, S, \nu)$ be measure preserving dynamical systems. A* **joining** *between μ and ν is a $T \times S$-invariant probability measure ρ on $X \times Y$ that projects respectively to μ and ν in both coordinates. The systems are called* **disjoint** *if $\mu \times \nu$ is the only joining.*

Suppose the systems are disjoint and $(Z, \mathcal{B}_Z, R, \rho)$ is another measure preserving dynamical systems that has both $(X, \mathcal{B}_X, T, \mu)$ and $(Y, \mathcal{B}_Y, S, \nu)$ as factors, with the factor maps respectively denoted by π_X and π_Y. Then $(\pi_X \times \pi_Y)_*\rho$ is a joining measure between μ and ν on $X \times Y$ and is thus equal to $\mu \times \nu$. This shows that for all $U \in \mathcal{B}_X$ and $V \in \mathcal{B}_Y$,

$$\begin{aligned} &\rho(\pi_X^{-1}(U) \cap \pi_Y^{-1}(V)) \\ =&\rho((\pi_X \times \pi_Y)^{-1}(U \times V)) = (\pi_X \times \pi_Y)_*\rho(U \times V) \\ =&(\mu \times \nu)(U \times V) = \mu(U)\nu(V). \end{aligned} \tag{4.20}$$

Since factor systems of a measure preserving dynamical system correspond to invariant σ-algebras (modulo null sets), (4.20) can be rephrased as:

Lemma 4.4.4. *For a measure preserving dynamical system $(X, \mathcal{B}, T, \mu)$, if for two T invariant σ-algebras $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{B}$, the factor systems $(X, \mathcal{A}_1, T, \mu)$ and $(X, \mathcal{A}_2, T, \mu)$ are disjoint, then $\mu(U_1 \cap U_2) = \mu(U_1)\mu(U_2)$ for all $U_i \in \mathcal{A}_i$, $i = 1, 2$.*

Pinsker [Pin60] proved that:

Theorem 4.4.5 (Pinsker)**.** *All K-mixing measure preserving dynamical systems are disjoint to all those of zero measure-theoretic entropy.*

In addition, he also introduced the so called Pinsker σ-algebra.

Definition 4.4.6. *Given a measure preserving dyamical system $(X, \mathcal{B}, T, \mu)$, the* **Pinsker σ-algebra** *is the collection $\mathcal{P} = \left\{U \in \mathcal{B} : h_\mu(T, \{U, U^c\}) = 0\right\}$.*

It follows from the subadditivity of entropy that $\mathcal{P}$ is a σ-subalgebra.

Lemma 4.4.7. *If the Pinsker σ-algebra is trivial, then $(X, \mathcal{B}, T, \mu)$ is K-mixing.*

Proof. Note that if $\mathcal{P}$ is trivial modulo μ, then for every T-invariant σ-subalgebra $\mathcal{A}$ that is non-trivial modulo μ, there is at least one $U \in \mathcal{A}$, such that $\mu(U) \neq 0$ and $\mu(U^c) \neq 0$. Then $U \notin \mathcal{P}$ and thus $h_\mu(T, \{U, U^c\}) > 0$. Therefore the dynamical system $(X, \mathcal{A}, T, \mu)$ has positive entropy. So $(X, \mathcal{B}, T, \mu)$ is K-mixing. ☐

Lemma 4.4.8. *For $k \in \mathbb{N}$, $(X, T, \mathcal{B}, \mu)$ and $(X, T^k, \mathcal{B}, \mu)$ have the same Pinsker σ-algebra.*

Proof. For the partition $\mathcal{U} = \{U, U^c\}$,

$$\begin{aligned}
&\frac{1}{kN} H_\mu(\bigvee_{n=0}^{kN-1} T^{-i}\mathcal{U}) \\
=&\frac{1}{kN} H_\mu(\bigvee_{j=0}^{k-1} T^{-j} \bigvee_{n=0}^{N-1} (T^k)^{-n}\mathcal{U}) \leq \frac{1}{kN}\sum_{j=0}^{k-1} H_\mu(T^{-j}\bigvee_{n=0}^{N-1}(T^k)^{-n}\mathcal{U}) \\
=&\frac{1}{kN}\cdot kH_\mu(\bigvee_{n=0}^{N-1}(T^k)^{-n}\mathcal{U}) = \frac{1}{N}H_\mu(\bigvee_{n=0}^{N-1}(T^k)^{-n}\mathcal{U}).
\end{aligned}$$

