Chapter 4

Entropy and mixing of affine automorphisms

In this chapter, 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 $T \in \text{Aut}(M)$, $T$ 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 $X \rightarrow X^n$ 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 $l^\infty$-metric $d^N(x, y) = \max_{n=0}^{N-1} d(x_n, y_n)$ on $X^n$, then

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

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

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

In the sequel, we will denote respectively by $B_\epsilon(x)$ and $B_{N,\epsilon}(x)$ the open balls centered at $x$ of radius $\epsilon$ 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

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

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

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

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 \( \Gamma \) acts freely and properly\(^1\) on a metric space \( (X, d) \) from the right by isometries. Let \( (X', d') \) be the quotient space \( X / \Gamma \) 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 \( \pi \) is an isometry between \( B_{\delta_x}(x) \) and its image \( \pi(B_{\delta_x}(x)) \). Because \( \Gamma \) 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 \( \delta \) in \( X \) is

\(^1\)An action \( X \curvearrowright \Gamma \) 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.
CHAPTER 4. ENTROPY AND MIXING

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 $\pi(z_1) \in \pi(B_{\frac{\delta}{2}}(x))$. Then there are $\gamma_1, \gamma_2 \in \Gamma$ such that $z_j \in B_{\frac{\delta}{2}}(x\gamma_j)$. Since $d(z_1, z_2) < \delta$, it follows $d(x\gamma_1, x\gamma_2) < \delta + \frac{2}{3}\delta_x \leq \delta_x$, or equivalently $d(x, x\gamma_2\gamma_1^{-1}) < \delta_x$. This contradicts the choice of $\delta_x$, by which $\pi$ 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 $x \in X$, $\pi$ 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 $\pi$ 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 $0 \leq n \leq N - 1$, $d(T^n x, T^n z) < \epsilon$ 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 $n \leq N - 1$,

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

Since $\pi$ 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 $S_{N, \epsilon}(X') \leq S_{N, \epsilon}(Y)$ where $Y \subseteq X$ is a sufficiently large compact set such that $\pi(Y) = X'$ (it suffices to choose, for example, $\bigcup_{i=1}^k B_{\frac{\delta}{2}}(x_i)$.) It follows that $h_{\text{top}}(T') \leq h_{\text{top}}(T)$.

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

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

for all compact subsets $Y$, and thus $h_{\text{top}}(T') \leq h_{\text{top}}(T')$. The proof is completed. \qed
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 $n \in \mathbb{N}$, $h_\mu(T^n|A) = nh_\mu(T|A)$.

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

Definition 4.1.6. For a finite measurable partition $Q$ of $X$, define the conditional information function

$$I_\mu(Q|A)(x) = -\log \mu^A_x([x]^Q),$$

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

$$H_\mu(Q|A) = \int_X I_\mu(Q|A)(x) d\mu(x).$$

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

The absolute and conditional entropies are related by

$$H_\mu(Q|A) = \int_X H_\mu^A(Q)d\mu(x). \quad (4.3)$$

The quantity $H_\mu(Q|A)$ is increasing in $Q$ and decreasing in $A$, where the orderings of partitions and $\sigma$-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 $Q$:

$$H_\mu(P \vee Q|A) \leq H_\mu(P|A) + H_\mu(Q|A). \quad (4.4)$$

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

$$H_\mu(P \vee Q|A) = H_\mu(P|Q \vee A) + H_\mu(Q|A), \quad (4.5)$$

where $Q \vee A$ is the $\sigma$-algebra generated by $\{Q \cap A : Q \in \mathcal{Q}, A \in \mathcal{A}\}$.
Notice that when $\mathcal{A}$ is trivial modulo $\mu$, $\mu^\mathcal{A}_x = \mu$ for $\mu$-a.e. $x$. It is clear that $0 \leq H_\mu(Q|A) \leq H_\mu(Q)$. From (4.3), it is not hard to see that $H_\mu(Q|A) \leq \log n$ if every atom of $\mathcal{A}$ intersects no more than $n$ atoms of $Q$. In particular, $H_\mu(Q) \leq \log(\#Q)$.

