



Ravotti, D. (2018). Mixing for suspension flows over skew-translations and time-changes of quasi-abelian filiform nilflows. *Ergodic Theory and Dynamical Systems*. <https://doi.org/10.1017/etds.2018.19>

Peer reviewed version

License (if available):  
Other

Link to published version (if available):  
[10.1017/etds.2018.19](https://doi.org/10.1017/etds.2018.19)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Cambridge University Press at <https://doi.org/10.1017/etds.2018.19> . Please refer to any applicable terms of use of the publisher.

## **University of Bristol - Explore Bristol Research**

### **General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# MIXING FOR SUSPENSION FLOWS OVER SKEW-TRANSLATIONS AND TIME-CHANGES OF QUASI-ABELIAN FILIFORM NILFLOWS

DAVIDE RAVOTTI

ABSTRACT. We consider suspension flows over uniquely ergodic skew-translations on a  $d$ -dimensional torus  $\mathbb{T}^d$ , for  $d \geq 2$ . We prove that there exists a set  $\mathcal{R}$  of smooth functions, which is dense in the space  $\mathcal{C}(\mathbb{T}^d)$  of continuous functions, such that every roof function in  $\mathcal{R}$  which is not cohomologous to a constant induces a mixing suspension flow. We also construct a dense set of mixing examples which is explicitly described in terms of their Fourier coefficients. In the language of nilflows on nilmanifolds, our result implies that, for every uniquely ergodic nilflow on a quasi-abelian filiform nilmanifold, there exists a dense subspace of smooth time-changes in which mixing occurs if and only if the time-change is not cohomologous to a constant. This generalizes a theorem by Avila, Forni and Ulcigrai (J. Diff. Geom., 2011) for the classical Heisenberg group.

## 1. INTRODUCTION

In this paper, we investigate the ergodic properties of a class of *parabolic* flows, i.e. flows for which the divergence of nearby points is polynomial in time. Examples of parabolic flows in the homogeneous set-up are given by the classical horocycle flow (more generally, by unipotent flows on semisimple Lie groups) and by nilflows on nilmanifolds. A *nilmanifold*  $M$  is the quotient of a nilpotent Lie group  $G$  by a lattice  $\Lambda$ , namely a discrete subgroup such that the quotient space  $M = \Lambda \backslash G$  has finite Haar measure. A flow on  $M$  induced by a left-invariant vector field is called a *nilflow*. The study of nilflows has also applications to number theory, for example to the distribution of fractional parts of polynomials and to estimates of theta sums (see, e.g., [8, 9, 6]). The ergodic properties of nilflows are well-understood: almost every nilflow, although it is uniquely ergodic, is not mixing; see §6.1 for a detailed discussion.

Very little is known for general smooth parabolic flows, even for smooth perturbations of homogeneous ones. We consider a simple class of smooth perturbations given by performing a *time-change*. Roughly speaking, a smooth time-change of a flow is obtained by keeping the orbits fixed and varying smoothly the speed of the points; a precise definition is given in §1.2. A natural question is which ergodic properties persist under perturbations: ergodicity is preserved by time-changes, but mixing is more delicate. The case of time-changes of the horocycle flow and of unipotent flows on semisimple Lie groups have been studied by many authors, including Marcus [12], Forni and Ulcigrai [7], Tiedra de Aldecoa [16], and Simonelli [14]; in this paper, building on a previous work by Avila, Forni and Ulcigrai [1], we address the question of mixing for time-changes of nilflows.

The simplest non-abelian nilpotent group is the Heisenberg group  $H$  consisting of  $3 \times 3$  upper triangular unipotent matrices, which is 3-dimensional and 2-step nilpotent. Let  $\{\varphi_t\}_{t \in \mathbb{R}}$  be a uniquely ergodic nilflow on a nilmanifold  $M = \Lambda \backslash H$ . There exists a cross-section  $\Sigma \subset M$  isomorphic to the 2-dimensional torus  $\mathbb{T}^2$  such that the Poincaré map  $T: \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is a uniquely ergodic skew-translation of the form  $T(x, y) = (x + \alpha, x + y + \beta)$ , for some  $\alpha, \beta \in \mathbb{R}$ , and  $\{\varphi_t\}_{t \in \mathbb{R}}$  is isomorphic to the suspension flow over  $(\mathbb{T}^2, T)$  with a constant roof function.

Performing a time-change of the nilflow  $\{\varphi_t\}_{t \in \mathbb{R}}$  amounts to changing the roof function. Clearly, constant time-changes do not produce mixing, therefore we want to rule out from

our analysis all time-changes which behave like constants. This class is given by time-changes *cohomologous to a constant*, see §1.1. Avila, Forni and Ulcigrai in [1] proved that there exists a set  $\mathcal{R}$  of smooth functions which is dense in  $\mathcal{C}(\mathbb{T}^2)$  such that for all positive  $\Psi \in \mathcal{R}$ , the suspension flow over  $(\mathbb{T}^2, T)$  with roof function  $\Psi$  is mixing if and only if  $\Psi$  is not cohomologous to a constant. Moreover, they showed that this condition can be checked explicitly. They also prove an analogous result for smooth time-changes of the original Heisenberg nilflow. In this paper, we generalize these results to higher dimensions, see Theorem 1.1 and 1.2.

**1.1. Suspension flows over skew-translations.** Let us consider a  $d$ -dimensional torus  $\mathbb{T}^d$  for some  $d \geq 2$ . We denote points in  $\mathbb{T}^d$  by row vectors  $\mathbf{x} = (x_1, \dots, x_d)$ . Let  $\text{Leb}_d$  be the  $d$ -dimensional Lebesgue measure. Let  $T: \mathbb{T}^d \rightarrow \mathbb{T}^d$  be a skew-translation of the form  $T\mathbf{x} = \mathbf{x}A + \mathbf{b}$ , where  $A = (a_{i,j})_{1 \leq i,j \leq d}$  is a  $d \times d$  upper-triangular unipotent matrix with integer coefficients such that  $A \neq \text{Id}$  and  $\mathbf{b} = (b_1, \dots, b_d) \in \mathbb{T}^d$ ; namely, for  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{T}^d$ , define

$$T(x_1, \dots, x_d) = (x_1, \dots, x_d) \begin{pmatrix} 1 & a_{1,2} & \cdots & a_{1,d} \\ & 1 & \ddots & \vdots \\ & & \ddots & a_{d-1,d} \\ & & & 1 \end{pmatrix} + (b_1, \dots, b_d). \quad (1.1)$$

We suppose that the skew-translation  $T: \mathbb{T}^d \rightarrow \mathbb{T}^d$  is ergodic (equivalently, uniquely ergodic [8]).

Any positive continuous function  $\Psi: \mathbb{T}^d \rightarrow \mathbb{R}_{>0}$  defines a *suspension flow* in the following way: consider the equivalence relation  $\sim_\Psi$  on  $\mathbb{T}^d \times \mathbb{R}$  generated by all the pairs  $\{(\mathbf{x}, \Psi(\mathbf{x})), (T\mathbf{x}, 0)\}$ . The suspension flow  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  on the space  $\mathbb{T}^d \times \mathbb{R} / \sim_\Psi$  is defined by moving at unit speed in the vertical direction for time  $t$ . If we denote by  $S_n(\Psi)$  the  $n$ -th Birkhoff sum along the orbits of  $T$ , namely

$$S_n(\Psi)(\mathbf{x}) := \begin{cases} \sum_{i=0}^{n-1} \Psi \circ T^i \mathbf{x} & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ -\sum_{i=-1}^n \Psi \circ T^i \mathbf{x} & \text{if } n < 0, \end{cases} \quad (1.2)$$

then we can write an explicit formula for the suspension flow

$$T_t^\Psi(\mathbf{x}, r) = \left( T^{n_t(\mathbf{x}, r)} \mathbf{x}, r + t - S_{n_t(\mathbf{x}, r)}(\Psi)(\mathbf{x}) \right), \quad (1.3)$$

where  $n_t(\mathbf{x}, r)$  denotes the maximum  $n \geq 0$  such that  $S_n(\Psi)(\mathbf{x}) \leq r + t$ .

For every roof function  $\Psi$ , unique ergodicity of  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  is equivalent to unique ergodicity of  $T$ ; on the other hand, mixing for the suspension flow is a delicate question. Suppose that there exists a measurable (resp. smooth) function  $u: \mathbb{T}^d \rightarrow \mathbb{R}$  such that

$$\Psi - \int_{\mathbb{T}^d} \Psi \, d\text{Leb}_d = u \circ T - u.$$

We say that  $u \circ T - u$  is a *measurable* (resp. *smooth*) *coboundary* for  $T$  and  $\Psi$  and the constant  $\int \Psi$  are *measurably* (resp. *smoothly*) *cohomologous w.r.t. T*. In this case, the map  $\zeta: (\mathbf{x}, r) \mapsto (\mathbf{x}, r + u(\mathbf{x}))$  defined on  $\mathbb{T}^d \times \mathbb{R}$  descends to the quotient and maps the invariant measure for  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  to the invariant measure of the suspension flow with constant roof function  $\int \Psi$ . In particular, any roof function  $\Psi$  cohomologous to a constant induces a non-mixing flow. Our main result, Theorem 1.1 below, shows that, within a dense subspace  $\mathcal{R}$ , the condition of not being cohomologous to a constant is also sufficient for mixing of the suspension flow.

Determining whether a function is a measurable coboundary is not, in general, effectively possible. An exception is the case of a 2-dimensional skew-translation treated in [1], where measurable coboundaries are explicitly characterized in terms of invariant distributions for the Heisenberg nilflow. At present, this result appears not to be generalizable to higher

dimensions, since it relies on sharp estimates on Weyl sums (see [6] and references therein), which are available only for degree two. However, exploiting the 2-dimensional case, we construct a dense and explicitly described set of mixing examples for a large class of higher dimensional skew-translations, which includes the ones arising from filiform nilflows, see §1.2.

**Theorem 1.1.** (a) *There exists a subspace  $\mathcal{R}$  of smooth functions, which is dense in  $\mathcal{C}(\mathbb{T}^d)$  w.r.t.  $\|\cdot\|_\infty$ , such that for all positive  $\Psi \in \mathcal{R}$  the suspension flow over  $(\mathbb{T}^d, T)$  with roof function  $\Psi$  is mixing if and only if  $\Psi$  is not cohomologous to a constant.*  
 (b) *If the entries above the diagonal are non-zero, namely if  $a_{i,i+1} \neq 0$  for  $i = 1, \dots, d-1$  in (1.1), then there exists a dense set  $\mathcal{M}$  of mixing examples which is explicitly described in terms of their Fourier coefficients.*

**1.2. Time-changes of quasi-abelian filiform nilflows.** From Theorem 1.1 we deduce an analogous statement for time-changes of *quasi-abelian filiform* nilflows, which are nilflows on a class of higher dimensional and higher step nilpotent groups. We recall all the relevant definitions, referring the reader to [10, §1] and to [3] for an introduction to homogeneous flows on Lie groups and for a background on nilpotent groups respectively.

Let  $G$  be a connected, simply connected Lie group and let  $\mathfrak{g}$  be its Lie algebra. For any vector  $\mathbf{w} \in \mathfrak{g}$  it is possible to define a vector field  $W$  over  $G$  in the following way: if  $L_g: h \mapsto gh$  denotes the left-multiplication by  $g \in G$ , we set  $W_g = (L_g)_* \mathbf{w}$ , where  $(L_g)_* = (L_g)_*(\text{Id}): \mathfrak{g} \rightarrow T_g G$  is the differential of  $L_g$  at the identity. The vector field  $W$  is left-invariant, i.e.  $(L_g)_* W = W$  for all  $g \in G$  and indeed the map  $\mathbf{w} \mapsto W$  is a bijection between  $\mathfrak{g}$  and the set of left-invariant vector fields over  $G$  [10, Proposition 1.72]. By a little abuse of notation, we will identify  $\mathbf{w}$  with  $W$ .

The flow  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  defined by  $\mathbf{w}$  is explicitly given by  $\varphi_t^{\mathbf{w}}(g) = g \exp(t\mathbf{w})$ . Given a lattice  $\Lambda \leq G$ , the flow  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  induces a flow on the quotient  $\Lambda \backslash G$ ; moreover, since  $G$  is unimodular [4, Proposition 9.20],  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  preserves both the Haar measure  $\mu$  on  $G$  and its push-forward onto the quotient  $\Lambda \backslash G$ . We will denote the latter with  $\mu$  as well and we will assume it is appropriately normalized, namely  $\mu(\Lambda \backslash G) = 1$ .

Denote by  $[\cdot, \cdot]_G$  the commutator in  $G$ . We recall that a group  $G$  is  $n$ -step nilpotent if  $G^{(n+1)} = \{\text{Id}\}$  and  $G^{(n)} \neq \{\text{Id}\}$ , where  $G^{(1)} = G$  and  $G^{(i+1)} = [G, G^{(i)}]_G$ . Setting  $\mathfrak{g}^{(1)} = \mathfrak{g}$  and  $\mathfrak{g}^{(i+1)} = [\mathfrak{g}, \mathfrak{g}^{(i)}]$ , where  $[\cdot, \cdot]$  denotes the Lie bracket in  $\mathfrak{g}$ , it is a well-known fact that  $\mathfrak{g}^{(i)}$  is the Lie algebra of  $G^{(i)}$ . In particular,  $\mathfrak{g}^{(n+1)} = \{0\}$  and  $\mathfrak{g}^{(n)} \neq \{0\}$ ; we say that  $\mathfrak{g}$  is a  $n$ -step nilpotent Lie algebra. Notice that in a  $n$ -step nilpotent algebra the centre is always nontrivial, more precisely  $\mathfrak{g}^{(n)} \subseteq \mathfrak{z}(\mathfrak{g})$ .

In this paper, we focus our attention on *quasi-abelian filiform* groups  $F_d$ , which we introduce through their Lie algebras. The quasi-abelian filiform algebra  $\mathfrak{f}_d$  of  $F_d$  is the  $(d+1)$ -dimensional nilpotent Lie algebra spanned by  $\mathcal{F}_d = \{\mathbf{f}_0, \dots, \mathbf{f}_d\}$  such that the only nontrivial brackets are  $[\mathbf{f}_0, \mathbf{f}_i] = \mathbf{f}_{i+1}$  for  $1 \leq i \leq d-1$ . Then,  $\mathfrak{f}_d$  is  $d$ -step nilpotent and we can represent it as a matrix algebra as

$$x\mathbf{f}_0 + \sum_{i=1}^d y_i \mathbf{f}_i \mapsto \begin{pmatrix} 0 & x & & & y_d \\ & 0 & \ddots & & \vdots \\ & & \ddots & & \\ & & & x & y_2 \\ & & & 0 & y_1 \\ & & & & 0 \end{pmatrix}.$$

We remark that  $F_1 \simeq \mathbb{R}^2$  and  $F_2$  is the Heisenberg group  $H$ .

Let  $F = F_d$  be a quasi-abelian filiform group and let  $\Lambda < F$  be a lattice; the quotient  $M = \Lambda \backslash F$  is said to be a *quasi-abelian filiform nilmanifold* and every flow  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  as above is called *quasi-abelian filiform nilflow*. Almost every quasi-abelian filiform nilflow is uniquely ergodic but not weak mixing, see §6.1.

