EXPONENTIAL MIXING OF NILMANIFOLD AUTOMORPHISMS

ALEXANDER GORODNIK* AND RALF SPATZIER**

Abstract. We study dynamical properties of automorphisms of compact nilmanifolds and prove that every ergodic automorphism is exponentially mixing and exponentially mixing of higher orders. This allows to establish probabilistic limit theorems and regularity of solutions of the cohomological equation for such automorphisms. Our method is based on the quantitative equidistribution results for polynomial maps combined with Diophantine estimates.

1. Introduction

Dynamics and ergodic theory of toral automorphisms have been well understood for quite some time. Ergodic toral automorphisms are always mixing and even Bernoulli [14], and have dense sets of periodic points [24]. However, unless they are hyperbolic, the toral automorphisms lack the specification property and, in particular, don't have Markov partitions [20]. Nonetheless, it is known that ergodic toral automorphisms satisfy the central limit theorem and its refinements [19, 17]. Regarding the quantitative aspects Lind established exponential mixing for ergodic toral automorphisms using Fourier analysis [21]. Surprisingly, some of these ergodic-theoretic properties turned out to be more delicate for automorphisms of compact nilmanifolds and still remained unexplored. In particular, the exponential mixing, which is one of the main results of this paper, has not been established and does not easily follow using the harmonic analysis on nilpotent Lie groups.

1.1. **Exponential mixing.** Let G be a simply connected nilpotent Lie group and Λ a discrete cocompact subgroup. The space $X = G/\Lambda$ is called a *compact nilmanifold*. An *automorphism* α of X is a diffeomorphism of X which lifts to an automorphism of G. We denote by μ the Haar probability measure on X. Then α preserves μ. The ergodic-theoretic properties of the dynamical system $\alpha \curvearrowright (X,\mu)$ have been studied by Parry [25]. He proved that an automorphism is ergodic if and only if the induced map on the maximal toral quotient is ergodic, and every ergodic automorphism satisfies the Kolmogorov property. In particular, it is mixing of all orders. In this paper we establish quantitative mixing properties of such automorphisms. We fix a right-invariant Riemannian metric on G which also defines a metric on X and denote by $C^\theta(X)$ the space of θ-Hölder functions on X.

Now we state the first main result of the paper.

* Supported in part by EPSRC grant EP/H000091/1 and ERC grant 239606.
** Supported in part by NSF grant DMS-0906085.

Theorem 1.1. *Let α be an ergodic automorphism of a compact nilmanifold X. Then there exists $\rho = \rho(\theta) \in (0,1)$ such that for all $f_0, f_1 \in C^\theta(X)$ and $n \in \mathbb{N}$,*

$$\int_X f_0(x) f_1(\alpha^n(x))\, d\mu(x) = \left(\int_X f_0\, d\mu\right)\left(\int_X f_1\, d\mu\right) + O\left(\rho^n \|f_0\|_{C^\theta}\|f_1\|_{C^\theta}\right).$$

The proof of Theorem 1.1 is based on an equidistribution result for the exponential map established in Section 2 (see Corollary 2.3 below), which is deduced from the work of Green and Tao [12]. This result shows that images of boxes under the exponential map are equidistributed in X provided that a certain Diophantine condition holds. We complete the proof of Theorem 1.1 in Section 3. The main idea is to relate the correlations $\langle f_0, f_1 \circ \alpha^n \rangle$ to averages along suitable foliations in X and apply the equidistribution result established in Section 2. In order to verify the Diophantine condition we use the Diophantine properties of algebraic numbers. This leads to the proof of Theorem 1.1 under an irreducibility condition, and the proof of the theorem in general uses an inductive argument.

We also establish multiple exponential mixing for ergodic automorphisms of compact nilmanifolds. For ergodic toral automorphisms, multiple exponential mixing was proved by Pène [26] and Dolgopyat [8].

Theorem 1.2. *Let α be an ergodic automorphism of a compact nilmanifolds X. Then there exists $\rho = \rho(\theta) \in (0,1)$ such that for all $f_0, \ldots, f_s \in C^\theta(X)$ and $n_0, \ldots, n_s \in \mathbb{N}$,*

$$\int_X \left(\prod_{i=0}^{s} f_i(\alpha^{n_i}(x))\right) d\mu(x) = \prod_{i=0}^{s}\left(\int_X f_i\, d\mu\right) + O\left(\rho^{\min_{i\neq j}|n_i - n_j|}\prod_{i=0}^{s}\|f_i\|_{C^\theta}\right).$$

The proof of Theorem 1.2 is given in Section 4. The first step of the proof is to establish an equidistribution result for images of exponential map in $X \times \cdots \times X$ (see Proposition 4.2). Then we approximate higher order correlations by averages of the exponential map. As in the proof of Theorem 1.1, we first consider the irreducible case and then deduce the theorem in general using an inductive argument.

1.2. **Probabilistic limit theorems.** It is well-known that the exponential mixing property is closely related to other chaotic properties of dynamical systems and, in particular, to the central limit theorem for observables $f \circ \alpha^n$. While one does not imply the other in general, the martingale differences approach [13, Ch. 5] usually allows to deduce the proof of the central limit theorem from quantitative equidistribution of unstable foliations. Using this approach, the central limit theorem and its generalisations have been established for ergodic toral automorphisms in [19, 17] and for ergodic automorphisms of 3-dimensional nilmanifolds in [4]. Here we extend these results to general nilmanifolds.

Theorem 1.3. *Let α be an ergodic automorphism of a compact nilmanifolds X and $f \in C^\theta(X)$ with $\int_X f\, d\mu = 0$ which is not a measurable coboundary (i.e., $f \neq \phi \circ \alpha - \phi$ for any measurable function ϕ on X). Then there exists $\sigma = \sigma(f) > 0$, the so-called* variance *of f, such that*

$$\mu\left(\left\{x \in X : \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(\alpha^i(x)) \in (a,b)\right\}\right) \to \frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-t^2/(2\sigma^2)}\, dt$$

as $n \to \infty$.

We also prove the *central limit theorem for subsequences*, and the *Donsker and Strassen invariance principles* for ergodic automorphisms of nilmanifolds. We refer to Section 6 for a detailed discussion of the results. The main ingredient of the proof is the exponential equidistribution of leaves of unstable foliations, which is established for this purpose in Section 5.

1.3. **Cohomological equation.** Let α be a measure-preserving transformation of a probability space (X, μ) and $f : X \to \mathbb{R}$ is a measurable function. The functional equation

$$f = \phi \circ \alpha - \phi, \qquad \phi : X \to \mathbb{R}, \tag{1.1}$$

is called the *cohomological equation*. This equation plays important role in many aspects of the theory of dynamical systems (for instance, existence of smooth invariant measures, existence of conjugacies, existence of isospectral deformations, rigidity of group actions). If a measurable solution ϕ of (1.1) exists, the function f is called a *measurable coboundary*. It is easy to see that a solution of (1.1) is unique (up to measure zero) up to an additive constant when α is ergodic with respect to μ.

We will apply the exponential mixing property to investigate regularity of solutions of the cohomological equation.

Theorem 1.4. *Let α be an ergodic automorphism of a compact nilmanifold X, and let $f \in C^\infty(X)$ be such that (1.1) has a measurable solution. Then there exists a C^∞ solution of (1.1).*

The first result of this type was proved by Livsic [22] for Anosov diffeomorphism and flows. More precisely, if α is an Anosov diffeomorphism and the given C^∞ function f is a measurable coboundary, then the cohomological equation (1.1) has a C^∞ solution ϕ. There are also versions of this result for Hölder functions and C^k functions. Recently, Wilkinson [35] has generalised Livsic' results to partially hyperbolic diffeomorphisms that satisfy the accessibility property. Automorphisms of nilmanifolds however do not have the accessibility property. In fact, the problem of regularity of solutions of the coboundary equation for ergodic toral automorphisms, which are not hyperbolic, turns out to be quite subtle [33, 16]. Veech [33] has constructed an example of $f \in C^1(\mathbb{T}^d)$ which sums to zero along periodic orbits, but the cohomological equation (1.1) has no C^1 solutions. By [33], if $f \in C^k(\mathbb{T}^d)$ with $k > d$ and (1.1) has a measurable solution, then there exists a solution in $C^{k-d}(\mathbb{T}^d)$. We are not aware of any results regarding regularity of solutions of (1.1) for a general ergodic toral automorphism when $f \in C^k(\mathbb{T}^d)$ with $k < d$.

Theorem 1.4 is proved in Section 7. We use a construction from Section 6 to show that there exists a square-integrable solution. Then we use a new method of proving smoothness as developed by Fisher, Kalinin and Spatzier in [10]: we consider the solution as a distribution on the space of Hölder functions and study its regularity along the stable, unstable and central foliations of α. While regularity along the first two foliations can be deduced using the standard contraction argument, the regularity along the central foliation is deduced from the exponential mixing property.

Combining Theorem 1.3 and 1.4, we observe that C^∞ functions with variance 0 are truly exceptional.

Corollary 1.5. *Let α be an ergodic automorphism of a compact nilmanifold X, and let $f \in C^\infty(X)$ have $\sigma(f) = 0$. Then f is a C^∞ coboundary.*

1.4. **Further generalisations.**

- We note that the results established here can be generalised to *affine diffeomorphisms* of a compact nilmanifold $X = G/\Lambda$. Those are diffeomorphisms $\sigma : X \to X$ that can be lifted to affine maps $\tilde{\sigma}$ of G, i.e., maps $\tilde{\sigma}$ that have constant derivatives with respect to a right invariant framing of G. Since every such diffeomorphism σ is of the form $\sigma(x) = g_0\alpha(x)$ for $g_0 \in G$ and an automorphism α of X, our method applies to such maps as well (see Remark 3.4 below).
- More generally, one may consider *infra-nilmanifolds* [6]. Let G be a simply connected nilpotent Lie group, C a compact subgroup of $\mathrm{Aut}(G)$, and Γ a discrete torsion-free subgroup of $G \rtimes C$ such that G/Γ is compact. The space $Y = G/\Gamma$ is called an infra-nilmanifold. By [1, Th. 1], the group $\Lambda = G \cap \Gamma$ has finite index in Γ. Hence, the infra-nilmanifold Y is finitely covered by the nilmanifold $X = G/\Lambda$. An affine diffeomorphism of Y is a diffeomorphism which lifts to an affine map of G. Every such diffeomorphism is of the form $g \mapsto g_0\alpha(g)$, where $g_0 \in G$ and α is an automorphism of G that preserves the orbits of Γ. By [7, Theorem 3.4], we must have $\alpha\Gamma\alpha^{-1} = \Gamma$. Since by [1, Prop. 2] Λ is the maximal normal nilpotent subgroup of Γ, we deduce that $\alpha(\Lambda) = \alpha\Lambda\alpha^{-1} = \Lambda$. Therefore, every affine diffeomorphism of Y lifts to an affine diffeomorphism of X, and our results can be generalised to this setting.
- Our techniques also allow to establish exponential mixing properties for $\mathbb{Z}^k$*-actions* by automorphisms of nilmanifolds when $k \geq 2$. Since this requires more delicate Diophantine estimates, we pursue this in a sequel paper [11]. This result has found a striking application to the problem of global rigidity of smooth actions. Given any C^∞-action of $\mathbb{Z}^k$, $k \geq 2$, on a nilmanifold that has sufficiently many Anosov elements, Fisher, Kalinin and the second author showed in [10] that this action is C^∞-conjugate to an affine action on the nilmanifold.
- In view of the works of Katznelson [14] and Parry [25], it is natural to ask whether ergodic automorphisms of compact nilmanifolds are *Bernoulli*. Surprisingly, we could not find this result in the literature, and in Section 8 we establish the Bernoulli property. While this easily follows from the works of Marcuard [23] and Rudolph [29], and the proof does not rely on the main ideas of this paper, we include this result in Section 8 to complete our discussion of ergodic properties of nilmanifold automorphisms.

Acknowledgements. We are indebted to J. Rauch for discussions concerning his regularity theorems with M. Taylor. Also we thank F. Ledrappier for the reference to Le Borgne's work which was crucial to our applications to the central limit theorem. A.G. would like to thank the University of Michigan for hospitality during his visit when the work on this project had started. R.S. thanks the University of Bristol for hospitality and support during this work.

2. Equidistribution of box maps

Let G be a simply connected nilpotent Lie group, Λ a discrete cocompact subgroup, and $X = G/\Lambda$ the corresponding nilmanifold equipped with the Haar probability measure μ. We fix a a right invariant Riemannian metric d on G which also defines a metric on X. Let $\mathcal{L}(G)$ be the Lie algebra of G and $\exp : \mathcal{L}(G) \to G$ the exponential map. The aim of this section is to investigate distribution of images of the maps

$$\mathbb{R}^k \to X : t \mapsto g_1 \exp(\iota(t)) g_2 \Lambda$$

with $g_1, g_2 \in G$ and an affine map $\iota : \mathbb{R}^k \to \mathcal{L}(G)$.

The lattice subgroup Λ defines a rational structure on $\mathcal{L}(G)$. Let $\pi : G \to G/G'$ denote the factor map, where G' is the commutator subgroup. We also have the corresponding map $D\pi : \mathcal{L}(G) \to \mathcal{L}(G/G')$. We fix an identification $G/G' \simeq \mathcal{L}(G/G') \simeq \mathbb{R}^l$ that respects the rational structures.

We call a *box map* an affine map

$$\iota : B := [0, T_1] \times \cdots \times [0, T_k] \to \mathcal{L}(G)$$

of the form

$$\iota : (t_1, \dots, t_k) \mapsto v + t_1 w_1 + \cdots + t_k w_k \tag{2.1}$$

with $v, w_1, \dots, w_k \in \mathcal{L}(G)$. We denote by

$$|B| := T_1 \cdots T_k$$

the volume of the box B and by

$$\min(B) := \min_{i=1,\dots,k} T_i,$$

the length of the shortest side of B.

Theorem 2.1. *There exist $L_1, L_2 > 0$ such that for every $\delta \in (0, 1/2)$ and every box map $\iota : B \to \mathcal{L}(G)$ as in (2.1), one of the following holds:*

(i) *For every Lipschitz function $f : X \to \mathbb{R}$, $u \in \mathcal{L}(G)$, and $g \in G$,*

$$\left| \frac{1}{|B|} \int_B f(\exp(u) \exp(\iota(t)) g\Lambda)\, dt - \int_X f\, d\mu \right| \le \delta \|f\|_{Lip}. \tag{2.2}$$

(ii) *There exists $z \in \mathbb{Z}^l \backslash \{0\}$ such that*

$$\|z\| \ll \delta^{-L_1} \quad \text{and} \quad |\langle z, D\pi(w_i) \rangle| \ll \delta^{-L_2}/T_i \quad \text{for all } i = 1, \dots, k. \tag{2.3}$$

Here and in the rest of the paper we explicitly list dependences of implied constants on relevant parameters. In particular, in (2.3) the implied constants are independent of the box map.

Proof: We suppose that (i) fails for some Lipschitz function f, $u \in \mathcal{L}(G)$, and $g \in G$. Then will show that (ii) holds. We pick $L \ge 2$ such that

$$\max\{\|u\|, \|v\|, T_1\|w_1\|, \dots, T_k\|w_k\|\} \le \delta^{-L}. \tag{2.4}$$

Making a linear change of variables in the integral (2.2), we arrange that $T_i \ge 1$ and $\|w_i\| \le 1$.