On the other hand,

$$\frac{1}{N}H_\mu(\bigvee_{n=0}^{N-1}(T^k)^{-n}\mathcal{U}) \leq \quad k\cdot\frac{1}{kN}H_\mu(\bigvee_{n=0}^{kN-1}T^{-i}\mathcal{U}).$$

By taking limit as $N \to \infty$, the above inequalities show that $h_\mu(T,\mathcal{U}) \leq h_\mu(T^k,\mathcal{U}) \leq kh_\mu(T,\mathcal{U})$. So $h_\mu(T,\mathcal{U}) = 0$ if and only if $h_\mu(T^k,\mathcal{U}) = 0$. The lemma follows. □

We are now ready to prove the key ingredient in Theorem 4.4.2.

Proposition 4.4.9. *In the setting of Theorem 4.3.1, suppose $\mathcal{A}$ is a T-invariant subalgebra of $\mathcal{B}$ that is invariant under both T and the $\mathbb{T}^d$-action $\{L_z : z \in \mathbb{T}^d\}$ on the principal $\mathbb{T}^d$-bundle X. Assume that $(X, \mathcal{A}, \mu)$ is ergodic for the $\mathbb{T}^d$-action, i.e. if $E \in \mathcal{A}$ is L_z-invariant modulo a null set with respect to μ for all $z \in \mathbb{T}^d$, then $\mu(E) \in \{0, 1\}$. There there exist a compact quotient group Y of $\mathbb{T}^d$ and an affine automorphism S of Y, in the form $Sy = c\Psi(y)$ where $c \in Y$ and $\Psi \in \mathrm{Aut}(Y)$, such that the measure preserving dynamical system $(X, \mathcal{A}, T, \mu)$ is isomorphic to $(Y, \mathcal{B}_Y, S, \mathbf{m}_Y)$, where $\mathcal{B}_Y$ is the Borel σ-algebra of Y.*

Proof. Consider an $\mathcal{A}$-measurable L^2-integrable function f. As $\mathcal{A} \subseteq \mathcal{B}$, $f \in L^2(\mathcal{B}, \mu)$ decomposes into the Fourier series $f(x) = \sum_{\xi\in\mathbb{Z}^d} \widehat{f}(\xi, x)$ along the $\mathbb{T}^d$-fibers, for which Lemma 3.4.2 holds. By the construction (3.12) and the $\mathbb{T}^d$-invariance of $\mathcal{A}$, each $\widehat{f}(\xi, x)$ is still $\mathcal{A}$-measurable. Suppose f_1, f_2 are both non-trivial $\mathcal{A}$-measurable fiberwise Fourier modes of the same frequency ξ, i.e. $f_i(x) = \widehat{f_i}(\xi, x)$, then the function $f_1\overline{f_2}$ is constant along the fibers, i.e. invariant under the $\mathbb{T}^d$-action, and at the same time $\mathcal{A}$-measurable. By the ergodicity of the $\mathbb{T}^d$-action on $(X, \mathcal{A}, \mu)$. The function $f_1\overline{f_2}$ is μ-almost everywhere a constant. In particular, for every Fourier mode, $|f| = (f\bar{f})^{\frac{1}{2}}$ is μ-a.e. constant. This shows $\overline{f_2}$ is a constant multiple of $\frac{1}{f_2}$ and therefore f_1 is a constant multiple of f_2.