We say that a flow  $\{\varphi_t^\alpha\}_{t \in \mathbb{R}}$  on  $M$  is a *continuous* (resp. *smooth*) *time-change* of  $\{\varphi_t\}_{t \in \mathbb{R}}$  if there exists a continuous (resp. smooth) function  $\alpha: M \rightarrow \mathbb{R}_{>0}$  such that  $\{\varphi_t^\alpha\}_{t \in \mathbb{R}}$  is

induced by  $\alpha \mathbf{w}$  (recall that  $\mathbf{w}$  is the vector field inducing  $\{\varphi_t\}_{t \in \mathbb{R}}$ ). We say that  $\alpha$  is the *infinitesimal generator* of the time-change. The time-change given by  $\alpha$  is cohomologous to a constant if there exists  $\beta: M \rightarrow \mathbb{R}$  such that

$$\alpha - \int_M \alpha d\mu = \mathbf{w}\beta.$$

As for suspension flows, time-changes cohomologous to constants are not mixing. From Theorem 1.1 we deduce the following result.

**Theorem 1.2.** *Let  $M = \Lambda \backslash F_d$  be a quasi-abelian filiform nilmanifold for some  $d \geq 2$  and consider a uniquely ergodic quasi-abelian filiform nilflow on  $M$ . There exists a set of smooth time-changes, which is dense in the set of continuous time-changes, such that every element is mixing if and only if it is not cohomologous to a constant.*

*Moreover, the set of mixing time-changes is dense in the set of continuous time-changes.*

**1.3. Outline of the paper.** Section 2 is devoted to explain the general strategy of the proof of Theorem 1.1-(a). First, we present a general mechanism that will allow us to reduce to consider a factor of the suspension flow for which the divergence of nearby points is of strictly higher order in the  $x_d$ -direction (we remark that  $T$  acts as a translation in this latter coordinate). This is obtained by applying inductively Proposition 2.1, whose proof is contained in §4. Then, we prove mixing for the new suspension flow by showing that there is stretch of Birkhoff sums of the roof function  $\Psi$  (see Theorem 2.4 in §5) and then using this stretch to show that segments in the  $x_d$ -direction get *sheared* along the flow direction, as we explain in §7. This is a well-known mechanism, often exploited in the context of parabolic flows to prove mixing. For example, it has been used by Marcus for time-changes of horocycle flows [12], by Sinai and Khanin [15], Fayad [5], Ulcigrai [17] and the author [13] for suspension flows over rotations and interval exchange transformations. In our case, shearing comes from the fact that the roof function is not cohomologous to a constant and a decoupling argument, which generalizes the one used by Avila, Forni and Ulcigrai in [1] (although, in our higher dimensional setting, an additional geometric localization argument is needed). In §3, we use Proposition 2.1 to construct a dense set of mixing examples on any dimension, starting from the 2-dimensional ones, hence proving Theorem 1.1-(b). These roof functions are explicitly characterized in terms of their Fourier coefficients (see Lemma 3.3, which generalizes a result by Katok [11, Theorem 11.25]). Finally, in Section 6, we prove Theorem 1.2 by constructing a cross-section for the quasi-abelian filiform nilflow such that, in appropriate coordinates, the Poincaré map is a skew-translation on  $\mathbb{T}^d$ , hence reducing the problem of time-changes of quasi-abelian filiform nilflows to the setting of Theorem 1.1.

**Acknowledgments.** I would like to thank Giovanni Forni and my supervisor Corinna Ulcigrai for several useful discussions. I also thank the referee for his/her attentive reading and helpful comments on a previous version of the paper. The research leading to these results has received funding from the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 335989.

## 2. PROOF OF THEOREM 1.1-PART A

In this section, we present the general structure of the proof of Theorem 1.1-(a), stating some intermediate results, which are proved in later sections.

Let  $T$  be a uniquely ergodic skew-translation as in (1.1). If we denote by  $E_j$  the image of the linear map  $(A - \text{Id})^j$ , we have a filtration of  $\mathbb{R}^d$  into rational subspaces

$$\mathbb{R}^d = E_0 > E_1 > \cdots > E_k > E_{k+1} = \{0\}.$$

Up to a linear isomorphism, we can assume that the basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$  of  $\mathbb{Z}^d$  is adapted to the filtration above, in particular  $\{\mathbf{e}_{d_0+1}, \dots, \mathbf{e}_d\}$  is a basis of  $E_k$ , where  $d - d_0 = \dim E_k$ . Since  $T$  is not a rotation,  $k \geq 1$  and  $1 \leq d_0 \leq d - 1$ . We remark that  $\mathbf{w} \in E_j$  for  $j \geq 1$  if and only if there exists  $\mathbf{v} \in E_{j-1}$  such that  $\mathbf{v}(A - \text{Id}) = \mathbf{w}$ , i.e.  $\mathbf{v}A = \mathbf{v} + \mathbf{w}$ . In particular,

for the basis elements  $\mathbf{e}_{d_0+i} \in E_k \cap \mathbb{Z}^d$ , for  $i = 1, \dots, d - d_0$ , there exists  $\mathbf{v} \in E_{k-1} \cap \mathbb{Z}^d$  such that  $\mathbf{v}A = \mathbf{v} + a\mathbf{e}_{d_0+i}$ , for some  $a \neq 0$ .

We want to reduce to the case  $d_0 = d - 1$ , that is  $\dim E_k = 1$ . In §2.1 we describe a general mechanism that allows us to deduce mixing from the assumption that a system with one less dimension is mixing. This motivates also the definition of the set  $\mathcal{R}$ , which is explained in §2.2. In §2.3 we prove  $\mathcal{R}$  is dense in  $\mathcal{C}(\mathbb{T}^d)$ . Finally, in §2.4 and §2.5 we apply inductively the result of §2.1 to reduce to the case  $d_0 = d - 1$  and then we conclude the proof of Theorem 1.1-(a).

**2.1. The wrapping mechanism.** Let  $\pi: \mathbb{T}^d \rightarrow \mathbb{T}^{d-1}$  be the projection given by suppressing the  $d$ -th coordinate. Then  $\pi$  gives a factor of  $(\mathbb{T}^d, T)$ ; more precisely, let  $\hat{A} = (a_{i,j})_{1 \leq i,j \leq d-1}$  be the  $(d-1) \times (d-1)$  matrix obtained by removing the last row and the last column from  $A$  and let  $\hat{\mathbf{b}} = \pi(\mathbf{b}) \in \mathbb{T}^{d-1}$ . Then, the skew-translation  $\hat{T}: \mathbb{T}^{d-1} \rightarrow \mathbb{T}^{d-1}$  defined by  $\hat{T}\mathbf{y} = \mathbf{y}\hat{A} + \hat{\mathbf{b}}$  makes the diagram

$$\begin{array}{ccc} \mathbb{T}^d & \xrightarrow{T} & \mathbb{T}^d \\ \pi \downarrow & & \downarrow \pi \\ \mathbb{T}^{d-1} & \xrightarrow{\hat{T}} & \mathbb{T}^{d-1} \end{array}$$

commute.

Let us denote also by  $\pi: \mathbb{T}^d \times \mathbb{R} \rightarrow \mathbb{T}^{d-1} \times \mathbb{R}$  the projection  $\pi(\mathbf{x}, r) = (\pi(\mathbf{x}), r)$ . Let  $\psi: \mathbb{T}^{d-1} \rightarrow \mathbb{R}_{>0}$  be a smooth function over  $\mathbb{T}^{d-1}$  and consider the roof function  $\psi \circ \pi$  over  $(\mathbb{T}^d, T)$  which is constant in the  $d$ -th coordinate. Then,  $\pi$  is a factor map of the suspension flow  $\{T_t^{\psi \circ \pi}\}_{t \in \mathbb{R}}$ , namely

$$(\pi \circ T_t^{\psi \circ \pi})(\mathbf{x}, r) = (\hat{T}_t^\psi \circ \pi)(\mathbf{x}, r). \quad (2.1)$$

As we discussed at the beginning of the section, there exists  $\mathbf{v} \in \mathbb{Z}^d$  such that  $\mathbf{v}A = \mathbf{v} + a\mathbf{e}_d$  for some  $a \neq 0$ . This means that the images of segments parallel to  $\mathbf{v}$  under  $T$  get sheared in direction  $\mathbf{e}_d$  and wrap around the circles parallel to  $\mathbf{e}_d$ . Exploiting this shearing effect along the fibers of the projection  $\pi$ , it is possible to “lift” mixing from the quotient to the original suspension flow, namely the following result.

**Proposition 2.1.** *Let  $\pi$  be the projection onto the first  $d-1$  coordinates and let  $\hat{T}: \mathbb{T}^{d-1} \rightarrow \mathbb{T}^{d-1}$  be the corresponding factor. Let  $\psi: \mathbb{T}^{d-1} \rightarrow \mathbb{R}_{>0}$  be a positive smooth function. If there exists  $\mathbf{v} \in \mathbb{Z}^d$  such that  $\mathbf{v}A = \mathbf{v} + a\mathbf{e}_d$  for some  $a \neq 0$ , then the suspension flow  $\{T_t^{\psi \circ \pi}\}_{t \in \mathbb{R}}$  over  $(\mathbb{T}^d, T)$  is mixing if and only if the suspension flow  $\{\hat{T}_t^\psi\}_{t \in \mathbb{R}}$  over  $(\mathbb{T}^{d-1}, \hat{T})$  is mixing.*

The proof of Proposition 2.1 is presented in §4.

**2.2. Definition of  $\mathcal{R}$ .** Generalizing the notation of §2.1, for each  $i = 1, \dots, d-1$ , denote by  $\pi_i: \mathbb{T}^d \rightarrow \mathbb{T}^i$  the projection onto the first  $i$  coordinates and by  $T_i: \mathbb{T}^i \rightarrow \mathbb{T}^i$  the corresponding factor map  $\pi_i \circ T = T_i \circ \pi_i$ . Let  $\mathcal{P}(d)$  be the space of trigonometric polynomials over  $\mathbb{T}^d$ . For any  $\Psi \in \mathcal{P}(d)$ , we can write

$$\Psi = \psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp,$$

where

$$\psi_{d-1}(\pi_{d-1}(\mathbf{x})) = \int_0^1 \Psi(\mathbf{x}) dx_d \quad \text{and} \quad \Psi_d^\perp(\mathbf{x}) = \Psi(\mathbf{x}) - \psi_{d-1}(\pi_{d-1}(\mathbf{x})).$$

The function  $\psi_{d-1} \circ \pi_{d-1}$  does not depend on the  $x_d$ -coordinate, thus we can see  $\psi_{d-1}$  as a trigonometric polynomial over  $\mathbb{T}^{d-1}$ . Inductively, we write

$$\Psi = \psi_{d_0} \circ \pi_{d_0} + \Psi_{d_0+1}^\perp \circ \pi_{d_0+1} + \dots + \Psi_d^\perp, \quad (2.2)$$

where

$$\psi_i \circ \pi_i = \int_0^1 \psi_{i+1} \circ \pi_{i+1} dx_{i+1} \quad \text{and} \quad \Psi_i^\perp \circ \pi_i = \psi_{i+1} \circ \pi_{i+1} - \psi_i \circ \pi_i.$$

The integral of  $\Psi_i^\perp$  in  $dx_i$  is equal to zero, hence we have the decomposition

$$\mathcal{P}(d) = \mathcal{P}(d_0) \oplus \bigoplus_{i=d_0+1}^d \mathcal{Q}(i), \quad \text{where} \quad \mathcal{Q}(i) = \left\{ \Psi \in \mathcal{P}(i) : \int_0^1 \Psi dx_i \equiv 0 \right\}. \quad (2.3)$$

Explicitly, let  $e(x) = \exp(2\pi i x)$  and consider a trigonometric polynomial of degree  $m$ ,

$$\Psi(\mathbf{x}) = \sum_{\mathbf{l} \in [-m, m]^d \cap \mathbb{Z}^d} c_{\mathbf{l}} e(\mathbf{l} \cdot \mathbf{x}) \in \mathcal{P}(d).$$

Then, denoting  $\mathbf{x}_i = \pi_i(\mathbf{x})$ , we have

$$\begin{aligned} \psi_{d_0}(\mathbf{x}_{d_0}) &= \sum_{\mathbf{l}_{d_0} \in [-m, m]^{d_0} \cap \mathbb{Z}^{d_0}} c_{(l_1, \dots, l_{d_0}, 0, \dots, 0)} e(\mathbf{l}_{d_0} \cdot \mathbf{x}_{d_0}) \quad \text{and} \\ \Psi_i^\perp(\mathbf{x}_i) &= \sum_{\mathbf{l}_i \in [-m, m]^i \cap \mathbb{Z}^i, l_i \neq 0} c_{(l_1, \dots, l_i, 0, \dots, 0)} e(\mathbf{l}_i \cdot \mathbf{x}_i), \end{aligned}$$

where the last sum is taken over all integer vectors  $\mathbf{l}_i = \pi_i(\mathbf{l}) \in [-m, m]^i \cap \mathbb{Z}^i$  such that the last component  $l_i \neq 0$ .

**Definition 2.2.** For each  $\Psi \in \mathcal{P}(d)$  consider the decomposition (2.2). We define the set  $\mathcal{R} = \mathcal{R}(T) \subset \mathcal{P}(d)$  associated to the skew-translation  $T$  by

$$\begin{aligned} \Psi \in \mathcal{R} \quad \text{iff} \quad & \Psi_i^\perp \text{ is a measurable coboundary for } T_i \text{ for all } i = d_0 + 2, \dots, d \\ & \text{and } \psi_{d_0} \text{ is smoothly cohomologous to a constant w.r.t. } T_{d_0}. \end{aligned}$$

**2.3. Density.** We now prove that  $\mathcal{R}$  is dense in  $\mathcal{C}(\mathbb{T}^d)$  w.r.t.  $\|\cdot\|_\infty$ . By (2.3), we have to show that the set of trigonometric polynomials which are smoothly cohomologous to a constant w.r.t.  $T_{d_0}$  is dense in  $\mathcal{P}(d_0)$  and that the set of measurable coboundaries for  $T_i$  in  $\mathcal{Q}(i)$  is dense in  $\mathcal{Q}(i)$  for all  $i = d_0 + 2, \dots, d$ . All factors  $T_{d_0}, \dots, T_{d-1}$  are uniquely ergodic skew-translations of the same form as  $T$ , hence it suffices to prove the following lemma; the proof follows the same ideas as a result by Katok [11, Proposition 10.13].

**Lemma 2.3.** *We have the following.*

- (i) *The set of trigonometric polynomials which are smoothly cohomologous to a constant w.r.t.  $T$  is dense in  $\mathcal{P}(d)$ .*
- (ii) *The set of smooth coboundaries for  $T$  in  $\mathcal{Q}(d)$  is dense in  $\mathcal{Q}(d)$ .*

*Proof.* We show (ii); the proof of (i) is analogous. Define  $P: \mathcal{Q}(d) \rightarrow \mathcal{Q}(d)$  by  $P\Psi_d^\perp = \Psi_d^\perp \circ T - \Psi_d^\perp$ ; it is sufficient to show that  $\mathcal{Q}(d) \subseteq \overline{\text{Im } P}$ , where the closure is w.r.t.  $\|\cdot\|_\infty$  in  $\mathcal{Q}(d)$ .