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

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

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

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

exists.  

**Definition 4.1.7.** For a countably generated $T$-invariant $\sigma$-subalgebra $\mathcal{A}$ and a $T$-invariant probability measure $\mu$, the conditional measure-theoretic entropy of $T$, with respect to $\mu$, and conditional to $\mathcal{A}$, is $h_\mu(T|A) = \sup_Q h_\mu(T, Q|A)$ 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 $\mu$, denoted by $h_\mu(T)$.

The measure-theoretic analogue to Lemma 4.1.5

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

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

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

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 $(Q_k)$ is an increasing sequence of finite measurable partitions, which together generates the $\sigma$-algebra $\mathcal{B}$, then

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

---

It follows from the Martingale Convergence Theorem that $h_\mu(T, Q|A) = H_\mu\left(Q\big|\bigvee_{n=1}^{\infty} T^{-n}Q\right) \vee A)$, where $\bigvee_{n=1}^{\infty} T^{-n}Q$ is the $\sigma$-algebra generated by all the $T^{-n}Q$'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

\[
h_{\mu}(T) = h_{\nu}(S) + h_{\mu}(T|_{\pi^{-1}A}).
\]

**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 \( \rho \) 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 \( m_{\mathbb{R}^d} \).

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

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

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 $h_{\text{top}}(T) = \sup_{k \to \infty} \lim_{n \to \infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k)$. Because $S_{N,\epsilon}(Y) \geq \frac{m_{\mathbb{R}^d}(Y_k)}{m_{\mathbb{R}^d}(B_{N,\epsilon}(x))}$, we see that
\[
\lim_{n \to \infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k) \geq \limsup_{N \to \infty} -\frac{1}{N} \log m_{\mathbb{R}^d}(B_{N,\epsilon}(x)), \forall k
\]
and $h_{\text{top}}(T) \geq h_{\text{vol}}(T)$ after taking limit in $k$ and $\epsilon$.

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 $S_{N,\epsilon}(Y_k) \leq S'_{N,\frac{1}{2}\epsilon}(Y_{k+1}) \leq \frac{m_{\mathbb{R}^d}(Y_k)}{m_{\mathbb{R}^d}(B_{N,\frac{1}{2}\epsilon}(x))}$, so
\[
\lim_{n \to \infty} \frac{1}{N} \log S_{N,\epsilon}(Y_k) \leq \limsup_{N \to \infty} -\frac{1}{N} \log m_{\mathbb{R}^d}(B_{N,\frac{1}{2}\epsilon}(x)), \forall k
\]
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 $\bigoplus \lambda V^\lambda_C$ of generalized eigenspaces, where $V^\lambda_C = V^\lambda_{\mathbb{C}}$. Then $\mathbb{R}^d = \bigoplus_{\lambda \geq 0} V^\lambda$, where
\[
V^\lambda = \begin{cases} 
\ker_{\mathbb{R}^d}(A - \lambda \text{Id})^d, & \text{if } \lambda \in \mathbb{R}; \\
(V^\lambda_C \oplus V^\lambda) \cap \mathbb{R}^d, & \text{if } \lambda \notin \mathbb{R}.
\end{cases}
\]

For each eigenvalue $\lambda$, write $|\lambda|_+ = \max(1, |\lambda|)$.

**Proposition 4.2.2.** If $\mathbb{R}^d = V^\lambda$ for some $\lambda \in \mathbb{C}\setminus\{0\}$, then there exists $K > 1$ that depends only on $A$ and $d$, such that for all $\epsilon > 0$,
\[
B_{K^{-1}N^{-(d-1)|\lambda|_+N\epsilon}}(x) \subseteq B_{N,\epsilon}(x) \subseteq B_{K^{-1}N^{(d-1)|\lambda|_+N\epsilon}}(x).
\]
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

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

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 $V^\lambda_C$ and $V^\bar{\lambda}_C$ 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 $V^\lambda_C$ and $V^\bar{\lambda}_C$, 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 $T^n y - T^n 0 = A^n y = J^n A^n_0 y$.