Denote by Ξ the set of $\xi \in \mathbb{Z}^d$ for which non-trivial $\mathcal{A}$-measurable fiberwise Fourier modes exist. If f_1, f_2 are now such modes for different frequencies $\xi_1, \xi_2 \in \Xi$, then $f_1 f_2$ is such a Fourier mode for frequency $\xi_1 + \xi_2$, and $\frac{1}{f_i}$ is such a mode for $-\xi_i$. It follows that Ξ is a subgroup of $\mathbb{Z}^d$. Let $H = \ker \Xi$, then H is a closed subgroup of Ξ and the Pontryagin dual Y of Ξ is isomorphic to the compact quotient group $\mathbb{T}^d/H$.

For every $\xi \in \Xi$, fix a non-trivial $\mathcal{A}$-measurable fiberwise Fourier mode $\phi_\xi \in L^2(\mathcal{A}, \mu)$ of constant modulus 1. Because Ξ is a copy of $\mathbb{Z}^p$ for some $p \leq d$, one can do this first for a set of generators $\xi_1, \cdots, \xi_p$, and then define $\phi_\xi = \prod_{i=1}^p \phi_{x_i}^{n_i}$ for all $\xi = \sum_{i=1}^p n_i \xi_i$, $n_i \in \mathbb{Z}$. This makes the $\xi \to \phi_\xi$ a group morphism. Given $x \in X$, let $y_x(\xi) = \phi_\xi(x)$. Then y_x is a group morphism from Ξ to $\{u \in \mathbb{C} : |u| = 1\}$ for μ-a.e. x. In other words, $y_x \in \widehat{\Xi} = Y$.

Since ϕ_ξ is a fiberwise Fourier mode with frequency ξ, by Lemma 3.4.2, $y_{Tx}(\xi) = \phi_\xi(Tx) = (T\phi_\xi)(x)$ is a fiberwise Fourier mode of frequency $A^{\mathrm{T}}\xi$ and modulus 1, and thus equals $c_\xi \phi_{A^{\mathrm{T}}\xi}(x) = c_\xi y_x(A^{\mathrm{T}}\xi)$ for a constant $c(\xi)$ of modulus 1. Thus $y_{Tx} = c_\xi y_x(A^{\mathrm{T}}\xi)$. Because both y_x and y_{Tx} are multiplicative characters of Ξ, it follows that c is also such a character; that is, $c \in Y$. Furthermore, the transformation $\Psi y \to y \circ A^{\mathrm{T}}$ is an automorphism of Y. Hence $y_{Tx} = S(y_x)$, where $S(y) = c\Psi(y)$ is an affine transform of the torus Y.

To conclude, we need to show that $\iota : x \to y_x$ is an isomorphism between the measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}_Y, \mathbf{m}_Y)$.

As Ξ is the Pontryagin dual of Y, the continuous functions on Y are densely spanned by $\{\tau_\xi : y \to y(\xi) : \xi \in \Xi\}$. For each τ_ξ, $\tau_\xi \circ \iota$ sends x to $y_x(\xi) = \phi_\xi(x)$, and is $\mathcal{A}$-measurable. Therefore ι is measurable.

The family $\{\tau_\xi : \xi \in \Xi\}$ forms an orthonormal basis of $L^2(\mathbf{m}_Y)$. At the same time $\tau_\xi \circ \iota = \phi_\xi$ is also an orthonormal family in $L^2(\mathcal{A}, \mu)$. It follows that $\tau \to \tau \circ \iota$ is an isometric map from $L^2(\mathbf{m}_Y)$ to $L^2(\mathcal{A}, \mu)$. This guarantees that $\iota_* \mu = \mathbf{m}_Y$.

It remains to show that $\tau \to \tau \circ \iota$ is surjective. In fact, for every $f \in L^2(\mathcal{A}, \mu)$, $\widehat{f}(\xi, x)$ is a constant multiple $a_{f,\xi}\phi_\xi(x)$ of $\phi_\xi(x) = \tau_\xi \circ \iota(x)$ where $\sum_\xi |a_{f,\xi}|^2 < \infty$. Hence $f = \sum_\xi (a_{f,\xi}\tau_\xi) \circ \iota \in \iota^*(L^2(\mathbf{m}_Y))$, which shows ι is an isomorphism between probability spaces. □