Suppose, by contradiction, that there exists  $\Phi \in \mathcal{Q}(d)$  and  $\Phi \notin \overline{\text{Im } P}$ . By Hahn-Banach Theorem, there exists  $\nu: \mathcal{Q}(d) \rightarrow \mathbb{R}$  linear and continuous such that  $\nu(\Phi) = 1$  and  $\nu|_{\overline{\text{Im } P}} = 0$ . We extend  $\nu$  to a functional  $\tilde{\nu}$  on all  $\mathcal{P}(d) = \mathcal{P}(d-1) \oplus \mathcal{Q}(d)$  by defining

$$\tilde{\nu}(\psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp) = \int_{\mathbb{T}^{d-1}} \psi_{d-1} d\text{Leb}_{d-1} + \nu(\Psi_d^\perp).$$

It is easy to check that  $\tilde{\nu}$  is again linear and continuous, hence it uniquely defines a measure on  $\mathbb{T}^d$ . For every  $\Psi_d^\perp \in \mathcal{Q}(d)$  we have

$$0 = \nu(P\Psi_d^\perp) = \nu(\Psi_d^\perp \circ T) - \nu(\Psi_d^\perp),$$

i.e.,  $\nu$  is  $T$ -invariant over  $\mathcal{Q}(d)$ . Therefore, for any  $\Psi = \psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp \in \mathcal{P}(d)$ ,

$$\tilde{\nu}(\Psi \circ T) = \int_{\mathbb{T}^{d-1}} \psi_{d-1} \circ T_{d-1} d\text{Leb}_{d-1} + \nu(\Psi_d^\perp \circ T) = \int_{\mathbb{T}^{d-1}} \psi_{d-1} d\text{Leb}_{d-1} + \nu(\Psi_d^\perp) = \tilde{\nu}(\Psi).$$

By unique ergodicity of  $T$ , we deduce that  $\tilde{\nu} = \text{Leb}_d$ . We conclude

$$\tilde{\nu}(\Phi) = \int_{\mathbb{T}^d} \Phi \, d\text{Leb}_d = 0,$$

in contradiction with  $\nu(\Phi) = 1$ .  $\square$

**2.4. Proof of Theorem 1.1-(a): step 1.** Using Proposition 2.1, we explain how to reduce the problem to the case of  $\dim E_k = 1$ , where, we recall,  $E_k$  is the image of  $(A - \text{Id})^k$  and  $(A - \text{Id})^{k+1} = 0$ . Let  $\Psi \in \mathcal{R}$ , and assume that it is not cohomologous to a constant w.r.t.  $T$ . If  $d_0 \leq d - 2$ , then, by definition of  $\mathcal{R}$ , the function  $\Psi_d^\perp$  is a measurable coboundary for  $T$ , i.e.  $\Psi_d^\perp = u \circ T - u$  for some measurable function  $u: \mathbb{T}^d \rightarrow \mathbb{R}$ . We claim that  $\psi_{d-1}$  is not cohomologous to a constant w.r.t. the factor map  $T_{d-1}$ . By contradiction, suppose that  $\psi_{d-1} - \int \psi_{d-1} = v \circ T_{d-1} - v$  for some  $v: \mathbb{T}^{d-1} \rightarrow \mathbb{R}$ . Then,

$$\begin{aligned} \Psi - \int_{\mathbb{T}^d} \Psi \, d\text{Leb}_d &= \psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp - \int_{\mathbb{T}^{d-1}} \psi_{d-1} \, d\text{Leb}_{d-1} \\ &= v \circ T_{d-1} \circ \pi_{d-1} - v \circ \pi_{d-1} + u \circ T - u = (v \circ \pi_{d-1} + u) \circ T - (v \circ \pi_{d-1} + u), \end{aligned}$$

in contradiction with the assumption on  $\Psi$ .

*Remark 1.* The differential of the map  $\zeta: (\mathbf{x}, r) \mapsto (\mathbf{x}, r + u(\mathbf{x}))$  defined on  $\mathbb{T}^d \times \mathbb{R}$  has determinant 1, hence  $\zeta$  preserves the  $(d + 1)$ -dimensional Lebesgue measure. Moreover, one can check that  $\zeta(\mathbf{x}, r + \Psi(\mathbf{x})) = \zeta(T\mathbf{x}, r)$ ; in particular,  $\zeta$  descends to the quotient spaces  $\zeta: \mathbb{T}^d \times \mathbb{R} / \sim_\Psi \rightarrow \mathbb{T}^d \times \mathbb{R} / \sim_{\psi_{d-1} \circ \pi_{d-1}}$  and hence maps the invariant measure for the suspension flow over  $(\mathbb{T}^d, T)$  with roof function  $\Psi = \psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp$  to the invariant measure of the one with roof function  $\psi_{d-1} \circ \pi_{d-1}$ . In particular, the former is mixing if and only if the latter is mixing.

By Remark 1 and by Proposition 2.1, mixing of  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  is equivalent to mixing of  $\{(T_{d-1})_t^{\psi_{d-1}}\}_{t \in \mathbb{R}}$ , where  $\psi_{d-1} \in \mathcal{R}(T_{d-1})$ . Iterating this process for all  $\Psi_i^\perp$  for  $i = d_0 + 2, \dots, d$ , we reduce to prove mixing for the suspension flow  $\{(T_{d_0+1})_t^{\psi_{d_0+1}}\}_{t \in \mathbb{R}}$  over  $(\mathbb{T}^{d_0+1}, T_{d_0+1})$  with roof function  $\psi_{d_0+1} \in \mathcal{R}(T_{d_0+1})$ . By construction, the map  $T_{d_0+1}$  is of the desired form.

**2.5. Proof of Theorem 1.1-(a): step 2.** We can now assume that the matrix  $A$  in the definition (1.1) of  $T$  satisfies  $d_0 = d - 1$ , i.e.  $\dim E_k = 1$ . Consider  $\Psi \in \mathcal{R}(T)$ , and assume that it is not cohomologous to a constant. Then, by definition of  $\mathcal{R}$ , we can write  $\Psi = \psi_{d-1} \circ \pi_{d-1} + \Psi_d^\perp$ , where  $\psi_{d-1}$  is smoothly cohomologous to a constant w.r.t.  $T_{d-1}$ . Thus, there exists a smooth function  $u: \mathbb{T}^{d-1} \rightarrow \mathbb{R}$  such that  $\psi_{d-1} - \int \psi_{d-1} = u \circ T_{d-1} - u$ .

We notice that  $\Psi_d^\perp$  is not a measurable coboundary for  $T$ . Indeed, if this were not the case and  $\Psi_d^\perp = v \circ T - v$  for some measurable function  $v: \mathbb{T}^d \rightarrow \mathbb{R}$ , we would have

$$\begin{aligned} \Psi - \int_{\mathbb{T}^d} \Psi \, d\text{Leb}_d &= \Psi - \int_{\mathbb{T}^d} \psi_{d-1} \circ \pi_{d-1} \, d\text{Leb}_d = \psi_{d-1} \circ \pi_{d-1} - \int_{\mathbb{T}^{d-1}} \psi_{d-1} \, d\text{Leb}_{d-1} + \Psi_d^\perp \\ &= u \circ T_{d-1} \circ \pi_{d-1} - u \circ \pi_{d-1} + v \circ T - v = (u \circ \pi_{d-1} + v) \circ T - (u \circ \pi_{d-1} + v), \end{aligned}$$

which is a contradiction since we are assuming that  $\Psi$  is not measurably cohomologous to a constant. The first step is to prove that the Birkhoff sums of  $\Psi_d^\perp$  grow in measure, namely the following result.

**Theorem 2.4.** *For any function  $\Psi^\perp \in \mathcal{Q}(d)$ , which is not a measurable coboundary for  $T$ , and any  $C > 1$  we have*

$$\lim_{n \rightarrow \infty} \text{Leb}_d (|S_n(\Psi^\perp)| < C) = 0.$$

From Theorem 2.4, using the fact that  $\psi_{d-1}$  is smoothly cohomologous to a constant, we deduce mixing. This final part follows more closely the ideas in [1], hence we leave the proof for the appendix.



**Theorem 2.5.** *Assume that  $d_0 = d - 1$ . Assume also that  $\Psi \in \mathcal{P}(d)$  is not a measurable coboundary for  $T$  and that the function  $\psi_{d-1}$  defined by (2.2) is smoothly cohomologous to a constant. Then, the suspension flow  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  is mixing.*

### 3. PROOF OF THEOREM 1.1-PART B

In this section we prove Theorem 1.1-(b) by constructing the set  $\mathcal{M} = \mathcal{M}(d)$ , dense in  $\mathcal{P}(d)$  w.r.t.  $\|\cdot\|_\infty$ , which consists of roof functions inducing a mixing suspension flow. We characterize smooth coboundaries for skew-translations in terms of their Fourier coefficients and we apply Proposition 2.1 inductively to produce mixing suspension flows in higher dimension, starting from the ones in dimension 2, see [1, §5].

If we denote again by  $\pi_i: \mathbb{T}^d \rightarrow \mathbb{T}^i$  the projection onto the first  $i$  coordinates and by  $\pi_i(\mathbf{x}) = \mathbf{x}_i$ , we have a sequence of factors

$$(\mathbb{T}^d, T) \mapsto (\mathbb{T}^{d-1}, T_{d-1}) \mapsto \cdots \mapsto (\mathbb{T}^2, T_2), \quad (3.1)$$

where  $T_i \mathbf{x}_i = \mathbf{x}_i A_i + \mathbf{b}_i$  and  $A_i = (a_{l,m})_{1 \leq l, m \leq i}$ .

**Definition 3.1.** Let  $\Psi \in \mathcal{P}(d)$  be written as

$$\Psi = \psi_2 \circ \pi_2 + \Psi_3^\perp \circ \pi_3 + \cdots + \Psi_d^\perp, \text{ with } \psi_2 \in \mathcal{P}(2), \text{ and } \Psi_i^\perp \in \mathcal{Q}(i), \text{ for } i = 3, \dots, d.$$

We say that  $\Psi \in \mathcal{M}(d)$  if  $\psi_2$  induces a mixing suspension flow for the 2-dimensional skew-translation  $(\mathbb{T}^2, T_2)$  and  $\Psi_i^\perp$  is a smooth coboundary for  $T_i$  for all  $i = 3, \dots, d$ .

Every function in  $\mathcal{M}(d)$  induces a mixing suspension flow by Remark 1 and Proposition 2.1 applied inductively in (3.1) up to the last factor. Moreover, the set of mixing roofs  $\psi_2$  is dense in  $\mathcal{P}(2)$  by [1] and, by Lemma 2.3, the set of smooth coboundaries  $\Psi_i^\perp$  for  $T_i$  is dense in  $\mathcal{Q}(i)$ . Therefore,  $\mathcal{M}(d)$  is dense in  $\mathcal{P}(d)$ , and hence in  $\mathcal{C}(\mathbb{T}^d)$ .

We now characterize the set  $\mathcal{M}(d)$  so that it is possible to effectively check if a trigonometric polynomial  $\Psi$  belongs to  $\mathcal{M}(d)$ . The case of  $\psi_2$  has already been treated in [1, §5]; let us analyze when  $\Psi_i^\perp \in \mathcal{Q}(i)$  is a smooth coboundary for  $T_i$ .

The following lemma is easy to be verified.

**Lemma 3.2.** *Let  $\mathcal{O}_i$  be the set of orbits of the action of the transpose  $A_i^T$  of  $A_i$  on  $\mathbb{Z}^i$  and for any  $\omega \in \mathcal{O}_i$  let*

$$\mathcal{H}_\omega = \bigoplus_{\mathbf{l} \in \omega} \mathbb{C}e(\mathbf{l} \cdot \mathbf{x}_i).$$

*The space  $L^2(\mathbb{T}^i)$  admits an orthogonal splitting*

$$L^2(\mathbb{T}^i) = \bigoplus_{\omega \in \mathcal{O}_i} \mathcal{H}_\omega,$$

*and all the components are  $T_i$ -invariant.*

Therefore, it is enough to investigate the existence of solutions  $u$  for the cohomological equation  $\Psi_i^\perp = u \circ T_i - u$  in each component  $\mathcal{H}_\omega$ . The following result is a generalisation in higher dimension of a theorem by Katok [11, Theorem 11.25].

**Lemma 3.3.** *Let  $\omega \in \mathcal{O}_i$ ; consider  $\mathbf{l}^{(0)} \in \omega$  and denote the elements of the orbit  $\omega$  by  $\mathbf{l}^{(k)} = \mathbf{l}^{(0)}(A_i^T)^k$  for  $k \in \mathbb{Z}$ . The function*

$$\Psi_i^\perp(\mathbf{x}_i) = \sum_{\mathbf{l} \in [-m, m]^i \cap \omega, \mathbf{l}_i \neq 0} c_{\mathbf{l}} e(\mathbf{l} \cdot \mathbf{x}_i) \in \mathcal{Q}(i) \cap \mathcal{H}_\omega$$

*is a smooth coboundary for  $T_i$  if and only if*

$$\sum_{k=1}^N c_{\mathbf{l}^{(k)}} e\left(-\sum_{j=0}^{k-1} \mathbf{l}^{(j)} \cdot \mathbf{b}_i\right) + c_{\mathbf{l}^{(0)}} + \sum_{k=1}^{N-1} c_{\mathbf{l}^{(-k)}} e\left(\sum_{j=1}^k \mathbf{l}^{(-j)} \cdot \mathbf{b}_i\right) = 0, \quad (3.2)$$

*where  $N \in \mathbb{N}$  is such that  $c_{\mathbf{l}^{(n)}} = 0$  for all  $n \geq N$ .*

*Proof.* There exists a smooth solution  $u$  to the cohomological equation  $\Psi_i^\perp = u \circ T_i - u$  if and only if for every  $\mathbf{l} \in [-m, m]^i \cap \omega$ ,  $l_i \neq 0$  we have

$$\sum_{\mathbf{l} \in [-m, m]^i \cap \omega, l_i \neq 0} c_{\mathbf{l}} e(\mathbf{l} \cdot \mathbf{x}_i) = \sum_{\mathbf{l} \in \omega} u_{\mathbf{l}} e(\mathbf{l} \cdot (\mathbf{x}_i A_i + \mathbf{b}_i)) - \sum_{\mathbf{l} \in \omega} u_{\mathbf{l}} e(\mathbf{l} \cdot \mathbf{x}_i),$$

where  $u_{\mathbf{l}}$  are the Fourier coefficients of  $u$ . Equating coefficients, we get

$$c_{\mathbf{l}} = u_{\mathbf{l}(A_i^T)^{-1}} e(\mathbf{l}(A_i^T)^{-1} \cdot \mathbf{b}_i) - u_{\mathbf{l}}$$

which implies, considering  $\mathbf{l}^{(0)} \in \omega$ ,

$$u_{\mathbf{l}^{(0)}} = u_{\mathbf{l}^{(-1)}} e(\mathbf{l}^{(-1)} \cdot \mathbf{b}_i) - c_{\mathbf{l}^{(0)}} \quad \text{and} \quad u_{\mathbf{l}^{(0)}} = (u_{\mathbf{l}^{(1)}} + c_{\mathbf{l}^{(1)}}) e(-\mathbf{l}^{(0)} \cdot \mathbf{b}_i).$$

Recursively, for all  $N \geq 1$  we obtain

$$\begin{aligned} u_{\mathbf{l}^{(0)}} &= u_{\mathbf{l}^{(-N)}} e\left(\sum_{k=1}^N \mathbf{l}^{(-k)} \cdot \mathbf{b}_i\right) - \sum_{k=1}^{N-1} c_{\mathbf{l}^{(-k)}} e\left(\sum_{j=1}^k \mathbf{l}^{(-j)} \cdot \mathbf{b}_i\right) - c_{\mathbf{l}^{(0)}}, \\ u_{\mathbf{l}^{(0)}} &= u_{\mathbf{l}^{(N)}} e\left(-\sum_{k=0}^{N-1} \mathbf{l}^{(k)} \cdot \mathbf{b}_i\right) + \sum_{k=1}^N c_{\mathbf{l}^{(k)}} e\left(-\sum_{j=0}^{k-1} \mathbf{l}^{(j)} \cdot \mathbf{b}_i\right). \end{aligned}$$

By assumption,  $a_{i,i+1} \neq 0$  and  $l_i \neq 0$ ; hence, for  $|N| \rightarrow \infty$ , we have  $\|\mathbf{l}^{(N)}\|_\infty \geq |l_{i-1} - a_{i,i+1} N l_i| \rightarrow \infty$ . Therefore, if a solution  $u$  exists, we have  $u_{\mathbf{l}^{(N)}} \rightarrow 0$ . We obtain two expressions for  $u_{\mathbf{l}^{(0)}}$

$$u_{\mathbf{l}^{(0)}} = \lim_{N \rightarrow \infty} -c_{\mathbf{l}^{(0)}} - \sum_{k=1}^{N-1} c_{\mathbf{l}^{(-k)}} e\left(\sum_{j=1}^k \mathbf{l}^{(-j)} \cdot \mathbf{b}_i\right), \quad u_{\mathbf{l}^{(0)}} = \lim_{N \rightarrow \infty} \sum_{k=1}^N c_{\mathbf{l}^{(k)}} e\left(-\sum_{j=0}^{k-1} \mathbf{l}^{(j)} \cdot \mathbf{b}_i\right),$$

which, equated, gives (3.2). We remark that the expressions above are finite sums, since there are only finitely many  $k$  such that  $c_{\mathbf{l}^{(k)}} \neq 0$ .