**Lower bound.** If $|y| < K^{-1} N^{-(d-1)} |\lambda|_+^{-N} \epsilon$, then for all $0 \leq n \leq N - 1$, $|A^n_0 y| = |\lambda|^n |y| < K^{-1} N^{-(d-1)} \epsilon$. Since $J$ is unipotent, all entries of $J$ are polynomials in $n$ of degree less than $d$, and $|J^n A^n_0 y| < \epsilon$ 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

$$|J^n y| < |\lambda|^{-n} \epsilon, \forall n = 0, \ldots, N - 1.$$ \hspace{1cm} (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

$$|y| \leq \|J^{-(N-1)}\| \cdot |J^{N-1} y| \leq K_2 N^{d-1} \cdot |\lambda|^{-(N-1)} \epsilon $$

$$= K_2 |\lambda| \cdot N^{d-1} |\lambda|_+^{-N} \epsilon.$$  

Here $K_2$ is a constant depending only on $J$, so $K = K_2 |\lambda|$ depends only on $A$. This proves the upper bound. \hfill \Box
We now return to the general case where \( \mathbb{R}^d \) is the direct sum of finitely many \( V^\lambda \)'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 \( x \in \mathbb{R}^d \),

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

for

\[
h = \sum_{\lambda} \log |\lambda|_+,
\]

(4.8)

where the sum is taken over all eigenvalues \( \lambda \) 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 \( V^\lambda, \Im \lambda \geq 0 \) are orthogonal to each other.

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

\[
\prod_{\Im \lambda \geq 0} B^T_{N,\epsilon}(x^\lambda) \subseteq B_{N,\epsilon}(x) \subseteq \prod_{\Im \lambda \geq 0} B^T_{N,\epsilon}(x^\lambda).
\]

(4.9)

By Proposition 4.2.2,

\[
K^{-1} (N^{-d} |\lambda|_+^{-N \epsilon} \epsilon)^d \leq \mathbf{m}_{V^\lambda}(B^T_{N,\epsilon}(x^\lambda)) \leq K (N^d |\lambda|_+^{-N \epsilon} \epsilon)^d,
\]

(4.10)

where \( d^\lambda = \dim V^\lambda \) and \( K \) depends only on \( A \) and \( d^\lambda \).

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

\[
K^{-d} N^{-\sum_{\Im \lambda \geq 0} (d^\lambda)^2} \prod_{\Im \lambda \geq 0} |\lambda|_+^{-N \epsilon} \epsilon^d \leq \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x))
\]

\[
\leq K^d N^\sum_{\Im \lambda \geq 0} (d^\lambda)^2 \prod_{\Im \lambda \geq 0} |\lambda|_+^{-N \epsilon} \epsilon^d.
\]

(4.11)

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

\[
K^{-1} N^{-d^2} \left( \prod_{\Im \lambda \geq 0} |\lambda|_+^{d^\lambda} \right)^{-N \epsilon} \leq \mathbf{m}_{\mathbb{R}^d}(B_{N,\epsilon}(x))
\]

\[
\leq K N^d \left( \prod_{\Im \lambda \geq 0} |\lambda|_+^{d^\lambda} \right)^{-N \epsilon}.
\]

(4.12)
where $K$ depends only on $A$ and $d$, and the product $\prod_{\Im \lambda \geq 0} |\lambda|^d_\Lambda$ is taken over all eigenvalues $\lambda$ 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 eigenvalues appear in conjugate pairs $\lambda, \bar{\lambda}$.

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 $A \in \text{GL}(d, \mathbb{R})$ and $b \in \mathbb{R}^d$, then $h_{\text{top}}(T)$ is given by (4.8).

**Proof.** This follows from Proposition 4.2.1 and Corollary 4.2.4. \qed

By the variational principle, we know that $h_{\text{m}}(T) \leq h_{\text{top}}(T)$. In the case of toral automorphisms, the inequality holds.