For $x_1, x_2, x_3 \in \mathcal{L}(G)$, we consider the map

$$P(x_1, x_2, x_3) := \exp(x_1)\exp(x_2 + x_3)\exp(-x_2)\exp(-x_1).$$

We note that G can be equipped with a structure of algebraic group so that exp is a polynomial isomorphism. Hence, the map P can be written as

$$P(x_1, x_2, x_3) = \exp(p_1(x_1, x_2, x_3)e_1 + \cdots + p_d(x_1, x_2, x_3)e_d)$$

for some polynomials p_i. Since $P(x_1, x_2, 0) = e$, these polynomials satisfy $p_i(x_1, x_2, 0) = 0$. Hence, assuming that $\|x_3\| \le 1$, we obtain

$$|p_i(x_1, x_2, x_3)| \ll (1 + \|x_1\|)^{\deg(p_i)}(1 + \|x_2\|)^{\deg(p_i)}\|x_3\|, \quad i = 1, \ldots, d.$$

Since in the neighborhood of the origin,

$$d(e, P(x_1, x_2, x_3)) \ll \max_{i=1,\ldots,d} |p_i(x_1, x_2, x_3)|,$$

we deduce that there exists $C_0 \ge 2$ such that for every $\epsilon \in (0, 1/2)$ and $x_1, x_2, x_3 \in \mathcal{L}(G)$ satisfying $\|x_1\|, \|x_2\| \le (k + 1)\epsilon^{-1}$ and $\|x_3\| \le k\epsilon^{C_0}$, we have

$$d(e, P(x_1, x_2, x_3)) \le \epsilon. \tag{2.5}$$

We set $s = \lceil \delta^{-CL} \rceil$, where $C \ge C_0$ is sufficiently large and will be specified later (see (2.7) and (2.12)–(2.13) below). Let

$$\mathcal{N} := \{(n_1, \ldots, n_k) : n_i = 0, \ldots, N_i - 1\},$$

where $N_i := \lceil T_i s \rceil \ge s$. We consider the polynomial map

$$p(n) := \exp(u)\exp\left(v + \sum_{i=1}^{k} \frac{n_i}{s} w_i\right) g, \quad n \in \mathcal{N}.$$

For $t_i \in [\frac{n_i}{s}, \frac{n_i+1}{s}]$, we apply (2.5) with

$$x_1 := u, \quad x_2 := v + \sum_{i=1}^{k} \frac{n_i}{s} w_i, \quad x_3 := \sum_{i=1}^{k} \left(t_i - \frac{n_i}{s}\right) w_i, \quad \epsilon = \delta^L.$$

It follows from (2.4) that

$$\begin{aligned}
&\|x_1\| \le \delta^{-L},\\
&\|x_2\| \le \delta^{-L} + \sum_{i=1}^{k} (N_i - 1)s^{-1}\|w_i\| \le \delta^{-L} + \sum_{i=1}^{k} T_i\|w_i\| \le (k + 1)\delta^{-L},\\
&\|x_3\| \le \sum_{i=1}^{k} s^{-1}\|w_i\| \le ks^{-1} \le k\delta^{CL}.
\end{aligned}$$

Hence, (2.5) gives

$$\begin{aligned}(2.6)\qquad & d\left(p(n)\Lambda, \exp(u)\exp\left(v+\sum_{i=1}^{k} t_i w_i\right) g\Lambda\right) \\ \le & d\left(e, \exp(u)\exp\left(v+\sum_{i=1}^{k} t_i w_i\right) g p(n)^{-1}\right) \\ = & d\left(e, \exp(u)\exp\left(v+\sum_{i=1}^{k} t_i w_i\right)\exp\left(v+\sum_{i=1}^{k}\frac{n_i}{s} w_i\right)^{-1}\exp(u)^{-1}\right) \le \delta^L.\end{aligned}$$

For $n=(n_1,\cdots,n_k)\in\mathcal{N}$, we set

$$B_n := \left[\frac{n_1}{s}, \frac{n_1+1}{s}\right]\times\cdots\times\left[\frac{n_k}{s}, \frac{n_k+1}{s}\right].$$

It follows from (2.6) that for every Lipschitz function f and $n\in\mathcal{N}$,

$$\left|f(p(n)\Lambda)|B_n| - \int_{B_n} f(\exp(u)\exp(\iota(t))g\Lambda)\,dt\right| \le \delta^L s^{-k}\|f\|_{Lip}.$$

We also observe that $B\supset \cup_{n\in\mathcal{N}} B_n$, and

$$\left|B - \left(\bigcup_{n\in\mathcal{N}} B_n\right)\right| \le k s^{-1} T_1\cdots T_k \le k s^{-k-1} N_1\cdots N_k.$$

Therefore, we deduce that

$$\begin{aligned}&\left|\sum_{n\in\mathcal{N}} f(p(n)\Lambda)|B_n| - \int_B f(\exp(u)\exp(\iota(t))g\Lambda)\,dt\right| \\ \le & \sum_{n\in\mathcal{N}}\left|f(p(n)\Lambda)|B_n| - \int_{B_n} f(\exp(u)\exp(\iota(t))g\Lambda)\,dt\right| + k s^{-k-1} N_1\cdots N_k\|f\|_{Lip} \\ \le & \left(\delta^L + k s^{-1}\right) s^{-k} N_1\cdots N_k\|f\|_{Lip},\end{aligned}$$

and

$$\begin{aligned}
&\left|\frac{1}{N_1\cdots N_k}\sum_{n\in\mathcal{N}} f(p(n)\Lambda) - \frac{1}{|B|}\int_B f(\exp(u)\exp(\iota(t))g\Lambda)\,dt\right| \\
\le &\left|\frac{1}{N_1\cdots N_k}\sum_{n\in\mathcal{N}} f(p(n)\Lambda) - \frac{s^k}{N_1\cdots N_k}\int_B f(\exp(u)\exp(\iota(t))g\Lambda)\,dt\right| \\
&+\left(\frac{1}{|B|}-\frac{s^k}{N_1\cdots N_k}\right)|B|\|f\|_{Lip} \\
\le &\left(\delta^L + ks^{-1}\right)\|f\|_{Lip} + \left(1-\frac{s^kT_1\cdots T_k}{N_1\cdots N_k}\right)\|f\|_{Lip} \\
\le &\left(\delta^L + ks^{-1}\right)\|f\|_{Lip} + \left(1-\frac{(N_1-1)\cdots(N_k-1)}{N_1\cdots N_k}\right)\|f\|_{Lip} \\
\le &\left(\delta^L + c_ks^{-1}\right)\|f\|_{Lip} \le (\delta^L + c_k\delta^{CL})\|f\|_{Lip}
\end{aligned}$$

with some $c_k > 0$. Here in the last line, we used that $N_i = \lceil T_i s\rceil \ge s = \lceil \delta^{-CL}\rceil$. We choose $C = C(k) > C_0 > 0$, so that

$$\delta^2 + c_k\delta^{CL} \le 3\delta/4. \tag{2.7}$$

Then since we are assuming that (2.2) fails, we deduce from the previous estimate that

$$\left|\frac{1}{N_1\cdots N_k}\sum_{n\in\mathcal{N}} f(p(n)\Lambda) - \int_X f\,d\mu\right| \ge (\delta-\delta^L-c_k\delta^{CL})\|f\|_{Lip} \ge \delta/4\|f\|_{Lip}. \tag{2.8}$$

Now we apply [12, Th. 8.6] to the polynomial map $p(n)$. Note that

$$\pi\left(p(n)\right) = D\pi\left(u+v+\sum_{i=1}^k \frac{n_i}{s}w_i\right) + \pi(g),$$

and

$$\pi\left(p(n)\right) - \pi\left(p(n-e_i)\right) = \frac{D\pi(w_i)}{s}.$$

By [12, Th. 8.6], there exist $L_1, L_2 > 0$ such that for every $\rho\in(0,1/2)$ and $N_1,\ldots,N_k \ge 1$, one of the following holds:

(i′) For every Lipschitz function $f : X\to\mathbb{R}$,

$$\left|\frac{1}{N_1\cdots N_k}\sum_{n\in\mathcal{N}} f(p(n)\Lambda) - \int_X f\,d\mu\right| \le \rho\|f\|_{Lip}. \tag{2.9}$$

(ii″) There exists $z\in\mathbb{Z}^l\backslash\{0\}$ such that

$$\|z\| \ll \rho^{-L_1} \quad\text{and}\quad \operatorname{dist}\left(\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle, \mathbb{Z}\right) \ll \rho^{-L_2}/N_i, \quad i=1,\ldots,k, \tag{2.10}$$

where the implied constants depend only on the degree of the polynomial map.

Comparing (2.8) and (2.9), we deduce that (ii$''$) holds with $\rho = \delta/4$, and there exists $z \in \mathbb{Z}^l\backslash\{0\}$ such that

$$\|z\| \ll \delta^{-L_1} \quad \text{and} \quad \operatorname{dist}\left(\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle, \mathbb{Z}\right) \ll \delta^{-L_2}/N_i, \quad i = 1, \dots, k. \tag{2.11}$$

Since $\|w_i\| \le 1$, we obtain

$$\left|\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle\right| \le \|z\|\|D\pi\|\|w_i\|s^{-1} \ll \delta^{-L_1+CL} \le \delta^{-L_1+C}. \tag{2.12}$$

Taking $C = C(L_1) > 0$ sufficiently large, the above estimate implies that

$$\left|\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle\right| \le 1/4. \tag{2.13}$$

Then

$$\operatorname{dist}\left(\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle, \mathbb{Z}\right) = \left|\left\langle z, \frac{D\pi(w_i)}{s}\right\rangle\right|,$$

and it follows from (2.11) that

$$|\langle z, D\pi(w_i)\rangle| \ll s\delta^{-L_2}/N_i \le \delta^{-L_2}/T_i, \quad i = 1, \dots, k.$$

Hence, (2.3) holds, as required. This completes the proof of the theorem. $\diamond$

We call a box map ι, defined as in (2.1), (c_1, c_2)-*Diophantine* if there exists at least one vector $w \in \Omega := [-1,1]D\pi(w_1) + \cdots + [-1,1]D\pi(w_k)$ such that

$$|\langle z, w\rangle| \ge c_1\|z\|^{-c_2} \quad \text{for all } z \in \mathbb{Z}^l\backslash\{0\}. \tag{2.14}$$

We emphasize that only one element of Ω has to satisfy the relevant Diophantine condition. This allows for the following remark which we will use later, e.g. in the proof of Theorem 3.1.

Remark 2.2. Let ι be a (c_1, c_2)-*Diophantine* box map, W the subspace spanned by the image of ι, and S a compact subset of $\mathrm{GL}(W)$. Then there exists a constant $c = c(S) > 0$, which only depends on S, such that for all $s \in S$, the box map $s \circ \iota$ is $(c\,c_1, c_2)$-Diophantine. Indeed, since S is compact, there exists $c = c(S) > 0$ such that for every $s \in S$,

$$[-1,1]D\pi(w_1) + \cdots + [-1,1]D\pi(w_k) \subset [-c^{-1}, c^{-1}]D\pi(sw_1) + \cdots + [-c^{-1}, c^{-1}]D\pi(sw_k).$$

If $w \in [-1,1]D\pi(w_1) + \cdots + [-1,1]D\pi(w_k)$ satisfies (2.14), then $cw \in [-1,1]D\pi(sw_1) + \cdots + [-1,1]D\pi(sw_k)$ and satisfies (2.14) with c_1 replaced by $c\,c_1$. Hence, the box map $s \circ \iota$ is $(c\,c_1, c_2)$-Diophantine.

The following corollary will play a crucial role in the next section.

Corollary 2.3. *Given $\theta, c_1, c_2 > 0$, there exists $\kappa = \kappa(c_2, \theta) > 0$ such that for every θ-Hölder function $f : X \to \mathbb{R}$, $u \in \mathcal{L}(G)$, (c_1, c_2)-Diophantine box map $\iota : B \to \mathcal{L}(G)$, and $x \in X$, we have*

$$\frac{1}{|B|}\int_B f(\exp(u)\exp(\iota(t))x)\,dt = \int_X f\,d\mu + O_{c_1,c_2}(\min(B)^{-\kappa}\|f\|_{C^\theta}).$$

Proof: We first give a proof assuming that the function f is Lipschitz.

We write the box map ι as

$$\iota(t) = v + t_1 w_1 + \cdots + t_k w_k, \quad t \in B = [0, T_1] \times \cdots \times [0, T_k]$$

with $v, w_1, \ldots, w_k \in W$ and $T_1, \ldots, T_k > 0$.

We take $\kappa, \epsilon > 0$ such that $\frac{-L_2\kappa+1}{L_1\kappa} > c_2$ and moreover $\frac{-L_2(\kappa+\epsilon)+1}{L_1\kappa} > c_2$, where L_1 and L_2 are as in Theorem 2.1. Let $\delta = \min(B)^{-\kappa}$. We first assume that $\min(B)$ is sufficiently large, so that $\delta < 1/2$. Then by Theorem 2.1, either

$$\left|\frac{1}{|B|}\int_B f(\exp(u)\exp(\iota(t))x)\,dt - \int_X f\,d\mu\right| \le \min(B)^{-\kappa}\|f\|_{Lip} \tag{2.15}$$

for all Lipschitz functions $f : X \to \mathbb{R}$, $u \in \mathcal{L}(G)$ and $x \in X$, or there exists $z \in \mathbb{Z}^l\backslash\{0\}$ such that

$$\|z\| \ll \min(B)^{L_1\kappa},$$
$$|\langle z, D\pi(w_i)\rangle| \ll \min(B)^{L_2\kappa}/T_i \le \min(B)^{L_2\kappa-1}, \quad i = 1, \ldots, k.$$

If the latter holds, then we deduce that there exists $z \in \mathbb{Z}^l\backslash\{0\}$ such that

$$\begin{aligned}|\langle z, D\pi(w_i)\rangle| &\ll \min(B)^{-L_2\epsilon}\min(B)^{L_2(\kappa+\epsilon)-1} \ll \min(B)^{-L_2\epsilon}\|z\|^{-\frac{-L_2(\kappa+\epsilon)+1}{L_1\kappa}} \\ &\le \min(B)^{-L_2\epsilon}\|z\|^{-c_2}\end{aligned}$$

for all $i = 1, \ldots, k$. Writing $w = \sum_{i=1}^k a_i D\pi(w_i)$ with $a_i \in [-1, 1]$, we also deduce that

$$|\langle z, D\pi(w)\rangle| \le \sum_{i=1}^k |\langle z, D\pi(w_i)\rangle| \ll \min(B)^{-L_2\epsilon}\|z\|^{-c_2}.$$

When $\min(B)$ is sufficiently large, this estimate contradicts (2.14). Hence, we conclude that when $\min(B) \ge T_0 = T_0(c_1, c_2)$, (2.15) holds and

$$\frac{1}{|B|}\int_B f(\exp(u)\exp(t)x)\,dt = \int_X f\,d\mu + O(\min(B)^{-\kappa}\|f\|_{Lip}).$$

It is also clear that this estimate holds in the range $[0, T_0]$ with the implicit constant depending on T_0, and this completes proof of the corollary for Lipschitz functions.

For Hölder functions, we use the following well-known approximation result. While we only use the estimate of the Lipschitz norm here, we will need this lemma in full in Section 7.

Lemma 2.4. *Given $\varepsilon > 0$ and $0 < \theta \le 1$, for any θ-Hölder function $f : X \to \mathbb{R}$, there is a C^∞ function $f_\varepsilon : X \to \mathbb{R}$ which satisfies the following bounds*

$$\|f_\varepsilon - f\|_{C^0} \le \varepsilon^\theta\|f\|_{C^\theta} \quad \text{and} \quad \|f_\epsilon\|_{Lip} \ll \epsilon^{-\dim(X)-1}\|f\|_{C^0}. \tag{2.16}$$

Furthermore, for all $l \in \mathbb{N}$,

$$\|f_\varepsilon\|_{C^l} \ll_l \varepsilon^{-\dim(X)-l}\|f\|_{C^0}. \tag{2.17}$$

Proof: Given a θ-Hölder function $f : X \to \mathbb{R}$, we set

$$f_\epsilon(x) := \int_G \phi_\epsilon(g^{-1}) f(gx)\, dm(g),$$

where m denotes the Haar measure on G, and ϕ_ϵ is a nonnegative function such that

$$\|\phi_\epsilon\|_{Lip} \ll \epsilon^{-\dim(X)-1}, \quad \int_G \phi_\epsilon\, dm = 1, \quad \mathrm{supp}(\phi_\epsilon) \subset B_\epsilon(e).$$

Then

$$\|f_\epsilon - f\|_{C^0} \le \max_{x\in X} \int_G \phi_\epsilon(g^{-1}) |f(gx) - f(x)|\, dm(g) \le \epsilon^\theta \|f\|_{C^\theta}.$$

For $x, y \in X$ satisfying $d(x, y) < \epsilon$, we can write $y = hx$ with $h \in B_\epsilon(e)$. Then

$$|f_\epsilon(x) - f_\epsilon(y)| \le \int_G |\phi_\epsilon(g^{-1}) - \phi_\epsilon(hg^{-1})| f(gx)|\, dm(g) \ll \epsilon^{-\dim(X)-1} \|f\|_{C^0}.$$

Hence,

$$\|f_\epsilon\|_{Lip} \ll \epsilon^{-\dim(X)-1} \|f\|_{C^0}.$$

We can further assume that ϕ_ε satisfies for all $l \in \mathbb{N}$,

$$\|\phi_\varepsilon\|_{C^l} \ll_l \varepsilon^{-\dim(X)-l} \|\phi\|_{C^l},$$

and it follows that

$$\|f_\varepsilon\|_{C^l} \ll_l \varepsilon^{-\dim(X)-l} \|f\|_{C^0},$$

as the lemma claims. $\diamond$

Returning to the proof of Corollary 2.3, we obtain

$$\begin{aligned}
\frac{1}{|B|} \int_B f(\exp(u)\exp(t)x)\, dt &= \frac{1}{|B|} \int_B f_\epsilon(\exp(u)\exp(t)x)\, dt + O(\epsilon^\theta \|f\|_{C^\theta}) \\
&= \int_X f_\epsilon\, d\mu + O\left(\min(B)^{-\kappa} \|f_\epsilon\|_{Lip} + \epsilon^\theta \|f\|_{C^\theta}\right) \\
&= \int_X f\, d\mu + O\left((\epsilon^{-\dim(X)-1} \min(B)^{-\kappa} + \epsilon^\theta) \|f\|_{C^\theta}\right).
\end{aligned}$$

To optimise the error term, we set $\epsilon = \min(B)^{-\kappa/(\dim(X)+\theta+1)}$. We readily obtain the corollary for Hölder functions. $\diamond$

We remark that the procedure just outlined applies quite generally, and allows to go from estimates for Lipschitz functions to ones for Hölder functions. In particular, exponential mixing for Lipschitz or even only smooth functions always implies exponential mixing for Hölder functions.

