Chapter 3

Ergodicity of affine automorphisms

3.1 Linear automorphisms of nilmanifolds

Definition 3.1.1. *An* **automorphism** *of a Lie group G is a diffeomorphism $\Psi : G \to G$ that is also a group isomorphism. An* **automorphism** *of a Lie algebra $\mathfrak{g}$ is a linear isomorphism $\psi : \mathfrak{g} \to \mathfrak{g}$ that also a Lie algebra morphism.*

Definition 3.1.2. *An* **automorphism** *(or* **linear automorphism***, when it is necessary to distinguish from affine automorphisms, which will defined later) of a compact nimanifold $M = G/\Gamma$ is a diffeomorphism $f : M \to M$ that lifts to an automorphism of the universal cover G of M. Here G is a simply connected nilpotent Lie group and Γ is a lattice in G.*

Definition 3.1.3. *An* **automorphism** *of a discrete group Γ is a group isomorphism $\Psi : \Gamma \to \Gamma$.*

Let G, $\mathfrak{g}$, Γ and M be as above. Their groups of automorphisms are respectively denoted by $\mathrm{Aut}(G)$, $\mathrm{Aut}(\mathfrak{g})$, $\mathrm{Aut}(\Gamma)$, $\mathrm{Aut}(M)$.

We know from Corollary 1.1.13 that the derivative $D_e\Psi$ of a Lie group automorphism is an automorphism of $\mathfrak{g} = \mathrm{Lie}(G)$. By the same corollary and the bijectivity of the exponential map, we also know that Ψ is determined by $D_e\Psi$. In addition, Corollary 1.1.16 shows that every $\psi \in \mathrm{Aut}(\mathfrak{g})$ is the derivative of some Ψ. To summarize, we have

Lemma 3.1.4. $D_e : \mathrm{Aut}(G) \to \mathrm{Aut}(\mathfrak{g})$ *is bijective.*

We now give a characterization of $\mathrm{Aut}(G)$, or more generally, all Lie group morphisms from G to itself.

Proposition 3.1.5. *Suppose $\Psi : G \to G$ is a Lie group morphism of a simply connected nilpotent Lie group G, and $\mathcal{X}$ is a Mal'cev basis of $\mathfrak{g}$ adapted to a filtration $\{\mathfrak{g}_i\}_{i=1}^{s+1}$. Then in the coordinate system (1.16),*

(1) Ψ is represented by the matrix $\psi = D_e\Psi$, which has the block form

$$\begin{pmatrix} \psi_1 & * & \cdots & * \\ & \psi_2 & & * \\ & & \ddots & \vdots \\ & & & \psi_r \end{pmatrix}, \tag{3.1}$$

where the i-th block corresponds to the $\mathbf{u}_i$ component of a vector $\mathbf{u}$ and is spanned by $X_{m-m_i+1}, \cdots, X_{m-m_{i+1}}$.

(2) $\psi_2, \cdots, \psi_r$ are determined by ψ_1.

(3) Every eigenvalue of ψ_i is the product of i eigenvalues of ψ_1, possibly with repetition.

Proof. (1) This is because of the fact that $D_e\Psi$ is a Lie algebra morphism and preserves each $\mathfrak{g}_{(i)}$ in the lower central series. (Recall that $X_{m-m_i+1}, \cdots, X_m$ is a basis of $\mathfrak{g}_{(i)}$.)

(2) Denote by π_i the natural projection from $\mathfrak{g}_{(i)}$ to $\mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)}$. Then claim is equivalent to that: for all $1 \le i \le s$ and any given $W \in \mathfrak{g}_{(i)}$, the projection $\pi_i(\psi W) \in \mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)}$ is determined by ψ_1. Because $\mathfrak{g}_{(i)}$ is spanned by vectors of the form $[Y_1, [Y_2, \cdots, [Y_{i-1}, Y_i] \cdots]]$, it suffices to prove for a vector W of this form. In this case $\psi W = [\psi Y_1, [\psi Y_2, \cdots, [\psi Y_{i-1}, \psi Y_i] \cdots]]$.

Observe that if any of the Y_j's is replaced by a translate $Y_j + V$ where $V \in \mathfrak{g}_{(2)}$, then the bracket vector $[Y_1, [Y_2, \cdots, [Y_{i-1}, Y_i] \cdots]]$ is translated by $[Y_1, [Y_2, \cdots, [V, \cdots [Y_{i-1}, Y_i] \cdots]]$, which is in $\mathfrak{g}_{(i+1)}$ by Lemma 1.4.3. Therefore $\pi_i W$ depends only on $\pi_1 Y_1, \cdots, \pi_1 Y_i$. Similarly, since ψ preserves $\mathfrak{g}_{(2)}$, $\pi_i(\psi W)$ depends only on $\pi_1(\psi Y_1), \cdots, \pi_1(\psi Y_i)$. Note that by construction, $\pi_1(\psi Y_j)$ is in turn determined by ψ_1 and $\psi_1 Y_j$. Thus when the Y_j's are given, $\pi_i(\psi W)$ depends only on ψ_1. This proves part (2).

(3) Remark that ψ_i is isomorphic to the morphism ψ induces on $\mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)}$. If λ is an eigenvalue of ψ_i, if and only if the subspace

$$E_i^\lambda := \{V \in \mathfrak{g}_{(i)} : (\psi - \lambda \mathrm{Id})^n V \in \mathfrak{g}_{(i+1)} \text{ for sufficiently large } n\},$$

which contains $\mathfrak{g}_{(i+1)}$, is strictly larger than $\mathfrak{g}_{(i+1)}$. Here the exponent n can be replaced with any positive integer greater than or equal to $d_i =$

$\dim(\mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)})$, but using an unspecified large exponent will be more convenient for the calculation below. Moreover, with Λ_1 denoting the set of eigenvalues of ψ_1,

$$\mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)} = \bigoplus_{\lambda\in\Lambda_1} (E_i^\lambda/\mathfrak{g}_{(i+1)}).$$

Because $\mathfrak{g}_{(i)}$ is spanned by vectors of the form $[Y_1, [Y_2, \cdots, [Y_{i-1}, Y_i] \cdots]]$, it is spanned by the subspaces

$$E_i^{\lambda_1,\cdots,\lambda_i} = \operatorname{span}\{[Y_1, [Y_2, \cdots, [Y_{i-1}, Y_i] \cdots]] : Y_j \in E_1^{\lambda_j}\},$$

where the $(\lambda_1, \cdots, \lambda_i) \in \Lambda_1^i$. Since ψ preserves the filtration $\mathfrak{g}_{(i)}$, all the E_i^λ's, and the $E_i^{\lambda_1,\cdots,\lambda_i}$'s, are ψ-invariant.

To conclude the proof, it suffices to prove that

$$E_i^{\lambda_1,\cdots,\lambda_i} \subseteq E_i^{\lambda_1+\cdots+\lambda_i}. \tag{3.2}$$

This will in turn follow inductively from the claim that

$$[E_1^\lambda, E_j^\mu] \subseteq E_{j+1}^{\lambda\mu}. \tag{3.3}$$

To show (3.3), let $X \in E_1^\lambda$, $Y \in E_j^\lambda$. Remark that

$$\begin{aligned}(\psi - \lambda\mu\mathrm{Id})[X, Y] =&\psi[X, Y] - \lambda\mu[X, Y] = [\psi X, \psi Y] - [\lambda X, \mu Y]\\ =&[(\psi - \lambda\mathrm{Id})X, \psi Y] + [\lambda X, (\psi - \mu\mathrm{Id})Y].\end{aligned} \tag{3.4}$$

The vectors $(\psi - \lambda\mathrm{Id})X$ and λX are still in E_1^λ, and ψY and $(\psi - \mu\mathrm{Id})Y$ remain in E_j^λ. Furthermore ψ commutes with $(\psi - \lambda\mathrm{Id})$. Therefore, by iterating (3.4) n times, we will see that

$$(\psi - \lambda\mu\mathrm{Id})^n[X, Y] = \sum_{k=0}^{n} \binom{n}{k} [\lambda^k(\psi - \lambda\mathrm{Id})^{n-k}X, \psi^{n-k}(\psi - \lambda\mathrm{Id})^k Y].$$

When n is sufficiently large, either k or $n-k$ is sufficiently large. So the k-th term belongs to either $[\mathfrak{g}_{(2)}, \mathfrak{g}_{(j)}]$ or $[\mathfrak{g}_{(1)}, \mathfrak{g}_{(j+1)}]$. By Lemma 1.4.3, both of these are in $\mathfrak{g}_{(j+2)}$ and therefore $(\psi-\lambda\mu\mathrm{Id})^n[X, Y] \in \mathfrak{g}_{(j+2)}$ for all sufficiently large n. In other words, $[X, Y] \in E_{j+1}^{\lambda\mu}$. This establishes (3.3), and (3.2) then follows. □

We will show below that $\mathrm{Aut}(\Gamma)$ is the same as $\mathrm{Aut}(M)$.

Lemma 3.1.6. $\mathrm{Aut}(\Gamma) \cong \mathrm{Aut}(M) \cong \{\Psi \in \mathrm{Aut}(G) : \Psi(\Gamma) = \Gamma\}$.

Proof. We first prove $\mathrm{Aut}(\Gamma) \cong \{\Psi \in \mathrm{Aut}(G) : \Psi(\Gamma) = \Gamma\}$. Clearly, if $\Psi \in \mathrm{Aut}(G)$ satisfies $\Psi(\Gamma) = \Gamma$ then the restriction of Ψ on Γ is an automorphism. We now show every element of $\mathrm{Aut}(\Gamma)$ can be uniquely realized in this way.

We can follow the proof of the "only if" part of Theorem 2.3.1 to construct a Mal'cev basis $\mathcal{X}$ of $\mathfrak{g}$ adapted to the lower central series $\{\mathfrak{g}_{(i)}\}$, that satisfies Claim 2.3.2. In particular, Γ is generated by $\gamma_j = \exp X_j$, $1 \le j \le m$. By Porposition 1.6.6, the map ϕ in (1.19) is a diffeomorphism from $\mathbb{R}^m$ to G. Suppose A is an automorphism of Γ, define $\Psi : G \to G$ by

$$\Psi(\phi(\mathbf{u})) = \exp(u_1 \exp^{-1} A(\gamma_1)) \cdots \exp(u_m \exp^{-1} A(\gamma_m)).$$

Then, by Proposition 1.6.6 again, Ψ is a polynomial map in terms of the coordinates (1.16) or (1.19).

By Claim 2.3.2.(2), $\Gamma = \phi(\mathbb{Z}^d)$. And because A is an automorphism of Γ, Ψ coincides with A on Γ. In particular $\Psi(\Gamma) = \Gamma$.

The map $\Psi(g)\Psi(h)\Psi(gh)^{-1}$ is polynomial and equals identity when $g, h \in \Gamma$. Since Γ is Zariski dense in G, so is $\Gamma \times \Gamma$ in $G \times G$. It follows that $\Psi(g)\Psi(h) = \Psi(gh)$ for all gh. Hence Ψ is a Lie group morphism extending A. One can similarly define a Lie group morphism $\widehat{\Psi}$ that is polynomial in the coordinates (1.16) and extends A^{-1} to G. Then $\Psi \circ \Psi'$ and $\widehat{\Psi} \circ \Psi$ are Lie group morphisms from G to itself and coincide with the identity map on Γ. Because both $\Psi \circ \widehat{\Psi}$ and $\widehat{\Psi} \circ \Psi$ are polynomial, this implies they agree with the identity map on G. Therefore Ψ is invertible and belong to $\mathrm{Aut}(G)$.