**Theorem 4.2.6.** If $Tx = Ax + b$ on $X = \mathbb{T}^d$, where $A \in \text{GL}(d, \mathbb{Z})$ and $b \in \mathbb{R}^d$, then $h_{\text{m}}(T) = h_{\text{top}}(T) = \sum \lambda \log |\lambda|_+$, where the sum is taken over all eigenvalues $\lambda$ of $A$, with multiplicities counted.

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

For $L \in \mathbb{N}$, let $Q = Q_L$ 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

$$h_{\text{m}}(T) \geq h_{\text{m}}(T, Q) = \lim_{N \to \infty} \frac{1}{n} H_{\text{m}}(\bigvee_{n=0}^{N-1} T^{-n} Q).$$

Note that if $\varepsilon \geq \frac{\sqrt{d}}{L}$, then for all atoms $P$ of $\bigvee_{n=0}^{N-1} T^{-n} Q$ and $x \in P$, $P \subset B_{N,\varepsilon}(x)$. When $L$ is sufficiently large, $\varepsilon$ is bounded by the constant $\varepsilon_0$ in Claim 4.1.4, and we know that $B_{N,\varepsilon}(x)$ is an isometrically projected copy of the Bowen ball $B_{N,\varepsilon} (\tilde{x}) \subset \mathbb{R}^d$ with respect to $\bar{T}$. It then follows from Corollary 4.2.4 that

$$m_{\mathbb{T}^d}(P) \leq m_{\mathbb{T}^d}(B_{N,\varepsilon}(x)) = m_{\mathbb{T}^d}(B_{N,\varepsilon}(\tilde{x})) \leq KN^d e^{-h_{\text{top}}(T)N} \varepsilon^d$$

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

$$I_{\text{m}}(\bigvee_{n=0}^{N-1} T^{-n} Q)(x) \geq Nh_{\text{top}}(T) - d^2 \log N - \log(K\varepsilon^d)$$
for all $x$ and thus
\[
\frac{1}{N} H_{m_{Td}}(\bigvee_{n=0}^{N-1} T^{-n} Q) \geq \frac{1}{N} (Nh_{\text{top}}(T) - d^2 \log N - \log(K e^d))
\]
\[
= h_{\text{top}}(T) - \frac{1}{N} (d^2 \log N + \log(K e^d)).
\]

By taking limit, we conclude that $h_{m_{Td}}(T) \geq h_{m_{Td}}(T, Q) \geq h_{\text{top}}(T)$. On the other hand, $h_{m_{Td}}(T) \leq h_{\text{top}}(T)$ 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
\[
\{(x_n)_{n \in \mathbb{Z}} : x_n \in \mathbb{T}^d, x_{n+1} = 2x_n, \forall n \in \mathbb{Z}\}
\]
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 $(X, B, T, \mu)$ and $(X_0, B_0, T_0, \mu_0)$ 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; $B$ is the product $\sigma$-algebra between $B_0$ and the Borel $\sigma$-algebra $B_{\mathbb{T}^d}$ of $\mathbb{T}^d$, $\mu$ is ergodic and $\mu$ is invariant under translations by $\mathbb{T}^d$ (i.e. uniform along fibers); and for some automorphism $A$ of $\mathbb{T}^d$ such that $Tzx = A(z)Tx$ 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:

\[ h_\mu(T|\pi^{-1}B_0) = h_{\text{top}}(A) = h_{m_{z\theta}}(A). \]

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 \( \psi \) 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 \( \sigma \)-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 \( x = (x_0, z) \), \( \mu_{x_0}^{\mathbb{T}^d} \) is the unique \( \mathbb{T}^d \)-invariant probability measure along the fiber \( \pi^{-1}(x_0) = \mathbb{T}^d x \) and hence coincides with \( (\iota_{x_0})_*m_{\mathbb{T}^d} \).