3. Mixing

In this section, we prove Theorem 1.1 on exponential mixing. Let us recall the statement:

Theorem 3.1. *Let α be an ergodic automorphism of a compact nilmanifold $X = G/\Lambda$. Then there exists $\rho = \rho(\theta) \in (0,1)$ such that for all θ-Hölder functions $f_0, f_1 : X \to \mathbb{R}$ and $n \in \mathbb{N}$,*

$$\int_X f_0(x) f_1(\alpha^n(x))\, d\mu(x) = \left(\int_X f_0\, d\mu\right)\left(\int_X f_1\, d\mu\right) + O\left(\rho^n \|f_0\|_{C^\theta} \|f_1\|_{C^\theta}\right).$$

We denote by μ the Haar probability measure on X, and by m the Haar measure on G which is normalised, so that $m(F) = 1$ where F is a fundamental domain for G/Λ.

Every automorphism β of G defines a Lie-algebra automorphism $D\beta : \mathcal{L}(G) \to \mathcal{L}(G)$ such that $\beta \circ \exp = \exp \circ D\beta$. If $\beta(\Lambda) \subset \Lambda$, then $D\beta$ preserves the rational structure of $\mathcal{L}(G)$ defined by Λ.

As in Section 2, we equip the group G with the structure of an algebraic group, so that exp is a polynomial isomorphism. More precisely, one can construct a basis, a so-called Malcev basis, $\{e_1, \ldots, e_d\}$ of $\mathcal{L}(G)_{\mathbb{Q}}$, such that the map

$$\mathbb{R}^d \to G : (t_1, \ldots, t_d) \mapsto \exp(t_1 e_1) \cdots \exp(t_d e_d)$$

is a polynomial isomorphism,

$$\Lambda = \exp(\mathbb{Z} e_1) \cdots \exp(\mathbb{Z} e_d),$$

and

$$F := \exp([0,1) e_1) \cdots \exp([0,1) e_d) \subset G$$

is a fundamental domain for G/Λ (see [5, 1.2.7, 5.1.6, 5.3.1]).

We present the proof of Theorem 3.1 in two stages: in Section 3.1, we give a proof assuming a suitable irreducibility condition, and in Section 3.2, we reduce the proof to the irreducible case using an inductive argument.

3.1. **Proof under an irreducibility assumption.** Let w be a (real or complex) eigenvector of $D\alpha$ acting on $\mathcal{L}(G) \otimes \mathbb{C}$ with eigenvalue λ such that $|\lambda| > 1$. Such an eigenvector exists by the following lemma.

Lemma 3.2. *If α is an ergodic automorphism of a nontrivial compact nilmanifold $X = G/\Lambda$, then $D\alpha$ has an eigenvalue λ with $|\lambda| > 1$.*

Proof: By [5, 5.4.13], $\Lambda G'/G'$ is a lattice in $G/G' \simeq \mathbb{R}^l$. The automorphism α defines a linear automorphism of the torus $T := G/(\Lambda G') \simeq \mathbb{R}^l/L$, where L is a lattice in $\mathbb{R}^l$, and there is an α-equivariant map $X \to T$ induced by π. Since $\alpha|_{\mathbb{R}^l}$ preserves the lattice L, it follows that the eigenvalues of $\alpha|_{\mathbb{R}^l}$ are algebraic integers. If we suppose that all these eigenvalues satisfy $|\lambda| \le 1$, then it follows from [9, Th. 1.31] that all the eigenvalues of $\alpha|_{\mathbb{R}^l}$ are roots of unity. Then the automorphism $\alpha|_T$ is not ergodic, and this contradicts ergodicity of α. Hence, $\alpha|_{\mathbb{R}^l}$ has an eigenvalue λ with $|\lambda| > 1$, and this implies that $D\alpha$ has such an eigenvalue as well. $\diamond$

Since $D\alpha$ preserves the rational structure on $\mathcal{L}(G)$ defined by the lattice Λ, we may choose the eigenvector w with coordinates in the algebraic closure $\overline{\mathbb{Q}}$. In the real case, we denote by W the corresponding one-dimensional eigenspace of $\mathcal{L}(G)$. In the complex case, we denote by W the two-dimensional subspace $\langle w, \bar{w}\rangle \cap \mathcal{L}(G)$, where $\bar{w}$ denotes the complex conjugate. We note that in a suitable basis

$$D\alpha|_W = r \cdot \omega \tag{3.1}$$

where $r = |\lambda| > 1$ and ω is a rotation by angle $\mathrm{Im}(\lambda)$.

In this subsection, we give a proof of Theorem 3.1 assuming that $D\pi(W)$ is not contained in any proper rational subspace of $\mathbb{R}^l$. This condition is used to guarantee existence of a "generic" vector in $D\pi(W)$ given by the following lemma.

Lemma 3.3. *Let $V \subset \mathbb{R}^l$ be a subspace defined over $\overline{\mathbb{Q}} \cap \mathbb{R}$ such that V is not contained in any proper subspace defined over $\mathbb{Q}$. Then there exists $w \in V \cap \overline{\mathbb{Q}}^l$ whose coordinates are real numbers linearly independent over $\mathbb{Q}$.*

Proof: Let $\{v_1, \ldots, v_s\}$ be a basis of V whose coordinates v_{ij} are in $\overline{\mathbb{Q}} \cap \mathbb{R}$. We denote by K the field generated by these coordinates. Clearly, K is a finite extension of $\mathbb{Q}$. We can pick $\alpha_1, \ldots, \alpha_s \in \overline{\mathbb{Q}} \cap \mathbb{R}$ which are linearly independent over K (for instance, we can take a sufficiently large finite extension K' of K and choose $\{\alpha_i\}$ from a basis of K' over K).

We set $w = \sum_{i=1}^{s} \alpha_i v_i$. Suppose that there exists $c \in \mathbb{Q}^l$ such that $c \cdot w = 0$. Then we have

$$c \cdot w = \sum_{j=1}^{l} c_j \left(\sum_{i=1}^{s} \alpha_i v_{ij} \right) = \sum_{i=1}^{s} \left(\sum_{j=1}^{l} c_j v_{ij} \right) \alpha_i = 0.$$

Now because $\sum_{j=1}^{l} c_j v_{ij}$ is in K, it follows that $\sum_{j=1}^{l} c_j v_{ij} = 0$ for all i, and $c \cdot V = 0$. Since V is not contained in any proper rational subspace, we conclude that $c = 0$, which concludes the proof. $\diamond$

As we remarked above, the subspace W is defined over $\overline{\mathbb{Q}}$. Moreover, since W is invariant under complex conjugation, it is defined over $\overline{\mathbb{Q}} \cap \mathbb{R}$. This implies that the subspace $D\pi(W)$ is also defined over $\overline{\mathbb{Q}} \cap \mathbb{R}$. Hence, by Lemma 3.3, $D\pi(W)$ contains a vector w whose coordinates are real algebraic numbers that are linearly independent over $\mathbb{Q}$. By [2, Th. 7.3.2], there exist $c_1, c_2 > 0$ (in fact, one can take any $c_2 > l - 1$) such that

$$|\langle z, w\rangle| \ge c_1 \|z\|^{-c_2} \quad \text{for all } z \in \mathbb{Z}^l \backslash \{0\}, \tag{3.2}$$

This will allow us to apply Corollary 2.3 to box maps $\mathbb{R}^{\dim(W)} \to W$.

Let $E \subset \mathcal{L}(G)$ be the preimage of the fundamental domain F under the exponential map. Since E is the image of $[0,1)^d$ under a polynomial isomorphism, it is a domain in $\mathcal{L}(G)$ with a piecewise smooth boundary. We fix a basis of $\mathcal{L}(G)$ which contains the basis of W and consider a tessellation of $\mathcal{L}(G)$ by cubes C of size ϵ with respect to this basis. Then

$$\left| E - \bigcup_{C \subset E} C \right| \ll \epsilon. \tag{3.3}$$

Using the above notation, we rewrite the original integral as

$$\int_X f_0(x) f_1(\alpha^n(x))\, d\mu(x) = \int_F f_0(g\Lambda) f_1(\alpha^n(g)\Lambda)\, dm(g) \tag{3.4}$$
$$= \int_E f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du,$$

where we used that the Haar measure on G is the image of a suitably normalised Lebesgue measure on $\mathcal{L}(G)$ under the exponential map [5, 1.2.10]. It follows from (3.3) that

$$\int_E f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du \tag{3.5}$$
$$= \sum_{C\subset E} \int_C f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du + O(\epsilon \|f_0\|_{C^0} \|f_1\|_{C^0}).$$

For every cube C in the above sum, we fix $u_C \in C$. Then for all $u \in C$,

$$|f_0(\exp(u)\Lambda) - f_0(\exp(u_C)\Lambda)| \le d(\exp(u), \exp(u_C)) \|f_0\|_{Lip} \ll \epsilon^\theta \|f_0\|_{C^\theta},$$

and

$$\int_C f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du \tag{3.6}$$
$$= f_0(\exp(u_C)\Lambda) \int_C f_1(\exp((D\alpha)^n u)\Lambda)\, du + O(\epsilon^\theta \|f_0\|_{C^\theta} \|f_1\|_{C^\theta}).$$

Since the cubes C are chosen in a compatible way with the subspace W, they can be written as $C = B' + B$ where B is a cube in W and B' is a cube in the complementary subspace. Given a cube $B \subset W$, we introduce a box map $\iota_B : \mathbb{R}^{\dim(W)} \to W$, defined with respect to the fixed basis of W, such that $\iota_B([0,\epsilon]^{\dim(W)}) = B$. Since ω is a rotation, it follows from Remark 2.2 that for some $c > 0$, each of the box maps

$$\mathbb{R}^{\dim(W)} \to W : t \mapsto v + \omega^n \iota_B(t), \quad v \in \mathcal{L}(G),$$

is $(c\,c_1, c_2)$-Diophantine. Therefore, applying Corollary 2.3, we obtain there exists $\kappa > 0$ such that for every $v \in \mathcal{L}(G)$,

(3.7)
$$\frac{1}{|B|} \int_B f_1(\exp(v + (D\alpha)^n b)\Lambda)\, db = \epsilon^{-\dim(W)} \int_{[0,\epsilon]^{\dim(W)}} f_1(\exp(v + (D\alpha)^n \iota_B(t))\Lambda)\, dt$$
$$= (r^n\epsilon)^{-\dim(W)} \int_{[0,r^n\epsilon]^{\dim(W)}} f_1(\exp(v + \omega^n \iota_B(t))\Lambda)\, dt$$
$$= \int_X f_1\, d\mu + O\left((r^n\epsilon)^{-\kappa} \|f_1\|_{C^\theta}\right).$$

Since this estimate is uniform over $v \in \mathcal{L}(G)$, we deduce that

$$\frac{1}{|C|} \int_C f_1(\exp((D\alpha)^n u)\Lambda)\, du = \frac{1}{|B'||B|} \int_{B'} \int_B f_1(\exp((D\alpha)^n b' + (D\alpha)^n b)\Lambda)\, db db'$$
$$= \int_X f_1\, d\mu + O\left((r^n\epsilon)^{-\kappa} \|f_1\|_{C^\theta}\right).$$

Combining the last estimate with (3.5) and (3.6), we deduce that

$$\int_E f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du = \left(\sum_{C\subset E} f_0(\exp(u_C)\Lambda)|C|\right)\int_X f_1\, d\mu + O\left(\left(\sum_{C\subset E}|C|(r^n\epsilon)^{-\kappa}+\epsilon^\theta\right)\|f_0\|_{C^\theta}\|f_1\|_{C^\theta}\right).$$

Since f_0 is θ-Hölder and $\mathrm{diam}(C)\ll\epsilon$, we obtain using (3.3),

$$\begin{aligned}\sum_{C\subset E} f_0(\exp(u_C)\Lambda)|C| &= \sum_{C\subset E}\int_C f_0(\exp(u)\Lambda)\, du + O(\epsilon^\theta\|f_0\|_{C^\theta})\\ &= \int_E f_0(\exp(u)\Lambda)\, du + O(\epsilon^\theta\|f_0\|_{C^\theta})\\ &= \int_X f_0\, d\mu + O(\epsilon^\theta\|f_0\|_{C^\theta}).\end{aligned} \tag{3.8}$$

Hence,

$$\int_E f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du = \left(\int_X f_1\, d\mu\right)\left(\int_X f_0\, d\mu\right) + O\left((r^n\epsilon)^{-\kappa}+\epsilon^\theta)\|f_0\|_{C^\theta}\|f_1\|_{C^\theta}\right).$$

To optimise the error term, we choose $\epsilon = r^{-n\kappa/(\kappa+\theta)}$. Then

$$\begin{aligned}\int_X f_0(x) f_1(\alpha^n(x))\, d\mu(x) &= \int_E f_0(\exp(u)\Lambda) f_1(\exp((D\alpha)^n u)\Lambda)\, du\\ &= \left(\int_X f_0\, d\mu\right)\left(\int_X f_1\, d\mu\right) + O\left(\rho^n\|f_0\|_{C^\theta}\|f_1\|_{C^\theta}\right),\end{aligned}$$

where $\rho = r^{-\kappa\theta/(\kappa+\theta)}\in(0,1)$. This proves Theorem 3.1 under the irreducibility assumption.

We also observe that Corollary 2.3 implies the following stronger version of estimate (3.7): for every $h\in G$, automorphism β of G such that $\beta = id$ on G/G', and $v\in\mathcal{L}(G)$,

$$\frac{1}{|B|}\int_B f_1(h\beta(\exp(v+(D\alpha)^n t))\Lambda)\, dt = \int_X f_1\, d\mu + O\left((r^n\epsilon)^{-\kappa}\|f_1\|_{C^\theta}\right).$$

Indeed, using that $\beta\circ\exp = \exp\circ D\beta$, we obtain

$$\begin{aligned}&\frac{1}{|B|}\int_B f_1(h\beta(\exp(v+(D\alpha)^n t))\Lambda)\, dt\\ =&(r^n\epsilon)^{-\dim(W)}\int_{[0,r^n\epsilon]^{\dim(W)}} f_1(\exp((D\beta)v+(D\beta)\omega^n\iota_B(t))\Lambda)dt.\end{aligned}$$

Since $(D\pi)(D\beta) = D\pi$, the box maps

$$t\mapsto (D\beta)v+(D\beta)\omega^n\iota_B(t)$$

are also $(c\,c_1, c_2)$-Diophantine, and the same estimate as in (3.7) holds. Therefore, the above argument implies that

$$\int_X f_0(x) f_1(h\,\beta(\alpha^n(x)))\,d\mu(x) = \left(\int_X f_0\,d\mu\right)\left(\int_X f_1\,d\mu\right) + O\left(\rho^n \|f_0\|_{C^\theta}\|f_1\|_{C^\theta}\right) \tag{3.9}$$

uniformly on $h \in G$ and automorphisms β which preserve Λ and act trivially on G/G'.

Remark 3.4. Let $\sigma : X \to X$ be an affine diffeomorphism of a compact nilmanifold X. Then $\sigma(x) = g_1\alpha(x)$ for $g_1 \in G$ and an automorphism α, and $\sigma^n(x) = g_n\alpha^n(x)$ for $g_n \in G$. Since the estimate (3.9) is uniform over $h \in G$, it also holds for affine diffeomorphisms. This allows to extend the main results of this paper to affine diffeomorphisms.

3.2. **Proof of mixing in general.** We prove Theorem 3.1 in general using induction on the dimension of the nilmanifold X.