Finally, suppose $\Psi, \Psi' \in \mathrm{Aut}(G)$ both preserve Γ and coincide with A on Γ. Because they are polynomial maps by Lemma 1.6.4 and Γ is Zariksi dense in G, $\Psi = \Psi'$. This shows the uniqueness of the realization of every $A \in \mathrm{Aut}(\Gamma)$ in $\mathrm{Aut}(G)$. The claim is established.

We then show $\mathrm{Aut}(M) \cong \{\Psi \in \mathrm{Aut}(G) : \Psi(\Gamma) = \Gamma\}$. This is because every element f of $\mathrm{Aut}(M)$ descends from an automorphism Ψ of G. It suffices to show what Ψ preserves Γ. Note that for two points $g\gamma$ and g in G that descends to the same point $g\Gamma \in M$, $\Psi(g\Gamma) = \Psi(g)\Psi(\gamma)$ and $\Psi(g)$ should project to the same point in M in order for f to be well-defined. Therefore $\Psi(\gamma) \in \Gamma$ for all $\gamma \in \Gamma$. ☐

From now on, we will always identify $\mathrm{Aut}(\Gamma)$ and $\mathrm{Aut}(M)$ with the above subgroup of $\mathrm{Aut}(G)$.

Example 3.1.7. For $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$,

$$\begin{aligned}\mathrm{Aut}(\mathbb{T}^d) \cong& \mathrm{Aut}(\mathbb{Z}^d) \cong \{A \in \mathrm{Aut}(\mathbb{R}^d) : A.\mathbb{Z}^d = \mathbb{Z}^d\} \\ =&\{A \in \mathrm{GL}(d, \mathbb{R}) : A.\mathbb{Z}^d = \mathbb{Z}^d\} = \mathrm{GL}(d, \mathbb{Z}).\end{aligned}$$

Since every Lie group morphism $\Psi : G \to G$ preserves the lower central series $\{G_{(i)}\}$. Hence for all indices $j < i$, Ψ induces an autormorphism $\Psi_{(j)}^{(i)}$of the quotient group $G_{(j)}/G_{(i)}$. It follows from the lemma above that all elements Ψ of $\mathrm{Aut}(M)$ preserves the subgroups $\Gamma_{(i)} \subseteq \Gamma$. Hence for all indices $j < i$, Ψ induces an autormorphism of the quotient group $G_{(j)}/G_{(i)}$, that preserves the lattice $\Gamma_{(j)}/\Gamma_{(i)}$, and hence an automorphism of the corresponding quotient, which is the compact nilmanifold $G_{(j)}/G_{(i)}\Gamma_{(j)}$.

Definition 3.1.8. *A number $\lambda \in \mathbb{C}$ is an* **algebraic integer** *if it is the root of an integer polynomial $p(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_0 \in \mathbb{Z}[x]$ with leading coefficient 1. It is an* **algebraic unit** *if in addition $a_k \in \{\pm 1\}$.*

Corollary 3.1.9. *If $\Psi \in \mathrm{Aut}(M)$, then all the eigenvalues of $\psi = D_e\Psi$ are algebraic units.*

Proof. By Proposition 3.1.5, it suffices to write $D_e\Psi$ in the form (3.1) and prove that the eigenvalues of each ψ_i are algebraic units. In this case, ψ_i is isomorphic to the isomorphism of the abelian Lie algebra $\mathfrak{g}_{(i)}/\mathfrak{g}_{(i+1)}$, which is identified with the abelian Lie group $G_{(i)}/G_{(i+1)}$ via the exponential map. The lattice $\Gamma_{(i)}/\Gamma_{(i+1)}$, which is the $\mathbb{Z}$-span of the basis vectors $X_{m-m_i+1}, \cdots, X_{m-m_{i+1}}$ of the i-th block of the current coordinate system, is ψ_i-invariant. Thus $\psi_i \in \mathrm{GL}(d_i, \mathbb{Z})$, which guarantees that it has only algebraic units as eigenvalues. □

A related fact to Corollary 3.1.9 is:

Lemma 3.1.10. *If $\Psi \in \mathrm{Aut}(M)$, then $D_e\Psi$ is a matrix with rational entries with respect to the $\mathbb{Q}$-structure on $\mathfrak{g}$ determined by Γ.*

Proof. It suffices to show that $D_e\Psi$ preserves the set of rational points of $\mathfrak{g}$. By Corollary 1.1.13, this is equivalent to showing that Ψ sends rational elements of G to rational elements. If g is rational, then $g^n \in \Gamma$ for some $n \in \mathbb{N}$ by Corollary 2.3.8. So $\Psi(g)$ is also rational because $\Psi(g)^n = \Psi(g^n)$ is also in Γ. □

In addition to linear automorphisms, we define a larger automorphism group.

Definition 3.1.11. *An affine automorphism of a compact nilmanifold $M = G/\Gamma$ is a diffeomorphism $T : M \to M$ of the form $Tx = g\Psi(x)$ where $g \in G$ and $\Psi \in \mathrm{Aut}(M)$.*

It is clear that affine automorphisms form a group, denoted by $\mathrm{Aff}(M)$, generated by linear automorphisms and translations, i.e. diffeomorphisms $T_g(x) = gx$ where $g \in G$.

Proposition 3.1.12. *Every affine automorphism of a compact manifold M preserves the Haar measure $\mathbf{m}_M$ on M.*

Proof. Suppose first $\Psi \in \mathrm{Aut}(M)$ and $\psi = D_e\Psi_i$. From the proof of Corollary 3.1.9 above, we know that every ψ_i in the form (3.1) has determinant ± 1. In particular, ψ sends the volume form dvol on $\mathfrak{g}$ to either dvol or $-$dvol. In consequence, ψ preserves the Haar measure $\mathbf{m}_{\mathfrak{g}}$, which then is identified with $\mathbf{m}_G$ and projects to $\mathbf{m}_M$.

Because every left translation L_g preserves $\mathbf{m}_G$, T_g preserves $\mathbf{m}_M$ for all $g \in G$. The porposition follows as every affine automorphism has the form $T_g \circ \Psi$.

□

Exercises

Exercise 3.1.1. Suppose $A \in \mathrm{SL}(d, \mathbb{Q})$ does not have roots of unity among its eigenvalues. By Exercise 1.5.4, the Lie group $G = \{g \in \mathrm{SL}(d, \mathbb{R}) : \lim_{n\to\infty} A^n g A^{-n} = \mathrm{Id}\}$ is nilpotent.

(1) Show that G is simply connected.

(2) Show that $\Gamma = G \cap \mathrm{SL}(d, \mathbb{Z})$ is a lattice in G.

(3) Suppose $B \in \mathrm{SL}(d, \mathbb{Z})$ commutes with A. Show that $g \to BgB^{-1}$ induces an automorphism of the compact nilmanifold G/Γ.

Exercise 3.1.2. Show that the automorphism group $\mathrm{Aut}(H_3)$ of the 3-dimensional Heisenberg Lie group is isomorphic to $\mathrm{GL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$.

Exercise 3.1.3. Show that the automorphism group $\mathrm{Aut}(M)$ of the 3-dimensional Heisenberg nilmanifold is isomorphic to $\mathrm{GL}(2, \mathbb{Z}) \ltimes \mathbb{Z}^2$.

Exercise 3.1.4. Show that: $\mathrm{Aff}(M) \cong \mathrm{Aut}(M) \ltimes \big(G/(\Gamma \cap C(G))\big)$. Here $C(G)$ denotes the center of G.

3.2 Basic facts from ergodic theory

We aim to study the ergodic theory of affine automorphisms of a compact nilmanifold M. For this purpose, wet first define some basic notions of ergodic theory.

Definition 3.2.1. *A* **topological dynamical system** *is a pair* (X, T) *where* X *is a compact metric space and* $T : X \to X$ *is a continuous map. A* **measure preserving dynamical system** $(X, T, \mathcal{B}, \mu)$ *is a topological dynamical system* (X, T) *equipped with a* T*-invariant* σ*-algebra* $\mathcal{B}$ *and a* T*-invariant* $\mathcal{B}$*-measurable probability measure* μ*.*

Example 3.2.2. By Proposition 3.1.12, if $M = G/\Gamma$ is a compact nilmanifold, $T \in \mathrm{Aff}(M)$, $\mathcal{B}_M$ is the Borel σ-algebra of M, and $\mathbf{m}_M$ is the unique left-invariant probability measure on M, then $(M, T, \mathcal{B}_M, \mathbf{m}_M)$ is a measure preserving dynamical system.

Here $\mathcal{B}$ is T-invariant if $T^{-1}B \in \mathcal{B}$ for all $B \in \mathcal{B}$. A set $B \in \mathcal{B}$ is T**-invariant modulo** μ if $T^{-1}B = B$ up to a null set, i.e. $\mu((T^{-1}B) \triangle B) = 0$.

Definition 3.2.3. *A measure preserving dynamical system* $(X, T, \mathcal{B}, \mu)$ *is:*

(1) **ergodic** *if for all* T*-invariant (modulo* μ*) subset* $A \in \mathcal{B}$*,* $\mu(A) = 0$ *or* 1*.*

(2) **weakly mixing** *if for all measurable subsets* $A, B \in \mathcal{B}$*,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)\right| = 0.$$

(3) **mixing** *if for all measurable subsets* $A, B \in \mathcal{B}$*,*

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

(4) **Kolmogrov mixing (K-mixing)** *if for every* T*-invariant* σ*-algebra* $\mathcal{A}$ *of* $\mathcal{B}$*, the Kolmogrov-Sinai entropy of* $(X, \mathcal{A}, T, \mu)$ *is positive unless* $\mu(A) \in \{0, 1\}$ *for all* $A \in \mathcal{A}$*.*

The definition of K-mixing uses the definition of the Kolmogrov-Sinai entropy. We will give the precise definition in future sections. [1]

For now, it suffices for the reader to know the following hierarchy:

Theorem 3.2.4. *ergodic* $\Leftarrow$ *weakly mixing* $\Leftarrow$ *mixing* $\Leftarrow$*Kolmogrov mixing*

These are standard facts from ergodic theory. The first two implications are easy to prove. The last one can be found in e.g. [Par81, §3.3].

[1] The definition of K-mixing given here is actually called **completely positive entropy** (CPE). The original definition of K-mixing is a different one. Two theorems, respectively due to Pinsker and Rokhlin-Sinai, together assert that CPE is equivalent to K-mixing.

For a measure preserving dynamical system $(X, T, \mathcal{B}, \mu)$, T induces an operator on functions on X, which we still denote by T, by $Tf = f \circ T$. The operator is isometric on every $L^p(\mu)$ because T preserves μ.

Denote the inner production on $L^2(\mu)$ by

$$\langle f_1, f_2 \rangle = \int f_1 \overline{f_2} \mathrm{d}\mu.$$

Because T is a unitary operator on $L^2(\mu)$, two eigenfunctions with distinct eigenvalues must be orthogonal.

The constant functions form a one dimensional (over $\mathbb{R}$ or $\mathbb{C}$ depending on the context) subspace of eigenfunctions for T, with 1 being the eigenvalue. Write $L^2_0(\mu)$ for the orthogonal complement of this subspace in $L^2(\mu)$.

Theorem 3.2.5. *The following are equivalent for a measure preserving dynamical system $(X, T, \mathcal{B}, \mu)$:*

(1) $(X, T, \mathcal{B}, \mu)$ is ergodic;

(2) If f is measurable and $Tf = f$ μ-a.e., then $f =$ constant μ-a.e.

(3) If $f \in L^2(\mu)$ and $Tf = f$ μ-a.e., then $f =$ constant μ-a.e.

(4) 1 is not an eigenvalue of T on $L^2_0(\mu)$.

(5) (von Neumann's L^2 Ergodic Theorem) $\frac{1}{N}\sum_{n=0}^{N-1} T^n f$ converges to the constant $\int f \mathrm{d}\mu$ in L^2-norm for all $f \in L^2(\mu)$.