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

\[ |h_\mu(T, Q_L|\mathcal{B}_0) - h_{m_{z\theta}}(A, Q_L)| < \delta. \tag{4.13} \]

Because

\[
H_\mu \left( \bigvee_{n=0}^{N-1} T^{-n}Q_L | B_0 \right) = \int_X H_{\mu_{x_0}^{\mathcal{B}_0}} \left( \bigvee_{n=0}^{N-1} T^{-n}Q_L \right) d\mu(x)
= \int_{X_0} H_{(\iota_{x_0})_*m_{\mathbb{T}^d}} \left( \bigvee_{n=0}^{N-1} T^{-n}Q_L \right) d\mu_0(x_0)
= H_{m_{\mathbb{T}^d}} \left( \bigvee_{n=0}^{N-1} T^{-n}Q_L \right) d\mu_0(x_0).
\]
we know

\[
h_{\mu}(T, Q_L|B_0) = \lim_{N \to \infty} \frac{1}{N} \int_{X_0} H_{\mu_d} \left( \sum_{n=0}^{N-1} T^{-n} Q_L \right) d\mu_0(x_0)
\]

\[
= \lim_{N \to \infty} \frac{1}{N} \int_{X_0} H_{\mu_d} \left( \bigvee_{n=0}^{N-1} T^{-n} Q_L \right) d\mu_0(x_0)
\]

(4.14)

Moreover,

\[
h_{\mu_d}(A, Q_L) = \lim_{N \to \infty} \frac{1}{N} H_{\mu_d} \left( \bigvee_{n=0}^{N-1} A^{-n} Q_L^d \right).
\]

(4.15)

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

\[
H(t_{x_0}^{-1} T^{-n} Q_L|A^{-n} Q_L^d) \leq \log 2^d, \quad H(A^{-n} Q_L^d|t_{x_0}^{-1} T^{-n} Q_L) \leq \log 2^d.
\]

(4.16)

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

\[
H_{\mu_d} \left( \bigvee_{n=0}^{N-1} t_{x_0}^{-1} T^{-n} Q_L \right) - H_{\mu_d} \left( \bigvee_{n=0}^{N-1} A^{-n} Q_L^d \right) \\
\leq H_{\mu_d} \left( \bigvee_{n=0}^{N-1} t_{x_0}^{-1} T^{-n} Q_L \bigvee_{n=0}^{N-1} A^{-n} Q_L^d \right) \\
\leq \sum_{n=0}^{N-1} H_{\mu_d} \left( t_{x_0}^{-1} T^{-n} Q_L \bigvee_{m=0}^{N-1} A^{-m} Q_L^d \right) \\
\leq \sum_{n=0}^{N-1} H_{\mu_d} \left( t_{x_0}^{-1} T^{-n} Q_L|A^{-n} Q_L^d \right) \leq N d \log 2.
\]

(4.17)

Similarly,

\[
H_{\mu_d} \left( \bigvee_{n=0}^{N-1} A^{-n} Q_L^d \right) - H_{\mu_d} \left( \bigvee_{n=0}^{N-1} t_{x_0}^{-1} T^{-n} Q_L \right) \leq N d \log 2.
\]

(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 ∈ \mathbb{N}, thanks to Lemma 4.1.8, \( n|h_\mu(T, Q_L|B_0) - h_{m_d}(A, Q_L)| < d \log 2 \) 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, \( \mu \) is not assumed to be \( T^d \)-invariant, but still is a \( T \)-invariant extension of \( \mu_0 \), then

\[
h_\mu(T) \leq h_{\mu_0}(T_0) + h_{\text{top}}(A).
\]

**Proof.** By Abramov-Rokhlin formula, it suffices to prove \( h_\mu(T|B_0) \leq h_{\text{top}}(A) \).

Without the assumption that \( \mu \) is uniform along fibers, the same proof of the theorem would yield instead that \( h_\mu(T, Q_L|B_0) \leq \lim_{N \to \infty} \frac{1}{N} \sup \nu H_\nu \left( \bigvee_{n=0}^{N-1} A^{-n} Q_L^{T^d} \right) \), (4.18)

where the supremum is taken over all probability measures \( \nu \) 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 \( Q_L^{T^d} \). Therefore \( B_{N,\epsilon}(x) \) intersects no more than \( 2^{Nd} \) atoms of \( \bigvee_{n=0}^{N-1} A^{-n} Q_L^{T^d} \). This shows the total number of atoms in \( \bigvee_{n=0}^{N-1} A^{-n} Q_L^{T^d} \) is bounded by \( 2^{Nd} S_{N,\epsilon} \), and thus by (4.19),

\[
h_\mu(T, Q_L|B_0) \leq \lim_{N \to \infty} \frac{1}{N} \log 2^{Nd} S_{N,\epsilon} = \lim_{N \to \infty} \frac{1}{N} \log S_{N,\epsilon} + d \log 2.
\]