Let $w \in \mathcal{L}(G) \otimes \mathbb{C}$ be an eigenvector of the automorphism $D\alpha$ with eigenvalue λ of maximal modulus. Since α is ergodic, $|\lambda| > 1$ by Lemma 3.2. We set $W = \mathcal{L}(G) \cap \langle w, \bar{w}\rangle$. Since $D\alpha|_W$ has eigenvalues λ and $\bar{\lambda}$, it follows either that $D\alpha|_{[W,W]}$ must have eigenvalues of modulus $|\lambda|^2 > |\lambda|$, or $[W, W] = 0$. Hence $\exp(W)$ is an abelian Lie subgroup of G. By [32, Ch. 3, Sec. 5], there exists a closed connected normal subgroup M containing $\exp(W)$ such that $M/(M\cap\Lambda)$ is compact, and for almost every $g \in G$, we have $\overline{\exp(W)g\Lambda} = Mg\Lambda$. Replacing the lattice Λ by $g\Lambda g^{-1}$, we may assume without loss of generality that

$$\overline{\exp(W)\Lambda} = M\Lambda. \tag{3.10}$$

Lemma 3.5. (i) *The group M is α-invariant.*

(ii) *Denoting by $\pi : M \to M/M'$ the factor map, $D\pi(W)$ is not contained in any proper rational subspace of $\mathcal{L}(M/M')$.*

(iii) $[G, M] < M'$.

Proof: We note that the group M can be described as the smallest closed connected normal subgroup containing $\exp(W)$ and intersecting Λ in a lattice ([32, Ch. 3, Sec. 5]). Equivalently, M is the smallest closed connected subgroup whose Lie algebra $\mathcal{L}(M)$ is an ideal in $\mathcal{L}(G)$ that contains W and is defined over $\mathbb{Q}$ with respect to the rational structure defined by Λ. To show that M is invariant under α, we observe that

$$\mathcal{L}(\alpha(M)) = D\alpha(\mathcal{L}(M))$$

also satisfies the above properties, and so does

$$\mathcal{L}(M \cap \alpha(M)) = \mathcal{L}(M) \cap D\alpha(\mathcal{L}(M)).$$

Therefore, $\alpha(M) = M$ by minimality of M proving (i).

To prove (ii), we consider the torus factor $M\Lambda/\Lambda \to T := M\Lambda/(\Lambda M')$ induced by the map π. If $\pi(W)$ is contained in a proper rational subspace of $\mathcal{L}(M/M')$, then the image of $D\pi(W)$ in T is not dense, which contradicts (3.10). This shows (ii).

Since the vector w has coordinates in $\overline{\mathbb{Q}}$, so does the vector $D\pi(w)$. For $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, we denote by $D\pi(w)^\sigma$ its Galois conjugate. Then $\left\langle D\pi(w)^\sigma : \sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})\right\rangle$ is a rational subspace, contains $D\pi(W)$ and, hence, cannot be a proper subspace. This shows that $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts transitively on the eigenvalues of $D\alpha$ in $V := \mathcal{L}(M/M')$. In particular, it follows that V does not contain any proper rational subspaces invariant under $D\alpha$. Now we consider the adjoint action Ad of G on V. Since G is nilpotent, the set V^G of G-fixed points in V is not trivial. Since V^G is $(D\alpha)$-invariant and rational, we conclude that $V^G = V$. This implies that every $g \in G$,

$$(\mathrm{Ad}(g) - id)(\mathcal{L}(M)) \subset \mathcal{L}(M)',$$

and the last claim of the lemma follows. $\diamond$

The nilmanifold $X = G/\Lambda$ fibers over the nilmanifold $Y = G/(M\Lambda)$ with fibers isomorphic to $Z = M\Lambda/\Lambda \simeq M/(M \cap \Lambda)$, and we have the disintegration formula

$$\int_X f\, d\mu = \int_Y \int_Z f(yz)\, d\mu_Z(z) d\mu_Y(y), \quad f \in C(X), \tag{3.11}$$

where μ_Y and μ_Z denote the normalised invariant measures on Y and Z respectively. Since the groups M and Λ are α-invariant, α defines transformations of Y and Z, and we obtain

$$\begin{aligned} \int_X f_0(x) f_1(\alpha^n(x))\, d\mu(x) &= \int_Y \left(\int_Z f_0(yz) f_1(\alpha^n(y)\alpha^n(z))\, d\mu_Z(z) \right) d\mu_Y(y) \\ &= \int_F \left(\int_Z f_0(gz) f_1(\alpha^n(g)\alpha^n(z))\, d\mu_Z(z) \right) dm_F(g), \end{aligned} \tag{3.12}$$

where $F \subset G$ is a bounded fundamental domain for $G/(M\Lambda)$, and m_F denotes the measure on F induced by μ_Y.

We claim that for some fixed $\rho \in (0,1)$ and every $g \in F$,

$$\begin{aligned} \int_Z f_0(gz) f_1(\alpha^n(g)\alpha^n(z))\, d\mu_Z(z) = &\left(\int_Z f_0(gz)\, d\mu_Z(z) \right) \left(\int_Z f_1(\alpha^n(g)z)\, d\mu_Z(z) \right) \\ &+ O(\rho^n \|f_0\|_{C^\theta} \|f_1\|_{C^\theta}) \end{aligned} \tag{3.13}$$

uniformly on $g \in F$. To prove the claim above, we write

$$\alpha^n(g) = am\lambda \quad \text{with } a \in F,\ m \in M,\ \lambda \in \Lambda.$$

Then

$$\int_Z f_0(gz) f_1(\alpha^n(g)\alpha^n(z))\, d\mu_Z(z) = \int_Z f_0(gz) f_1(am\beta(\alpha^n(z)))\, d\mu_Z(z),$$

where β denotes the transformation of Z induced by the automorphism $m \mapsto \lambda m \lambda^{-1}$, $m \in M$. We note that β acts trivially on M/M' by Lemma 3.5. Let

$$\phi_0(z) := f_0(gz) \quad \text{and} \quad \phi_1(z) := f_1(az) \quad \text{with } z \in Z.$$

Since $g, a \in F$, we have

$$\|\phi_0\|_{C^\theta} \ll \|f_0\|_{C^\theta} \quad \text{and} \quad \|\phi_1\|_{C^\theta} \ll \|f_1\|_{C^\theta},$$

and since $a(M\Lambda) = \alpha^n(g)(M\Lambda)$,

$$\int_Z \phi_1\, d\mu_Z = \int_Z f_1(\alpha^n(g)z)\, d\mu_Z(z).$$

Therefore, it follows from (3.9) that there exists $\rho \in (0,1)$ such that

$$\begin{aligned}&\int_Z \phi_0(z)\phi_1(m\beta(\alpha^n(z))))\, d\mu_Z(z)\\ &=\left(\int_Z \phi_0\, d\mu_Z\right)\left(\int_Z \phi_1\, d\mu_Z\right) + O(\rho^n \|\phi_0\|_{C^\theta}\|\phi_1\|_{C^\theta})\\ &=\left(\int_Z f_1(gz)\, d\mu_Z(z)\right)\left(\int_Z f_0(\alpha^n(g)z)\, d\mu_Z(z)\right) + O(\rho^n \|f_0\|_{C^\theta}\|f_1\|_{C^\theta})\end{aligned}$$

uniformly over $g, a \in F$, $m \in M$, and automorphisms β of Z which act trivially on M/M'. This proves the claim (3.13), and we conclude that

$$\int_X f_0(x)f_1(\alpha^n(x))\, d\mu(x) = \int_Y \bar{f}_0(y)\bar{f}_1(\alpha^n(y))\, d\mu_Y(y) + O(\rho^n \|f_0\|_{C^\theta}\|f_1\|_{C^\theta}), \tag{3.14}$$

where the functions $\bar{f}_i : Y \to \mathbb{R}$ are defined by $y \mapsto \int_Z f_i(yz)\, d\mu_Z(z)$. We note that

$$\int_Y \bar{f}_i\, d\mu_Y = \int_X f_i\, d\mu.$$

Since $\dim(Y) < \dim(X)$, Theorem 3.1 follows from (3.14) by induction on dimension.

4. Multiple mixing

In this section, we prove Theorem 1.2 on multiple exponential mixing. Let us recall the statement:

Theorem 4.1. *Let α be an ergodic automorphism of a compact nilmanifolds $X = G/\Lambda$. Then there exists $\rho = \rho(\theta) \in (0,1)$ such that for all θ-Hölder function $f_0, \dots, f_s : X \to \mathbb{R}$ and $n_0, \dots, n_s \in \mathbb{N}$,*

$$\int_X \left(\prod_{i=0}^{s} f_i(\alpha^{n_i}(x))\right) d\mu(x) = \prod_{i=0}^{s}\left(\int_X f_i\, d\mu\right) + O\left(\rho^{\min_{i\neq j}|n_i - n_j|}\prod_{i=0}^{s}\|f_i\|_{C^\theta}\right).$$

We note that alternately one can also deduce Theorem 4.1 from Dolgopyat's work on multiple mixing [8, Theorem 2], and Corollary 5.2 below. However, our approach, which is a simple variation of the basic argument from Section 3, allows the treatment to be self-contained.

Without loss of generality, we may assume that $n_0 = 0$ and $0 < n_1 < \cdots < n_s$.

As a preparation for the proof, we establish a result regarding equidistribution of images of box maps that generalises Corollary 2.3. We call a box map, defined as in (2.1), c_0-*bounded* if $\|w_i\| \le c_0$ for all $i = 1, \dots, k$.

Proposition 4.2. *Given $c_0, c_1, c_2, \theta > 0$, there exists $\kappa = \kappa(c_2, \theta) > 0$ such that for all θ-Hölder functions $f_1, \dots, f_s : X \to \mathbb{R}$, $u_1, \dots, u_s \in \mathcal{L}(G)$, automorphisms $\beta_1, \dots, \beta_s$ of G such that $\beta_i = id$ on G/G', $0 < r_1 < \cdots < r_s$, c_0-bounded and (c_1, c_2)-Diophantine box* *maps* $\iota_1, \dots, \iota_s : B \to \mathcal{L}(G)$*, and* $x_1, \dots, x_s \in X$*, we have*

$$\frac{1}{|B|}\int_B \left(\prod_{i=1}^s f_i(\exp(u_i)\beta_i(\exp(\iota_i(r_i t)))x_i)\right) dt = \prod_{i=1}^s \left(\int_X f_i\, d\mu\right) + O_{c_0,c_1,c_2}\left(\sigma(B, r_1, \dots, r_s)^{-\kappa} \prod_{i=1}^s \|f_i\|_{C^\theta}\right),$$

where $\sigma(B, r_1, \dots, r_s) = \min\{\min(r_1 B), r_s r_{s-1}^{-1}, \dots, r_2 r_1^{-1}\}$*.*

Proof: We first note that using the approximation argument as in the proof of Corollary 2.3, one can reduce the proof of the proposition to the case when the functions are Lipschitz. Since this part is very similar to the proof of Corollary 2.3, we omit details, and assume right away that the f_i's are Lipschitz.

The proof involves applying Theorem 2.1 to the nilmanifold $X^s = G^s/\Lambda^s$. Let $L_1, L_2 > 0$ be the constants from this theorem. To simplify notation, we write $\sigma = \sigma(B, r_1, \dots, r_s)$. Let $\delta = \sigma^{-\kappa}$ where $\kappa > 0$ is chosen so that $\frac{-\kappa(L_1+L_2)+1}{L_1\kappa} > c_2$ and moreover $\frac{-(\kappa+\epsilon)(L_1+L_2)+1}{L_1\kappa} > c_2$ for some fixed $\epsilon > 0$. First, we assume that σ is sufficiently large so that $\delta \in (0, 1/2)$.

We write the box maps ι_i as

$$\iota_i(t) = v_i + t_1 w_i^{(1)} + \cdots + t_k w_i^{(k)}, \quad t \in B = [0, T_1] \times \cdots \times [0, T_k],$$

with $v_i, w_i^{(j)} \in \mathcal{L}(G)$ and $T_1, \dots T_k, > 0$ and set

$$\begin{aligned} &f = f_1 \otimes \cdots \otimes f_s : X^s \to \mathbb{R},\\ &u = (u_1, \dots, u_s) \in \mathcal{L}(G)^s,\\ &\iota : B \to \mathcal{L}(G)^s : t \mapsto (D\beta_1\iota_1(r_1 t), \dots, D\beta_s\iota_s(r_s t)),\\ &x = (x_1, \dots, x_s) \in X^s. \end{aligned}$$

Then

$$\int_B \left(\prod_{i=1}^s f_i(\exp(u_i)\beta_i(\exp(\iota_i(r_i t))))x_i\right) db = \int_B f(\exp(u)\exp(\iota(t))x)\, dt.$$

Applying Theorem 2.1, we deduce that for every $\delta \in (0, 1/2)$, either

$$\left|\frac{1}{|B|}\int_B f(\exp(u)\exp(\iota(t)))\, dt - \int_X f\, d\mu\right| \le \delta\|f\|_{Lip}, \tag{4.1}$$

or there exists $(z_1, \dots, z_s) \in (\mathbb{Z}^l)^s \backslash \{0\}$ such that

$$\|z_1\|, \dots, \|z_s\| \ll \delta^{-L_1} = \sigma^{\kappa L_1} \tag{4.2}$$

and

$$\left|\sum_{j=1}^s r_j \left\langle z_j, D\pi D\beta_j(w_j^{(i)})\right\rangle\right| \ll \delta^{-L_2}/\min(B) = \sigma^{\kappa L_2}/\min(B) \quad \text{for all } i = 1, \dots, k. \tag{4.3}$$

We note that since $\beta_j = id$ on G/G', we have $D\pi D\beta_j = D\pi$.

Suppose that (4.2)–(4.3) holds. Since $\|w_j^{(i)}\| \le c_0$ by assumption, using the triangle inequality we deduce that

$$\left|\left\langle z_s, D\pi(w_s^{(i)})\right\rangle\right| \ll \sigma^{\kappa L_2}/\min(r_s B) + \sum_{j=1}^{s-1} \sigma^{\kappa L_1}/(r_s r_j^{-1}) \le \sigma^{\kappa(L_1+L_2)-1}$$

for all $i = 1, \ldots, k$. Then by (4.2),

$$\left|\left\langle z_s, D\pi(w_s^{(i)})\right\rangle\right| \ll \sigma^{-(L_1+L_2)\epsilon}\sigma^{(\kappa+\epsilon)(L_1+L_2)-1} \ll \sigma^{-(L_1+L_2)\epsilon}\|z\|^{-\frac{-(\kappa+\epsilon)(L_1+L_2)+1}{L_1\kappa}}$$
$$\le \sigma^{-(L_1+L_2)\epsilon}\|z\|^{-c_2}.$$

Since the box map ι_s is (c_1, c_2)-Diophantine, there exists $w_s \in \sum_{i=1}^k [-1,1] D\pi(w_s^{(i)})$ which satisfies (2.14). On the other hand, it follows from the previous estimate that

$$|\langle z_s, D\pi(w_s)\rangle| \le \sum_{i=1}^{k} \left|\left\langle z_s, D\pi(w_s^{(i)})\right\rangle\right| \ll \sigma^{-(L_1+L_2)\epsilon}\|z\|^{-c_2}.$$

When σ is sufficiently large, this estimate contradicts (2.14), unless $z_s = 0$. Hence, we deduce that $z_s = 0$.

Now we repeat the above argument and deduce from (4.2)–(4.3) that

$$\left|\left\langle z_{s-1}, D\pi(w_{s-1}^{(i)})\right\rangle\right| \ll \sigma^{\kappa L_2}/\min(r_{s-1} B) + \sum_{j=1}^{s-2} \sigma^{\kappa L_1}/(r_{s-1} r_j^{-1}) \le \sigma^{\kappa(L_1+L_2)-1}$$

for all $i = 1, \ldots, k$, and ultimately that $z_{s-1} = 0$, when σ is sufficiently large. Hence, we conclude that $(z_1, \ldots, z_s) = 0$ when $\sigma \ge \sigma_0 = \sigma_0(c_0, c_1, c_2)$. Therefore, in this range (4.1) holds with $\delta = \sigma^{-\kappa}$. This proves the claim of the proposition for sufficiently large σ. It is also clear that this estimate holds in the range $[0, \sigma_0]$ with the implicit constant depending on σ_0. This completes the proof of the proposition. ⋄

4.1. **Multiple mixing under irreducibility assumption.** In this section, we prove Theorem 4.1 under the irreducibility condition as in Section 3.1. Namely, W denotes a $(D\alpha)$-invariant subspace of $\mathcal{L}(G)$ such that $D\pi(W)$ is not contained in a proper rational subspace and (3.1) holds.