Corollary 4.4.10. *In the setting of Theorem 4.3.1, if T_0 is K-mixing and T is weakly mixing, then T is K-mixing.*

Proof. Let $\mathcal{P} \subseteq \mathcal{B}$ be the Pinsker σ-algebra in $(X, T, \mathcal{B}, \mu)$. We show first that $\mathcal{P}$ is $\mathbb{T}^d$-invariant, i.e. preserved by L_z for all $z \in \mathbb{T}^d$. Because rational points are dense, it suffices to show this when $z \in \mathbb{Q}^d/\mathbb{Z}^d \subset \mathbb{T}^d$. In this case, there exists n such that $A^k z = z$. Thus for $P \in \mathcal{P}$, $T^k \circ L_z = L_{A^k z} \circ T^n = L_z \circ T^k$. In other words, L_z is an automorphism of the measure preserving dynamical system $(X, T^n, \mathcal{B}, \mu)$, and must preserve the Pinsker σ-algebra of T^n, which is the same as that of T by Lemma 4.4.8.

Since $(X, T, \mathcal{B}_0, \mu) \cong (X_0, T_0, \mathcal{B}_0, \mu)$ is K-mixing, and $(X, T, \mathcal{P}, \mu)$ has zero entropy, by Lemma 4.4.4 and Theorem 4.4.5, for all $U \in \mathcal{P}$ and $V \in \mathcal{B}_0$, $\mu(U \cap V) = \mu(U)\mu(V)$. In particular, by letting $U = V$, we conclude that every $U \in \mathcal{P} \cap \mathcal{B}_0$ has measure 0 or 1. Remark that having $U \in \mathcal{P} \cap \mathcal{B}_0$ is the same as having a $\mathcal{P}$-measurable subset that is invariant under the $\mathbb{T}^d$-action by translation. Thus $\mathbb{T}^d$ acts on $\mathcal{P}$ ergodically.

By Proposition 4.4.9, the factor system $(X, T, \mathcal{P}, \mu)$ of $(X, T, \mathcal{B}, \mu)$ is isomorphic to $(Y, S, \mathcal{B}_Y, \mathbf{m}_Y)$, where Y is a quotient torus Y of $\mathbb{T}^d$, and S is affine automorphism. Because $(X, T, \mathcal{B}, \mu)$ is weakly mixing, so is S. Viewing Y as $\mathbb{T}^p$ where $p \leq d$, and write $Sy = c\Psi(y)$ where $c \in \mathbb{T}^p$ and $\Psi \in \mathrm{GL}(p, \mathbb{Z})$. On the one hand, because $(X, T, \mathcal{P}, \mu)$ has zero entropy, $h_{\mathbf{m}_Y}(S) = 0$, and therefore all eigenvalues of Ψ are bounded by 1 in absolute value by Theorem 4.3.3. Since the determinant of Ψ is ± 1, in fact all the eigenvalues must have absolute value 1. Because the eigenvalues are algebraic units and always appear together with Galois conjugates, we conclude that Ψ only have roots of unity among its eigenvalues. On the other hand, since S is weakly mixing, by Theorem 4.4.1, the eigenvalues Ψ include no roots of unity. It then follows that $p = 0$ and Y is trivial.

We have thus proved the Pinsker σ-algebra $\mathcal{P}$ is trivial modulo μ, which is equivalent to completely positive entropy. Hence T is K-mixing. □

Corollary 4.4.10 is sufficient to produce Theorem 4.4.2.

Proof of Theorem 4.4.2. Again let T_j be the factor induced by T on $M_j = G/G_{(j+1)}\Gamma$, and $\mathcal{B}_j$ be the Borel σ-algebra on M_j. Then $(M_0, T_0, \mathcal{B}_0, \mathbf{m}_{M_0})$ is a trivial dynamical system whose Pinsker σ-algebra is trivial. So it should be regarded as K-mixing. Since T is weakly mixing, every T_j is also weakly mixing. By Corollary 4.4.10, if T_j is K-mixing, then so is T_{j+1}. By induction, $T = T_s$ is K-mixing. □