By letting \( L \) tend to \( \infty \), we obtain that

\[
h_\mu(T|B_0) \leq h_{\text{top}}(A) + d \log 2.
\]

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 \( \infty \), using Lemmas 4.1.5 and 4.1.8.

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

\[
h_{m_d}(T) = h_{\text{top}}(T) = \sum \lambda \log |\lambda|_+,
\]

where the sum is taken over all eigenvalues \( \lambda \) of \( D_e A \), with multiplicities counted.
Proof. We first show that \( h_{m_{\mathcal{M}}}(T) = \sum_{\lambda} \log |\lambda|_+ \). Recall that \( M \) has the structure of a tower of principal torus bundles
\[
M = M_s \to M_{s-1} \to \cdots \to M_1 \to M_0 = \{ \text{pt} \},
\]
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 \( z \in G^{(i)}/G^{(i+1)} \), \( T_i(zx) = T x + A_i(z) \). Since the restriction of \( A_i \) to \( G^{(i)}/G^{(i+1)} \) is given by the matrix \( \psi_i \) in Proposition 3.1.5, which is from \( \text{GL}(d_i, \mathbb{Z}) \) 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, \( h_{m_{\mathcal{M}}}(T) = h_{m_{\mathcal{M}_{i+1}}}(T_{i+1}) + h_{\text{top}}(\psi_i) \). Thus
\[
h_{m_{\mathcal{M}}}(T) = h_{m_{\mathcal{M}}(s)}(T_s) = \sum_{i=1}^{s} h_{\text{top}}(\psi_i) = \sum_{i=1}^{s} \sum_{\lambda: \text{eigenvalue of } \psi_i \text{ counting multiplicities}} \log |\lambda|_+ \]
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 \( h_{\text{top}}(T) = h_{m_{\mathcal{M}}}(T) \). By the variational principle, \( h_{\text{top}}(T) \geq h_{m_{\mathcal{M}}}(T) \). On the other hand, for all \( T \)-invariant probability measures \( \mu \), we can show \( h_{\mu}(T) \leq h_{m_{\mathcal{M}}}(T) \) by the argument above and Proposition 4.3.2. So the variational principle also implies \( h_{\text{top}}(T) \leq h_{m_{\mathcal{M}}}(T) \). \( \square \)

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 \( \tilde{T}: x \to h(x).Tx \) 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 $A_1 \in \text{GL}(d_1, \mathbb{Z})$, 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)$\iff$(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) $\iff$ $T \times T$ is ergodic $\iff$ $T_1 \times T_1$ is ergodic $\iff$ (2).

(2)$\rightarrow$(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)$\rightarrow$(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 $\mu$ and $\nu$ is a $T \times S$-invariant probability measure $\rho$ on $X \times Y$ that projects respectively to $\mu$ and $\nu$ 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 $\pi_X$ and $\pi_Y$. Then $(\pi_X \times \pi_Y)_* \rho$ is a joining measure between $\mu$ and $\nu$ on $X \times Y$ and is thus equal to $\mu \times \nu$. This shows that for all $U \in \mathcal{B}_X$ and $V \in \mathcal{B}_Y$,

$$\rho(\pi_X^{-1}(U) \cap \pi_Y^{-1}(V)) = \rho((\pi_X \times \pi_Y)^{-1}(U \times V)) = (\pi_X \times \pi_Y)_* \rho(U \times V) = (\mu \times \nu)(U \times V) = \mu(U) \nu(V).$$

(4.20)

Since factor systems of a measure preserving dynamical system correspond to invariant $\sigma$-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 $\sigma$-algebras $A_1, A_2 \subseteq \mathcal{B}$, the factor systems $(X, A_1, T, \mu)$ and $(X, A_2, T, \mu)$ are disjoint, then $\mu(U_1 \cap U_2) = \mu(U_1)\mu(U_2)$ for all $U_i \in A_i$, $i = 1, 2$.