On the other hand, if (3.2) holds, defining  $u_{\mathbf{l}}$  as above gives us the Fourier coefficients of the solution  $u$  to the cohomological equation.  $\square$

**Example 3.4.** Consider, for example, a uniquely ergodic skew shift over  $\mathbb{T}^3$  of the form

$$T(x, y, z) = (x, y, z)A + (b_x, b_y, b_z), \quad \text{with} \quad A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

First, consider the quotient system  $T_2(x, y) = (x, y)A_2 + (b_x, b_y)$ , where  $A_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ . Any function

$$\psi_2(x, y) = c_0 + \sum_{0 \leq |k| \leq m} c_{(k,1)} e(kx + y)$$

which satisfies

$$\sum_{0 \leq |k| \leq m} c_{(k,1)} e\left(\frac{k - k^2}{2} b_x - k b_y\right) \neq 0,$$

induces a mixing suspension flow over the quotient system  $(\mathbb{T}^2, T_2)$ , as shown in [1, §2.4].

Straightforward computations give us

$$\sum_{j=0}^{k-1} (A^T)^j = \sum_{j=0}^{k-1} \begin{pmatrix} 1 & 0 & 0 \\ j & 1 & 0 \\ j^2 + j & 2j & 1 \end{pmatrix} = \begin{pmatrix} k & 0 & 0 \\ -\frac{k^2 - k}{2} & k & 0 \\ \frac{k^3 - k}{3} & k^2 - k & k \end{pmatrix}$$

and

$$\sum_{j=1}^k (A^T)^{-j} = \sum_{j=1}^k \begin{pmatrix} 1 & 0 & 0 \\ -j & 1 & 0 \\ j^2 - j & -2j & 1 \end{pmatrix} = \begin{pmatrix} k & 0 & 0 \\ -\frac{k^2 + k}{2} & k & 0 \\ -\frac{k^3 - k}{3} & -k^2 - k & k \end{pmatrix},$$

for all  $k \geq 1$ . Fix  $\mathbf{l}^{(0)} = (0, 0, 1)$ , then  $\mathbf{l}^{(k)} = (k^2 + k, 2k, 1)$ . By Lemma 3.3, any function

$$\Psi_3^\perp(x, y, z) = \sum_{0 \leq |k| \leq m} c_{(k^2+k, 2k, 1)} e((k^2 + k)x + 2ky + z)$$

satisfying

$$\sum_{0 \leq |k| \leq m} c_{(k^2+k, 2k, 1)} e\left(-\left(\frac{k^3 - k}{3}b_x + (k^2 - k)b_y + kb_z\right)\right) = 0 \quad (3.3)$$

is a smooth coboundary for  $T$ . Proposition 2.1 implies that  $\Psi = \psi_2 + \Psi_3^\perp \in \mathcal{M}(3)$  induces a mixing suspension flow over  $(\mathbb{T}^3, T)$ .

#### 4. PROOF OF PROPOSITION 2.1

We show that, under the assumption of Proposition 2.1, if the quotient suspension flow  $\{\widehat{T}_t^\psi\}_{t \in \mathbb{R}}$  is mixing, then  $\{T_t^{\psi \circ \pi}\}_{t \in \mathbb{R}}$  is mixing.

**4.1. Preliminaries.** Let us denote  $\Psi := \psi \circ \pi$ , which we remark is constant along the  $x_d$ -coordinate, and assume  $\int \Psi = 1$ . Let  $c$  and  $C$  be its minimum and maximum respectively. Consider  $Q = \prod_{j=1}^d [w_j, w'_j] \times [q_1, q_2]$  and  $R = \prod_{j=1}^d [v_j, v'_j] \times [r_1, r_2]$  two cubes in  $\{(\mathbf{x}, r) : \mathbf{x} \in \mathbb{T}^d \text{ and } 0 \leq r < \Psi(\mathbf{x})\}$ ; it is sufficient to prove mixing for sets of this form. Denote by  $\widehat{Q} = \pi(Q)$ ,  $\widehat{R} = \pi(R)$  the corresponding cubes in the quotient system, namely  $\widehat{Q} = \prod_{j=1}^{d-1} [w_j, w'_j] \times [q_1, q_2]$  and  $\widehat{R} = \prod_{j=1}^{d-1} [v_j, v'_j] \times [r_1, r_2]$ . For any  $\varepsilon > 0$ , define

$$\begin{aligned} \widehat{Q}_{-\varepsilon} &= \prod_{j=1}^{d-1} [w_j + \varepsilon, w'_j - \varepsilon] \times [q_1 + \varepsilon, q_2 - \varepsilon] \subset \widehat{Q}, \\ \widehat{R}_{-\varepsilon} &= \prod_{j=1}^{d-1} [v_j + \varepsilon, v'_j - \varepsilon] \times [r_1 + \varepsilon, r_2 - \varepsilon] \subset \widehat{R}. \end{aligned}$$

Let  $\mathbf{v} \in \mathbb{Z}^d$  be such that  $\mathbf{v}A = \mathbf{v} + a\mathbf{e}_d$ , with  $a \neq 0$ . Up to changing  $\mathbf{v}$  with  $\mathbf{v} - (\mathbf{v} \cdot \mathbf{e}_d)\mathbf{e}_d$  and up to rescaling, we can assume that  $\mathbf{v} \cdot \mathbf{e}_d = 0$  and the coordinates of  $\mathbf{v}$  are coprime. Denote by  $\partial_{\mathbf{v}}$  the directional derivative along  $\mathbf{v}$ , namely, if  $f \in \mathcal{C}^1(\mathbb{T}^d)$ , let  $\partial_{\mathbf{v}}f = \nabla f \cdot \mathbf{v}$ , where  $\nabla f$  is the gradient of  $f$ . Fix  $\varepsilon > 0$  and choose  $0 < \varepsilon_0 < 1$  such that  $3(dC + 1)\varepsilon_0 < \varepsilon$ . Recalling (1.2), let  $S_n(\partial_{\mathbf{v}}\Psi)$  be the Birkhoff sum up to  $n$  of the derivative of  $\Psi$  along  $\mathbf{v}$ . By Birkhoff Ergodic Theorem, since  $T$  is uniquely ergodic and  $\partial_{\mathbf{v}}\Psi$  has zero average, there exists  $N \geq 1$  such that for all  $n \geq N$  we have

$$\frac{1}{n} S_n(\partial_{\mathbf{v}}\Psi)(\mathbf{x}) \leq \frac{ac}{2C}\varepsilon_0, \quad (4.1)$$

for all  $\mathbf{x} \in \mathbb{T}^d$ .

For every  $\mathbf{x}, \mathbf{x}' \in \mathbb{T}^d$ , from the definition (1.3) of  $n_t(\mathbf{x})$  it follows immediately that

$$n_t(\mathbf{x})c \leq S_{n_t(\mathbf{x})}(\Psi)(\mathbf{x}) \leq t < S_{n_t(\mathbf{x}')+1}(\Psi)(\mathbf{x}') \leq (n_t(\mathbf{x}') + 1)C \leq 2n_t(\mathbf{x}')C,$$

for all  $t > C$ . Then, we have  $n_t(\mathbf{x})/n_t(\mathbf{x}') \leq 2C/c$ . Choose  $\bar{t} > 0$  such that for all  $t \geq \bar{t}$

$$\begin{aligned} \text{(i)} \quad & t \geq (N + 1)C \quad \text{so that} \quad n_t(\mathbf{x}) > t/C - 1 \geq N; \\ \text{(ii)} \quad & \frac{\|\mathbf{v}\|C}{a(t - C)} \leq \varepsilon_0 \quad \text{so that} \quad \frac{\|\mathbf{v}\|}{n_t(\mathbf{x})a} < \frac{\|\mathbf{v}\|C}{a(t - C)} \leq \varepsilon_0; \\ \text{(iii)} \quad & \left| \text{Leb}\left(\widehat{T}_t^\psi(\widehat{R}_{-\varepsilon_0}) \cap \widehat{Q}_{-\varepsilon_0}\right) - \text{Leb}(\widehat{R}_{-\varepsilon_0}) \text{Leb}(\widehat{Q}_{-\varepsilon_0}) \right| \leq \varepsilon_0. \end{aligned} \quad (4.2)$$

The third condition above is guaranteed by mixing of the suspension flow  $\{\widehat{T}_t^\psi\}_{t \in \mathbb{R}}$  on the quotient  $\mathbb{T}^{d-1}$ .

**4.2. Wrapping segments.** We now consider segments of length less than  $\varepsilon_0$  parallel to  $\mathbf{v}$  contained in  $R$  and we study their evolution after sufficiently large time  $t$ . Recalling (2.1), fix  $t \geq \bar{t}$  and consider a point  $\mathbf{r} = (\mathbf{x}, r) \in R$  such that  $\mathbf{r} + (n_t(\mathbf{r})a)^{-1}\mathbf{v} \in R$ . Let  $\gamma_{\mathbf{r}}(s) = \mathbf{r} + s\mathbf{v}$ , with  $0 \leq s \leq \bar{s} = (n_t(\mathbf{r})a)^{-1}$ , be the segment parallel to  $\mathbf{v}$  starting from  $\mathbf{r}$  of length  $\bar{s}$ . Condition (4.2)-(ii) ensures that the length of  $\gamma_{\mathbf{r}}$  is less than  $\varepsilon_0$  so that, by hypothesis, it is all contained in  $R$ , see Figure 1. We will prove that, if there exists a point of  $\pi \circ T_t^\Psi(\gamma_{\mathbf{r}})$  which is contained in  $\widehat{Q}_{-\varepsilon_0}$ , then all the curve is contained in  $\widehat{Q}$ .

Let us denote by  $\Gamma_{\mathbf{r}}(s) = T_t^\Psi(\gamma_{\mathbf{r}}(s))$  the image of  $\gamma_{\mathbf{r}}(s)$  under  $T_t^\Psi$  and let us compute its tangent vector  $\partial_s \Gamma_{\mathbf{r}}(s)$  at a generic point. For almost every  $s$ , the value  $n_t(\gamma_{\mathbf{r}}(s))$  is locally constant; from the definition (1.3), we get

$$\begin{aligned} \partial_s \Gamma_{\mathbf{r}}(s) &= \partial_s \left( T^{n_t(\gamma_{\mathbf{r}}(s))} \gamma_{\mathbf{r}}(s), t - S_{n_t(\gamma_{\mathbf{r}}(s))}(\Psi)(\gamma_{\mathbf{r}}(s)) \right) \\ &= \left( \mathbf{v} A^{n_t(\gamma_{\mathbf{r}}(s))}, - \sum_{j=0}^{n_t(\gamma_{\mathbf{r}}(s))-1} \nabla \Psi \circ T^j(\gamma_{\mathbf{r}}(s)) \cdot \mathbf{v} A^j \right) \\ &= \left( \mathbf{v} + n_t(\gamma_{\mathbf{r}}(s)) a \mathbf{e}_d, - \sum_{j=0}^{n_t(\gamma_{\mathbf{r}}(s))-1} (\partial_{\mathbf{v}} \Psi) \circ T^j(\gamma_{\mathbf{r}}(s)) \right), \end{aligned} \quad (4.3)$$

where we used the fact that the partial derivative of  $\Psi = \psi \circ \pi$  in the  $d$ -th variable is zero, since  $\Psi$  is constant along the  $x_d$ -coordinate.

We first show that the function  $s \mapsto n_t(\gamma_{\mathbf{r}}(s))$  is constant. In order to do this, we estimate the maximal distance in the  $t$ -coordinate between two points in the curve  $\Gamma_{\mathbf{r}}(s)$ . By definition and (4.3), it equals

$$\max_{0 \leq s', s'' \leq \bar{s}} |(\Gamma_{\mathbf{r}}(s') - \Gamma_{\mathbf{r}}(s'')) \cdot \mathbf{e}_{d+1}| \leq \int_0^{\bar{s}} \left| \sum_{j=0}^{n_t(\gamma_{\mathbf{r}}(s))-1} (\partial_{\mathbf{v}} \Psi) \circ T^j(\gamma_{\mathbf{r}}(s)) \right| ds.$$

From the choice of  $N$ , (4.1) and (4.2)-(i), it follows

$$\begin{aligned} \max_{0 \leq s', s'' \leq \bar{s}} |(\Gamma_{\mathbf{r}}(s') - \Gamma_{\mathbf{r}}(s'')) \cdot \mathbf{e}_{d+1}| &\leq \int_0^{\bar{s}} |S_{n_t(\gamma_{\mathbf{r}}(s))}(\partial_{\mathbf{v}} \Psi)(\gamma_{\mathbf{r}}(s))| ds \\ &\leq \frac{ac}{2C} \varepsilon_0 \int_0^{\bar{s}} n_t(\gamma_{\mathbf{r}}(s)) ds \leq \frac{c}{2C} \varepsilon_0 \frac{\max_s n_t(\gamma_{\mathbf{r}}(s))}{n_t(\mathbf{r})} \leq \varepsilon_0. \end{aligned}$$

In a similar way, using (4.2)-(ii), the maximal distance in any other coordinate  $x_i$  for  $1 \leq i \leq d-1$  between two points in  $\Gamma_{\mathbf{r}}(s)$  can be bounded by

$$\max_{0 \leq s', s'' \leq \bar{s}} |(\Gamma_{\mathbf{r}}(s') - \Gamma_{\mathbf{r}}(s'')) \cdot \mathbf{e}_i| \leq \|\mathbf{v}\| \int_0^{\bar{s}} ds = \frac{\|\mathbf{v}\|}{n_t(\mathbf{r})a} \leq \varepsilon_0.$$

In particular, if  $\pi(\gamma_{\mathbf{r}}(s)) \in \widehat{T}_{-t}^\psi(\widehat{Q}_{-\varepsilon_0})$  for some  $0 \leq s \leq \bar{s}$ , then  $\pi \circ \Gamma_{\mathbf{r}}(s) \subset \widehat{Q}$  and therefore we deduce that  $n_t$  is constant along  $\gamma_{\mathbf{r}}(s)$  and equal to  $n_t(\mathbf{r})$ , see Figure 1.