Proof. **(1)⇒(2)** It suffices to prove for real valued functions. Suppose $f : X \to \mathbb{R}$ satsifies $Tf = f$ a.e. but is not constant a.e.. Then there exists Y such that $A = \{x \in X : f(x) \leq R\}$ has measure $\mu(A) \in (0, 1)$. However by the T-invariance of f, $\mu((T^{-1}A) \triangle A) = 0$. Hence A is T-invariant modulo μ. This shows T is not ergodic.

(2)⇒(3) is evident.

(3)⇒(1) Note that if A is T-invariant modulo μ, then $f = 1_A$ satifies $Tf = f$ μ-a.e. So assuming (3), one will have either $1_A = 0$ or $1_A = 1$ a.e., or equivalently $\mu(A) = 0$ or $\mu(A) = 1$.

(3)⇔(4) is obvious, because it is already known that constant functions are eigenfunctions with eigenvalue 1.

(5)⇒(4) Suppose for contradiction that f is a non-constant function from $L^2_0(\mu)$ that is an eigenfunction for eigenvalue 1. That is, $Tf = f$ a.e. Then $\frac{1}{N}\sum_{n=0}^{N-1} T^n f = f$. So by the claim (5), $f = \int f \mathrm{d}\mu$ is a constant a.e, contradicting the hypothesis on f.

(3)⇒(5) Let $V \subseteq L^2(\mu)$ be the closure of the subspace $\{Th - h : h \in L^2(\mu)\}$, and let $V^\perp$ be its orthogonal complement. Then since $L^2(\mu)$ is a Hilbert space, $L^2(\mu) = V \oplus V^\perp$.

We first show that for $g \in V$, $\frac{1}{N}\sum_{n=0}^{N-1}\langle T^n f, g\rangle \to 0$. It suffices to show this for $g = Th - h$ where $h \in L^2(\mu)$. In this case

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\langle T^n f, g\rangle =& \frac{1}{N}\sum_{n=0}^{N-1}(\langle T^n f, Th\rangle - \langle T^n f, h\rangle)\\ =& \frac{1}{N}\sum_{n=0}^{N-1}(\langle T^n f, Th\rangle - \langle T^{n+1} f, Th\rangle)\\ =& \frac{1}{N}(\langle f, Th\rangle - \langle T^{N-1} f, h\rangle \to 0.\end{aligned}$$

We then show that for $g \in V^\perp$, $\frac{1}{N}\sum_{n=0}^{N-1}\langle T^n f, g\rangle = \langle f, g\rangle$. Because $\langle T^{n+1} f - T^n f, g\rangle = 0$, we have $\langle T^n f, g\rangle = \langle f, g\rangle$ for all n. The claim follows.

Combining the two previous facts, we know that $\frac{1}{N}\sum_{n=0}^{N-1}\langle T^n f, g\rangle \to \langle \pi_{V^\perp} f, g\rangle$ where $\pi_{V^\perp}$ is the orthogonal projection from $L^2(\mu)$ to $V^\perp$. In other words, $\frac{1}{N}\sum_{n=0}^{N-1} T^n f$ weakly converges to $\pi_{V^\perp} f$ in L^2. Since

$$\left\| T(\frac{1}{N}\sum_{n=0}^{N-1} T^n f) - \frac{1}{N}\sum_{n=0}^{N-1} T^n f \right\|_{L^2(\mu)} = \left\| \frac{1}{N}(T^N f - f) \right\|_{L^2(\mu)} \to 0,$$

this implies $\pi_{V^\perp} f$ is T-invariant, and hence by property (3) a constant function a.e. for all $f \in L^2$. In particular, this shows $V^\perp = \{\text{constant functions}\}$ and thus $V = L_0^2(\mu)$.

For a function of the form $f = Th - h$, $\frac{1}{N}\sum_{n=0}^{N-1} T^n f = \frac{1}{N}(T^N f - f) \to 0$. Because such functions are dense in $V = L_0^2(\mu)$, for every $f \in L_0^2(\mu)$, $\frac{1}{N}\sum_{n=0}^{N-1} T^n f \to 0$ in L^2 norm.

Finally, for all $f \in L^2(\mu)$, $f - \int f \mathrm{d}\mu \in L_0^2(\mu)$. So

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n f = \int f \mathrm{d}\mu + \frac{1}{N}\sum_{n=0}^{N-1} T^n (f - \int f \mathrm{d}\mu) \to \int f \mathrm{d}\mu + 0 = \int f \mathrm{d}\mu$$

in $L^2(\mu)$. □

The most important theorem in ergodic theory is Birkhoff's ergodic theorem, which strengthens L^2 ergodic theorem.

Theorem 3.2.6 (Birkhoff's Pointwise Ergodic Theorem)**.** *If $(X, T, \mathcal{B}, \mu)$ is ergodic and $f \in L^1(\mu)$, then $\frac{1}{N}\sum_{n=0}^{N-1} T^n f$ pointwisely converges to $\int f \mathrm{d}\mu$ at μ-almost every point.*

All T-invariant $\mathcal{B}$-measurable probability measures form a convex set of which the ergodic measures are the extremal points. When this set degenerates into a single point, we have the following definition

Definition 3.2.7. *Given $(X, \mathcal{B})$, a continuous map $T : X \to X$ is* **uniquely ergodic** *if there is only one T-invariant $\mathcal{B}$-measurable probability measure on X.*

The next theorem says every point in this convex set can be uniquely written as a convex combination of the extremal points.

Theorem 3.2.8 (Ergodic decomposition)**.** *In a measure preserving dynamical system $(X, T, \mathcal{B}, \mu)$, there is a measurable map that assignes to μ-almost every $x \in X$ an ergodic T-invariant probability measure $\mu_x^{\mathcal{E}}$, called the* **ergodic component of** *μ , such that $\mu = \int \mu_x^{\mathcal{E}} \mathrm{d}\mu$.*

The construction of $\mu_x^{\mathcal{E}}$ is not hard. First let $\mathcal{A} \subseteq \mathcal{B}$ be the σ-subalgebra of all subsets that are T-invariant modulo μ. Then for every $f \in L^2(\mathcal{B}, \mu)$ there is a **conditional expectation** function $\mathbb{E}(f|\mathcal{A}) : X \to \mathbb{C}$ such that:

1. $\mathbb{E}(f|\mathcal{A})$ is $\mathcal{A}$-measurable;
2. $\int_B \mathbb{E}(f|\mathcal{A})(x)\mathrm{d}\mu(x) = \int_B f(x)\mathrm{d}\mu(x)$;
3. $\mathbb{E}(f|\mathcal{A})(x) \geq 0$ if $f \geq 0$.

The function $\mathbb{E}(f|\mathcal{A})$ is in fact the orthogonal projection of f from the Hilbert space $L^2(\mathcal{B}, \mu)$ to the closed subspace $L^2(\mathcal{A}, \mu)$. For each given x, the correspondence $I_x^{\mathcal{A}}(f) = \mathbb{E}(f|\mathcal{A})(x)$ from $L^2(\mathcal{B}, \mu)$ to $\mathbb{C}$ is a linear functional and one can show that $|\mathbb{E}(f|\mathcal{A})(x)| \leq \|f\|_{L^\infty}$. In particular, one can restrict $I_x^{\mathcal{A}}$ to $C(X) \subseteq L^2(\mathcal{B}, \mu)$ and get a bounded linear functional on $C(X)$. Because X is compact, by Riesz representation theorem, there is a measure $\mu_x^{\mathcal{A}}$ such that $\mathbb{E}(f|\mathcal{A})(x) = I_x^{\mathcal{A}}(f) = \int f \mathrm{d}\mu_x^{\mathcal{A}}$ for all $f \in C(X)$. The ergodic component is given by $\mu_x^{\mathcal{E}} = \mu_x^{\mathcal{A}}$. In general, for all Lebesgue probability measure spaces $(X, \mathcal{B}, \mu)$ and countably generated σ-subalgebra $\mathcal{A} \subseteq \mathcal{B}$, one can construct conditional measures $\mu_x^{\mathcal{A}}$ in this way (See, for example, [?EW11, Theorem 5.14]).

We next state some equivalence definitions of weak mixing and mixing without proofs.

Theorem 3.2.9. *The following are equivalent for a measure preserving dynamical system $(X, T, \mathcal{B}, \mu)$:*

(1) $(X, T, \mathcal{B}, \mu)$ is weakly mixing;

(2) For all functions $f, g \in L^2_0(\mu)$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \left|\langle T^n f, g\rangle\right| = 0$.
(3) $(X \times X, T \times T, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$ *is ergodic;*
(4) $(X \times X, T \times T, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$ *is weakly mixing;*
(5) T *does not have eigenvalues on* $L^2_0(\mu)$.

Theorem 3.2.10. *The following are equivalent for a measure preserving dynamical system* $(X, T, \mathcal{B}, \mu)$*:*
(1) $(X, T, \mathcal{B}, \mu)$ *is mixing;*
(2) For all functions $f, g \in L^2_0(\mu)$, $\lim_{N\to\infty} \left|\langle T^n f, g\rangle\right| = 0$.

Theorems 3.2.8-3.2.10 can be found in, for instance, [Par81, Ch. 1 & 3].

Definition 3.2.11. *A measure preserving dynamical system* $(Y, S, \mathcal{A}, \nu)$ *is a* **factor** *of another measure preserving dyamical system* $(X, T, \mathcal{B}, \mu)$ *if there is a measurable surjective map* $\pi : (X, \mathcal{B}) \to (Y, \mathcal{A})$ *satisfying* $\pi \circ T = S \circ \pi$ *and* $\pi_*\mu = \nu$.

Notice that each factor system can be realized within X by restricting to a σ-subalgebra: $(Y, S, \mathcal{A}, \nu)$ is isomorphic to $(X, T, \pi^{-1}\mathcal{A}, \mu)$. It follows easily that if T is ergodic (resp. weakly mixing, mixing) then so is S.

Lemma 3.2.12. *If* $\pi : (X, \mathcal{B}) \to (Y, \mathcal{A})$ *is a measurable surjective map between two measurable spaces where* X *and* Y *are compact metric spaces. If homeomorphisms* $T : X \to X$ *and* $S : Y \to Y$ *satsify* $\pi \circ T = S \circ \pi$*. Then for every ergodic* S*-invariant probability measure* ν*, there is an ergodic* T*-invariant probability measure* μ *such that* $\pi_*\mu = \nu$.

Proof. By Birkhoff's Ergodic Theorem, ν-a.e. y is generic for ν in the sense that $\frac{1}{N}\sum_{n=0}^{N-1} \delta_{S^n y}$ converges to ν in weak-* topology. For every x that lifts y, a subsequence of $\frac{1}{N}\sum_{n=0}^{N-1} \delta_{S^n y}$ weak-* converges because the set of probability measures on a compact metric space is compact in the weak-* topology. Any such sequential limit is a T-invariant probability measure that converges to ν, and so are all its ergodic components because of the ergodicity of ν. Let μ be such an ergodic component. □

Exercises

Exercise 3.2.1. Show that by Definition 3.2.3, weak mixing implies ergodicity.

Exercise 3.2.2. Show that the translation $Tx = x + b$ of the torus $\mathbb{T}^d$, where $b = (b_1, \cdots, b_d) \in \mathbb{T}^d$, is ergodic with respect to the Haar measure $\mathbf{m}_{\mathbb{T}^d}$ if and only if $1, b_1, \cdots, b_d$ are linearly independent over $\mathbb{Q}$. Moreover, show that T is ergodic if and only if it is uniquely ergodic.

Exercise 3.2.3. Show that the translation in Exercise 3.2.3 is not weakly mixing.