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 $\sigma$-algebra.

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

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

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

**Proof.** Note that if $\mathcal{P}$ is trivial modulo $\mu$, then for every $T$-invariant $\sigma$-subalgebra $\mathcal{A}$ that is non-trivial modulo $\mu$, 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 $\sigma$-algebra.
CHAPTER 4. ENTROPY AND MIXING

91

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

\[
\frac{1}{kN} H_\mu \left( \bigvee_{n=0}^{kN-1} T^{-n} \mathcal{U} \right) = \frac{k}{kN} H_\mu \left( \bigvee_{j=0}^{k-1} \bigvee_{n=0}^{N-1} (T^k)^{-n} \mathcal{U} \right) \leq \frac{k}{kN} \sum_{j=0}^{k-1} H_\mu \left( \bigvee_{n=0}^{N-1} (T^k)^{-n} \mathcal{U} \right) = \frac{1}{kN} \cdot kH_\mu \left( \bigvee_{n=0}^{N-1} (T^k)^{-n} \mathcal{U} \right) = \frac{1}{N} H_\mu \left( \bigvee_{n=0}^{N-1} (T^k)^{-n} \mathcal{U} \right).
\]

On the other hand,

\[
\frac{1}{N} H_\mu \left( \bigvee_{n=0}^{N-1} (T^k)^{-n} \mathcal{U} \right) \leq k \cdot \frac{1}{kN} H_\mu \left( \bigvee_{n=0}^{kN-1} T^{-n} \mathcal{U} \right).
\]

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 \( A \) is a \( T \)-invariant subalgebra of \( B \) that is invariant under both \( T \) and the \( T^d \)-action \( \{ L_z : z \in \mathbb{T}^d \} \) on the principal \( \mathbb{T}^d \)-bundle \( X \). Assume that \((X, A, \mu)\) is ergodic for the \( T^d \)-action, i.e. if \( E \in A \) is \( L_z \)-invariant modulo a null set with respect to \( \mu \) for all \( z \in \mathbb{T}^d \), then \( \mu(E) \in \{0,1\} \). 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 \( \Psi \in \text{Aut}(Y) \), such that the measure preserving dynamical system \((X, A, T, \mu)\) is isomorphic to \((Y, B_Y, S, \mu_Y)\), where \( B_Y \) is the Borel \( \sigma \)-algebra of \( Y \).

Proof. Consider an \( A \)-measurable \( L^2 \)-integrable function \( f \). As \( A \subseteq B \), \( f \in L^2(B, \mu) \) decomposes into the Fourier series \( f(x) = \sum_{\xi \in \mathbb{Z}^d} \hat{f}(\xi, x) \) along the \( \mathbb{T}^d \)-fibers, for which Lemma 3.4.2 holds. By the construction (3.12) and the \( T^d \)-invariance of \( A \), each \( \hat{f}(\xi, x) \) is still \( A \)-measurable. Suppose \( f_1, f_2 \) are both non-trivial \( A \)-measurable fiberwise Fourier modes of the same frequency \( \xi \), i.e. \( f_1(x) = \hat{f}_1(\xi, x) \), then the function \( f_1f_2 \) is constant along the fibers, i.e. invariant under the \( T^d \)-action, and at the same time \( A \)-measurable. By the ergodicity of the \( T^d \)-action on \((X, A, \mu)\). The function \( f_1f_2 \) is \( \mu \)-almost everywhere a constant. In particular, for every Fourier
mode, \(|f| = (f\hat{f})^{1/2}\) is \(\mu\)-a.e. constant. This shows \(f_2\) is a constant multiple of \(\frac{1}{f_2}\) and therefore \(f_1\) is a constant multiple of \(f_2\).