As in (3.4), we obtain

$$\int_X f_0(x) \left(\prod_{i=1}^{s} f_i(\alpha^{n_i}(x))\right) d\mu(x) = \int_E f_0(\exp(u)\Lambda) \left(\prod_{i=1}^{s} f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right) du.$$

As in Section 3.1, we tessellate the region E by cubes C of size ϵ which are compatible with the subspace W and get

$$\begin{aligned}(4.4)\qquad & \int_E f_0(\exp(u)\Lambda)\left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du \\ &= \sum_{C\subset E} f_0(\exp(u_C)\Lambda)\int_C \left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du + O\left(\epsilon^\theta \prod_{i=0}^s \|f_i\|_{C^\theta}\right),\end{aligned}$$

where $u_C \in C$. Each cube C can be written as $C = B' + B$ where B is a cube in W and B' is a cube in the complementary subspace. For every cube B, we take a box map $\iota_B : \mathbb{R}^{\dim(W)} \to W$ such that $\iota_B([0,\epsilon]^{\dim(W)}) = B$. Because ω is a rotation, there exists $c_0 > 0$ such that each of the box maps

$$\mathbb{R}^{\dim(W)} \to W : t \mapsto v + \omega^n \iota_B(t), \quad v \in \mathcal{L}(G),\ n \in \mathbb{N},$$

is c_0-bounded. It was also observed in Section 3.1 that each of these maps is (c_1, c_2)-Diophantine. Hence, Proposition 4.2 implies that there exists $\kappa \in (0,1)$ such that uniformly on $v_1, \ldots, v_s \in \mathcal{L}(G)$,

$$\begin{aligned}(4.5)\qquad & \frac{1}{|B|}\int_B \left(\prod_{i=1}^s f_i(\exp(v_i + (D\alpha)^{n_i}b)\Lambda)\right)db \\ &= \epsilon^{-\dim(W)}\int_{[0,\epsilon]^{\dim(W)}} \left(\prod_{i=1}^s f_i(\exp(v_i + r^{n_i}\omega^{n_i}\iota_B(t))\Lambda)\right)dt \\ &= \prod_{i=1}^s \left(\int_X f_i\,d\mu\right) + O\left(\sigma^{-\kappa}\prod_{i=1}^s \|f_i\|_{C^\theta}\right),\end{aligned}$$

where $\sigma = \min\{\epsilon r^{n_1}, r^{n_2-n_1}, \ldots, r^{n_s-n_{s-1}}\}$. Since this estimate is uniform over v_i's, we conclude that

$$\begin{aligned} & \frac{1}{|C|}\int_C \left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du \\ &= \frac{1}{|B'||B|}\int_{B'}\int_B \left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}b' + (D\alpha)^{n_i}b)\Lambda)\right)dbdb' \\ &= \prod_{i=1}^s \left(\int_X f_i\,d\mu\right) + O\left(\sigma^{-\kappa}\prod_{i=1}^s \|f_i\|_{C^\theta}\right).\end{aligned}$$

Now it follows from (4.4) that

$$\begin{aligned} & \int_E f_0(\exp(u)\Lambda)\left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du \\ &= \left(\sum_{C\subset E} f_0(\exp(u_C)\Lambda)|C|\right)\prod_{i=1}^s \left(\int_X f_i\,d\mu\right) + O\left((\sigma^{-\kappa} + \epsilon^\theta)\prod_{i=0}^s \|f_i\|_{C^\theta}\right),\end{aligned}$$

and by (3.8),

$$\int_E f_0(\exp(u)\Lambda)\left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du = \prod_{i=0}^s\left(\int_X f_i\,d\mu\right) + O\left((\sigma^{-\kappa}+\epsilon^\theta)\prod_{i=0}^s \|f_i\|_{C^\theta}\right).$$

Finally, taking $\epsilon = r^{-\kappa n_1/(\theta+\kappa)}$, we obtain

$$\begin{aligned}&\int_X f_0(x)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(x))\right)d\mu(x)\\ &=\int_E f_0(\exp(u)\Lambda)\left(\prod_{i=1}^s f_i(\exp((D\alpha)^{n_i}u)\Lambda)\right)du\\ &=\prod_{i=0}^s\left(\int_X f_i\,d\mu\right)+O\left(\min\{r^{\theta n_1/(\theta+\kappa)}, r^{n_2-n_1},\ldots,r^{n_s-n_{s-1}}\}^{-\kappa}\prod_{i=1}^s\|f_i\|_{C^\theta}\right).\end{aligned}$$

This completes the proof of Theorem 4.1 under the irreducibility assumption.

The proof of the general case will be given in the following section using an inductive argument. For this purpose, we note that the above argument gives the following stronger result: there exists $\rho \in (0,1)$ such that for every $h_1,\ldots,h_1 \in G$ and automorphisms $\beta_1,\ldots,\beta_s$ of G which preserve Λ that act trivially on G/G', we have

$$\begin{aligned}&\int_X f_0(x)\left(\prod_{i=1}^s f_i(h_i\beta_i(\alpha^{n_i}(x)))\right)d\mu(x)\\ &=\prod_{i=0}^s\left(\int_X f_i\,d\mu\right)+O\left(\rho^{\min\{n_1,n_2-n_1,\ldots,n_s-n_{s-1}\}}\prod_{i=0}^s\|f_i\|_{C^\theta}\right)\end{aligned}\tag{4.6}$$

uniformly over h_i's and β_i's. Indeed, Proposition 4.2 implies that in (4.5) we have, more generally,

$$\begin{aligned}&\frac{1}{|B|}\int_B\left(\prod_{i=1}^s f_i(h_i\beta_i(\exp(v_i+(D\alpha)^{n_i}b))\Lambda)\right)db\\ &=\prod_{i=1}^s\left(\int_X f_i\,d\mu\right)+O\left(\sigma^{-\kappa}\prod_{i=1}^s\|f_i\|_{C^\theta}\right),\end{aligned}$$

and the rest of the proof can be carried out as well.

4.2. **Proof of multiple mixing in general.** We use notation introduced in Section 3.2. In particular, W denotes a $(D\alpha)$-invariant subspace of $\mathcal{L}(G)$, and we arrange that $\exp(W)\Lambda = M\Lambda$ where M is closed connected normal α-invariant subgroup containing $\exp(W)$ such that $M/(M\cap\Lambda)$ is compact.

The nilmanifold $X = G/\Lambda$ fibers in α-invariant fashion over the nilmanifold $Y = G/(M\Lambda)$ with fibers isomorphic to $Z = M\Lambda/\Lambda \simeq M/(M\cap\Lambda)$, and the disintegration formula (3.11) holds. Using this disintegration formula, we obtain, similarly to (3.12),

$$
\begin{aligned}
(4.7)\qquad &\int_X f_0(x)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(x))\right)\,d\mu(x)\\
&=\int_Y\left(\int_Z f_0(yz)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(y)\alpha^{n_i}(z))\right)\,d\mu_Z(z)\right)\,d\mu_Y(y)\\
&=\int_F\left(\int_Z f_0(gz)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(g)\alpha^{n_i}(z))\right)\,d\mu_Z(z)\right)\,dm_F(g).
\end{aligned}
$$

We claim that there exists $\rho\in(0,1)$ such that for every $g\in F$,

$$
\begin{aligned}
(4.8)\qquad &\int_Z f_0(gz)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(g)\alpha^{n_1}(z))\right)\,d\mu_Z(z)\\
&=\left(\int_Z f_0(gz)\,d\mu_Z(z)\right)\prod_{i=1}^s\left(\int_Z f_i(\alpha^{n_i}(g)z)\,d\mu_Z(z)\right)+O\left(\rho^n\prod_{i=0}^s\|f_i\|_{C^\theta}\right)
\end{aligned}
$$

uniformly on $g\in F$. To prove this claim, we write $\alpha^{n_i}(g)=a_im_i\lambda_i$ with $a_i\in F$, $m_i\in M$, and $\lambda_i\in\Lambda$. Then

$$
\int_Z f_0(gz)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(g)\alpha^{n_i}(z))\right)\,d\mu_Z(z)=\int_Z f_0(gz)\left(\prod_{i=1}^s f_i(a_im_i\beta_i(\alpha^{n_i}(z)))\right)\,d\mu_Z(z),
$$

where β_i denotes the transformation of Z induced by the automorphism $m\mapsto\lambda_im\lambda_i^{-1}$, $m\in M$. Note that by Lemma 3.5 the automorphism β_i is trivial on M/M'. Let

$$
\phi_0(z):=f_0(gz)\quad\text{and}\quad\phi_i(z):=f_i(a_iz),\ i=1,\dots,s,\quad\text{with } z\in Z.
$$

Since g and a_i's belong to the compact set F,

$$
\|\phi_i\|_{C^\theta}\ll\|f_i\|_{C^\theta},\quad i=0,\dots s,
$$

and since $a_i(M\Lambda)=\alpha^{n_i}(g)(M\Lambda)$,

$$
\int_Z\phi_i\,d\mu_Z=\int_Z f_i(\alpha^{n_i}(g)z)\,d\mu_Z(z),\quad i=1,\dots,s.
$$

Applying the estimate (4.6), we deduce that for some $\rho\in(0,1)$,

$$
\begin{aligned}
&\int_Z\phi_0(z)\left(\prod_{i=1}^s\phi_i(m_i\beta_i(\alpha^{n_i}(z)))\right)\,d\mu_Z(z)\\
&=\prod_{i=1}^s\left(\int_Z\phi_i\,d\mu_Z\right)+O\left(\rho^n\prod_{i=0}^s\|\phi_i\|_{C^\theta}\right)\\
&=\left(\int_Z f_0(gz)\,d\mu_Z(z)\right)\prod_{i=1}^s\left(\int_Z f_i(\alpha^{n_i}(g)z)\,d\mu_Z(z)\right)+O\left(\rho^n\prod_{i=0}^s\|f_i\|_{C^\theta}\right).
\end{aligned}
$$

This implies the claim (4.8). Now combining (4.8) with (4.7), we deduce that

$$\int_X f_0(x)\left(\prod_{i=1}^s f_i(\alpha^{n_i}(x))\right) d\mu(x) = \int_Y \bar{f}_0(y)\left(\prod_{i=1}^s \bar{f}_i(\alpha^{n_i}(y))\right) d\mu_Y(y) + O\left(\rho^n \prod_{i=1}^s \|f_i\|_{C^\theta}\right). \tag{4.9}$$

where the functions $\bar{f}_i : Y \to \mathbb{R}$ are defined by $y \mapsto \int_Z f_i(yz)\, d\mu_Z(z)$. Clearly,

$$\int_Y \bar{f}_i\, d\mu_Y = \int_X f_i\, d\mu.$$

Since $\dim(Y) < \dim(X)$, Theorem 4.1 now follows from (4.9) by induction on dimension.

5. Equidistribution of unstable manifolds

In this section we prove an equidistribution result for unstable manifolds. Besides its own intrinsic interest, we will use this later in our treatment of probabilistic limit theorems in Section 6.

Let α be an ergodic automorphism of a compact nilmanifold $X = G/\Lambda$. We denote by $W^\alpha \subset \mathcal{L}(G)$ the unstable subspace of $D\alpha$, namely, the subspace of $\mathcal{L}(G)$ spanned by Jordan subspaces of $D\alpha$ with eigenvalues λ satisfying $|\lambda| > 1$. Note that since $[W^\alpha, W^\alpha] \subset W^\alpha$, $\exp(W^\alpha)$ is a Lie subgroup of G. We decompose W^α as a direct sum $W^\alpha = \oplus_{i=1}^\ell W_i^\alpha$, so that $D\alpha|_{W_i^\alpha}$ acts as a (real) Jordan block. Namely, each subspace W_i^α has a basis $\{w_1, \dots, w_s\}$ such that

$$\begin{aligned}(D\alpha)w_i &= \lambda w_i + w_{i+1}, \quad i < s,\\ (D\alpha)w_s &= \lambda w_s,\end{aligned} \tag{5.1}$$

where λ is a real eigenvalue of $D\alpha$, or a basis $\{w_1, w_1', \dots, w_s, w_s'\}$ such that

$$(D\alpha)w_i = aw_i + bw_i' + w_{i+1},\ \ (D\alpha)w_i' = -bw_i + aw_i' + w_{i+1}', \quad i < s, \tag{5.2}$$

$$(D\alpha)w_s = aw_s + bw_s',\ \ (D\alpha)w_s' = -bw_s + aw_s', \tag{5.3}$$

where $\lambda = a + bi$ is a complex eigenvalue of $D\alpha$. We order the subspaces W_i^α with respect to the size of $|\lambda|$. Then

$$[W^\alpha, W_i^\alpha] \subset \oplus_{j>i} W_j^\alpha. \tag{5.4}$$

For each i, we define a map $\psi_i : \mathbb{R}^{\dim(W_i^\alpha)} \to \exp(W_\alpha)$ which is either

$$\psi_i : (t_1, \dots, t_s) \mapsto \exp(t_1 w_1) \cdots \exp(t_s w_s)$$

in the real case, or

$$\psi_i : (t_1, t_1' \dots, t_s, t_s') \mapsto \exp(t_1 w_1 + t_1' w_1') \cdots \exp(t_s w_s + t_s' w_s')$$

in the complex case. Let $\psi : \mathbb{R}^{\dim(W^\alpha)} \to \exp(W^\alpha)$ be the product of the maps ψ_i. It follows from (5.4) that ψ is a diffeomorphism and that the image of the Lebesgue measure gives the Haar measure on $\exp(W^\alpha)$ [5, 1.2.7].

Theorem 5.1. *Let α be an ergodic automorphism of a compact nilmanifold $X = G/\Lambda$. Then there exist $\kappa = \kappa(\theta) > 0$ and $\rho = \rho(\theta) \in (0,1)$ such that for every box $B \subset \mathbb{R}^{\dim(W^\alpha)}$, θ-Hölder function $f : X \to \mathbb{R}$, $h \in G$, and $g \in G$, we have*

$$\frac{1}{|B|}\int_B f(\alpha^n(h\psi(b))g\Lambda)\,db = \int_X f\,d\mu + O(\min(B)^{-\kappa}\rho^n\|f\|_{C^\theta}).$$

Proof: We give a proof using an inductive argument similar to the proof of exponential mixing in Section 3.

Let $W = W^\alpha \cap \langle w, \bar{w}\rangle$ where w is the eigenvector of $D\alpha$ in W^α_ℓ. More explicitly, $W = \langle w_s\rangle$ or $W = \langle w_s, w_s'\rangle$ with notation (5.1)–(5.2). As in Section 3.2, we deduce that there exists a closed normal subgroup M of G containing $\exp(W)$ such that $M/(M\cap\Lambda)$ is compact and for almost all $g \in G$,

$$\overline{\exp(W)g\Lambda} = Mg\Lambda. \tag{5.5}$$

The map $\psi : \mathbb{R}^{\dim(W^\alpha)} \to \exp(W^\alpha)$ can be written as a product $\psi = \xi\cdot\eta$ with $\xi : \mathbb{R}^{\dim(W^\alpha)-\dim(W)} \to \exp(W^\alpha)$ and $\eta : \mathbb{R}^{\dim(W)} \to \exp(W)$, where $\eta : t \mapsto \exp(tw_s)$ or $\eta : (t,t') \mapsto \exp(tw_s + t'w_s')$ and ξ is the product of the remaining exponential maps appearing in ψ. Then

$$\int_B f(\alpha^n(h\psi(b))g\Lambda)\,db = \int_C\int_D f(\alpha^n(h\xi(u)\eta(v))g\Lambda)\,dudv,$$

where C is a box in $\mathbb{R}^{\dim(W^\alpha)-\dim(W)}$ and D is a box in $\mathbb{R}^{\dim(W)}$ such that $B = C\times D$.

We first show that images of the map η are equidistributed in a suitable sense. Namely, we claim that there exists $\rho\in(0,1)$ such that for every $h \in G$ and every $g\Lambda \in X$ such that (5.5) holds,

$$\frac{1}{|D|}\int_D f(\alpha^n(h\eta(t))g\Lambda)\,dt = \int_Z f(\alpha^n(h)gm\Lambda)\,\mu_Z(m) + O(\rho^n\|f\|_{C^\theta}), \tag{5.6}$$

where μ_Z denotes the invariant normalised measure on the nilmanifold $Z = M/(M\cap\Lambda)$. Let $F_0 \subset G$ be a bounded subset such that $G = F_0\Lambda$. Then there exists a bounded subset F of G such that $G = FM(g_0\Lambda g_0^{-1})$ for all $g_0 \in F_0$. Indeed, we can take $F = F_0F_0^{-1}$. We note that in (5.6) we may assume that $g \in F_0$, and to simplify notation, we replace Λ by $g\Lambda g^{-1}$. Then (5.6) holds with $g = e$. We note that our estimates below (in particular, (5.8)) are uniform on g. Indeed, we use the equidistribution of (c_1,c_2)-Diophantine box maps in the proof, and the constants c_1 and c_2 will not depend on g because they are determined by the projections to the torus $G/(G'g\Lambda g^{-1}) = G/(G'\Lambda)$.