3.3 Ergodicity of affine automorphisms of tori

Notation 3.3.1. *In the future, when we say an affine automorphism T of a nilmanifold M is ergodic, it will be by defaut with respect to the probability Haar measure $\mathbf{m}_M$. For simplicity, the measures $L^2(\mathbf{m}_M)$ and $L^2_0(\mathbf{m}_M)$ will be denoted by $L^2(M)$ and $L^2_0(M)$.*

In this section, we will give necessary and sufficient conditions for an affine automorphism f of $\mathbb{T}^d$ to be ergodic.

Recall that $T \in \mathrm{Aff}(\mathbb{T}^d)$ has the form $Tx = Ax + b$ where A is an element of $\mathrm{Aut}(\mathbb{T}^d)$, which by Example 3.1.7 can be identified with $\mathrm{GL}(d, \mathbb{Z})$, and $b \in \mathbb{R}^d$. Note that b can be chosen from $\mathbb{T}^d$ instead of $\mathbb{R}^d$.

As $\det A = 1$, all eigenvalues of A are algebraic units. We first show an sufficient condition:

Lemma 3.3.2. *If $A \in \mathrm{GL}(d, \mathbb{Z})$ does not have roots of unity among its eigenvalues, then for all $b \in \mathbb{R}^d$, the affine automprhism $Tx = Ax + b$ of $\mathbb{T}^d$ is ergodic.*

Proof. By Theorem 3.2.5 it suffices to show that 1 is not among the eigenvalues of f on $L^2_0(\mathbb{T}^d)$.

The Fourier modes $\{e(\xi \cdot x) : \xi \in \mathbb{Z}^d\}$ form an orthonormal basis of $L^2(\mathbb{T}^d)$.[2] Those with $\xi \neq 0$ form an orthonormal basis of $L^2_0(\mathbb{T}^d)$. Remark that

$$e(\xi \cdot Tx) = e(\xi b)e(\xi \cdot Ax) = e(A^{\mathrm{T}}\xi \cdot x). \tag{3.5}$$

Assume $f(x) \in L^2_0(\mathbb{T}^d)$ is a non-zero eigenfunction with eigenvalue 1; that is $Tf = f$. Take the Fourier series $f(x) = \sum_{\xi \in \mathbb{Z}^d \setminus \{0\}} \widehat{f}(\xi)e(\xi \cdot x)$ in L^2, then

$$\begin{aligned} Tf(x) &= \sum_{\xi \in \mathbb{Z}^d \setminus \{0\}} \widehat{f}(\xi)e(\xi b)e(A^{\mathrm{T}}\xi \cdot x) \\ &= \sum_{\xi \in \mathbb{Z}^d \setminus \{0\}} \widehat{f}((A^{\mathrm{T}})^{-1}\xi)e((A^{\mathrm{T}})^{-1}\xi \cdot b)e(\xi \cdot x). \end{aligned} \tag{3.6}$$

Comparing Tf with f, we know that

$$\widehat{f}(\xi) = \widehat{f}((A^{\mathrm{T}})^{-1}\xi)e((A^{T})^{-1}\xi \cdot b) \tag{3.7}$$

[2]Here and in the remainder of these notes, $e(z)$ will denote the function $e^{2\pi i z}$, which can be regarded as a function from either $\mathbb{R}$ or $\mathbb{T}^1$ to the unit circle in $\mathbb{C}$.

for all $\xi \in \mathbb{Z}^d \backslash \{0\}$. In particular, $|\widehat{f}(\xi)| = |\widehat{f}((A^{\mathrm{T}})^{-1}\xi)|$, and thus $|\widehat{f}(\xi)| = |\widehat{f}((A^{\mathrm{T}})^{-n}\xi)|$ for all $n \in \mathbb{N}$ by iteration.

Since f is non-trivial, we can fix a $\xi^* \neq 0$ such that $\widehat{f}(\xi^*) \neq 0$. Because $\|f\|_{L^2} = \sum_\xi |\widehat{f}(\xi)|^2$, there are only finitely many ξ with $|\widehat{f}(\xi)| \geq |\widehat{f}(\xi^*)|$. In particular, $(A^{\mathrm{T}})^{-n}\xi^*$ is in this finite set for all $n \in \mathbb{N}$. So $(A^{\mathrm{T}})^{-n}\xi^* = (A^{\mathrm{T}})^{-m}\xi^*$ for some m, n with $m > n$. This shows ξ^* is a fixed vector of $(A^{\mathrm{T}})^{m-n}$. So 1 is an eigenvalue of $(A^{\mathrm{T}})^{m-n}$, and thus of A^{m-n}. It follows that among the eigenvalues of A, there must be an $(m-n)$-th root of unity. □

The following is a necessary condition, weaker than the sufficient condition above.

Lemma 3.3.3. *If an affine automprhism $Tx = Ax + b$ of $\mathbb{T}^d$ is ergodic, then $A \in \mathrm{GL}(d, \mathbb{Z})$ does not have any roots of unity other than 1 among its eigenvalues.*

Proof. Assume for contradiction that $u \neq 1$ is a primitive k-th roots of unity, where $k > 1$, and at the same time an eigenvalue of A. Then u is also an eigenvalue of A^{T}.

Let $V = \bigoplus_{\lambda \neq 1} \ker(A^{\mathrm{T}} - \lambda \mathrm{Id})^d$ be the direct sum of all generalized eigenspaces whose corresponding eigenvalue is not 1. Then V is a rational subspace of $\mathbb{R}^d$. Moreover, the eigenspace $\ker(A^{\mathrm{T}} - u\mathrm{Id})$ of A^{T} corresponding to the eigenvalue u is a subset of

$$V \cap \Big(\ker\big((A^{\mathrm{T}})^k - \mathrm{Id}\big) \backslash \bigcup_{j=1}^{k-1} \ker\big((A^{\mathrm{T}})^j - \mathrm{Id}\big) \Big). \tag{3.8}$$

So (3.8) is non-empty. Because $A^{\mathrm{T}} \in \mathrm{GL}(d, \mathbb{Z})$, every $\ker((A^{\mathrm{T}})^j - \mathrm{Id})$ is a rational vector subspace, and thus one can find a non-zero rational vector ξ^* in (3.8) above. By rescaling, one may assume $\xi^* \in \mathbb{Z}^d \backslash \{0\}$.

Then $(A^{\mathrm{T}})^k\xi^* = \xi^*$, but ξ^*, $A^{\mathrm{T}}\xi^*$, $\cdots$, $(A^{\mathrm{T}})^{k-1}\xi^*$ are all distinct. The vector $\sum_{j=0}^{k-1}(A^{\mathrm{T}})^j\xi^*$ is a fixed vector of A^{T} in V. Since 1 is not an eigenvector of A^{T} on V by construction, $\sum_{j=0}^{k-1}(A^{\mathrm{T}})^j\xi^* = 0$.

We will define Fourier coefficients at fequencies ξ^*, $A^{\mathrm{T}}\xi^*$, $\cdots$, $(A^{\mathrm{T}})^{k-1}\xi^*$ by $\widehat{f}((A^{\mathrm{T}})^j\xi^*) = e(\sum_{l=0}^{j-1}(A^{\mathrm{T}})^l\xi \cdot b)$. Note that this definition holds for $j = k$ as well, in which case $\widehat{f}((A^{\mathrm{T}})^k\xi^*) = \widehat{f}(\xi^*) = 1 = e(\sum_{l=0}^{k-1}(A^{\mathrm{T}})^l\xi \cdot b)$. By this definition, (3.7) holds at the frequency $\xi = (A^{\mathrm{T}})^j\xi^*$ for $j = 0, \cdots, k-1$. Therefore the function $f(x) = \sum_{j=0}^{k-1}\widehat{f}((A^{\mathrm{T}})^j\xi^*)e((A^{\mathrm{T}})^j\xi^* \cdot b)$ satisfies $Tf = f$. Moreover, since these k frequecies are distinct and $|\widehat{f}((A^{\mathrm{T}})^j\xi^*)| = 1$ for each j, f is a non-trivial function in $L^2_0(\mathbb{T}^d)$. Thus 1 is an eigenvalue for T in $L^2_0(\mathbb{T}^d)$, and by Theorem 3.2.5, T is not ergodic. □

Lemma 3.3.2 and Lemma 3.3.3 leave to us the case when 1 is the only root of unity among the eigenvalues of A. We will focus on this case below.

Lemma 3.3.4. *If the eigenvalues of A include 1 but no other roots of unity, then an affine automorphism $Tx+b$ of $\mathbb{T}^d$ is not ergodic if and only if there exists $\xi \in \mathbb{Z}^d \backslash \{0\}$ such that $A^{\mathrm{T}}\xi = \xi$ and $\xi \cdot b \in \mathbb{Z}$.*

Proof. By proof of , T fails to be ergodic if and only if there exists an L^2 function f satisfying (3.7) for all $\xi \in \mathbb{Z}^d \backslash \{0\}$. This can happen only if for each ξ with non-vanishing $\widehat{f}(\xi)$, $(A^{\mathrm{T}})^k\xi = \xi$ for some $k \in \mathbb{N}$. Because the eigenvalues of A^{T}, which are the same as those of A, does not have roots of unity of higher orders, we must have $k = 1$ and $A^{\mathrm{T}}\xi = \xi$ for all ξ with $\widehat{f}(\xi) \neq 0$. For these ξ, (3.7) becomes $\widehat{f}(\xi) = \widehat{f}(\xi)e(\xi \cdot b)$, and thus $e(\xi \cdot b) = 1$. Equivalently $\xi \cdot b \in \mathbb{Z}$.

Conversely, if $A^{\mathrm{T}}\xi = \xi$ and $\xi \cdot b$, then $f(x) = e(\xi \cdot x)$ is a non-trivial solution to (3.7). □

Summarizing the lemmas above, we have the following criterion for ergodicity:

Theorem 3.3.5. *Suppose $Tx = Ax + b$ is an affine automorphism of $\mathbb{T}^d$ where $A \in \mathrm{GL}(d, \mathbb{Z})$ and $b \in \mathbb{R}$. Then T is ergodic if and only if:*

A does not have roots of unity other than 1 among its eigenvalues, and for every $\xi \in \mathbb{Z}^d \backslash \{0\}$ such that $A^{\mathrm{T}}\xi = \xi$, $\xi \cdot b \notin \mathbb{Z}$.

Example 3.3.6. When $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $b = (\alpha, 0)$, T is the affine automorphism of $\mathbb{T}^2$ given by $T(x, y) = (x + \alpha, y + x)$ where $x, y \in \mathbb{R}/\mathbb{Z}$. The only eigenvalue of $A^{\mathrm{T}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is 1, and the eigenvectors in $\mathbb{Z}^d \backslash \{0\}$ are $\xi = (q, 0)$. By Theorem 3.3.5, T is ergodic if and only if for every $q \in \mathbb{Z} \backslash \{0\}$, $q\alpha \notin \mathbb{Z}$, i.e. α is irrational.

In this example, T projects to the rotation $Sx = x + \alpha$ of the circle $\mathbb{T}^1$. By Exercise 3.2.3, S is ergodic if and only if α is irrational. In other words, the ergodicity of T is equivalent to that of S.