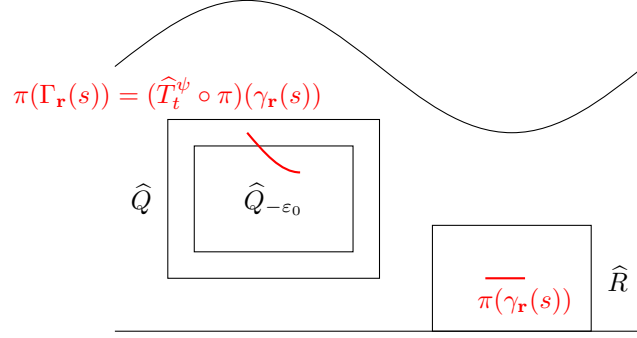
Since  $n_t(\gamma_{\mathbf{r}}(s)) = n_t(\mathbf{r})$ , by (4.3) the speed in the  $x_d$ -coordinate is constant and equal to  $n_t(\mathbf{r})a$ . Moreover, the distance in the  $x_d$ -coordinate of the endpoints of  $\Gamma_{\mathbf{r}}(s)$  is equal to

$$\left| \int_0^{\bar{s}} n_t(\gamma_{\mathbf{r}}(s)) a ds \right| = \left| \int_0^{\bar{s}} n_t(\mathbf{r}) a ds \right| = |\bar{s} n_t(\mathbf{r}) a| = 1.$$

**4.3. Final estimates.** In order to estimate the measure of  $R \cap T_{-t}^\Psi(Q)$ , we want to apply Fubini's Theorem and integrate along each circle parallel to  $\mathbf{v}$ . Indeed, the torus  $\mathbb{T}^d$  is a circle bundle over a closed submanifold  $W$  isomorphic to a  $(d-1)$ -dimensional torus with fibers parallel to  $\mathbf{v} \in \mathbb{Z}^d$ . Let us consider the corresponding decomposition of the Lebesgue measure as product measure, namely  $d \text{Leb}_d = d\mathbf{v} \wedge \omega$ , where  $\omega$  is a volume form over  $W$ .

Let

$$S = \left\{ \mathbf{r} \in R : \pi(\mathbf{r}) \in \widehat{R}_{-\varepsilon_0} \cap \widehat{T}_{-t}^\psi(\widehat{Q}_{-\varepsilon_0}) \right\}.$$



**Figure 1.** Quotient system  $(\mathbb{T}^{d-1}, \widehat{T})$ : if some point of the curve  $\pi(\Gamma_{\mathbf{r}}(s))$  is contained in  $\widehat{Q}_{-\varepsilon_0}$ , then the whole curve is contained in  $\widehat{Q}$ .

We want to consider all segments  $\gamma_{\mathbf{r}}(s)$  that contain at least one point in  $S$ . For any point  $\mathbf{w} \in W$ , let us partition the fiber  $\mathbf{w} + [0, 1)\mathbf{v}$  into segments of length  $\bar{s}$ ; more precisely define

$$\mathbf{r}_0(\mathbf{w}) = \mathbf{w}, \quad \mathbf{r}_1(\mathbf{w}) = \mathbf{r}_0 + (n_t(\mathbf{r}_0)a)^{-1}\mathbf{v}, \dots, \quad \mathbf{r}_{i+1}(\mathbf{w}) = \mathbf{r}_i + (n_t(\mathbf{r}_i)a)^{-1}\mathbf{v},$$

up to the largest  $i$  such that  $\sum_i (n_t(\mathbf{r}_i)a)^{-1} < 1$ , and let  $R_-(t)$  be the union for  $\mathbf{w} \in W$  of all segments  $\gamma_{\mathbf{r}_i(\mathbf{w})}(s)$  which contain at least one point in  $S$ . Notice that  $\text{Leb}_d(S) \leq \text{Leb}_d(R_-(t))$ ; moreover, recalling the definition of  $R$ , by Fubini's Theorem,

$$\begin{aligned} \text{Leb}_d(S) &= \text{Leb}_{d-1} \left( \widehat{R}_{-\varepsilon_0} \cap \widehat{T}_{-t}^\psi(\widehat{Q}_{-\varepsilon_0}) \right) |v'_d - v_d|, \\ \text{Leb}_d(R_-(t)) &= \int_{\mathbb{T}^d} \mathbb{1}_{R_-(t)} d\mathbf{v} \wedge \omega = \int_W \left( \sum_{i: \gamma_{\mathbf{r}_i(\mathbf{w})} \subset R_-(t)} (n_t(\mathbf{r}_i)a)^{-1} \right) \omega(\mathbf{w}). \end{aligned} \quad (4.4)$$

By definition of  $R_-(t)$  and  $S$ , we have  $R_-(t) \subset R$ , thus

$$\begin{aligned} \text{Leb}_d(T_t^\Psi(R) \cap Q) &= \int_{\mathbb{T}^d} (\mathbb{1}_R \circ T_{-t}^\Psi) \cdot \mathbb{1}_Q d\text{Leb}_d \geq \int_{\mathbb{T}^d} \mathbb{1}_{R_-(t)} \cdot (\mathbb{1}_Q \circ T_t^\Psi) d\text{Leb}_d \\ &= \int_W \left( \sum_{i: \gamma_{\mathbf{r}_i(\mathbf{w})} \subset R_-(t)} \int_0^{\bar{s}} \mathbb{1}_Q \circ T_t^\Psi \circ \gamma_{\mathbf{r}_i(\mathbf{w})}(s) ds \right) \omega(\mathbf{w}). \end{aligned} \quad (4.5)$$

For each curve  $\gamma_{\mathbf{r}_i(\mathbf{w})} \subset R_-(t)$ , by definition of  $R_-(t)$ , there exists a point  $\gamma_{\mathbf{r}_i(\mathbf{w})}(s)$  contained in  $S$ , so that  $\pi(\Gamma_{\mathbf{r}_i(\mathbf{w})}(s)) \in \widehat{Q}$ . Hence, the point  $\Gamma_{\mathbf{r}_i(\mathbf{w})}(s) \in Q$  if and only if its  $x_d$ -coordinate  $\Gamma_{\mathbf{r}_i(\mathbf{w})}(s) \cdot \mathbf{e}_d$  is in  $[w_d, w'_d]$ . Since the speed of  $\Gamma_{\mathbf{r}_i(\mathbf{w})}$  in this latter direction is constant and equal to  $\bar{s}^{-1} = n_t(\mathbf{r}_i(\mathbf{w}))a$ , we get

$$\int_0^{\bar{s}} \mathbb{1}_Q \circ T_t^\Psi \circ \gamma_{\mathbf{r}_i(\mathbf{w})}(s) ds = \bar{s} |w'_d - w_d|. \quad (4.6)$$

Combining (4.6) with (4.5) and (4.4), we obtain

$$\begin{aligned} \text{Leb}_d(T_t^\Psi(R) \cap Q) &\geq \text{Leb}_d(R_-(t)) |w'_d - w_d| \geq \text{Leb}_d(S) |w'_d - w_d| \\ &= \text{Leb}_{d-1} \left( \widehat{R}_{-\varepsilon_0} \cap \widehat{T}_{-t}^\psi(\widehat{Q}_{-\varepsilon_0}) \right) |v'_d - v_d| |w'_d - w_d|. \end{aligned}$$

The area of a face of  $Q$  is less than  $C = \max \Psi > 1$ , thus we can bound  $\text{Leb}_{d-1}(\widehat{Q}_{-\varepsilon_0}) \geq \text{Leb}_{d-1}(\widehat{Q}) - (3d)C\varepsilon_0$ , where  $3d$  is the number of faces of  $Q$ . Using (4.2)-(iii), we get

$$\begin{aligned} \text{Leb}_d(T_t^\Psi(R) \cap Q) &\geq (\text{Leb}_{d-1}(\widehat{R}_{-\varepsilon_0}) \text{Leb}_{d-1}(\widehat{Q}_{-\varepsilon_0}) - \varepsilon_0) |v'_d - v_d| |w'_d - w_d| \\ &\geq ((\text{Leb}_{d-1}(\widehat{R}) - 3dC\varepsilon_0)(\text{Leb}_{d-1}(\widehat{Q}) - 3dC\varepsilon_0) - \varepsilon_0) |v'_d - v_d| |w'_d - w_d| \\ &\geq \text{Leb}_d(R) \text{Leb}_d(Q) - 3dC(\text{Leb}_d(R) + \text{Leb}_d(Q))\varepsilon_0 - \varepsilon_0 \geq \text{Leb}_d(R) \text{Leb}_d(Q) - \varepsilon, \end{aligned}$$

by the choice of  $\varepsilon$ . The other inequality can be derived in a similar way: one considers  $R_+(t)$  instead of  $R_-(t)$ , where  $R_+(t)$  is defined analogously to  $R_-(t)$  as the union of the segments  $\gamma_{\mathbf{r}}(s)$  which contain at least one point that belongs to  $S' = R \cap \pi^{-1}(\widehat{T}_{-t}^\Psi(\widehat{Q}))$ ; then, one notices that

$$R_+(t) \subset R_{+\varepsilon} \cap \pi^{-1}(\widehat{T}_{-t}^\Psi(\widehat{Q}_{+\varepsilon})),$$

where

$$\widehat{Q}_{+\varepsilon} = \prod_{j=1}^{d-1} [w_j - \varepsilon, w'_j + \varepsilon] \times [q_1 - \varepsilon, q_2 + \varepsilon] \supset \widehat{Q},$$

and similarly for  $R_{+\varepsilon}$ . Finally, it is sufficient to estimate  $\text{Leb}_d(R \cap T_{-t}^\Psi(Q)) = \text{Leb}_d(S' \cap T_{-t}^\Psi(Q)) \leq \text{Leb}_d(R_+(t) \cap T_{-t}^\Psi(Q_{+\varepsilon}))$  by applying Fubini's Theorem as above. The proof is therefore complete.

## 5. PROOF OF THEOREM 2.4

We now suppose that  $E_k = \text{Im}(A - \text{Id})^k = \langle \mathbf{e}_d \rangle$  and  $\text{Im}(A - \text{Id})^{k+1} = \{0\}$ . Let  $\Psi^\perp \in \mathcal{Q}(d)$  and, denoting  $e(x) = \exp(2\pi i x)$ , write

$$\Psi^\perp(\mathbf{x}) = \sum_{\mathbf{l} \in [-m, m]^d \cap \mathbb{Z}^d} c_{\mathbf{l}} e(\mathbf{l} \cdot \mathbf{x}). \quad (5.1)$$

Let us assume that  $\Psi^\perp \in \mathcal{Q}(d)$  is not a measurable coboundary for  $T$ ; we prove that Birkhoff sums  $S_n(\Psi^\perp)$  of  $\Psi^\perp$  grow in measure. In order to do this, we first apply a classical Gottschalk-Hedlund argument to prove that they grow in average (Lemma 5.1) and then a decoupling result (Lemma 5.2), which generalizes [1, Lemma 5] to higher dimension. The key observation is that, due to the form of the skew-translation  $T$ , for large  $N \geq 1$  the divergence of nearby points happens mostly in the  $x_d$ -direction, namely it is of higher order than in the other coordinates.

Denote by  $\widehat{\mathbf{x}} := \pi(\mathbf{x}) \in \mathbb{T}^{d-1}$  the projection of  $\mathbf{x} \in \mathbb{T}^d$  onto the first  $d-1$  coordinates; the projection  $\pi$  gives a factor  $(\mathbb{T}^{d-1}, \widehat{T})$  of  $(\mathbb{T}^d, T)$ .

*Remark 2.* For any  $N \geq 1$ , we can express the  $N$ -th iterate of  $T$  as  $T^N \mathbf{x} = \mathbf{x}A^N + \mathbf{b}(N)$ , where  $\mathbf{b}(N) = (b_1(N), \dots, b_d(N)) = \sum_{i=0}^{N-1} \mathbf{b}A^i$  and  $A^N = (a_{i,j}(N))_{i,j}$  is an upper triangular unipotent matrix. For any  $N \geq k+1$ , we can write  $A^N = (\text{Id} + (A - \text{Id}))^N = \sum_{i=0}^k \binom{N}{i} (A - \text{Id})^i$ . It follows that each nonzero entry  $a_{i,j}(N)$  is a polynomial in  $N$  of degree  $\leq k$ . Moreover, since  $E_k = \langle \mathbf{e}_d \rangle$ , the only terms  $a_{i,j}(N)$  of order  $O(N^k)$  are in the last column, namely for  $j = d$ . With this notation, we have

$$T^N \mathbf{x} = T^N(\widehat{\mathbf{x}}, x_d) = (\widehat{T}^N \widehat{\mathbf{x}}, x_d + x_{d-1}a_{d-1,d}(N) + \dots + x_1 a_{1,d}(N) + b_d(N)).$$

**Lemma 5.1** ([1, Corollary 1]). *For any  $C > 1$  we have*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \text{Leb}_d(|S_n(\Psi^\perp)| < C) = 0. \quad (5.2)$$

*In particular, for any  $\varepsilon > 0$  there exist arbitrarily long arithmetic progressions  $\{i\overline{m}\}_{i=1}^\ell$  such that  $\text{Leb}_d(|S_{i\overline{m}}(\Psi^\perp)| < C) < \varepsilon$ .*

*Proof.* The proof of the first statement is the same as in [1, Corollary 1]. For the reader's convenience, we present a proof of the second part. Fix  $\varepsilon > 0$  and let

$$B_\varepsilon = \{n \in \mathbb{N} : \text{Leb}_d(|S_n(\Psi^\perp)| < C) \geq \varepsilon\} \subset \mathbb{N}.$$

By (5.2),  $B_\varepsilon$  has zero density, see, e.g., [2, Theorem 2.8.1].

Let us consider  $\ell \geq 1$ ,  $0 < \delta < 2/(\ell^2 + \ell)$  and  $N_0 \geq 1$  such that for all  $N \geq N_0$  we have  $\#\{n \in B_\varepsilon : n \leq N\} \leq \delta N$ . Fix  $N \geq N_0$ ; we want to find  $\bar{n} \leq N$  such that  $\bar{n}, 2\bar{n}, \dots, \ell\bar{n} \in \mathbb{N} \setminus B_\varepsilon$ . Equivalently, if we denote by  $B_\varepsilon/j := \{b/j : b \in B_\varepsilon\} \subset \mathbb{Q}$ , we look for  $1 \leq \bar{n} \leq N$  such that

$$\bar{n} \notin \{1, \dots, N\} \cap \frac{B_\varepsilon}{j} \quad \text{for all } j = 1, \dots, \ell.$$

We estimate the cardinality

$$\begin{aligned} \#\left(\{1, \dots, N\} \setminus \bigcup_{j=1}^{\ell} \{1, \dots, N\} \cap \frac{B_\varepsilon}{j}\right) &\geq N - \sum_{j=1}^{\ell} \#\left(\{1, \dots, N\} \cap \frac{B_\varepsilon}{j}\right) \\ &\geq N - \sum_{j=1}^{\ell} \#\left(\{1, \dots, jN\} \cap B_\varepsilon\right) \geq N \left(1 - \frac{\ell(\ell+1)}{2}\delta\right) > 0, \end{aligned}$$

by the choice of  $\delta$ . In particular, the set  $\{1 \leq \bar{n} \leq N : j\bar{n} \notin B_\varepsilon, \text{ for } j = 1, \dots, \ell\}$  is not empty and the claim follows.  $\square$

**5.1. Decoupling.** The following is our decoupling result.

**Lemma 5.2.** *Let  $C > 1$  and  $\varepsilon > 0$ . There exist  $C' > 1$  and  $\varepsilon' > 0$  such that for all  $n \geq 1$  satisfying  $\text{Leb}_d(|S_n(\Psi^\perp)| < C') < \varepsilon'$  there exists  $N_0 \geq 1$  such that for all  $N \geq N_0$  we have*

$$\text{Leb}_d(|S_N(\Psi^\perp) \circ T^n - S_N(\Psi^\perp)| < 2C) < \varepsilon. \quad (5.3)$$

*Proof.* First of all, by the cocycle relation for Birkhoff sums, we notice that  $S_N(\Psi^\perp) \circ T^n - S_N(\Psi^\perp) = S_{N+n}(\Psi^\perp) - S_n(\Psi^\perp) - S_N(\Psi^\perp) = S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp)$ . We want to compare  $|S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp)|$  with  $|S_n(\Psi^\perp)|$ , which, by hypothesis, is larger than  $C'$  up to a set of measure at most  $\varepsilon'$ , the latter constants still to be determined.