Denote by \(\Xi\) 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 \(f_1f_2\) is such a Fourier mode for frequency \(\xi_1 + \xi_2\), and \(\frac{1}{f_1f_2}\) is such a mode for \(-\xi_2\). It follows that \(\Xi\) is a subgroup of \(\mathbb{Z}^d\). Let \(H = \ker \Xi\), then \(H\) is a closed subgroup of \(\Xi\) and the Pontryagin dual \(Y\) of \(\Xi\) is isomorphic to the compact quotient group \(T^d/H\).

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 \(\Xi\) is a copy of \(\mathbb{Z}^p\) for some \(p \leq d\), one can do this first for a set of generators \(\xi_1, \ldots, \xi_p\), and then define \(\phi_\xi = \prod_{i=1}^p \phi_{\xi_i}^{n_i}\) for all \(\xi = \sum_{i=1}^p n_i \xi_i, n_i \in \mathbb{Z}\). This makes the \(\xi \mapsto \phi_\xi\) a group morphism. Given \(x \in X\), let \(y_x(\xi) = \phi_\xi(x)\). Then \(y_x\) is a group morphism from \(\Xi\) to \(\{u \in \mathbb{C} : |u| = 1\}\) for \(\mu\)-a.e. \(x\). In other words, \(y_x \in \widehat{\Xi} = Y\).

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

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

As \(\Xi\) is the Pontryagin dual of \(Y\), the continuous functions on \(Y\) are densely spanned by \(\{\tau_\xi : y \mapsto y(\xi) : \xi \in \Xi\}\). For each \(\tau_\xi, \tau_\xi \circ \iota\) sends \(x\) to \(y_x(\xi) = \phi_\xi(x)\), and is \(\mathcal{A}\)-measurable. Therefore \(\iota\) 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 \(\tau \mapsto \tau \circ \iota\) 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 \(\tau \mapsto \tau \circ \iota\) is surjective. In fact, for every \(f \in L^2(\mathcal{A}, \mu), \hat{f}(\xi, x)\) 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 \(\iota\) is an isomorphism between probability spaces. \(\square\)

**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 \(\sigma\)-algebra in \((X, T, \mathcal{B}, \mu)\). We show first
that $\mathcal{P}$ is $T^d$-invariant, i.e. preserved by $L_z$ for all $z \in T^d$. Because rational points are dense, it suffices to show this when $z \in \mathbb{Q}^d/\mathbb{Z}^d \subset T^d$. In this case, there exists $n$ such that $A^k z = z$. Thus for $P \in \mathcal{P}$, $T^k \circ L_z = L_{A^k z} \circ T^n = L_z \circ T^n$. 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 $\sigma$-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 $T^d$-action by translation. Thus $T^d$ 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 $(Y, S, \mathcal{B}_Y, m_Y)$, where $Y$ is a quotient torus $Y$ of $T^d$, and $S$ is affine automorphism. Because $(X, T, \mathcal{B}, \mu)$ is weakly mixing, so is $S$. Viewing $Y$ as $T^p$ where $p \leq d$, and write $S_y = c \Psi(y)$ where $c \in T^p$ and $\Psi \in GL(p, \mathbb{Z})$. On the one hand, because $(X, T, \mathcal{P}, \mu)$ has zero entropy, $h_{m_Y}(S) = 0$, and therefore all eigenvalues of $\Psi$ are bounded by 1 in absolute value by Theorem 4.3.3. Since the determinant of $\Psi$ is $\pm 1$, 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 $\Psi$ only have roots of unity among its eigenvalues. On the other hand, since $S$ is weakly mixing, by Theorem 4.4.1, the eigenvalues $\Psi$ include no roots of unity. It then follows that $p = 0$ and $Y$ is trivial.

We have thus proved the Pinsker $\sigma$-algebra $\mathcal{P}$ is trivial modulo $\mu$, 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 $\sigma$-algebra on $M_j$. Then $(M_0, T_0, \mathcal{B}_0, m_{M_0})$ is a trivial dynamical system whose Pinsker $\sigma$-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.