In general, suppose 1, but no other roots of unity, is among the eigenvalues of A. The corresponding generalized eigenspace is $V^1 = \ker(A - \mathrm{Id})^d$. Using Jordan canonical form, we know that there exists $1 \leq k \leq d$ such that

$$\{0\} \subsetneq \ker(A - \mathrm{Id}) \subsetneq \ker(A - \mathrm{Id})^2 \cdots \subsetneq \ker(A - \mathrm{Id})^k = V^1. \tag{3.9}$$

Recall that $\mathbb{C}^d = \bigoplus_\zeta V_{\mathbb{C}}^\zeta$ where ζ ranges over all eigenvalues of A and $V_{\mathbb{C}}^\zeta = \ker_{\mathbb{C}^d}(A-\zeta\mathrm{Id})^d$ is the generalized eigenspace. The subspace $V_{\mathbb{C}}^{\neq 1} = \bigoplus_{\zeta\neq 1} V_{\mathbb{C}}^\zeta = \ker_{\mathbb{C}^d}\prod_{\zeta\neq 1}(A-\zeta\mathrm{Id})^d$ is defined over $\mathbb{Q}$ as 1 is rational. Let $V^{\neq 1} = \mathbb{R}^d \cap V_{\mathbb{C}}^{\neq 1}$, which is an A-invariant rational subspace of $\mathbb{R}^d$. The subspace $V = V^{\neq 1} \oplus \ker(A-\mathrm{Id})^{k-1}$ is again an A-invariant rational subspace of $\mathbb{R}^d$, and $W_1 = \mathbb{R}^d/V = V^1/\ker(A-\mathrm{Id})^{k-1}$. Because $(A-\mathrm{Id})v \in \ker(A-\mathrm{Id})^{k-1}$ for all $v \in V^1$, A projects to the identity map on W_1. In fact V is the largest A-invariant subspace for which A induces the identity map on the quotient.

The affine transform $x \to Ax + b$ projects to a translation $x \to x + b_1$ on W_1. Where b_1 is the projection of V in W_1. Because V is a rational subspace, $\Lambda_1 = \mathbb{Z}^d/(\mathbb{Z}^d \cap V)$ is a lattice in W_1, and $Y_1 = W_1/\Lambda_1$ is a torus, on which T induces the rotation $S_1(x) = x + b_1$. The torus X_1 is the largest quotient torus of $\mathbb{T}^d$ on which T induces a rotation. $(Y_1, S_1, \mathbf{m}_{Y_1})$ is a factor dynamical system of $(\mathbb{T}^d, T, \mathbf{m}_{\mathbb{T}^d})$, which we call the **maximal rotation factor**.

Corollary 3.3.7. *An affine toral automorphism $Tx = Ax + b$ is ergodic if and only if the eigenvalues of A include no non-trivial roots of unity and the maximal rotation factor of T is ergodic.*

Proof. The "only if" direction follows from Theorem 3.3.5 and the ergodicity of factors. For the opposite statement, notice that under the assumptions, by Theorem 3.3.5 the map T can be non-ergodic only if for some $\xi \in \mathbb{Z}^d\backslash\{0\}$, $A^{\mathrm{T}}\xi = \xi$ and $\xi \cdot b \in \mathbb{Z}$.

We adapt the notations above. For $v \in V_{\mathbb{C}}^\zeta$ where $\zeta \neq 1$, $\xi \cdot v = A^{\mathrm{T}}\xi \cdot v = \xi \cdot Av = \bar{\lambda}(\xi \cdot v)$, and thus $\xi \cdot v = 0$.

On the other hand, one can see using Jordan canonical forms that for $v \in \ker(A-\mathrm{Id})^{k-1}$ there exists $w \in \ker(A-\mathrm{Id})^k = V^1$ such that $v = (A-\mathrm{Id})w$. Thus $\xi \cdot v = (A^{\mathrm{T}} - \mathrm{Id})\xi \cdot w = 0 \cdot w = 0$.

It follows that ξ, as an element of $(\mathbb{R}^d)^*$, vanishes on V and descends to a non-trivial linear functional ξ_1 on W_1. Moreover, ξ assumes integer values on Λ_1. Thus ξ_1 is a Fourier frequency from the torus Y_1 and $\xi \cdot b = \xi_1 \cdot b_1$. Thus $\xi_1 \cdot b_1 \neq 0$. Again by Theorem 3.3.5 , S_1 is not ergodic, which yields the desired contradiction. □

Exercises

Exercise 3.3.1. If $T(x_1, x_2, \cdots, x_d)$ equals

$$(x_1 + \alpha_1, x_2 + x_1 + \alpha_2, x_3 + x_2 + \alpha_3, \cdots, x_d + x_{d-1} + \alpha_d)$$

on $\mathbb{T}^d$, show that T is ergodic if and only if α_1 is irrational.

3.4 Ergodicity and unique ergodicity on principal bundles

In this section, suppose $(X_0, T_0, \mathcal{B}_0, \mu_0)$ is an ergodic measure preserving dynamical system, K is a second countable compact abelian group, and X is a left principal K-bundle over X_0 equipped with the product σ-algebra $\mathcal{B} = \mathcal{B}_0 \times \mathcal{B}_K$, where $\mathcal{B}_K$ is the Borel σ-algebra of K. We remark that the expression $\mathcal{B}_0 \times \mathcal{B}_K$ is indeed an abuse of notation as X is not a direct product between X_0 and K. To define the product σ-algebra, we actually should look at local charts U of X_0, the preimage $\pi^{-1}(U)$ of which is homeomorphic to $U \times K$. On each of these preimages, we can define a product σ-algebra $\mathcal{B}_{\pi^{-1}(U)}$ which is identified with $\mathcal{B}_0|_U \times \mathcal{B}_K$ through the homeomorphism. The global product σ-algebra $\mathcal{B}$ is the σ-algebra generated by all the $\mathcal{B}_{\pi^{-1}(U)}$'s.

Let $T : X \to X$ be a **measurable extension** of T_0, i.e. a measurable map such that $\pi \circ T = T_0 \circ \pi$, where $\pi : X \to X_0$ is the bundle projection. In addition, assume that T acts on the fibers by an automorphism A of K, i.e.

$$T(zx) = A(z)Tx, \forall z \in K, \mu\text{-a.e. } x \in X. \tag{3.10}$$

Note $\pi^{-1}(\mathcal{B}_0)$ is the σ-algebra on X consisting of all unions of K-fibers. By abusing notation, we denote it indifferently by $\mathcal{B}_0$. Then for every probability measure ν and ν-almost all x, $\nu_x^{\mathcal{B}_0}$ is supported on a single fiber, which is the K-orbit of x.

The group K acts freely on the principal bundle X, write $L_z(x) = zx$ for $z \in K$ and $x \in X$. For every x, the Haar measure $\mathbf{m}_K$ on K pushes forward to a K-invariant measure on the orbit $K.x$. By invariance of $\mathbf{m}_K$, this measure depends only on the orbit $K.x$ but not on the choice of x within the orbit, and hence is a function on the underlying quotient X_0. For each $x_0 \in X_0$, which represents a K-orbit in X, denote this measure by $\mathbf{m}_{K,x_0}$. Then the measure

$$\mu = \int_{X_0} \mathbf{m}_{K,x_0} \mathrm{d}\mu_0(x_0) \tag{3.11}$$

is equal to the direct product $\mu_0 \times \mathbf{m}_K$ on local charts of the principal bundle structure.

For every $z \in K$, as each $\mathbf{m}_{K,x_0}$ is translation invariant, $(L_z)_*\mu = \mu$. For $f \in L^2(\mu)$, define $L_z f = f \circ L_z$. Then L_z is an isometry of $L^2(\mu)$

Lemma 3.4.1. *The measure μ in (3.11) is T-invariant.*

Proof. $T_*\mu = \int_{X_0} T_*\mathbf{m}_{K,x_0} \mathrm{d}\mu_0(x_0)$. As $(T_0)_*\mu_0 = \mu_0$, it suffices to show $T_*\mathbf{m}_{K,x_0} = \mathbf{m}_{K,T_0x_0}$. Consider the fiber $\pi^{-1}(x_0)$ above x_0 in X, which is the K-orbit that x_0 represent, then $T(\pi^{-1}(x_0))$ is the fiber $\pi^{-1}(T_0x_0)$ as $\pi \circ T = T_0 \circ \pi$, and $T_*\mathbf{m}_{K,x_0}$ is a probability measure on this fiber. It remains to show $T_*\mathbf{m}_{K,x_0}$ is the unique K-invariant probability measure on TF_0. To see this, note that for $z \in K$, $T \circ L_z = L_{A(z)} \circ T$. Hence $(L_z)_*T_*\mathbf{m}_{K,x_0} = T_*(L_A^{-1}(z))_*\mathbf{m}_{K,x_0} = T_*\mathbf{m}_{K,x_0}$. The lemma is proved. □

Write $\widehat{K}$ for the Pontryagin dual of K, which is the group of all continuous characters $\xi : K \to \mathbb{T}^1$. Then $\widehat{K}$ is a countable discrete abelian group and A induces a transpose A^{T} on $\widehat{K}$ by $A^{\mathrm{T}}(\xi) \cdot z = \xi \cdot A(z)$

For any function $f \in L^2(\mu)$, because f is L^2-integrable with respect to $\mathbf{m}_{K,x_0}$ along μ_0-almost every fiber $\pi^{-1}(x_0)$, one can perform a Fourier series decomposition along the fibers by letting

$$\widehat{f}(\xi, x) = \int_{z\in K} f(zx)e(-\xi \cdot z)\mathrm{d}z, \forall \xi \in \widehat{K}. \tag{3.12}$$

Then, in $L^2(\mu)$-sense, for $x \in M$ and $z \in K$,

$$f(zx) = \sum_{\xi\in\widehat{K}} e(\xi \cdot z)\widehat{f}(\xi, x). \tag{3.13}$$

In particular,

$$f(x) = \sum_{\xi\in\widehat{K}} \widehat{f}(\xi, x). \tag{3.14}$$

Because the decomposition into eigenfunctions is unique, $x \to \widehat{f}(\xi, x)$ is an eigenfunction for the K-action in the sense that

$$\widehat{f}(\xi, zx) = e(\xi \cdot z)\widehat{f}(\xi, x). \tag{3.15}$$

Lemma 3.4.2. *Assuming* (3.10)*, then in the fiberwise Fourier series above,* $\widehat{Tf}(\xi, x) = \widehat{f}((A^{\mathrm{T}})^{-1}\xi, Tx)$.

Proof. Notice that $Tf(x) = \sum_{\xi\in\widehat{K}} \widehat{f}(\xi, Tx)$. By (3.15),

$$\widehat{f}(\xi, T(zx)) = \widehat{f}(\xi, A(z)Tx) = e(\xi \cdot A(z))\widehat{f}(\xi, Tx) = e(A^{\mathrm{T}}\xi \cdot z))\widehat{f}(\xi, Tx).$$

So $\widehat{f}(A^{\mathrm{T}}\xi, x) = \widehat{Tf}(\xi, Tx)$. The lemma follows by a coordinate change in ξ. □

We now introduce an important property, essentially due to Furstenberg [Fur61], and extended by Parry [Par69b].

Theorem 3.4.3. *In the above setting, assume that for all $k \in \mathbb{N}$ and $\xi \in \widehat{K}$ such that $(A^{\mathrm{T}})^k\xi = \xi$, $A^{\mathrm{T}}\xi = \xi$.*

Then the implications (1) ⇒ (2) ⇒ (3) hold for the following properties:

(1) μ is the unique T-invariant measure projecting to μ_0;

(2) There do not exist a non-trivial continuous character $\xi : K \to \mathbb{T}^1$ and a measurable assignment to μ_0-a.e. $x_0 \in X_0$ of a $(\ker\xi)$-orbit $E_{x_0} \in \pi^{-1}(x_0)/\ker\xi$ inside the K-orbit $\pi^{-1}(x_0)$, such that $A^{\mathrm{T}}\xi = \xi$ and $TE_{x_0} = E_{T_0x_0}$ for μ_0-a.e. x_0;

(3) μ is ergodic.

Moreover, if in addition the automorphism A in (3.10) is unipotent, i.e. $(A - \mathrm{Id})^n = 0$ for some n, then the implication (3) ⇒ (1) is also true.

When using the notation $A - \mathrm{Id}$, we are thinking of K as an additive abelian group.