We denote by  $\mathbf{a}_j(N)$  the transpose of the  $j$ -th column of  $A^N$  and the translation vector by  $\mathbf{b}(N) = (b_1(N), \dots, b_d(N))$ . Let  $\hat{\mathbf{a}}_d(N) = \pi(\mathbf{a}_d(N)) = (a_{1,d}(N), \dots, a_{d-1,d}(N))$  be the vector obtained from  $\mathbf{a}_d(N)$  by suppressing the last coordinate  $a_{d,d}(N) = 1$ .

From (5.1), write

$$\Psi^\perp(\mathbf{x}) = \sum_{0 < |l| \leq m} c_l(\hat{\mathbf{x}}) e(lx_d).$$

Using Remark 2, we can express the Birkhoff sum of  $\Psi^\perp$  as

$$S_n(\Psi^\perp)(\mathbf{x}) = \sum_{r=0}^{n-1} \sum_{0 < |l| \leq m} c_l(\hat{T}^r \hat{\mathbf{x}}) e(l(x_d + \hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d(r) + b_d(r))) = \sum_{0 < |l| \leq m} c_{l,n}(\hat{\mathbf{x}}) e(lx_d), \quad (5.4)$$

where we have denoted

$$c_{l,n}(\hat{\mathbf{x}}) = \sum_{r=0}^{n-1} c_l(\hat{T}^r \hat{\mathbf{x}}) e(l(\hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d(r) + b_d(r))).$$

Therefore, we can write

$$(S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp))(\mathbf{x}) = \sum_{0 < |l| \leq m} c_{l,n,N}(\hat{\mathbf{x}}) e(lx_d),$$

where

$$c_{l,n,N}(\hat{\mathbf{x}}) = c_{l,n}(\hat{T}^N \hat{\mathbf{x}}) e(l(\hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d(N) + b_d(N))) - c_{l,n}(\hat{\mathbf{x}}). \quad (5.5)$$

We will now estimate the measure of the set where the modulus of the coefficients  $c_{l,n,N}$  is comparable to  $c_{l,n}$ . The idea is the following: we first partition  $\mathbb{T}^{d-1}$  into sets on which the

coefficients  $c_{l,n}$  and  $c_{l,n} \circ \widehat{T}^N$  are almost constant. We then show that on a large set there are no cancellations for  $c_{l,n,N}$  by using the fact that the factor  $e(l(\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N)))$  is of higher order, namely  $O(N^k)$ .

Let  $n \geq 1$  be fixed. The functions  $c_{l,n}$  are uniformly continuous, hence let  $\delta > 0$  be such that if  $\|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\| \leq \delta$  then  $|c_{l,n}(\widehat{\mathbf{x}}) - c_{l,n}(\widehat{\mathbf{x}}')| \leq 1/4$ . By Remark 2,

$$\|\widehat{T}^N \widehat{\mathbf{x}} - \widehat{T}^N \widehat{\mathbf{x}}'\|_\infty \leq \|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\|_\infty \|\widehat{A}^N\|_\infty = \|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\|_\infty O(N^{k-1}).$$

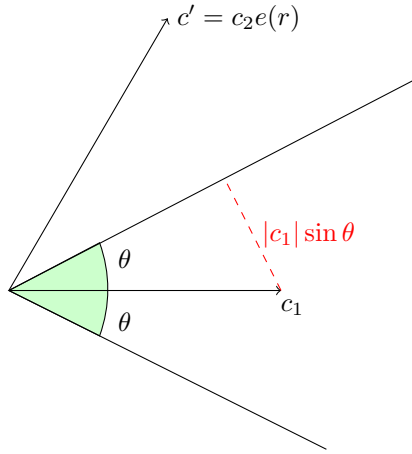
Let  $N_0 \geq 1$  be such that for all  $N \geq N_0$ , if  $\|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\|_\infty \leq (N^{k-1} \log N)^{-1}$  then the term above is less than  $\delta$ , so that

$$\left| c_{l,n}(\widehat{T}^N \widehat{\mathbf{x}}) - c_{l,n}(\widehat{T}^N \widehat{\mathbf{x}}') \right| \leq 1/4. \quad (5.6)$$

Partition  $\mathbb{T}^{d-1}$  into cubes with edges of length  $L = (N^{k-1} \log N \sqrt{d-1})^{-1}$  and one face  $F$  orthogonal to  $\widehat{\mathbf{a}}_d(N)$ . If  $\widehat{\mathbf{x}}$  and  $\widehat{\mathbf{x}}'$  are in one of such cubes, which we will denote by  $Q$ , then  $\|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\|_\infty \leq \sqrt{d-1}L$  and so (5.6) holds. Fix  $Q$  and let  $\overline{\mathbf{x}}$  be one of its vertices. Let  $c_1 = c_{l,n}(\overline{\mathbf{x}})$  and  $c_2 = c_{l,n}(\widehat{T}^N \overline{\mathbf{x}})e(lb_d(N))$ ; then for any  $\widehat{\mathbf{x}} \in Q$ , by (5.5) and (5.6),

$$\begin{aligned} |c_{l,n,N}(\widehat{\mathbf{x}})| &\geq \left| c_{l,n}(\widehat{T}^N \overline{\mathbf{x}})e(l(\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N) + b_d(N))) - c_{l,n}(\overline{\mathbf{x}}) \right| \\ &\quad - \left| c_{l,n}(\widehat{T}^N \widehat{\mathbf{x}}) - c_{l,n}(\widehat{T}^N \overline{\mathbf{x}}) \right| \cdot |e(l(\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N) + b_d(N)))| - |c_{l,n}(\widehat{\mathbf{x}}) - c_{l,n}(\overline{\mathbf{x}})| \\ &\geq |c_2 e(l\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N)) - c_1| - \frac{1}{2}. \end{aligned}$$

Call  $\theta_1, \theta_2$  the argument of  $c_1, c_2 \in \mathbb{C}$  respectively; fix  $\theta \in (0, \frac{\pi}{2})$ . If  $r \in \mathbb{R}$  is such that  $\theta_2 + 2\pi r \notin [\theta_1 - \theta, \theta_1 + \theta] + 2\pi\mathbb{Z}$ , then  $|c_2 e(r) - c_1| > |c_1| \sin \theta$ , see Figure 2.



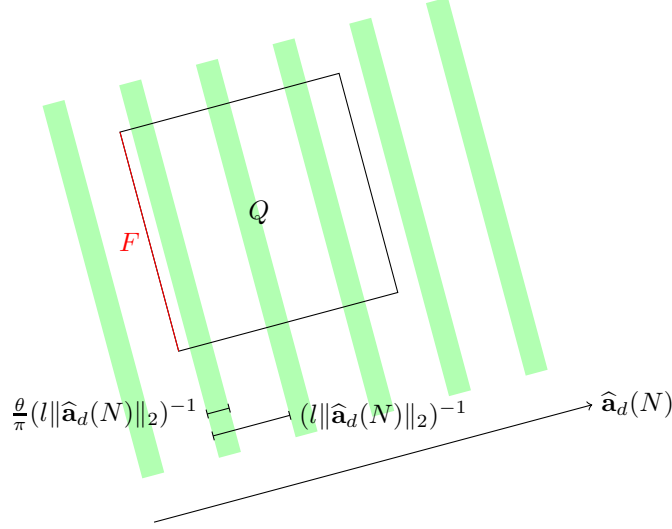
**Figure 2.** Any point  $c' \in \mathbb{C}$  outside the cone of  $1/2$ -angle  $\theta$  about the line  $\mathbb{R}c_1$  has distance from  $c_1$  larger than the distance of  $c_1$  from the boundary of the cone.

Thus, in our case,  $|c_2 e(l\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N)) - c_1| \leq |c_1| \sin \theta$  implies  $\theta_2 + (2\pi l)\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N) \in [\theta_1 - \theta, \theta_1 + \theta] + 2\pi\mathbb{Z}$ ; in particular,  $l\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N)$  belongs to an interval mod  $\mathbb{Z}$  of size  $\theta/\pi$ . The level sets of the linear functional  $\widehat{\mathbf{x}} \mapsto (2\pi l)\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N)$  are affine  $(d-2)$ -dimensional sets orthogonal to  $\widehat{\mathbf{a}}_d(N)$  and hence parallel to a face  $F$  of  $Q$ , see Figure 3.

Therefore,

$$\text{Leb} \left( \widehat{\mathbf{x}} \in Q : \theta_2 + (2\pi l)\widehat{\mathbf{x}} \cdot \widehat{\mathbf{a}}_d(N) \in [\theta_1 - \theta, \theta_1 + \theta] + 2\pi\mathbb{Z} \right) \leq \text{Leb}(F) \frac{\theta}{\pi} \left( L + \frac{1}{l \|\widehat{\mathbf{a}}_d(N)\|_2} \right).$$





**Figure 3.** In color, the set of  $\mathbf{x}$  such that  $\theta_2 + 2\pi l\hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d \in [\theta_1 - \theta, \theta_1 + \theta] + 2\pi\mathbb{Z}$ .

By Remark 2,  $\|\hat{\mathbf{a}}_d(N)\|_2 = O(N^k)$ ; since  $L = O(1/(N^{k-1} \log N))$ , we get

$$\begin{aligned} \text{Leb}(\hat{\mathbf{x}} \in Q : |c_2 e(l\hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d(N)) - c_1| \leq |c_1| \sin \theta) &\leq \frac{\theta}{\pi} \text{Leb}(Q) \left(1 + \frac{1}{l\|\hat{\mathbf{a}}_d(N)\|_2 L}\right) \\ &= \frac{\theta}{\pi} \text{Leb}(Q) \left(1 + O\left(\frac{\log N}{N}\right)\right). \end{aligned}$$

On the complement of this set,

$$|c_{l,n,N}(\hat{\mathbf{x}})| \geq |c_2 e(l\hat{\mathbf{x}} \cdot \hat{\mathbf{a}}_d(N)) - c_1| - \frac{1}{2} > |c_1| \sin \theta - \frac{1}{2} \geq |c_{l,n}(\hat{\mathbf{x}})| \sin \theta - \frac{3}{4};$$

hence

$$\begin{aligned} \limsup_{N \rightarrow \infty} \text{Leb} \left( |c_{l,n,N}(\hat{\mathbf{x}})| \leq |c_{l,n}(\hat{\mathbf{x}})| \sin \theta - \frac{3}{4} \right) &\leq \limsup_{N \rightarrow \infty} \sum_{Q \subset \Sigma} \left(1 + O\left(\frac{\log N}{N}\right)\right) \frac{\theta}{\pi} \text{Leb}(Q) \\ &= \limsup_{N \rightarrow \infty} \frac{\theta}{\pi} \left(1 + O\left(\frac{\log N}{N}\right)\right) = \frac{\theta}{\pi}. \end{aligned}$$

We have obtained an estimate of the measure of the set where the coefficients  $c_{l,n,N}$  are small compared to  $c_{l,n}$ ; outside this set we can estimate  $|S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp)|$  thanks to the hypothesis on  $|S_n(\Psi^\perp)|$  as follows.

Let us add all these estimates as  $0 < |l| \leq m$ , where we recall  $m$  is the degree of the trigonometric polynomial  $\Psi^\perp$ . Choose  $C' \geq 9m^2$ ; pick  $\theta \in (0, \frac{\pi}{2})$  such that  $1/\sqrt{C'} \leq \sin \theta \leq \sqrt{2/C'}$ . Clearly,  $\theta/\pi < \sin(\theta/2) = \sqrt{(1 - \cos \theta)/2} \leq (\sin \theta)/\sqrt{2} \leq 1/\sqrt{C'}$ . Outside a set of measure at most  $2m(\theta/\pi) \leq 2m/\sqrt{C'}$ , we have

$$\sum_{0 < |l| \leq m} |c_{l,n,N}(\hat{\mathbf{x}})| \geq \sum_{0 < |l| \leq m} |c_{l,n}(\hat{\mathbf{x}})| \sin \theta - \frac{6m}{4} \geq \left| \sum_{0 < |l| \leq m} c_{l,n}(\hat{\mathbf{x}}) e(lx_d) \right| \frac{1}{\sqrt{C'}} - \frac{\sqrt{C'}}{2}.$$

We apply the following result.

**Lemma 5.3** ([1, Lemma 4]). *For each  $m \geq 1$  and for any norm  $\|\cdot\|_m$  on  $\mathbb{C}^{2m}$ , there exists constants  $D_m$  and  $d_m > 0$  such that, if  $\mathbf{c} = (c_{-m}, \dots, c_{-1}, c_1, \dots, c_m) \in \mathbb{C}^{2m}$  has unit norm*

$\|\mathbf{c}\|_m = 1$ , then for every  $\delta > 0$ ,

$$\text{Leb} \left( \left| \sum_{0 < |l| \leq m} c_l e(lx) \right| < \delta \right) < D_m \delta^{d_m}.$$

Hence, in our case, there exist constants  $D_m, d_m > 0$  such that for every  $\delta > 0$  and for fixed  $\widehat{\mathbf{x}} \in \mathbb{T}^{d-1}$  the measure of the set of  $x_d \in \mathbb{T}$  where  $|(S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp))(\widehat{\mathbf{x}}, x_d)| < \delta \sum_{0 < |l| \leq m} |c_{l,n,N}(\widehat{\mathbf{x}})|$  is less than  $D_m \delta^{d_m}$ . By Fubini's Theorem, choosing  $\delta = 4C/\sqrt{C'}$ , outside a subset of  $\mathbb{T}^d$  of measure less than  $D_m \delta^{d_m}$  the following estimate holds:

$$|(S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp))(\mathbf{x})| \geq \frac{4C}{\sqrt{C'}} \sum_{0 < |l| \leq m} |c_{l,n,N}(\widehat{\mathbf{x}})|.$$

Thus, on a set of measure at least  $1 - 2m/\sqrt{C'} - D_m(4C/\sqrt{C'})^{d_m}$ , we have

$$\begin{aligned} |(S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp))(\mathbf{x})| &\geq \frac{4C}{\sqrt{C'}} \left( \left| \sum_{0 < |l| \leq m} c_{l,n}(\widehat{\mathbf{x}}) e(lx_d) \right| \frac{1}{\sqrt{C'}} - \frac{\sqrt{C'}}{2} \right) \\ &= \frac{4C}{\sqrt{C'}} \left( \frac{|S_n(\Psi^\perp)(\mathbf{x})|}{\sqrt{C'}} - \frac{\sqrt{C'}}{2} \right). \end{aligned}$$

Let us enlarge  $C'$  if necessary and choose  $\varepsilon' > 0$  such that

$$\frac{2m}{\sqrt{C'}} + D_m \left( \frac{4C}{\sqrt{C'}} \right)^{d_m} + 2\varepsilon' < \varepsilon.$$