Next we write $\alpha^n(h) = am\lambda$ with $a \in F$, $m \in M$, and $\lambda\in\Lambda$. Then

$$\int_D f(\alpha^n(h\eta(t))\Lambda)\,dt = \int_D f(am\beta(\alpha^n(\eta(t)))\Lambda)\,dt, \tag{5.7}$$

where β denotes the automorphism of M defined by $m \mapsto \lambda m\lambda^{-1}$. We note that β acts trivially on M/M' by Lemma 3.5. To analyse (5.7), we apply Corollary 2.3 to the nilmanifold $Z = M/(M\cap\Lambda)$. Setting $\phi(z) := f(az)$, $z\in Z$, we get

$$\int_D f(am\beta(\alpha^n(\eta(t)))\Lambda)\,dt = \int_D \phi(m\beta(\alpha^n(\eta(t)))\Lambda)\,dt,$$

and since $a \in F$, we have
$$\|\phi\|_{C^\theta} \ll \|f\|_{C^\theta}.$$
For the next computation, let us assume that $\dim(W) = 2$. When $\dim(W) = 1$, the proof is similar and simpler. We observe that $D\alpha|_W = r\omega$ where $r > 1$ and ω is a rotation of W, so that
$$\beta(\alpha^n(\eta(t,t'))) = \exp(r^n(t(D\beta\omega^n w_s) + t'(D\beta\omega^n w'_s))).$$
Making a change of variables,
$$\int_D \phi(m\beta(\alpha^n(\eta(t)))\Lambda)\, dt = r^{-2n} \int_{r^n D} \phi(m \exp(\iota_n(t))\Lambda)\, dt,$$
where ι_n denotes the box map $(t,t') \mapsto t(D\beta\omega^n w_s)+t'(D\beta\omega^n w'_s)$. We note that $(D\pi)(D\beta) = D\pi$. Since $\overline{\exp(W)\Lambda} = M\Lambda$, it follows that $D\pi(W)$ is not contained in any proper rational subspace. In particular, it follows from Lemma 3.3 and [2, Th. 7.3.2] that $D\pi(W)$ contains a vector w satisfying the Diophantine condition (2.14). Since ω is an isometry, this implies that the box map ι_n is (c_1, c_2)-Diophantine where c_1, c_2 are uniform in n and β (see Remark 2.2). Therefore, Corollary 2.3 implies that there exists $\kappa > 0$ such that

$$\frac{1}{|r^n D|} \int_{r^n D} \phi(m\beta(\exp(\iota_n(t)))\Lambda)\, dt = \int_Z \phi\, d\mu_Z + O(\min(r^n D)^{-\kappa} \|\phi\|_{C^\theta}) \tag{5.8}$$

This shows that
$$\begin{aligned}\frac{1}{|D|}\int_D f(am\beta(\alpha^n(\eta(t)))\Lambda)\, dt &= \int_Z f(az)\, d\mu_Z(z) + O(\min(D)^{-\kappa} r^{-\kappa n}\|f\|_{C^\theta})\\ &= \int_Z f(\alpha^n(h)z)\, d\mu_Z(z) + O(\min(B)^{-\kappa}\rho^n\|f\|_{C^\theta})\end{aligned}$$
with $\rho = r^\kappa \in (0,1)$. This proves (5.6).

Next, we apply the above argument inductively. For a Hölder function f on $X = G/\Lambda$, we define a function $\bar{f}$ on $\bar{X} = G/M\Lambda$ by
$$\bar{f}(gM\Lambda) := \int_{M/(M\cap\Lambda)} f(gm\Lambda) d\mu_Z(m).$$
Clearly,
$$\|\bar{f}\|_{C^\theta} \le \|f\|_{C^\theta}.$$
Let $\bar{G} = G/M$, $\bar{\Lambda} = (M\Lambda)/M$, and $p : G \to \bar{G}$ be the projection map. Then $\bar{X} \simeq \bar{G}/\bar{\Lambda}$. We note that $Dp(W^\alpha)$ is precisely the unstable space of $D\alpha$ acting on $\mathcal{L}(\bar{G})$. It follows from (5.6) that there exists $\rho \in (0,1)$ such that
$$\frac{1}{|B|}\int_B f(\alpha^n(h\psi(b))g\Lambda)\, db = \frac{1}{|B|}\int_B \bar{f}(\alpha^n(\bar{h}\bar{\psi}(b))\bar{g}\bar{\Lambda})\, db + O(\min(B)^{-\kappa}\rho^n\|f\|_{C^\theta}),$$
where $\bar{\psi}$ is the product of the maps of the form
$$\bar{\psi}_i : (t_1, \dots, t_s) \mapsto \exp(t_1\bar{w}_1)\cdots\exp(t_s\bar{w}_s),$$
or
$$\bar{\psi}_i : (t_1, t'_1 \dots, t_s, t'_s) \mapsto \exp(t_1\bar{w}_1 + t'_1\bar{w}'_1)\cdots\exp(t_s\bar{w}_s + t'_s\bar{w}'_s).$$
with $\bar{w}_i = Dp(w_i)$ and $\bar{w}'_i = Dp(w'_i)$, $\bar{h} = p(h)$ and $\bar{g} = p(g)$. In this product we may skip terms with $\bar{w}_i = 0$ or $\bar{w}'_i = 0$ (note that if $\bar{w}_i = 0$, then $\bar{w}'_i = 0$ and conversely). Then the

relations (5.1)–(5.2) are still satisfied. In particular, the last exponential in the obtained product corresponds to the subspace $Dp(W^\alpha)\cap\langle w,\bar{w}\rangle$ where w is an eigenvector of $D\alpha$ in $\mathcal{L}(\bar{G})$ with the eigenvalue of maximal modulus. Now we can again apply the argument as in the proof of (5.6) reducing the number of terms in the product defining $\bar{\psi}$. Repeating the same argument repeatedly, we deduce that for some $\rho\in(0,1)$ and $\kappa>0$,

$$\frac{1}{|B|}\int_B f(\alpha^n(h\psi(b))g\Lambda)\,db=\int_{M/(M\cap\Lambda)} f(\alpha^n(h)gm\Lambda)\,d\mu_Z(m)+O(\min(B)^{-\kappa}\rho^n\|f\|_{C^\theta}),$$

where M is a closed normal α-invariant subgroup containing $\exp(W^\alpha)$ such that $M/(M\cap\Lambda)$ is compact. We observe that $D\alpha$ acting on $\mathcal{L}(G/M)$ has no eigenvalues with absolute value greater than one. Since α is ergodic, it follows from Lemma 3.2 that $M=G$. This proves the theorem for the set of $g\in G$ that satisfy (5.5) at every inductive step, with the estimate which is uniform over g. Since this set has full measure, we conclude that the estimate holds for all g completing the proof of the theorem. ⋄

The following corollary will be used in the proof of the limit theorems in the next section.

Corollary 5.2. *Let Ω be a domain in W^α with a piecewise smooth boundary. Then there exist $\kappa=\kappa(\theta)>0$ and $\rho=\rho(\theta)\in(0,1)$ such that for every θ-Hölder function $f:X\to\mathbb{R}$, $g\in G$ and $\epsilon>0$, we have*

$$\int_\Omega f(\alpha^n(\psi(b))g\Lambda)\,db=|\Omega|\int_X f\,d\mu+O\left((|\partial_\epsilon\Omega|+\epsilon^{-\kappa}\rho^n|\Omega|)\|f\|_{C^\theta}\right),$$

where $\partial_\epsilon\Omega$ denotes the ϵ-neighbourhood of the boundary of Ω.

Proof: We tessellate W^α by cubes B of size ϵ. Then

$$\left|\Omega-\bigcup_{B\subset\Omega}B\right|\le|\partial_\epsilon\Omega|,$$

and

$$\int_\Omega f(\alpha^n(\psi(b))g\Lambda)\,db=\sum_{B\subset\Omega}\int_B f(\alpha^n(\psi(b))g\Lambda)\,db+O(|\partial_\epsilon\Omega|\|f\|_{C^0}).$$

By Theorem 5.1, for some $\kappa>0$ and $\rho\in(0,1)$,

$$\int_B f(\alpha^n(\psi(b))g\Lambda)\,db=|B|\int_X f\,d\mu+O(|B|\epsilon^{-\kappa}\rho^n\|f\|_{C^\theta}).$$

Therefore,

$$\int_\Omega f(\alpha^n(\psi(b))g\Lambda)\,db=\left(\sum_{B\subset\Omega}|B|\right)\int_X f\,d\mu+O\left(\left(|\partial_\epsilon\Omega|+\sum_{B\subset\Omega}|B|\epsilon^{-\kappa}\rho^n\right)\|f\|_{C^\theta}\right)$$

$$=|\Omega|\int_X f\,d\mu+O((|\partial_\epsilon\Omega|+|\Omega|\epsilon^{-\kappa}\rho^n)\|f\|_{C^\theta}).$$

This completes the proof of corollary. ⋄

6. Central limit theorem and invariance principles

Let us first review the terminology regarding the central limit theorem and other probabilistic limit theorems. Let $\alpha : X \to X$ be a measure-preserving map of a probability space (X, μ). For a function $f : X \to \mathbb{R}$, we consider a sequence of observables $f \circ \alpha^n$. If the dynamical system $\alpha \curvearrowright X$ is sufficiently chaotic, this sequence is expected to behave similarly to a sequence of independent random variables. We set

$$S_n(f, x) = \sum_{i=0}^{n-1} f(\alpha^i(x)),$$

and for simplicity assume that $\int_X f\, d\mu = 0$.

The sequence $f \circ \alpha^n$ satisfies the *central limit theorem* if there exists $\sigma > 0$ such that $n^{-1/2} S_n(f, \cdot)$ converges in distribution to the normal law with mean 0 and variance σ^2. More generally, the sequence $f \circ \alpha^n$ satisfies the *central limit theorem for subsequences* if there exists $\sigma > 0$ such that for every increasing sequence of measurable functions $k_n(x)$ taking values in $\mathbb{N}$ such that for almost all x, $\lim_{n\to\infty} \frac{k_n(x)}{n} = c$ for some fixed constant $0 < c < \infty$, the sequence $n^{-1/2} S_{k_n(\cdot)}(f, \cdot)$ converges in distribution to the normal law with mean 0 and variance σ^2/c. We define $S_t(f, x)$ for all $t \geq 0$ by linear interpolation of its values at integral points. The sequence $f \circ \alpha^n$ satisfies the *Donsker invariance principle* if there exists $\sigma > 0$ such that the sequence of random functions $(n\sigma^2)^{-1/2} S_{nt}(f, \cdot) \in C([0, 1])$ converges in distribution to the standard Brownian motion in $C([0, 1])$. The sequence $f \circ \alpha^n$ satisfies the *Strassen invariance principle* if there exists $\sigma > 0$ such that for almost every x, the sequence of functions $(2n\sigma^2 \log\log n)^{-1/2} S_{nt}(f, x)$ is relatively compact in $C([0, 1])$ and its limit set is precisely the set of absolutely continuous functions g on $[0, 1]$ such that $g(0) = 0$ and $\int_0^1 g'(t)^2\, dt \leq 1$. This is a strong version of the law of the iterated logarithm.

In this section we establish the above limit theorems for sequences generated by ergodic automorphisms of compact nilmanifolds. In the case of toral automorphism, these theorems have been established by LeBorgne [17] using the method of martingale differences, and we follow a similar approach. We shall use the following general result:

Theorem 6.1. *Let $(X, \mathcal{B}, \mu, \alpha)$ be an invertible ergodic dynamical system and $f \in L^2(X)$ such that $\int_X f\, d\mu = 0$. Let $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{B}$ such that $\mathcal{A}_n = \alpha^{-n}(\mathcal{A})$ is a non-increasing sequence of σ-algebras satisfying*

$$\sum_{n>0} \|E(f \mid \mathcal{A}_n)\|_2 < \infty \quad \text{and} \quad \sum_{n<0} \|f - E(f \mid \mathcal{A}_n)\|_2 < \infty. \tag{6.1}$$

Then

(i) $\sigma^2 = \int_X f^2\, d\mu + 2\sum_{j=1}^{\infty} \int_X (f \circ \alpha^j) f\, d\mu$ *is finite.*
(ii) $\sigma = 0 \Leftrightarrow f$ *is an L^2 coboundary* $\Leftrightarrow f$ *is a measurable coboundary.*
(iii) *If $\sigma > 0$, then $f \circ \alpha^n$ satisfies the central limit theorem, the central limit theorem of subsequences, and the Donsker and Strassen invariance principles.*

It is well-known (see, for instance, [34, Theorem 4.13]) that under the assumption (6.1) the function f has a decomposition $f = (\phi \circ \alpha - \phi) + \psi$ with $\phi, \psi \in L^2(X)$, where $\psi \circ \alpha^n$ is a reverse martingale difference with respect to the σ-algebras $\mathcal{A}_n$, and $\sigma = \|\psi\|_2$. In particular, $\sigma < \infty$ and if $\sigma = 0$, then f is an L^2 coboundary. On the other hand, if f is a measurable coboundary, then ψ is also a measurable coboundary, and it follows from [31] that $\psi = 0$, so that $\sigma = 0$. For (iii) we refer to [13, Ch. 5].

The following is the main result of this section:

Theorem 6.2. *Let α be an ergodic automorphism of a compact nilmanifold X, and let f be a Hölder function on X which has zero integral and is not a measurable coboundary. Then the sequence $f \circ \alpha^n$ satisfies the central limit theorem, the central limit theorem of subsequences, and the Donsker and Strassen invariance principles.*

To find the sub-σ-algebra $\mathcal{A}$ suitable for Theorem 6.1, we use the results of Section 5 combined with the works of Lind [21] and Le Borgne [17]. We call a measurable partition $\mathcal{P}$ of X *δ-fine* if the diameter of any set in $\mathcal{P}$ is at most δ. We say that a partition generates under α if the σ-algebra generated by all $\alpha^n(\mathcal{P})$ with $n \in \mathbb{Z}$ is the Borel σ-algebra of X modulo null sets. Given a partition $\mathcal{P}$ and $x \in X$, we denote by $\mathcal{P}(x)$ the element of the partition that contains x. Given integers $k \le l$, we denote by $\mathcal{P}_k^l$ the partition generated by $\alpha^{-k}(\mathcal{P}), \dots, \alpha^{-l}(\mathcal{P})$. We also set $\mathcal{P}_k^\infty(x) = \cap_{l \ge k} \mathcal{P}_k^l(x)$.

Proposition 6.3. *Let $\mathcal{P}$ be a finite measurable partition of X such that for every $P \in \mathcal{P}$,*

- *P is the closure of its interior,*
- *the boundary of P is piecewise smooth,*
- *the diameter of P is at most δ.*

Then if δ is sufficiently small,

(i) *the partition $\mathcal{P}$ generates under α,*
(ii) *for almost every x, the atoms $\mathcal{P}_0^\infty(x)$ are contained in the stable manifolds $\mathcal{W}^s(x)$ of x, and the diameter of $\mathcal{P}_0^\infty(x)$ in $\mathcal{W}^s(x)$ is bounded,*
(iii) *for almost every $x \in X$, the atoms $\mathcal{P}_0^\infty(x)$ have non-empty interior in the stable manifolds $\mathcal{W}^s(x)$.*

Proof of (i)–(ii). The proof follows that of [21, Th. 1] almost completely albeit with some differences in the final argument involving isometries. We will show that α almost surely separates points, i.e., that for some null set X_0 in X, if $x, y \in X \backslash X_0$, then for some n, the points $\alpha^n(x)$ and $\alpha^n(y)$ belong to different elements of the partition $\mathcal{P}$. It then follows from Rohklin's work [28] that $\mathcal{P}$ generates under α.

There exist $c_0 > 1$ and $\delta_0 > 0$ such that for every $w \in \mathcal{L}(G)$ satisfying $\|w\| < \delta_0$ and $x \in X$,

$$c_0^{-1} \|w\| \le d(x, \exp(w)x) \le c_0 \|w\|. \tag{6.2}$$

We assume that δ is sufficiently small, so that $\|D\alpha\| c_0 \delta < \delta_0$, and if p and q belong to the same element P of the partition, then $q = \exp(w)p$ with $\|w\| < \delta_0$. Since $\operatorname{diam}(P) \le \delta$, we have $\|w\| \le c_0\delta$. We observe that

$$d(\alpha^n(p), \alpha^n(q)) = d(\alpha^n(p), \exp((D\alpha)^n w)\alpha^n(p)). \tag{6.3}$$

Suppose that $\|(D\alpha)^n w\| \to \infty$ as $n \to \infty$. We pick the greatest $n \ge 0$ such that $\|(D\alpha)^n w\| \le c_0 \delta$. Then

$$c_0\delta < \|(D\alpha)^{n+1} w\| \le \|D\alpha\| c_0 \delta < \delta_0,$$

and it follows from (6.2)–(6.3) that $d(\alpha^{n+1}(p), \alpha^{n+1}(q)) > \delta$. Hence, $\alpha^{n+1}(p)$ and $\alpha^{n+1}(q)$ belong to different elements of the partition.