Proof. **(1)⇒(2)**: Suppose there exist ξ and the measurable assignment $x_0 \to E_{x_0}$ as in (2). The probability Haar measure $\mathbf{m}_{\ker\xi}$ on the compact abelian subgroup $\ker\xi$ pushes forward to a $\ker\xi$-invariant measure $\mathbf{m}_{\ker\xi,x_0}$ on $E_{x_0} \subset \pi^{-1}(x_0)$.

Since $A^{\mathrm{T}}\xi = \xi$, $A\ker\xi = \ker\xi$ and $A^{-1}\ker\xi = \ker\xi$. In other words, A is an automorphism of the compact abelian group $\ker\xi$ and hence preserves $\mathbf{m}_{\ker\xi}$.

It follows that, because for each $z \in \ker\xi$ and $x \in \pi^{-1}(x_0)$, $T(zx) = A(z)Tx$ where $A(z) \in \ker\xi$, the pushforward $T_*\mathbf{m}_{\ker\xi,x_0}$ is translation invariant under elements of $\ker\xi$. Since it is supported on the orbit $TE_{x_0} = E_{T_0x_0}$. Thus $T_*\mathbf{m}_{\ker\xi,x_0} = \mathbf{m}_{\ker\xi,T_0x_0}$.

Define $\mu' = \int_{X_0} \mathbf{m}_{\ker\xi,x_0}\mathrm{d}\mu_0(x_0)$. Then μ' projects to μ_0 and

$$\begin{aligned} T_*\mu' &= \int_{X_0} T_*\mathbf{m}_{\ker\xi,x_0}\mathrm{d}\mu_0(x_0) = \int_{X_0} \mathbf{m}_{\ker\xi,T_0x_0}\mathrm{d}\mu_0(x_0) \\ &= \int_{X_0} \mathbf{m}_{\ker\xi,x_0}\mathrm{d}(T_0)_*\mu_0(x_0) = \int_{X_0} \mathbf{m}_{\ker\xi,x_0}\mathrm{d}\mu_0(x_0) = \mu'. \end{aligned}$$

However, $(\mu')_x^{\mathcal{B}_0} = \mathbf{m}_{\ker\xi,x_0} \neq \mathbf{m}_{K,x_0} = \mu_x^{\mathcal{B}_0}$. So μ is not the unique T-invariant measure projecting to μ_0, contradicting statement (1).

(2)⇒(3): Suppose T is μ is not ergodic and there is a non-constant function $f \in L^2(\mu)$ such that $Tf = f$ almost everywhere[3]. Decompose f as in

[3] The argument here still works if f is a non-constant eigenfunction of T.

(3.12), then there is a non-zero ξ such that $\widehat{f}(\xi, x) \neq 0$ in $L^2(\mu)$. By Lemma 3.4.2, $\widehat{f}(A^{\mathrm{T}}\xi, x) = h(\xi, Tx)$ has the same norm in $L^2(\mu)$. Because the Fourier modes $\widehat{f}(\xi, x)$ are orthogonal for different ξ's, and f is L^2-integrable, ξ must have a finite orbit under A^{T}. By assumption $A^T\xi = \xi$.

By (3.15), $|\widehat{f}(\xi, x)|$ is constant along each K-orbit, and thus factors through a function a on X_0. Moreover, a is T_0-invariant modulo μ_0 as $\widehat{f}(\xi, x)$ is T-invariant in $L^2(\mu)$. So by ergodicity of (T_0, μ_0), a is μ_0-almost everywhere constant. Thus without loss of generality, we may assume $|\widehat{f}(\xi, x)| = 1$ for μ-almost all $x \in X$.

For each x_0, define $E_{x_0} = \{x \in \pi^{-1}(x_0) : \widehat{f}(\xi, x) = 1\} \subset \pi^{-1}(x)$. Thanks to (3.15), E_{x_0} is a $\ker \xi$-orbit. Moreover, $TE_{x_0} = E_{T_0 x_0}$ by the T-invariance of the function $\widehat{f}(\xi, x)$. This contradicts claim (2) and therefore we have showed the implication (2)⇒(3).

(3)⇒(1) when A is unipotent: We first prove this in the special case where $A = \mathrm{Id}$. Suppose a T-invariant probability measure ν projects to μ_0 on X_0. Decompose $\nu = \int \nu_x^{\mathcal{B}_0} \mathrm{d}\nu(x)$ into conditional measures along fibers. Because $x \to \nu_x^{\mathcal{B}_0}$ is $\mathcal{F}$-measurable, $\nu_x^{\mathcal{B}_0}$ depends only on the fiber, i.e. the projection x_0 of x in X_0. In this case we write $\nu_{x_0}^{\mathcal{B}_0} = \nu_x^{\mathcal{B}_0}$. Then

$$\nu = \int_X \nu_{\pi(x)}^{\mathcal{B}_0} \mathrm{d}\nu(x) = \int_X \nu_{x_0}^{\mathcal{B}_0} \mathrm{d}\pi_*\nu(x_0) = \int_{X_0} \nu_{x_0}^{\mathcal{B}_0} \mathrm{d}\mu_0(x_0).$$

Consider the average

$$\bar{\nu} = \int_K (L_z)_*\nu \mathrm{d}\mathbf{m}_K(z) = \int_{x_0 \in X_0} \int_{z \in K} (L_z)^* \nu_{x_0}^{\mathcal{B}_0} \mathrm{d}\mathbf{m}_K(z) \mathrm{d}\mu_0(x_0).$$

Because $\int_{z\in K} (L_z)_* \nu_{x_0}^{\mathcal{B}_0} \mathrm{d}\mathbf{m}_K(z)$ is K-invariant on the K-orbit over x_0, it is equal to $\mathbf{m}_{K,x_0}$, thus $\bar{\nu} = \mu$. Because $A = \mathrm{Id}$, T commutes with the translation L_z by $z \in K$. Hence $(L_z)_*\nu$ is also T-invariant for all $z \in K$. So μ is an average of T-invariant probability measures $(L_z)_*\nu$. Since μ is ergodic, these measures coincide for $\mathbf{m}_K$-almost all z. That is, $(L_z)_*\nu = \nu$ for almost all z, and $\mu = \bar{\nu} = \nu$. The unique ergodicity is established.

Now assume $(A - \mathrm{Id})^n = 0$ instead. When $n = 1$ this was proved above. For induction, suppose the claim is known when $(A - \mathrm{Id})^{n-1} = 0$. Then A preserves the close subgroup $\ker(A - \mathrm{Id})$ and induces an automorphism A_1 on the quotient group $K_1 = K/\ker(A - \mathrm{Id})$. Because for all $z \in K$, $(A - \mathrm{Id})^{n-1}(z) \in \ker(A - \mathrm{Id})$, $(A_1 - \mathrm{Id})^{n-1} = 0$.

Let $X_1 = X/\ker(A - \mathrm{Id})$ be the space of $\ker(A - \mathrm{Id})$-orbits on X. Then X_1 is an intermediate K_1-principal bundle over X_0. Because $T(zx) =$

$A(z).Tx = z.Tx$ for all $z \in \ker(A - \mathrm{Id})$, T sends $\ker(A - \mathrm{Id})$-orbits to $\ker(A - \mathrm{Id})$-orbits, and thus descends to a map T_1 on X_1. The map T_1 further projects to T_0. Let $\pi_1 : X \to X_1$ and $\pi_0 : X_1 \to X_0$ be the bundle projections and $\mu_1 = (\pi_1)_*\mu$. Then μ is an ergodic T_1-invariant measure and projects to μ_0.

For $x_1 \in X_1$ and $z_1 \in K_1$, choose any lifts x and z respectively in X and K. The equality $T(zx) = A(z).Tx$ yields $T_1(z_1x_1) = A_1(z_1).T_1(x_1)$ after projection. So by inductive assumption, μ_1 is the unique T_1-invariant measure projecting to μ_0.

Now X is a $\ker(A-\mathrm{Id})$-principal bundle over X_1. Since $T(zx) = z.Tx$ for all $z \in \ker(A - \mathrm{Id})$, by the special case $A = \mathrm{Id}$, μ is the unqiue T_1-invariant measure projecting to μ_1.

For a T-invariant measure μ' projecting to μ_0, let $\mu'_1 = (\pi_1)_*\mu'$ be its projection on the intermediate bundle X_1. Because $(\pi_0)_*\mu'_1 = \pi_*\mu' = \mu_0$, μ'_1 must be μ_1 and it in turn follows that $\mu' = \mu$. This establishes the proof of statement (1). □

Corollary 3.4.4. *An affine toral automorphism $Tx = Ax + b$ on $\mathbb{T}^d$ is uniquely ergodic if and only if $A \in \mathrm{GL}(d, \mathbb{Z})$ is unipotent and T is ergodic.*

Proof. Suppose first A is unipotent and T is ergodic, i.e. its only eigenvalue is 1. Define the maximal rotation factor Y_1 as in the last paragraphs of §3.3. Let k be the length of the flag as in (3.9). Then $\mathbb{R}^d = V^1 = \ker(A - \mathrm{Id})^k$. For each $1 \le j \le k$, $\ker(A - \mathrm{Id})^{k-j}$ is a rational subspace. More generally, we can define a torus $Y_j = W_j/\Lambda_j$, where $W_j = \mathbb{R}^d/\ker(A - \mathrm{Id})^{k-j}$ is a rational subspace, and $\Lambda_j = \mathbb{Z}^d/(\mathbb{Z}^d \cap \ker(A - \mathrm{Id})^{k-1})$. T projects to an affine automorphism T_j on Y_j, which is ergodic because T is. Note that $Y_k = \mathbb{T}^d$. Then each Y_{j+1} is a torus principal bundle over Y_j where the fiber is parametrized by a torus quotient of the vector space $\ker(A-\mathrm{Id})^{k-j}/\ker(A-\mathrm{Id})^{k-j-1}$. Note that the map A_{j+1} induced by A on $\mathbb{R}^d/\ker(A - \mathrm{Id})^{k-j-1}$ restricts to the identity map on $\ker(A - \mathrm{Id})^{k-j}/\ker(A - \mathrm{Id})^{k-j-1}$. Thus for all $z \in \ker(A - \mathrm{Id})^{k-j}/\ker(A - \mathrm{Id})^{k-j-1}$ and $x \in Y_{j+1}$, $T_{j+1}(x + z) = A_{j+1}(x+z)+b = T_{j+1}x+z$. So by the implication (3)⇒(1) in Theorem 3.4.3, if T_j is uniquely ergodic, then T_{j+1} is uniquely ergodic as T_{j+1} is ergodic. Becaue T_1 is an ergodic rotation of Y_1, it is uniquely ergodic (Exercise 3.2.3). It follows by induction that all the T_j's, including $T = T_k$, are uniquely ergodic.

Conversely, suppose T is uniquely ergodic. Then it clearly must also be ergodic with respect to $\mathbf{m}_{\mathbb{T}^d}$ (otherwise $\mathbf{m}_{\mathbb{T}^d}$ splits into distinct ergodic components). Assume A is not unipotent, then the subspace $V^1 = \ker(A -$

Id$)^k$ in (3.9) is a proper rational A-invariant subspace of $\mathbb{R}^d$. The affine tranform $x \to Ax + b$ projects from $\mathbb{R}^d$ to an affine transform $x \to A'x + b'$ on $W = \mathbb{R}/V^1$ where A' has no eigenvalue 1. Moreover $\Lambda = \mathbb{Z}^d/(\mathbb{Z}^d \cap V^1)$ is a lattice in W, and $x \to A'x + b'$ induces an affine automorphism S of the torus $Y = W/\Lambda$. Thanks to Lemma 3.2.12, it suffices to prove that S is not uniquely ergodic. Equivalently, we need to show: if $A \in \mathrm{GL}(d, \mathbb{Z})$ does not have eigenvalue 1, then the affine automorphism $Tx = Ax + b$ is not uniquely ergodic.