Let  $n \geq 1$  such that  $|S_n(\Psi^\perp)| \geq C'$  up to a set of measure  $\varepsilon'$ , by Corollary 5.1. Outside a set of measure less than  $\varepsilon$ , we conclude

$$|(S_n(\Psi^\perp) \circ T^N - S_n(\Psi^\perp))(\mathbf{x})| \geq 2C.$$

□

**5.2. Conclusion of the proof of Theorem 2.4.** Lemma 5.1 implies that  $\liminf_{n \rightarrow \infty} \text{Leb}_d(|S_n(\Psi^\perp)| < C) = 0$ ; let  $L$  be the limsup and assume by contradiction that it is different from 0. Choose  $\varepsilon > 0$  and  $\ell \geq 1$  such that

$$\frac{1}{\ell} + \frac{\ell+1}{2}\varepsilon < \frac{L}{2},$$

and consider  $C' > 1$  and  $\varepsilon' > 0$  given by Lemma 5.2. By Lemma 5.1, there exists an arithmetic progression  $\{i\bar{n}\}_{i=1}^\ell$  of length  $\ell$  such that  $\text{Leb}_d(|S_{i\bar{n}}(\Psi^\perp)| < C') < \varepsilon'$ . By Lemma 5.2, let  $N_0(i) \geq 1$  be such that the conclusion (5.3) is satisfied with  $n = i\bar{n}$ ; let  $\bar{N}_0$  be the maximum of all  $N_0(i)$  for  $i = 1, \dots, \ell$ . Choose  $N \geq \bar{N}_0$  such that  $\text{Leb}_d(|S_N(\Psi^\perp)| < C) \geq \frac{L}{2}$ . Since  $T$  is measure-preserving, for  $1 \leq j < i \leq \ell$  we get

$$\begin{aligned} &\text{Leb}_d(T^{-i\bar{n}}\{|S_N(\Psi^\perp)| < C\} \cap T^{-j\bar{n}}\{|S_N(\Psi^\perp)| < C\}) \\ &\leq \text{Leb}_d(|S_N(\Psi^\perp) \circ T^{i\bar{n}} - S_N(\Psi^\perp) \circ T^{j\bar{n}}| < 2C) \\ &= \text{Leb}_d(|S_N(\Psi^\perp) \circ T^{(i-j)\bar{n}} - S_N(\Psi^\perp)| < 2C), \end{aligned}$$

which is less than  $\varepsilon$  by Lemma 5.2. Thus by the inclusion-exclusion principle,

$$\begin{aligned} \text{Leb}_d \left( \bigcup_{i=1}^\ell T^{-i\bar{n}}\{|S_n(\Psi^\perp)| < C\} \right) &\geq \sum_{i=1}^\ell \text{Leb}_d(T^{-i\bar{n}}\{|S_n(\Psi^\perp)| < C\}) - \\ &- \sum_{1 \leq j < i \leq \ell} \text{Leb}_d \left( T^{-i\bar{n}}\{|S_n(\Psi^\perp)| < C\} \cap T^{-j\bar{n}}\{|S_n(\Psi^\perp)| < C\} \right) \geq \ell \frac{L}{2} - \frac{\ell(\ell+1)}{2}\varepsilon. \end{aligned}$$

This implies  $L/2 \leq 1/\ell + \varepsilon(\ell+1)/2$ , in contradiction with the initial choice of  $\ell$  and  $\varepsilon$ . Thus  $L = 0$ , which settles the proof.

## 6. PROOF OF THEOREM 1.2

In this section, we prove Theorem 1.2 by reducing the problem to the setting of suspension flows over skew-translations as in Theorem 1.1. Recalling the definitions and notation of §1.2, let  $F := F_d$  be a quasi-abelian filiform group,  $M = \Lambda \backslash F$  a quasi-abelian filiform nilmanifold and  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  a quasi-abelian filiform nilflow, where  $\mathbf{w} = w_0 \mathbf{f}_0 + \cdots + w_d \mathbf{f}_d \in \mathfrak{f} = \mathfrak{f}_d$ .

**6.1. Preliminaries.** We first show that, although almost every quasi-abelian filiform nilflow is uniquely ergodic, they are not weak mixing; we remark that the same argument applies to general nilflows. Indeed, each nilflow has a factor which is isomorphic to a rotation on a torus and furthermore unique ergodicity of the latter is equivalent to unique ergodicity of the former, see, e.g., [4, p. 344].

**Lemma 6.1.** *The abelianization  $\text{ab}: F \rightarrow F/F^{(2)}$  induces a factor of  $(M, \{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}})$  which is isomorphic to a linear flow  $\{\overline{\varphi}_t\}_{t \in \mathbb{R}}$  on  $\mathbb{T}^2$ .*

*Proof.* We have the following diagram

$$\begin{array}{ccc}
 & F & \\
 & \parallel & \\
 & \Lambda F^{(2)} & \\
 \Lambda & & F^{(2)} \\
 \parallel & & \parallel \\
 & \Lambda \cap F^{(2)} &
 \end{array}$$

where the double line denotes a normal subgroup (indeed,  $F^{(2)}$  is characteristic and  $\Lambda F^{(2)}$  is normal since it contains  $F^{(2)}$ ). The quotient  $F/F^{(2)}$  is an abelian group which is isomorphic to  $\mathbb{R}^2$ ; the abelianized lattice  $\Lambda/\Lambda \cap F^{(2)} \simeq \Lambda F^{(2)}/F^{(2)}$  is isomorphic to  $\mathbb{Z}^2$ , so that the quotient  $F/\Lambda F^{(2)}$  is a 2-dimensional torus and we obtain an exact sequence

$$0 \rightarrow \Lambda \backslash \Lambda F^{(2)} \rightarrow M \rightarrow \mathbb{T}^2 \rightarrow 0,$$

which expresses  $M$  as a bundle over the torus  $\mathbb{T}^2$  with fibers isomorphic to  $\Lambda \backslash \Lambda F^{(2)}$ . The differential of the induced projection  $\overline{\text{ab}}: M \rightarrow \mathbb{T}^2$  on  $M$  maps the vector field  $\mathbf{w} = w_0 \mathbf{f}_0 + \cdots + w_d \mathbf{f}_d \in \mathfrak{f}$  to a vector field on  $\mathbb{T}^2$ , which gives the linear flow  $\overline{\varphi}_t(x_0, x_1) = (x_0, x_1) + t(w_0, w_1)$ .  $\square$

**Theorem 6.2** (see, e.g., [4, p. 344]). *With the notation above, the following are equivalent:*

- (i)  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  is uniquely ergodic,
- (ii)  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  is ergodic with respect to the induced Haar measure  $\mu$  on  $M$ ,
- (iii)  $\{\overline{\varphi}_t\}_{t \in \mathbb{R}}$  is an irrational linear flow.

In order to assure that the quasi-abelian filiform nilflow  $\{\varphi_t^{\mathbf{w}}\}_{t \in \mathbb{R}}$  is uniquely ergodic, it is therefore sufficient to assume that  $w_0$  and  $w_1$  are rationally independent, which is a generic condition with respect to the Lebesgue measure on the Lie algebra  $\mathfrak{f}$  of  $F$ . For simplicity of notation, we will suppress the dependence on  $\mathbf{w}$ , writing  $\varphi_t$  instead of  $\varphi_t^{\mathbf{w}}$ .

For any quasi-abelian filiform nilflow, there are nontrivial eigenfunctions for the Koopman operator  $U_{\varphi_t}$  arising from the ones for the rotation  $\overline{\varphi}_t$  on the torus  $\mathbb{T}^2$ : let  $f \in L^2(\mathbb{T}^2)$  be a nontrivial eigenfunction for  $U_{\overline{\varphi}_t}$ , then the pull-back  $(\overline{\text{ab}})^* f \in L^2(M)$  is an eigenfunction for  $U_{\varphi_t}$ , since

$$U_{\varphi_t}((\overline{\text{ab}})^* f) = (f \circ \overline{\text{ab}}) \circ \varphi_t = f \circ \overline{\varphi}_t \circ \overline{\text{ab}} = (U_{\overline{\varphi}_t} f) \circ \overline{\text{ab}} = f \circ \overline{\text{ab}} = (\overline{\text{ab}})^* f.$$

We are interested in the time-changes of  $\{\varphi_t\}_{t \in \mathbb{R}}$ . Let  $\Sigma \subset M$  be a cross-section for the nilflow and let  $\Psi: \Sigma \rightarrow \mathbb{R}_{>0}$  be the first return time function. Let  $P: \Sigma \rightarrow \Sigma$  be the Poincaré map  $P\mathbf{x} = \varphi_{\Psi(\mathbf{x})}(\mathbf{x})$  for all  $\mathbf{x} \in \Sigma$ . The nilflow  $\{\varphi_t\}_{t \in \mathbb{R}}$  on  $M$  is isomorphic to the suspension flow over  $(\Sigma, P)$  with roof function  $\Psi$ . For any time-change, the cross-section  $\Sigma$  and the Poincaré map  $P$  remain the same; on the other hand, one can check that the first return time map  $\Psi^\alpha$  for the new flow with infinitesimal generator  $\alpha$  is given by

$$\Psi^\alpha(\mathbf{x}) = \int_0^{\Psi(\mathbf{x})} (\alpha \circ \varphi_t)(\mathbf{x}) dt.$$

The time-change  $\{\varphi_t^\alpha\}_{t \in \mathbb{R}}$  is then isomorphic to the suspension flow  $\{P_t^{\Psi^\alpha}\}_{t \in \mathbb{R}}$  over  $(\Sigma, P)$  with roof function  $\Psi^\alpha$ .

We prove Theorem 1.2 by choosing a cross-section  $\Sigma$  for the nilflow  $\{\varphi_t\}_{t \in \mathbb{R}}$  such that, in appropriate coordinates,  $\Sigma \simeq \mathbb{T}^d$  and the Poincaré map is a skew-translation as in Theorem 1.1. Moreover, the first return time is constant for all points in  $\Sigma$ ; see Lemma 6.5 below.

**6.2. Exponential coordinates and lattices.** Let us recall (see, e.g., [3, Theorem 1.2.1]) that for any connected, simply connected nilpotent Lie group  $G$  the exponential map  $\exp: \mathfrak{g} \rightarrow G$  is an analytic diffeomorphism and the following *Baker-Campbell-Hausdorff formula* holds:

$$\exp(\mathbf{v}) \exp(\mathbf{w}) = \exp\left(\mathbf{v} + \mathbf{w} + \frac{1}{2}[\mathbf{v}, \mathbf{w}] + \dots\right) \quad \text{for any } \mathbf{v}, \mathbf{w} \in \mathfrak{g}. \quad (6.1)$$

We can use the exponential map to transfer coordinates from  $\mathfrak{f}$  to  $F$ , so that we can cover the group with a single chart. In these coordinates, usually called the *exponential coordinates*, the multiplication law becomes the Baker-Campbell-Hausdorff (BCH) product  $\exp(\mathbf{v} * \mathbf{w}) = \exp(\mathbf{v}) \exp(\mathbf{w})$ . Therefore, we can safely identify  $F \simeq (\mathbb{R}^{d+1}, *)$ .

It is possible to characterize lattices in quasi-abelian filiform groups using exponential coordinates. It is well-known that, for any co-compact lattice  $\Lambda$ , one can choose coordinates so that  $\Lambda \simeq \mathbb{Z}^{d+1}$  (see, e.g., [3, Theorem 5.1.6]). However, for completeness and for the reader's convenience, we present a proof that provides new coordinates via a Lie algebra automorphism, hence preserving the Lie brackets.

Let us first state an auxiliary lemma. Denote by  $\text{Ad}: F \rightarrow \text{GL}(\mathfrak{f})$  the adjoint representation and by  $\mathfrak{ad}: \mathfrak{f} \rightarrow \mathfrak{gl}(\mathfrak{f})$  its differential.

**Lemma 6.3.** *For any  $\mathbf{v}, \mathbf{w} \in \mathfrak{f}$  we have that*

$$(-\mathbf{w}) * \mathbf{v} * \mathbf{w} = \left( \sum_{j=0}^{d-1} \frac{\mathfrak{ad}(\mathbf{w})^j}{j!} \right) \mathbf{v} = \mathbf{v} + [\mathbf{w}, \mathbf{v}] + \frac{1}{2}[\mathbf{w}, [\mathbf{w}, \mathbf{v}]] + \dots.$$

*In particular, if  $\mathbf{v}$  and  $\mathbf{w}$  commute with  $[\mathbf{v}, \mathbf{w}]$ , we have that  $\exp([\mathbf{v}, \mathbf{w}]) = [\exp(\mathbf{v}), \exp(\mathbf{w})]_F$ .*

*Proof.* We compute  $(\text{Ad} \circ \exp(\mathbf{w}))(\mathbf{v}) = \exp(-\mathbf{w})\mathbf{v} \exp(\mathbf{w})$ . By the commutation rule  $\text{Ad} \circ \exp = \exp \circ \mathfrak{ad}$ , it equals

$$(\exp \circ \mathfrak{ad}(\mathbf{w}))(\mathbf{v}) = \mathbf{v} + [\mathbf{w}, \mathbf{v}] + \frac{1}{2}[\mathbf{w}, [\mathbf{w}, \mathbf{v}]] + \dots.$$

We remark that, since  $F$  is  $d$ -step nilpotent,  $\mathfrak{ad}(\mathbf{w})^j = 0$  if  $j \geq d$ . Applying  $\exp$  to both sides, we conclude

$$\exp(-\mathbf{w}) \exp(\mathbf{v}) \exp(\mathbf{w}) = \exp\left[\left(\sum_{j=0}^{d-1} \frac{\mathfrak{ad}(\mathbf{w})^j}{j!}\right) \mathbf{v}\right].$$

If  $\mathbf{v}$  and  $\mathbf{w}$  commute with  $[\mathbf{v}, \mathbf{w}]$ , we have explicitly

$$\exp(-\mathbf{w}) \exp(\mathbf{v}) \exp(\mathbf{w}) = \exp(\mathbf{v} + [\mathbf{w}, \mathbf{v}]) = \exp(\mathbf{v}) \exp([\mathbf{w}, \mathbf{v}]),$$

from which we get  $\exp([\mathbf{v}, \mathbf{w}]) = [\exp(\mathbf{v}), \exp(\mathbf{w})]_F$ . □

If the integer  $E_1$  divides  $E_2$  we write  $E_1 \mid E_2$ .

**Lemma 6.4.** *Let  $\Lambda \leq F$  be a co-compact lattice in the  $d + 1$ -dimensional quasi-abelian filiform group  $F = F_d$  equipped with the exponential coordinates. Then, there exist  $1 = E_1 | E_2 | \cdots | E_d \in \mathbb{N}$ , with  $i! | E_i$ , such that, up to an automorphism of  $F$ ,*

$$\Lambda = \left\{ x\mathbf{f}_0 + \sum_{i=1}^d \frac{y_i}{E_i} \mathbf{f}_i : x, y_i \in \mathbb{Z} \right\}.$$

*Proof.* Let  $\pi_i$  be the canonical projection of  $F = F_d$  onto  $F/F^{(i)}$ . The image  $\pi_2(\Lambda) \subset F/F^{(2)}$  is a lattice in  $\mathbb{R}^2$ , hence there exist  $\mathbf{v}_0, \mathbf{v}_1 \in \Lambda$  such that  $\pi_2(\mathbf{v}_0), \pi_2(\mathbf{v}_1)$  generate  $\pi_2(\Lambda)$ . We can suppose that the first component of  $\mathbf{v}_0$  in the basis  $\mathcal{F}_d = \{\mathbf{f}_0, \dots, \mathbf{f}_d\}$  is different from zero.

We first show that for every  $1 \leq i \leq d$  there exists  $\mathbf{v}_i \in \Lambda \cap F^{(i)} \setminus F^{(i+1)}$ . By induction, suppose there exists  $\mathbf{v}_{i-1} \in \Lambda \cap F^{(i-1)} \setminus F^{(i)}$  for  $i \geq 2$ . Then, by Lemma 6.3,

$$[\pi_{i+1}(\mathbf{v}_0), \pi_{i+1}(\mathbf{v}_{i-1})] = \pi_{i+1}([\mathbf{v}_0, \mathbf{v}_{i-1}]) \in (\Lambda \cap F^{(i)})/F^{(i+1)},$$

since it belongs to the centre of  $F/F^{(i+1)}$ . It is also different from zero, as  $\mathbf{v}_{i-1} \notin F^{(i)}$ . Thus, there exists  $\mathbf{v}_i \in \Lambda \cap F^{(i)} \setminus F^{(i+1)}$  such that  $\pi_{i+1}(\mathbf{v}_i) = \pi_{i+1}([\mathbf{v}_0, \mathbf{v}_{i-1}])$ , hence the claim.