A similar argument also applies when $\|(D\alpha)^n w\| \to \infty$ as $n \to -\infty$. Therefore, it remains to consider the case when $w \in E^{iso}$ which is the span of eigenspaces of $D\alpha$ with eigenvalues of modulus one. We adapt Lind's idea [21] for this situation. Let K denote the closed group of isometries generated by $\beta := D\alpha|_{E^{iso}}$. Then β acts ergodically on K by translations. Since α is mixing, the product $\alpha \times \beta$ acts ergodically on $X \times K$. It follows from ergodicity and Fubini's theorem that there exists a null set $X_0 \subset X$ and $k \in K$ such that the sequence $(\alpha^n(x), \beta^n k)$ is dense in $X \times K$ for every $x \in X \backslash X_0$. Then the sequence $(\alpha^n(x), \beta^n)$ is also dense in $X \times K$.

Now suppose that $p, q \in X \backslash X_0$ and $q = \exp(w)p$ for some nonzero $w \in E^{iso}$. Given an element $P \in \mathcal{P}$, we set

$$P(w, \epsilon) = \{x \in P : d(\exp(w)x, P) > \epsilon\}.$$

When $\epsilon > 0$ is sufficiently small, this set has a nonempty interior. Hence, for every $p \in X \backslash X_0$, there exists n such that

$$\alpha^n(p) \in P(w, \epsilon) \quad \text{and} \quad d(\exp(w), \exp((D\alpha)^n w)) < \epsilon/2.$$

Then

$$\begin{aligned} d(\alpha^n(q), P) &= d(\exp((D\alpha)^n w)x, P) \\ &\geq d(\exp(w)x, P) - d(\exp((D\alpha)^n w)x, \exp(w)x) > \epsilon/2. \end{aligned}$$

In particular, $\alpha^n(p) \in P$ and $\alpha^n(q) \notin P$. This proves that $\mathcal{P}$ generates under α. The part (ii) can be proved by the same argument. $\diamond$

To prove Proposition 6.3(iii), we follow Le Borgne's approach [17] for toral automorphisms. We pick $c, r_0 \in (0, 1)$ such that the map α^{-n} expands the distance on $\mathcal{W}^s$ by at least $c\, r_0^{-n}$ for $n \geq 0$, and take $r \in (r_0, 1)$. Let

$$V_n := \{x \in X : \mathcal{P}_0^\infty(x) \supset B_{r^n/c}(x) \cap \mathcal{W}^s(x)\}.$$

Proposition 6.3(iii) immediately follows from the following lemma.

Lemma 6.4. *$\mu(X \backslash V_n) \ll r^n$.*

Proof: Let

$$W_n := \{y \in X : d(\alpha^j(y), \partial\mathcal{P}(\alpha^j(y))) \geq r_0^j r^n / c^2 \text{ for all } j \geq 0\}.$$

If y is in W_n, then $\mathcal{P}(\alpha^j(y))$ contains the ball in $\mathcal{W}^s(\alpha^j(y))$ of radius $r_0^j r^n/c^2$. Hence, $\alpha^{-j}(\mathcal{P}(\alpha^j(y)))$ contains the ball in $\mathcal{W}^s(y)$ of radius r^n/c. Since

$$\mathcal{P}_0^\infty(y) = \bigcap_{j \geq 0} \alpha^{-j}(\mathcal{P}(\alpha^j(y))),$$

we conclude that $V_n \supset W_n$.

To prove the lemma, it suffices to estimate $\mu(X \backslash W_n)$. It follows from our assumption on the partition $\mathcal{P}$ that

$$\mu(\{y \in X : d(y, \partial\mathcal{P}(y)) \leq \epsilon\}) \ll \epsilon,$$

and since α is measure-preserving, for every $j \geq 0$,

$$\mu(\{y \in X : d(\alpha^j(y), \partial\mathcal{P}(\alpha^j(y))) \leq r_0^j r^n / c\}) \ll r_0^j r^n.$$

Hence,

$$\mu(X \backslash W_n) \ll \sum_{j \geq 0} r_0^j r^n \ll r^n,$$

which implies the lemma. $\diamond$

We also mention an alternative way to construct a suitable sequence of σ-algebras, which was used, for instance, in [3, 18]. We define a new partition

$$\tilde{\mathcal{P}}(x) = \mathcal{P}(x) \cap \mathcal{W}^s_\delta(x),$$

where $\mathcal{W}^s_\delta(x)$ is the δ-neighbourhood of x in the stable manifold, and set

$$\tilde{\mathcal{P}}^\infty_n(x) = \bigcap_{j \geq n} \alpha^{-j}(\tilde{\mathcal{P}}(\alpha^j(x))).$$

Then the property (ii) is automatically satisfied, and one just needs to check (iii). However, it seems that the result regarding generating partitions, generalising [21] to nilmanifolds, might be useful for other applications.

Proof of Theorem 6.2. Let $\mathcal{A}$ be the σ-algebra generated by the partition $\mathcal{P}^\infty_0$ and $\mathcal{A}_n = \alpha^{-n}(\mathcal{A}) = \mathcal{P}^\infty_n$. It is clear that the sequence $\mathcal{A}_n$ is non-increasing. To prove the theorem, it suffices to check the conditions (6.1). Since the partition $\mathcal{P}^\infty_n$ is measurable in the sense of [28], for almost every x,

$$E(f|\mathcal{A}_n)(x) = \int_{\mathcal{P}^\infty_n(x)} f(y)\, dm_{\mathcal{P}^\infty_n(x)}(y),$$

where $m_{\mathcal{P}^\infty_n(x)}$ is the conditional probability measure on $\mathcal{P}^\infty_n(x)$.

To verify the second part of (6.1), we observe that when $\mathcal{P}^\infty_0(\alpha^n(x)) \subset \mathcal{W}^s(x)$,

$$\mathrm{diam}(\mathcal{P}^\infty_n(x)) = \mathrm{diam}(\alpha^{-n}(\mathcal{P}^\infty_0(\alpha^n(x))))$$

decays exponentially as $n \to -\infty$ uniformly on x. Since the function f is θ-Hölder, it follows that for some $\tau \in (0,1)$,

$$\|f - E(f \mid \mathcal{A}_n)\|_2 \ll \tau^{-n} \|f\|_{C^\theta} \quad \text{and} \quad \sum_{n<0} \|f - E(f \mid \mathcal{A}_n)\|_2 < \infty.$$

To check the other condition in (6.1), we observe that by Lemma 6.4,

$$\int_{X \backslash \alpha^{-n}(V_n)} |E(f|\mathcal{A}_n)|^2 \, d\mu \ll r^n \|f\|^2_{C^0}. \tag{6.4}$$

On the other hand, for $x \in \alpha^{-n}(V_n)$,

$$B_{r^n/c}(\alpha^n(x)) \cap \mathcal{W}^s(\alpha^n(x)) \subset \mathcal{P}(\alpha^n(x)) \quad \text{and} \quad B_{r^n r_0^{-n}}(x) \cap \mathcal{W}^s(x) \subset \alpha^{-n}(\mathcal{P}(\alpha^n(x))).$$

Since the diameter of $\mathcal{P}(x)$ is at most δ, as soon as $r^n r_0^{-n} > \delta$, we get that $\mathcal{P}(x) \subset B_{r^n/r_0^n}(x)$. Hence, by Proposition 6.3, for almost every $x \in \alpha^{-n}(V_n)$,

$$\mathcal{P}_0^\infty(x) = \bigcap_{n=0}^{\infty} \alpha^{-n}(\mathcal{P}(\alpha^n(x))) \cap \mathcal{W}^s(x) = \bigcap_{n=0}^{\left\lceil \frac{\log \delta}{\log(r/r_0)} \right\rceil} \alpha^{-n}(\mathcal{P}(\alpha^n(x))) \cap \mathcal{W}^s(x). \tag{6.5}$$

Thus, $\mathcal{P}_0^\infty(x)$ is the intersection of the stable manifold of x with at most finitely many sets whose boundaries consist of finitely many piecewise smooth submanifolds. Then the right hand side of (6.5) equals to $exp(\Omega_x)x$, and hence

$$\mathcal{P}_0^\infty(x) = \exp(\Omega_x)x, \tag{6.6}$$

where Ω_x is a domain in the unstable subspace $W = W^{\alpha^{-1}}$ of $D(\alpha^{-1})$ in $\mathcal{L}(G)$ whose boundary is piecewise smooth and depends smoothly on x. In particular, $|\partial_\epsilon \Omega_x| \ll \epsilon$ uniformly on $x \in X$. It follows from (6.6) that

$$\mathcal{P}_n^\infty(x) = \alpha^{-n}(\mathcal{P}_0^\infty(\alpha^n(x))) = \exp((D\alpha)^{-n}\Omega_x)x.$$

Then by [3, Prop. 4.3],

$$m_{\mathcal{P}_n^\infty(x)} = \frac{1}{m_x(\mathcal{P}_n^\infty(x))} m_x|_{\mathcal{P}_n^\infty(x)},$$

where m_x is the Haar measure on $\exp(W)x$. Now we apply Corollary 5.2. It follows from the definition of V_n that for $x \in \alpha^{-n}(V_n)$, we have $|\Omega_x| \gg r^n$. Hence, by Corollary 5.2, for every $x \in \alpha^{-n}(V_n)$ and $\epsilon > 0$,

$$\begin{aligned}\frac{1}{m_x(\mathcal{P}_n^\infty(x))} \int_{\mathcal{P}_n^\infty(x)} f(y)\, dm_x(y) &= O\left(\left(\frac{|\partial_\epsilon \Omega|}{|\Omega|} + \epsilon^{-\kappa}\rho^n\right) \|f\|_{C^\theta}\right) \\ &= O\left((\epsilon r^{-n} + \epsilon^{-\kappa}\rho^n)\|f\|_{C^\theta}\right),\end{aligned}$$

where $\rho \in (0, 1)$. We take $\epsilon = (r^n\rho^n)^{1/(\kappa+1)}$. If we also take r sufficiently close to 1, then this quantity decays exponentially as $n \to \infty$. Then

$$\int_{\alpha^{-n}(V_n)} |E(f|\mathcal{A}_n)|^2\, d\mu \ll \tau^n \|f\|_{C^\theta}^2$$

for some $\tau \in (0, 1)$. Combining this estimate with (6.4), we deduce the first part of (6.1). Now the theorem follows from Theorem 6.1. $\diamond$

7. Cohomological equation

In this section we apply exponential mixing to establish regularity of solutions of the cohomological equation. We recall that for ergodic systems the solution is unique up to a constant, up to measure zero.

Theorem 7.1. *Let α be an ergodic automorphism of a compact nilmanifold X and $f \in C^\infty(X)$ such that $f = \phi \circ \alpha - \phi$ for some measurable function ϕ. Then ϕ is almost everywhere equal to a C^∞ function.*

The method of proof of Theorem 7.1 applies to other classes of homogeneous partially hyperbolic systems for which exponential mixing holds. For instance, we may consider an ergodic partially hyperbolic left translation on the homogeneous space G/Γ, where G is connected semisimple Lie group and Γ is a cocompact irreducible lattice. This dynamical system is also exponentially mixing for Hölder functions [15, Appendix], and the argument of Theorem 7.1 applies. For $X = \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$, an analogous result for Hölder functions f was established in [18]. Furthermore, we get both Hölder and smooth versions of Theorem 7.1 for compact G/Γ and G semi simple from Wilkinson's general result for accessible partially hyperbolic diffeomorphisms [35, Theorem A] under the additional assumption that the left translation projected to any factor of G does not belong to a compact subgroup.

Before starting the proof, we need to develop some language and review a result on regularity of distributions. Let M be a compact manifold. We fix a Riemannian metric on M, and denote by $C^\theta = C^\theta(M)$ the space of θ-Hölder functions on M. We let $(C^\theta)^*$ be the dual space to C^θ. Note that any smooth function on M naturally belongs to any C^θ. Hence any element in $(C^\theta)^*$ defines a distribution on smooth functions on M. Conversely, $(C^\theta)^*$ is the space of distributions (dual to C^∞ functions) which extend to continuous linear functionals on C^θ. As for notation, we will write the pairing $D(g) = \langle D, g\rangle$ for $D \in (C^\theta)^*$ and $g \in C^\theta$.

Let $\mathcal{F}$ be a C^∞ foliation on M, and consider a C^∞ vector field V tangent to $\mathcal{F}$. Given a distribution D on M, define the derivative $V(D)$ by evaluating on C^∞ test functions g as follows: $\langle V(D), g\rangle = -\langle D, V(g)\rangle$ where $V(g)$ denotes the directional derivative of g along V.

Given smooth vector fields $V_1, \dots, V_r$, we call $V_{i_1}, V_{i_2} \dots V_{i_m} D$ the partial derivatives of order m of D. Suppose that we can cover M with open sets $\mathcal{U}$ such that we can find smooth vector fields $V_1, \dots, V_r$ which span the tangent spaces to $\mathcal{F}$ at any point of $\mathcal{U}$. Suppose moreover that all partial derivatives of any order m, $V_{i_1}, V_{i_2} \dots V_{i_m} D$ of a distribution D belong to $(C^\theta)^*$, for all such choices of $\mathcal{U}$ and $V_1, \dots V_r$. Then for any other C^∞ vector fields $V_1', \dots V_r'$ tangent to $\mathcal{F}$, the partial derivatives $V_{i_1}', V_{i_2}' \dots V_{i_m}' D$ also belong to $(C^\theta)^*$ as follows from a partition of unity argument. Thus we can say that partials along $\mathcal{F}$ of a distribution belong to $(C^\theta)^*$, without any reference to a particular set of vector fields.[1]

The following result is inspired by results of Rauch and Taylor in [27], and was known to Rauch for the case of C^∞ foliations. We are not aware of a simple reference. It is also a straight-forward consequence of a similar much more technical result for Hölder foliations proved in [10], namely that the wavefront set of a distribution for which the partial derivatives of all orders along a single foliation belong to the dual of Hölder functions is co-normal to the foliation. We refer to [27, 10] for more details.

Corollary 7.2 ([10])**.** *Let $\mathcal{F}_1, \dots, \mathcal{F}_r$ be C^∞ foliations on a compact manifold M whose tangent spaces span the tangent spaces to M at all points. Consider a distribution D defined by integration against an L^1 function ϕ. Assume that any partial derivative of D of any order along the foliations $\mathcal{F}_1, \dots, \mathcal{F}_r$ belongs to $(C^\theta)^*$ for all $\theta > 0$. Then ϕ is C^∞.*

[1]In our application we will have globally defined vector fields for which the partials exist for all orders, and we will not need this comment.

We are now ready to tackle the proof of Theorem 7.1. Let us first give an outline of the argument. Using Theorem 6.1, we first show in Lemma 7.3 that the function ϕ has to be in $L^2(X)$. Then we describe ϕ as distribution. We consider three dynamically defined foliations for α: the unstable foliation $\mathcal{W}^u$, the stable foliation $\mathcal{W}^c$, and the central foliation $\mathcal{W}^c$. The unstable foliation is tangent to the right invariant distribution on X corresponding to the sum of all generalized eigenspaces with eigenvalues $|\lambda| > 1$, the stable foliation is tangent to the right invariant distribution on X corresponding to the sum of all generalized eigenspaces with eigenvalues $|\lambda| < 1$, and the central foliation is tangent to the right invariant distribution on X corresponding to the sum of all generalized eigenspaces with eigenvalues $|\lambda| = 1$. Note that these distributions are integrable as is easily seen by taking Lie brackets. We show that the distribution derivatives of ϕ along the foliations $\mathcal{W}^s$, $\mathcal{W}^u$, $\mathcal{W}^c$ of α define distributions on Hölder functions. This is established in Lemmas 7.4 and 7.5. Since all these foliations are smooth, Corollary 7.2 shows that the function ϕ is C^∞.

We now establish Lemmas 7.3, 7.4 and 7.5 which will finish the proof of Theorem 7.1.