In this case, note that $A - \mathrm{Id}$ is an invertible matrix. For $v = -(A - \mathrm{Id})^{-1}b \in \mathbb{R}^d$, $v = Av + b$ and therefore v projects to a fixed point of T in $\mathbb{T}^d$. The point mass at this point is a T-invariant probability measure other than $\mathbf{m}_{\mathbb{T}^d}$. So T is not uniquely ergodic. □

Exercises

Exercise 3.4.1. Let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be the continuous map

$$T(x, y) = (x + \alpha, y + x^2) \mod \mathbb{Z}^2,$$

where α is irrational and the torus $\mathbb{T}^2$ is parametrized by $[0, 1)^2$. Show that T is ergodic.

Exercise 3.4.2. Prove Corollary 3.3.7 using Theorem 3.4.3.

3.5 Ergodicity and unique ergodicity of affine nilmanifold automorphisms

Suppose $M = G/\Gamma$ is a compact manifold. Let $M_1 = G/G_{(2)}\Gamma$ be the horizontal torus of M, which is a quotient of M. Let $T \in \mathrm{Aff}(M)$ be an affine automorphism of M, then $Tx = bA(x)$ where $b \in G$ and $A \in \mathrm{Aut}(M)$. By Lemma 3.1.6, A can be viewed as an automorphism of G that preserves Γ. Since A preserves both $G_{(2)}$ and Γ, it projects to a linear automorphism $A_1 \in \mathrm{Aut}(M_1)$. Thus T projects to the affine transform $T_1x = g_1A_1(x)$ where g_1 is the projection of g in $G/G_{(2)}$. Moregenerally, for each $1 \leq j \leq s$, T projects to $T_i \in \mathrm{Aut}(M_i)$, where $M_i = G/G_{(i+1)}\Gamma$, and $T_ix = g_iA_i(x)$, where g_i and A_i are respectively the projections of g and A_i to $G/G_{(i+1)}$.

The main result of this section is:

Theorem 3.5.1 (Parry [Par69a])**.** *T is ergodic if and only if T_1 is.*

Because T_1 is a factor of T, the "only if" part is a general property of factor maps. So only the "if" direction needs to be proved. We now assume T_1 is ergodic. Then no eigenvalue of T_1 on $L^2_0(M_1)$ is equal to 1. The theorem would follow from the following fact:

Proposition 3.5.2. *If T_1 is ergodic, then every eigenfunction of T on $L^2_0(M)$ has the form $f_1 \circ \pi$ where f_1 is an eigenfunction of T_1 on $L^2_0(M_1)$ and $\pi : M \to M_1$ is the projection.*

We now reduce the proposition to a simpler setting.

Recall that $M = M_s$ is a principal $\mathbb{T}^{d_s}$-bundle over M_{s-1} where $d_s = \dim G_{(s)}$. More precisely, the torus group $G_{(s)}/\Gamma_{(s)} \cong \mathbb{T}^{d_s}$ freely acts on M_s with the quotient space modulo the orbit equivalence relation being M_{s-1}, where $\Gamma_{(s)} = G_{(s)} \cap \Gamma$. Below, identify $G_{(s)}$ with $\mathbb{R}^{d_s}$ and $\Gamma_{(s)}$ with $\mathbb{Z}^{d_s}$.

Any function $f \in L^2(M)$ can be decomposed into a Fourier series along the $\mathbb{T}^{d_s}$-fibers by (3.12), where the Fourier frequencies are from $\mathbb{Z}^{d_s} = \widehat{\mathbb{T}^{d_s}}$. Note that $f \in L^2_0(M)$ if and only if $\widehat{f}(0, x) = 0$.

For $z \in G_{(s)}$,

$$T(zx) = bA(zx) = bA(z).A(x) = A(z)b.A(x) = A(z).Tx \tag{3.16}$$

because A preserves $G_{(s)}$ and $A(z) \in G_{(s)}$ commutes with b. Thus Lemma 3.4.2 applies in this seeting.

For $g \in G$, let $(L_g f)(x) = f(gx)$. Because f is bounded, by Luzin's Theorem, the map $g \to L_g f$ from G to $L^2_0(M)$ is continuous for a given f.

Proof of Proposition 3.5.2. First of all, let $n \in \mathbb{N}$ be such that the order of every root of unity among the eigenvalues of $D_e A \in \mathrm{Aut}(\mathfrak{g})$ divides n. Then $D_e A^n$ has no root of unity other than 1 among its eigenvalues. Suppose $f \in L^2_0(M)$ is an eigenfunction of T, then it is also an eigenvalue of T^n. If the proposition holds for T^n, f factors through an eigenfunction f_1 of T_1^n on M_1. Because f is an eigenfunction of T and T projects to T_1 on M_1, f_1 must be an eigenfunction of T_1. Hence the proposition also holds for T. Therefore, one may assume $D_e A$ has no root of unity other than 1.

The statement holds trivially if G is abelian. Suppose for induction that G is s-step nilpotent and the proposition is true for all cases where the step of nilpotency is bounded by $s-1$. In particular, every eigenfunction of T_{s-1} in $L^2_0(M_{s-1})$ factors through an eigenfunction of T_1 in on $L^2_0(M_1)$. Hence the problem reduces to show that every eigenfunction f of T factors through an eigenfunction of T_{s-1}.

Suppose for the sake of contradiction that f is not a constant and let θ be the eigenvalue such that $Tf = \theta f$. Then it follows the proof of implication

(2)⇒(3) in Theorem 3.4.3 that $A^{\mathrm{T}}\xi = \xi$ and $\widehat{f}(\xi, x) \neq 0$ in L^2 for some $\xi \in \mathbb{Z}^{d_s}\backslash\{0\}$. So by Lemma3.4.2, $\widehat{f}(\xi, Tx) = \lambda \widehat{f}(\xi, x)$. In other words, $x \to \widehat{f}(\xi, x)$ is a non-constant eigenfunction for T.

Let $\ker_{\mathbb{T}^{d_s}} \xi$ be the kernel of ξ on $\mathbb{T}^{d_s}$, which is a closed subgroup of dimension $d_s - 1$ and the projection of $\ker \xi \in \mathbb{R}^{d_s} \cong G_{(s)}$. Let $G' = G/\ker \xi$ and $M' = M/\ker_{\mathbb{T}^{d_s}} \xi$, then G' is a simply connected nilpotent Lie group by Corollary 1.5.14, and $M' = G'/\Gamma'$ is a compact nilmanifold where $\Gamma' = \Gamma/(\Gamma \cap \ker \xi)$. The lower central series of G' is $G'_{(i)} = G_{(i)}/\ker \xi$, in particular, $G'_{(s)} \cong \mathbb{R}$. By (3.15), $x \to \widehat{f}(\xi, x) = 0$ is constant along orbits of $\ker_{\mathbb{T}^{d_s}} \xi$, thus $\widehat{f}(\xi, x)$ factors through a function $f'(x)$ in $L^2(M')$. M' is a principal $\mathbb{T}^1$-bundle over M_{s-1} where the fiber $\mathbb{T}^1$ is given by $\mathbb{T}^{d_s}/\ker_{\mathbb{T}^{d_s}} \xi$. So by (3.15), for every $z' \in \mathbb{T}^1$ and almost all $x' \in M'$, $f'(z'x') = e(z')f'(x')$.

Therefore, after replacing M and f respectively by M' and f', we may assume that $d_s = 1$ and a non-zero T-invariant function $f \in L^2_0(M)$ satisfies

$$f(zx) = e(z)f(x), \forall z \in G_{(s)} \cong \mathbb{R}. \tag{3.17}$$

It remains to show in this case that f factors through a function on M_{s-1}. Observe that under these conditions, $|f|$ is T-invariant and is constant along $G_{(s)}/\Gamma_{(s)}$-fibers. Thus $|f|$ factors through a T_{s-1}-invariant function on M_{s-1}, and therefore by inductive hypothesis, through a T_1-invariant function on M_1. As T_1 is ergodic, $|f|$ is constant. By replacing f with $\frac{f}{|f|}$, it may also be assumed that $|f| = 1$.

Let $H = \{g \in G_{(s-1)} : f \text{ is an eigenfunction of } L_g\}$ and let $\lambda(h)$ be the corresponding eigenvalue for $h \in H$. Then H is obviously a subgroup, and λ is a group morphism from H to the unit circle in $\mathbb{C}$. Moreover, H is closed thanks to the continuity above. By (3.17), $G_{(s)} \subseteq H$ and $\lambda(z) = e(z)$ for $z \in G_{(s)} \cong \mathbb{R}$, where $\mathbb{Z} \subseteq \mathbb{R}$ is identified with $\Gamma_{(s)} \cong \mathbb{Z}$. In addition, H is a normal subgroup of $G_{(s-1)}$. To see this, suppose $g \in G_{(s-1)}$ and $h \in H$. Then because $h^{-1}ghg^{-1} \in G_{(s)}$,

$$L_{ghg^{-1}}f = (L_{h^{-1}ghg^{-1}} \circ L_h)f = L_{h^{-1}ghg^{-1}}(\lambda(h)f) = e(h^{-1}ghg^{-1})\lambda(h)f,$$

f is an eigenfunction of $L_{ghg^{-1}}$ with eigenvalue $e(h^{-1}ghg^{-1})\lambda(h)$.

Remark that the group morphism λ can be extended to a function on G by $\lambda(g) = \langle L_g f, f\rangle$. A priori, the resulting function does not need to be group morphism any more. However, for $g \in G$ and $h \in H$, we still have

$$\lambda(hg) = \langle L_g L_h f, f\rangle = \langle \lambda(h) L_g f, f\rangle = \lambda(h)\lambda(g). \tag{3.18}$$

Furthermore, as L_g is unitary, $\lambda(g^{-1}) = \langle L_{g^{-1}}f, f\rangle = \langle f, L_g(f)\rangle = \overline{\lambda(g)}$. Therefore

$$\lambda(gh) = \overline{\lambda(h^{-1}g^{-1})} = \overline{\lambda(h^{-1})\lambda(g^{-1})} = \lambda(h)\lambda(g). \tag{3.19}$$

These also show that, if $g, k \in G$ are such that $gk \in H$, then

$$\lambda(gk) = \lambda(g^{-1})^{-1}\lambda(k). \tag{3.20}$$

In particular, we can apply (3.20) to gh and $g^{-1}h^{-1}$ where $g \in G$ and $h \in H$, because $ghg^{-1}h^{-1} \in H$ in this case. This would yield

$$\begin{aligned}\lambda(ghg^{-1}h^{-1}) =& \lambda(h^{-1}g^{-1})^{-1}\lambda(g^{-1}h^{-1}) \\ =& \lambda(h^{-1})^{-1}\lambda(g^{-1})^{-1}\lambda(g^{-1})\lambda(h^{-1}) = 1.\end{aligned} \tag{3.21}$$

In other words, $[G, H] \subseteq H \cap G_{(s)}$ is in contained in $\ker\lambda$ and

$$f(gx) = f(x), \forall g \in [G, H]. \tag{3.22}$$

Comparing (3.17) to (3.22), we see that $[G, H] \subsetneq G_{(s)}$. Since $G_{(s)} \cong \mathbb{R}$, it follows that $[G, H]$ is discrete. In particular, if H^0 denotes the connected component of H (which is a Lie subgroup as H is closed), then $[G, H^0]$ is trivial. Therefore, H^0 is in the center Z of G.

Decompose the complexified abelian Lie algebra $(\mathfrak{g}_{(s-1)}/\mathfrak{g}_{(s)}) \otimes \mathbb{C}$ as a direct sum $\bigoplus_\zeta V_{\mathbb{C}}^\zeta$ of generalized eigenspaces of the linear tranform induced by D_eA, where ζ is the corresponding eigenvalue. Then $\mathfrak{g}_{(s-1)}/\mathfrak{g}_{(s)} = \bigoplus_\zeta V^\zeta$ where $V^\zeta = (V_{\mathbb{C}}^\zeta \oplus V_{\mathbb{C}}^{\bar{\zeta}}) \cap (\mathfrak{g}_{(s-1)}/\mathfrak{g}_{(s)})$ and the sum runs over a subset of indices ζ that contains exactly one value from each imaginary comjugate pairs of eigenvalues. Let $\widetilde{V}^\zeta$ be the preimage of V^ζ in $\mathfrak{g}_{(s-1)}$. Then $\mathfrak{g}_{(s)} \subseteq \widetilde{V}^\zeta$. Take $Y \in \widetilde{V}^\zeta$ and $g = \exp Y$.

Suppose first $|\zeta| < 1$, then as $n \to \infty$, $(D_eA)^nY \to \mathfrak{g}_{(s)}$. So there are $W_n \in \mathfrak{g}$ and $Y_n \in \mathfrak{g}_{(s)}$ such that $W_n \to 0$ and $(D_eA)^nY = W_n + Y_n$. Since $[W_n, Y_n] = 0$, $A^n(g) = \exp(W_n + Y_n) = g_nw_n$ where $w_n = \exp W_n \to 0$, and $y_n = \exp Y_n \in G_{(s)}$.

For almost every $x \in M$,

$$\begin{aligned}T^n(gx) =& b \cdot A(b) \cdot \cdots \cdot A^{n-1}(b)A^n(gx) \\ =& Q_n(b)A^n(g)A^n(x) \\ =& Q_n(b)A^n(g)Q_n(b)^{-1}T^n(x) \\ =& Q_n(b)A^n(g)Q_n(b)^{-1}A^n(g)^{-1}A^n(g)T^n(x) \\ =& Q_n(b)A^n(g)Q_n(b)^{-1}A^n(g)^{-1}g_nw_nT^n(x)\end{aligned}$$

where $Q_n(b) = b{\cdot}A(b)\cdot\cdots A^{n-1}(b)$. Because $Q_n(b)A^n(g)Q_n(b)^{-1}A^n(g)^{-1}g_n \in G_{(s)} \subseteq H$ and f is assumed to be T-invariant, we have

$$f(x) = f(T^n x);$$

$$L_g f(x) = f(T^n(gx)) = \lambda\Big(Q_n(b)A^n(g)Q_n(b)^{-1}A^n(g)^{-1}g_n\Big)f(w_nT^nx).$$

In consequence,

$$\begin{aligned}|\langle f, L_g f\rangle| =& \left|\int f(T^n x)\overline{f(w_nT^nx)}\mathrm{d}\mathbf{m}_M(x)\right|\\ =& \left|\int f(x)\overline{f(w_nx)}\mathrm{d}\mathbf{m}_M(x)\right| = |\langle f, L_{w_n}f\rangle|.\end{aligned}$$

Since $L_{w_n}f$ is continuous in w_n, as $n\to\infty$, $L_{w_n}f$ tends to f in $L^2(M)$.So by taking limit we obtain that $|\langle f, L_g f\rangle| = \|f\|^2_{L^2(M)} = \|L_gf\|^2_{L^2(M)}$.This can happen if and only if $L_gf = \eta f$ for some η from the unit circle in $\mathbb{C}$, i.e. when f is an eigenvector of L_g. Thus $g\in H$. Because this argument applies to $\exp tY$ for all $t\in\mathbb{R}$, we know that $Y\in\mathfrak{h}$.

When $|\zeta| > 1$, $|\zeta^{-1}| < 1$ and ζ^{-1} is an eigenvalue of the linear map induced by D_eA^{-1} on $\mathfrak{g}_{(s-1)}/\mathfrak{g}_{(s)}$, with Y being the eigenvector. So one can work with T^{-1} in stead of T, and prove that $g\in H$.

Finally, assume $\zeta = 1$, then $(\zeta-1)^n\pi(Y) = 0$ for some $n\in\mathbb{N}$ where $\pi(Y)\in V^\zeta$ is the projection of Y in $\mathfrak{g}_{(s-1)}/\mathfrak{g}_{(s)}$. So $(D_eA-1)^nY\in\mathfrak{g}_{(s)}\subseteq\mathfrak{h}$. We claim in this case that $Y\in\mathfrak{h}$ and hence $g\in H$. For this, it suffices to show that if a vector $W\in\mathfrak{h}$ satisfies $(D_eA)W - W\in\mathfrak{h}$, then $W\in\mathfrak{h}$.

For such a vector W, let $w = \exp W\in G_{(s-1)}$, then

$$\begin{aligned}A(w)\cdot w^{-1} =& \exp((D_eA)W)w^{-1} = \exp\big(W + ((D_eA)W - W)\big)w^{-1}\\ =& \exp((D_eA)W - W)w\cdot w^{-1} = \exp((D_eA)W - W)\in H.\end{aligned}$$

Here we used that $(D_eA)W - W\in\mathfrak{h}\subseteq\mathfrak{z}$ and hence $[W, (D_eA)W - W] = 0$. In addition, because $A(w)\in G_{(s-1)}$, $bA(w)b^{-1}A(w)^{-1}\in G_{(s)}\subseteq H$. In this case, notice

$$\begin{aligned}L_wf(x) =& (L_wTf)(x) = Tf(wx) = f(bA(w)A(x))\\ =& f(bA(w)b^{-1}A(w)^{-1}A(w)w^{-1}w\cdot bA(x))\\ =& \lambda(bA(w)b^{-1}A(w)^{-1})\lambda(A(w)w^{-1})f(w\cdot Tx)\\ =& \lambda(bA(w)b^{-1}A(w)^{-1})\lambda(A(w)w^{-1})\cdot(TL_wf)(x).\end{aligned}$$

Therefore, $L_w f$ is also an eigenfunction of T, and so is $(L_w f)\bar{f}$. Moreover, for all $z \in G_{(s)}$,

$$\begin{aligned}\big((L_w f)\bar{f}\big)(zx) = f(wzx)\overline{f(zx)} = f(zwx)\overline{f(zx)} = \lambda(z)\overline{\lambda(z)}f(wx)\overline{f(x)} \\ =\big((L_w f)\bar{f}\big)(x).\end{aligned}$$

Thus $(L_w f)\bar{f}$ factors through an invariant function of T_{s-1} on M_{s-1}, and therefore by inductive hypothesis, factors through an eigenfunction ϕ of T_1 of on M_1.

If the eigenvalue for ϕ is not equal to 1, then

$$\langle L_w f, f\rangle = \int (L_w f)\bar{f}\mathrm{d}\mathbf{m}_M = \int \phi \mathrm{d}_{M_1} = \langle \phi, 1\rangle = 0.$$

Otherwise, because T_1 is ergodic, ϕ is a constant c and $L_w f = c(\bar{f})^{-1} = cf$. In paricular, $|c| = 1$ as L_w is an isometry. In this case, $\langle L_w f, f\rangle = c$. To summarize, $|\langle L_w f, f\rangle|$ is either 0 or 1. Remark that this is not only true for w but for all $w_t = \exp tW$. Because $|\langle L_w f, f\rangle|$ is continuous in w, $|\langle L_w f, f\rangle| = |\langle L_{w_0} f, f\rangle| = |\langle f, f\rangle|$. This happens only when ϕ is a constant c, in which case $L_w f = cf$ and $w \in H$.

In summary, we have proved that $\mathfrak{h} \subseteq \mathfrak{g}_{(s-1)}$ contains the subspace $\bigoplus_{\substack{|\zeta|\neq 1 \\ \text{or } \zeta=1}} \widetilde{V}^{\zeta}$. Recall that $H^0 \subseteq Z$, thus $\mathfrak{h} \subseteq \mathfrak{z} \cap \mathfrak{g}_{(s-1)}$ and the automorphism ϕ of $\mathfrak{g}_{(s-1)}/(\mathfrak{z} \cap \mathfrak{g}_{(s-1)})$ induced by $D_e A$ has only eigenvalues that are not equal to 1 but of modulus 1. Observe that $\mathfrak{s}_{(s)} \subseteq \mathfrak{z}$, hence $\mathfrak{g}_{(s-1)}/(\mathfrak{z} \cap \mathfrak{g}_{(s-1)})$ is abelian and can be identified with $G_{(s-1)}/(Z \cap G_{(s-1)})$. With this identification, A induces a map, which acts as ϕ, on the vector space $G_{(s-1)}/(Z \cap G_{(s-1)})$, while preserving the $\Gamma_{(s-1)}/(Z \cap \Gamma_{(s-1)})$, which is a lattice by Proposition 2.3.10. This shows that all eigenvalues of ϕ are algebraic units. It is a basic fact from number theory that if all Galois conjugates of an algebraic unit ζ has modulus 1, then ζ is a root of unity.[4] However, all eigenvalues of ϕ are eigenvalues of $D_e A$, which earlier simplicification include no non-trivial roots of unity. Therefore $\mathfrak{g}_{(s-1)}/(\mathfrak{z} \cap \mathfrak{g}_{(s-1)})$ is trivial and $\mathfrak{g}_{(s-1)} \subset \mathfrak{z}$. This makes the step of nilpotency $s-1$ instead of s, a contradiction. The proof is completed. □

Since we have a complete criterion for ergodicity of affine nilmanifold automorphisms, it is natural to ask about unique ergodicity.

[4]To see this, prove first that such algebraic units in every given number field form a finite group.

Theorem 3.5.3 (Parry [Par69a]). *Let $Tx = bA(x)$ be an affine automorphism of a compact nilmanifold $M = G/\Gamma$, then the following are equivalent:*

(1) T is uniquely ergodic;

(2) The induced affine automorphism T_1 on the horizontal torus is unique ergodic;

(3) D_eA is unipotent and T is ergodic.

Proof. **(1)⇒(1):** Notice that by Lemma 3.2.12 for every ergodic T_1-invariant probability measure there is at least one ergodic T-invariant probability measure that lifts it. Therefore, if T is uniquely ergodic, then so is T_1.

(2)⇒(3): First, if T_1 is unique ergodic then it is ergodic by Corollary 3.4.4. It then follows T is ergodic by Theorem 3.5.1.

Again by Corollary 3.4.4, the automorphism A_1 induced by A on $G/G_{(2)} \cong \mathbb{R}^{d_1}$ is unipotent, i.e. only has eigenvalue 1. As A_1 can be identified with D_eA_1, by Proposition 3.1.5, D_eA also only has eigenvalue 1.

(3)⇒(1): We inductively prove this in terms of the step of nilpotency. When $s = 1$, $T = T_1$ and the claim is empty. Suppose the claim is known for $(s-1)$-step nilmanifolds and T_1 is uniquely ergodic, then the projected map T_{s-1} on $M_{s-1} = G/G_{(s)}\Gamma$ is uniquely ergodic.

Because A acts on $G_{(s)}$ by the matrix $D_eA|_{\mathfrak{g}(s)}$, which only has eigenvalue 1 and is unipotent, it follows from (3.16) and the implication (3)⇒(1) in Theorem 3.4.3 that T is uniquely ergodic. □

Exercises

Exercise 3.5.1. For all automorphisms T of the three dimensional Heisenberg nilmanifold (classified in Exercise 3.1.3), either show T is ergodic or describe the ergodic decomposition of $\mathbf{m}_M$.

Exercise 3.5.2. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function, and T be the diffeomorphism of the 3-dimensional Heisenberg group defined by $T((x, y, z)\Gamma) = (\alpha, \beta, h(x, y))((x, y, z)\Gamma)$. Show that T preserves $\mathbf{m}_M$ and is ergodic with respect to $\mathbf{m}_M$ if and only if $T_1(x, y) = (x + \alpha, y + \beta)$ is ergodic on $\mathbb{T}^2$.