Lemma 7.3. *The function ϕ in Theorem 7.1 is in L^2.*

Proof: Recall that along the proof of Theorem 6.2 we have verified the conditions of Theorem 6.1. Hence, the lemma follows from part (ii) of this theorem. $\diamond$

Define the distributions P^+ and P^- by evaluating them on test functions $g \in C^\infty(X)$ by

$$P^+(g) = \sum_{i=0}^{\infty} \langle f \circ \alpha^i, g\rangle \qquad \text{and} \qquad P^-(g) = \sum_{i=1}^{\infty} \langle f \circ \alpha^{-i}, g\rangle.$$

Note that $\int_X f\, d\mu = 0$ since f is an L^2 coboundary. Hence, by exponential mixing (Theorem 1.1), these sums converge as long as the test function g is Hölder, and $P^+, P^- \in (C^\theta)^*$. Moreover, since $\langle \phi \circ \alpha^i, g\rangle \to 0$ as $i \to \pm\infty$. we get by a telescoping-sum argument that

$$P^+(g) = \sum_{i=0}^{\infty} \langle f \circ \alpha^i, g\rangle = \sum_{i=0}^{\infty} \langle \phi \circ \alpha^{i+1} - \phi \circ \alpha^i, g\rangle = \lim_{N\to\infty} \langle \phi \circ \alpha^N - \phi, g\rangle = -\langle \phi, g\rangle.$$

Similarly, we see that $P^-(g) = \langle \phi, g\rangle$. Hence, the distribution $P^+ = -P^-$ is given by integration against the L^2-function ϕ. We will use this to show that ϕ is smooth.

According to Corollary 7.2, it suffices to show that partial derivatives of all orders of the distribution $P^+ = -P^-$ along any of the three foliations $\mathcal{W}^s$, $\mathcal{W}^u$ and $\mathcal{W}^c$ belong to $(C^\theta)^*$ for any $\theta > 0$. We will show this in the next two lemmas.

Lemma 7.4. *Partial derivatives of all orders of the distribution $P^+ = -P^-$ along $\mathcal{W}^s$ and $\mathcal{W}^u$ belong to $(C^\theta)^*$ for any $\theta > 0$.*

Proof: Let V be a right invariant vector field tangent to $\mathcal{W}^s$ and g a C^∞ test function. Then

$$\langle V(P^+), g\rangle = -\langle P^+, V(g)\rangle = -\sum_{i=0}^{\infty} \langle f \circ \alpha^i, V(g)\rangle = \sum_{i=0}^{\infty} \langle V(f \circ \alpha^i), g\rangle.$$

The derivative $V(f\circ\alpha^i)$ decays exponentially fast since V is tangent to W^s. Hence,

$$|\langle V(P^+), g\rangle| \ll \|g\|_{C^0},$$

and in particular, $V(P^+)\in (C^\theta)^*$ for all $\theta>0$. Since $P^+=-P^-$, an analogous proof shows that $V(P^+)$ lies in the dual of Hölder functions for all vector fields V tangent to $\mathcal{W}^u$. A similar argument also applies to higher order derivatives along vector fields tangent to $\mathcal{W}^s$ or $\mathcal{W}^u$. We refer for the details to [10, Lemma 5.1]. $\diamond$

Finally, we show that partials of all orders of $P^+=-P^-$ along $\mathcal{W}^c$ are distributions on Hölder functions. This argument uses exponential decay very strongly, and was first discovered in [10]. For a detailed account we refer to [10, Lemma 5.1].

Lemma 7.5. *Partial derivatives of all orders of the distribution $P^+=-P^-$ along $\mathcal{W}^c$ belong to $(C^\theta)^*$ for any $\theta>0$.*

Proof: Let V be a right invariant vector field tangent to $\mathcal{W}^c$, and let g be a C^∞ function. Then the partial derivative of P^+ along V is given by

$$\langle V(P^+), g\rangle = \sum_{i=0}^{\infty}\langle V(f\circ\alpha^i), g\rangle = -\sum_{i=0}^{\infty}\langle f\circ\alpha^i, V(g)\rangle, \tag{7.1}$$

and we have estimates for all of these expressions in terms of the Hölder norm of $V(g)$, due the exponential mixing of α. We will show that this distribution extends to Hölder functions g by approximating g by smooth functions g_ε and carefully balancing the speed of the approximation with the loss of exponential decay due to the growth of the C^l-norm of g_ε. More precisely, we shall show that there exists $\xi=\xi(\theta)\in(0,1)$ such that for every $g\in C^\theta(X)$ and sufficiently large i,

$$|\langle V(f\circ\alpha^i), g\rangle| \ll \xi^i\cdot\|f\|_{C^1}\|g\|_{C^\theta}. \tag{7.2}$$

It would follow from (7.1) and (7.2) that $V(P^+)\in(C^\theta)^*$.

We recall from Lemma 2.4 that for $\varepsilon>0$, there is a C^∞ function g_ε such that

$$\|g_\varepsilon - g\|_{C^0}\le \varepsilon^\theta\|g\|_{C^\theta} \quad\text{and}\quad \|g_\varepsilon\|_{C^2}\ll \varepsilon^{-m-2}\|g\|_{C^0} \tag{7.3}$$

where $m=\dim(X)$. We first estimate $|\langle V(f\circ\alpha^i), g_\epsilon\rangle|$. By the exponential mixing (Theorem 1.1) and since V is bounded, we have for some $\rho\in(0,1)$,

$$|\langle f\circ\alpha^i, V(g_\varepsilon)\rangle| \ll \rho^i\|f\|_{C^1}\|V(g_\varepsilon)\|_{C^1} \ll \rho^i\|f\|_{C^1}\,\varepsilon^{-m-2}\|g\|_{C^0}. \tag{7.4}$$

On the other hand, we can estimate $|\langle f\circ\alpha^i, V(g-g_\varepsilon)\rangle| = |\langle V(f\circ\alpha^i), g-g_\varepsilon\rangle|$ as follows. First, we note that the derivatives $V(f\circ\alpha^i)$ by the chain rule are composites of derivatives of f and derivatives of α^i along $\mathcal{W}^c$. The latter grows at most polynomially since $\mathcal{W}^c$ is the central foliation for α. Hence, for any $\eta>0$, there is $i_\eta\in\mathbb{Z}_+$ such that

$$\|D(\alpha^i|_{\mathcal{W}^c})\| < (1+\eta)^i \quad \text{for all } i\ge i_\eta.$$

Hence, for all $i>i_\eta$, we get the estimate

$$\|V(f\circ\alpha^i)\|_{C^0}\le (1+\eta)^i\|f\|_{C^1},$$

and

$$|\langle V(f\circ\alpha^i), g-g_\varepsilon\rangle| \le \|V(f\circ\alpha^i)\|_{C^0}\,\|g-g_\varepsilon\|_{C^0} \le (1+\eta)^i\|f\|_{C^1}\,\varepsilon^\theta\|g\|_{C^\theta}. \tag{7.5}$$

We have exponential decay with respect to i in (7.4), but exponential growth in (7.5) at first sight. However, choosing ε carefully depending on i, we can still achieve exponential decay in (7.5), and hence for $|\langle V(f\circ\alpha^i), g\rangle|$. More precisely, we take $\varepsilon = \rho^{\frac{i}{\theta+m+2}}$. Then we obtain from (7.4) that

$$|\langle f\circ\alpha^i, V(g_\varepsilon)\rangle| \ll \left(\rho^{\frac{\theta}{\theta+m+2}}\right)^i\ \|f\|_{C^1}\,\|g\|_{C^0},$$

and from (7.5) that for $i > i_\eta$,

$$|\langle V(f\circ\alpha^i), g-g_\varepsilon\rangle| \le \left((1+\eta)\rho^{\frac{\theta}{\theta+m+2}}\right)^i\|f\|_{C^1}\|g\|_{C^\theta}.$$

Now we choose $\eta>0$ so that $\xi := (1+\eta)\rho^{\frac{\theta}{\theta+m+2}} < 1$. Finally, we obtain from the last two inequalities that for $i > i_\eta$,

$$|\langle V(f\circ\alpha^i), g\rangle| \ll \xi^i\cdot\|f\|_{C^1}\|g\|_{C^\theta}.$$

This proves (7.2) and shows that $V(P^+)$ extends to a continuous linear functional on the space of θ-Hölder functions. A similar argument shows that higher order derivatives of P^+ along the central foliation define distributions dual to Hölder functions. For the details we refer to [10, Lemma 5.1].

We also mention that one can give a different argument for the estimate (7.2) using the linearity of α along the foliation $\mathcal{W}^c$. First, we can assume that $\int_X g\,d\mu = 0$ because for constant g, the estimate (7.2) follows from integration by parts. Then we can write $V(f\circ\alpha^i) = V_i(f)\circ\alpha^i$ for another differential operator V_i. One can show that the Hölder norm of $V_i(f)$ grows polynomially. Hence, (7.2) can be deduced from the exponential mixing. The argument that we presented above is more versatile, and it applies to (nonlinear) diffeomorphisms satisfying the exponential mixing property. $\diamond$

This finishes the proof of Theorem 7.1.

8. Bernoulli property

Here we show that ergodic automorphisms on compact nilmanifolds are Bernoulli combining results from [14], [23], and [29]. It was already shown in [25] that such automorphisms satisfy the Kolmogorov property.

Theorem 8.1. *Ergodic automorphisms on compact nilmanifolds are Bernoulli.*

Proof: Let α be an ergodic automorphism of a compact nilmanifold $X = G/\Lambda$. We will argue by induction on the dimension of X. We note that when X is a torus, this result was established by Katznelson in [14], and this forms the base of induction. Let Z be the centre of G. It follows from [5, 5.2.3] that $Z\Lambda$ is a closed subgroup of G. Then $\alpha \curvearrowright X$ is measurably isomorphic to a skew product with the base $\alpha \curvearrowright Y = G/(Z\Lambda)$ and fibers isomorphic to the torus $T = Z\Lambda/\Lambda$, where the action on the fibers is by affine linear maps $t\mapsto z_y + \alpha(t)$, $z_y\in Z$. We consider two cases.

First, suppose that the automorphism α acts ergodically on the torus T. Then it follows from Marcuard's theorem [23, Theorem 4] that $\alpha \curvearrowright X$ is measurably isomorphic to the direct product of the systems $\alpha \curvearrowright Y$ and $\alpha \curvearrowright T$. Hence, it follows from the inductive assumption that $\alpha \curvearrowright X$ is measurably isomorphic to the product of two Bernoulli maps, and thus Bernoulli.

Second, suppose that the action of α on the torus $T = Z\Lambda/\Lambda$ is not ergodic. Then T contains a nontrivial subtorus $T_0 = Z_0\Lambda/\Lambda$, where Z_0 is a closed connected subgroup of Z, on which α acts isometrically, and $\alpha \curvearrowright X$ is measurably isomorphic to a skew product with the base $\alpha \curvearrowright G/(Z_0\Lambda)$ and the fibers isomorphic to torus T_0, where the action on the fibers is by affine linear maps $t \mapsto z_y + \alpha(t)$, $z_y \in Z_0$. We note that this is an isometric extension of the base, and by the inductive assumption, the base is Bernoulli. Hence, we can apply Rudolph's theorem [29] which shows that weakly mixing isometric extensions of Bernoulli maps are Bernoulli. $\diamond$

References

[1] L. Auslander, *Bieberbach's Theorem on Space Groups and Discrete Uniform Subgroups of Lie Groups.* Ann. of Math. 71 (1960), 579–590.

[2] E. Bombieri and W. Gubler, *Heights in Diophantine geometry.* New Mathematical Monographs, 4. Cambridge University Press, Cambridge, 2006.

[3] J.-P. Conze and S. Le Borgne, *Méthode de martingales et flot géodśique sur une surface de courbure constante négative.* Ergodic Theory Dynam. Systems 21 (2001), no. 2, 421–441.

[4] J.-P. Conze and S. Le Borgne, *Le TCL pour une classe d'automorphismes non hyperboliques de variétés.* Preprint, 2002.

[5] L. Corwin and F. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part I. Basic theory and examples.* Cambridge Studies in Advanced Mathematics, 18. Cambridge University Press, Cambridge, 1990.

[6] K. Dekimpe, *What is . . . an infra-nilmanifold endomorphism?* Notices Amer. Math. Soc. 58 (2011), no. 5, 688–689.

[7] K. Dekimpe, *What an infra-nilmanifold endomorphism really should be.* arXiv:1008.4500.

[8] D. Dolgopyat, *Limit theorems for partially hyperbolic systems.* Trans. Amer. Math. Soc. 356 (2004), no. 4, 1637–1689.

[9] G. Everest and T. Ward, *Heights of polynomials and entropy in algebraic dynamics.* Universitext. Springer-Verlag London, Ltd., London, 1999.

[10] D. Fisher, B. Kalinin and R. J. Spatzier, *Global Rigidity of Anosov actions of Higher Rank Anosov Actions on Tori and Nilmanifolds.* To appear in J. Amer. Math. Soc.; arXiv:1110.0780.

[11] A. Gorodnik and R. Spatzier. *Mixing Properties of Commuting Nilmanifold Automorphisms.* Preprint.

[12] B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds.*, Ann. of Math. (2) 175 (2012), no. 2, 465–540.

[13] P. Hall and C. C. Heyde, *Martingale limit theory and its application.* Academic Press, New York, 1980.

[14] Y. Katznelson, *Ergodic automorphisms of T^n are Bernoulli shifts.* Israel J. Math. 10 (1971), 186–195.

[15] D. Y. Kleinbock and G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces.* Sinai's Moscow Seminar on Dynamical Systems, 141–172, Amer. Math. Soc. Transl. Ser. 2, 171, Amer. Math. Soc., Providence, RI, 1996.

[16] S. Le Borgne, *Un probleme de regularite dans l'equation de cobord.* Fascicule de probabilites (Rennes, 1998), 10 pp., Publ. Inst. Rech. Math. Rennes, 1998, Univ. Rennes I, Rennes, 1998.

[17] S. Le Borgne, *Limit theorems for non-hyperbolic automorphisms of the torus.* Israel J. Math. 109 (1999), 61–73.
[18] S. Le Borgne, *Principes d'invariance pour les flots diagonaux sur $SL(d,R)/SL(d,Z)$.* Ann. Inst. H. Poincaré Probab. Statist. 38 (2002), no. 4, 581–612.
[19] V. P. Leonov, *On the central limit theorem for ergodic endomorphisms of compact commutative groups.* Dokl. Akad. Nauk SSSR 135 (1960), 258–261.
[20] D. A. Lind, *Ergodic group automorphisms and specification.* Ergodic theory (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1978), pp. 93–104, Lecture Notes in Math., 729, Springer, Berlin, 1979.
[21] D. A. Lind, *Dynamical properties of quasihyperbolic toral automorphisms.* Ergodic Theory Dynamical Systems 2 (1982), no. 1, 49–68.
[22] A. N. Livsic, *Cohomology of dynamical systems.* Izv. Akad. Nauk SSSR Ser. Mat. 36 (1972), 1296–1320.
[23] J.-C. Marcuard, *Produits semi-directs et application aux nilvariétés et aux tores en théorie ergodique.* Bull. Soc. Math. France 103 (1975), no. 3, 267–287.
[24] B. Marcus, *A note on periodic points for ergodic toral automorphisms.* Monatsh. Math. 89 (1980), no. 2, 121–129.
[25] W. Parry, *Ergodic properties of affine transformations and flows on nilmanifolds.* Amer. J. Math. 91 (1969), 757–771.
[26] F. Pène, *Averaging method for differential equations perturbed by dynamical systems.* ESAIM Probab. Statist. 6 (2002), 33–88.
[27] J. Rauch and M. Taylor, *Regularity of functions smooth along foliations, and elliptic regularity.* J. Funct. Anal. 225 (2005), no. 1, 74–93.
[28] V. A. Rokhlin, *On the fundamental ideas of measure theory.* American Mathematical Society Translations, Series 1 10 (1962), 1–54.
[29] D. Rudolph, *Classifying the isometric extensions of a Bernoulli shift.* J. Analyse Math. 34 (1978), 36–60.
[30] D. Rudolph, *Ergodic behaviour of Sullivan's geometric measure on a geometrically finite hyperbolic manifold.* Ergodic Theory Dynam. Systems 2 (1982), no. 3-4, 491–512.
[31] P. Samek and D. Volný, *Uniqueness of a martingale-coboundary decomposition of stationary processes.* Comment. Math. Univ. Carolin. 33 (1992), no. 1, 113–119.
[32] A. Starkov, *Dynamical systems on homogeneous spaces.* Translations of Mathematical Monographs, 190. American Mathematical Society, Providence, RI, 2000.
[33] W. Veech, *Periodic points and invariant pseudomeasures for toral endomorphisms.* Ergodic Theory Dynam. Systems 6 (1986), no. 3, 449–473.
[34] M. Viana, *Stochastic Dynamics of Deterministic Systems*, Brazilian Math. Colloquium 1997, IMPA.
[35] A. Wilkinson, *The cohomological equation for partially hyperbolic diffeomorphisms.* To appear in Asterisque; arXiv:0809.4862.

Department of Mathematics, University of Bristol, Bristol, BS8 1TW, U.K.
E-mail address: `a.gorodnik@bristol.ac.uk`

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109.
E-mail address: `spatzier@umich.edu`