If  $d = 1$ , the group  $F_1$  is abelian and isomorphic to  $\mathbb{R}^2$  and the conclusion follows. Suppose  $d \geq 2$  and let  $\mathbf{v}_0, \mathbf{v}_1 \in \Lambda$  as above. Consider  $\mathbf{v}_{d-1} \in \Lambda \cap F^{(d-1)}$ ; by Lemma 6.3, we have  $[\mathbf{v}_0, \mathbf{v}_{d-1}], [\mathbf{v}_1, \mathbf{v}_{d-1}] \in \Lambda \cap F^{(d)}$ . The latter is isomorphic to a discrete subgroup of  $\mathbb{R}$ , thus the two vectors are rationally dependent. This implies that the first coordinate of  $\mathbf{v}_0$  and  $\mathbf{v}_1$  are rationally dependent. Up to replace  $\mathbf{v}_1$  with a vector of the form  $(-\mathbf{v}_0) * \cdots * (-\mathbf{v}_0) * \mathbf{v}_1 * \cdots * \mathbf{v}_1 \in \Lambda$ , we can suppose that the first coordinate of  $\mathbf{v}_1$  is zero.

Define  $\ell: F \rightarrow F$  as the unique group automorphism such that  $\ell(\mathbf{v}_0) = \mathbf{f}_0$  and  $\ell(\mathbf{v}_1) = \mathbf{f}_1$ . Then,  $\mathbf{f}_0$  and  $\mathbf{f}_1$  generate the projected lattice  $\ell(\Lambda)/F^{(2)}$  and moreover, by Lemma 6.3,  $\ell(\Lambda)$  contains

$$(-\mathbf{f}_1) * (-\mathbf{f}_0) * \mathbf{f}_1 * \mathbf{f}_0 = -\mathbf{f}_1 * \left( \mathbf{f}_1 + \sum_{i=2}^d \frac{1}{i!} \mathbf{f}_i \right) = \sum_{i=2}^d \frac{1}{i!} \mathbf{f}_i.$$

Inductively, by replacing  $\mathbf{f}_1$  above with  $\sum_{i \geq 2} (i!)^{-1} \mathbf{f}_i$  and so on, it is easy to see that  $\ell(\Lambda)$  contains the lattice generated by  $\frac{1}{i!} \mathbf{f}_i$ , hence

$$\ell(\Lambda) = \mathbb{Z} \times \frac{1}{E_1} \mathbb{Z} \times \cdots \times \frac{1}{E_d} \mathbb{Z},$$

for some integers  $E_1 = 1, E_2, \dots, E_d$  such that  $i! | E_i$ . Moreover, for all  $1 \leq i \leq d$ ,

$$\left( -\frac{1}{E_i} \mathbf{f}_i \right) * (-\mathbf{f}_0) * \left( \frac{1}{E_i} \mathbf{f}_i \right) * \mathbf{f}_0 = \frac{1}{E_i} \mathbf{f}_{i+1} + \text{terms in } F^{(i+2)},$$

hence  $E_i | E_{i+1}$ .  $\square$

We consider the new basis  $\mathcal{F}'_d = \{\mathbf{f}'_0, \dots, \mathbf{f}'_d\}$ , where  $\mathbf{f}'_0 = \mathbf{f}_0$  and  $\mathbf{f}'_i = (1/E_i) \mathbf{f}_i$  for  $i = 1, \dots, d$ . In this way, we have  $\Lambda = (\mathbb{Z}^{d+1}, *) \leq F$  and the only nontrivial brackets are  $[\mathbf{f}'_0, \mathbf{f}'_i] = (E_{i+1}/E_i) \mathbf{f}'_{i+1}$ .

**6.3. Reduction to suspension flows.** Let  $\mathbf{w} = (w_0, \dots, w_d) \in \mathfrak{f}$  be a vector inducing a uniquely ergodic nilflow on  $M = \Lambda \backslash F$ ; equivalently, by Theorem 6.2, such that  $w_0/w_1 \notin \mathbb{Q}$ . Define the smooth submanifold

$$\Sigma = \{\Lambda(0, x_1, \dots, x_d) : x_i \in \mathbb{R}, 1 \leq i \leq d\}.$$

Since the ideal generated by  $\mathbf{f}_1, \dots, \mathbf{f}_d$  is abelian, the submanifold  $\Sigma$  is isomorphic to a torus  $\mathbb{T}^d$  via the map

$$\begin{aligned} \varsigma: \mathbb{R}^d / \mathbb{Z}^d &\rightarrow \Sigma \\ \mathbf{x} = (x_1, \dots, x_d) &\mapsto \Lambda(0, x_1, \dots, x_d). \end{aligned}$$

**Lemma 6.5.** *The first return time to  $\Sigma$  is constant for any point of  $\Sigma$ ; the Poincaré map  $P: \Sigma \rightarrow \Sigma$  is given by*

$$P \circ \varsigma(\mathbf{x}) = \varsigma(\mathbf{x}A + \mathbf{b})$$

for some  $\mathbf{b} \in \mathbb{T}^d$  and an upper triangular  $d \times d$  matrix  $A = (a_{i,j})$ , with  $a_{i,j} = E_j / (E_i \cdot (j-i)!)$  for  $1 \leq i \leq j \leq d$ .

*Proof.* Let  $\varsigma(\mathbf{x}) = \Lambda(0, \mathbf{x}) \in \Sigma$ . By definition, we have

$$\varphi_{1/w_0}(\Lambda(0, \mathbf{x})) = \Lambda(0, x_1, \dots, x_d) * \left(1, \frac{w_1}{w_0}, \dots, \frac{w_d}{w_0}\right).$$

Since  $\Lambda = \Lambda(-1, 0, \dots, 0)$ , by Lemma 6.3 we get

$$\begin{aligned} \varphi_{1/w_0}(\Lambda(0, \mathbf{x})) &= \Lambda(-1, 0, \dots, 0) * (0, x_1, \dots, x_d) * \left(1, \frac{w_1}{w_0}, \dots, \frac{w_d}{w_0}\right) \\ &= \Lambda\left(\sum_{j=0}^d \frac{\mathfrak{a}\mathfrak{d}(1, 0, \dots, 0)^j}{j!} (0, \mathbf{x})\right) * (-1, 0, \dots, 0) * \left(1, \frac{w_1}{w_0}, \dots, \frac{w_d}{w_0}\right). \end{aligned}$$

Therefore, defining  $(0, b_1, \dots, b_d) = (-1, 0, \dots, 0) * (1, w_1/w_0, \dots, w_d/w_0)$ , we obtain

$$\begin{aligned} \varphi_{1/w_0}(\Lambda(0, \mathbf{x})) &= \Lambda\left(0, x_1, \dots, \sum_{i=0}^{j-1} \frac{1}{(j-i)!} \frac{E_j}{E_i} x_i, \dots\right) * (0, b_1, \dots, b_d) \\ &= \Lambda\left(0, x_1 + b_1, \dots, \sum_{i=0}^{j-1} \frac{1}{(j-i)!} \frac{E_j}{E_i} x_i + b_j, \dots\right). \end{aligned}$$

The set of return times to  $\Sigma$  is a subset of the set of the return times of the projected linear flow on the abelianization  $F/F^{(2)} \simeq \mathbb{T}^2$ , which is  $(1/w_0)\mathbb{Z}$ . The equation above shows that  $1/w_0$  is indeed a return time, hence it is the first return time to  $\Sigma$ , and the Poincaré map is of the requested form.  $\square$

We showed that any uniquely ergodic nilflow  $\{\varphi_t\}_{t \in \mathbb{R}}$  is isomorphic to a suspension flow over a skew-translation  $(\mathbb{T}^d, T)$  with constant roof function  $\Psi \equiv 1$ . As discussed in §6.1, given the infinitesimal generator  $\alpha$  of a time-change  $\{\varphi_t^\alpha\}_{t \in \mathbb{R}}$ , the new roof function  $\Psi^\alpha = R(\alpha)$  is given by

$$\Psi^\alpha(\mathbf{x}) = \int_0^1 (\alpha \circ \varphi_t)(\mathbf{x}) dt.$$

The map  $R: \mathcal{C}^\infty(M) \rightarrow \mathcal{C}^\infty(\mathbb{T}^d)$ ,  $R(\alpha) = \Psi^\alpha$  is linear, surjective and continuous w.r.t.  $\|\cdot\|_\infty$ , thus  $R^{-1}(\mathcal{R})$  is a dense set of infinitesimal generators. Theorem 1.2 now follows from Theorem 1.1.

## 7. APPENDIX: PROOF OF THEOREM 2.5

The proof of this result follows closely the argument by Avila, Forni and Ulcigrai in [1]: we outline the main ideas, referring the reader to the cited article for the details. We use the same notation as in §5.

**7.1. Shearing.** We briefly explain the shearing phenomenon that produces mixing; a similar mechanism was used by many authors in different contexts, see [12, 15, 5, 17, 13]. We want to apply the following criterion, see [5] and [17, §1.3.2] for details.

**Lemma 7.1 (Mixing Criterion).** *The suspension flow  $\{T_t^\Psi\}_{t \in \mathbb{R}}$  is mixing if for any cube  $Q = \prod_{i=1}^d [w_i, w'_i] \times [0, h]$ , with  $0 < h < \min \Psi$ , any  $\varepsilon > 0$  and  $\delta > 0$  there exists  $t_0 \geq 0$  such that for all  $t \geq t_0$  there exists a measurable set  $\widehat{X}(t) \subset \mathbb{T}^{d-1}$  and for each  $\widehat{\mathbf{x}} = \pi(\mathbf{x}) \in \widehat{X}(t)$  there exists a partition  $\mathcal{P}_m(t, \widehat{\mathbf{x}})$  into intervals  $J \subset \{\widehat{\mathbf{x}}\} \times \mathbb{T}$  such that*

$$\text{Leb}_d(\mathbb{T}^d \setminus \cup_{\widehat{\mathbf{x}} \in \widehat{X}(t)} \mathcal{P}_m(t, \widehat{\mathbf{x}})) \leq \delta, \quad (7.1)$$

and for all  $\widehat{\mathbf{x}} \in \widehat{X}(t)$  and all  $J = \{\widehat{\mathbf{x}}\} \times [a, b] \in \mathcal{P}_m(t, \widehat{\mathbf{x}})$ ,

$$\text{Leb}_1(J \cap T_{-t}^\Psi(Q)) \geq (1 - \varepsilon)(b - a) \text{Leb}_d(Q). \quad (7.2)$$

In order to apply Lemma 7.1, we will construct a partition of intervals  $J$  in the  $x_d$ -direction most of which becomes *sheared* for sufficiently large  $t$ . More precisely, for any  $J = \{\widehat{\mathbf{x}}\} \times [a, b]$ , we define the *stretch* of  $S_n(\Psi)$  over  $J$  as

$$\Delta S_n(\Psi)(J) = \max_{\mathbf{x} \in J} S_n(\Psi)(\mathbf{x}) - \min_{\mathbf{x} \in J} S_n(\Psi)(\mathbf{x}).$$

We will prove that, for a set of intervals  $J$  whose measure is large in  $\mathbb{T}^d$ , the stretch  $\Delta S_n(\Psi)(J)$  is large for all  $n$  of the form  $n = n_t(\mathbf{x})$  for some  $\mathbf{x} \in J$  and large  $t$ . This would imply that the image of  $J$  after time  $t$  can be written as the union of curves  $\gamma_i = T_t^\Psi(J_i)$ , for subintervals  $J_i \subset J$ , which project over intervals in the  $x_d$ -direction and on which the derivative  $\partial_d S_n(\Psi)$  of  $S_n(\Psi)$  w.r.t.  $x_d$  is large. The base points of these curves, i.e. the intersections  $\gamma_i \cap \mathbb{T}^d \times \{0\}$ , shadow with good approximation an orbit under  $T$ , hence, by unique ergodicity, are uniformly distributed in  $\mathbb{T}^d$ ; this leads to the mixing estimate.

**7.2. Stretch of Birkhoff sums for continuous time.** Recall that for  $\mathbf{x} = (\widehat{\mathbf{x}}, x_d) \in \mathbb{T}^d$  we denote  $n_t(\mathbf{x}) = \max\{n : S_n(\Psi)(\mathbf{x}) \leq t\}$ ; let

$$\underline{n}_t(\widehat{\mathbf{x}}) = \min\{n_t(\widehat{\mathbf{x}}, x_d) : x_d \in \mathbb{T}\}.$$

The following lemma ensures that the Birkhoff sums  $S_n(\Psi^\perp)$  grow in measure not only as  $n$  tends to infinity (see Theorem 2.4), but also when  $t$  tends to infinity. The proof uses the assumption that  $\psi$  is smoothly cohomologous to a constant.

**Lemma 7.2.** *For all  $C > 1$ , let*

$$\widehat{X}(t, C) = \{\widehat{\mathbf{x}} \in \mathbb{T}^{d-1} : \text{there exists } x_d \in \mathbb{T} \text{ s.t. } |S_{\underline{n}_t(\widehat{\mathbf{x}})}(\Psi^\perp)(\widehat{\mathbf{x}}, x_d)| > C\}.$$

*Then*

$$\lim_{t \rightarrow \infty} \text{Leb}(\mathbb{T}^{d-1} \setminus \widehat{X}(t, C)) = 0.$$

*Proof.* Let us assume by contradiction that there exist  $C > 1$ ,  $\delta > 0$  and an increasing sequence  $\{t_j\}_{j \in \mathbb{N}}$ , with  $t_j \rightarrow \infty$ , such that  $\text{Leb}(\mathbb{T}^{d-1} \setminus \widehat{X}(t_j, C)) \geq \delta$  for all  $j \in \mathbb{N}$ . If  $\widehat{\mathbf{x}} \notin \widehat{X}(t_j, C)$ , for all  $x_d \in \mathbb{T}$  we have  $|S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi^\perp)(\widehat{\mathbf{x}}, x_d)| \leq C$ ; thus, by Fubini's Theorem,

$$\text{Leb}\left\{\mathbf{x} \in \mathbb{T}^d : \left|S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi^\perp)(\widehat{\mathbf{x}}, x_d)\right| \leq C\right\} \geq \text{Leb}(\mathbb{T}^{d-1} \setminus \widehat{X}(t_j, C)) \geq \delta.$$

As we want to get a contradiction with Theorem 2.4, we look for a sequence  $\{\underline{n}_{t_j}(\widehat{\mathbf{x}})\}_{j \in \mathbb{N}}$  not depending on the point  $\widehat{\mathbf{x}}$ . Since  $\psi$  is smoothly cohomologous to the constant  $\int \Psi$ , there exists a smooth function  $u : \Sigma \rightarrow \mathbb{R}$  such that  $\psi - \int \Psi = u \circ T - u$ . Let  $\mathbf{y}$  be the point in  $\mathbb{T}^d$  for which  $\underline{n}_{t_j}(\widehat{\mathbf{x}}) = n_{t_j}(\mathbf{y})$ . We have

$$\begin{aligned} S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi)(\mathbf{y}) &= S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi^\perp)(\mathbf{y}) + S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\psi)(\mathbf{y}) \\ &= S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi^\perp)(\mathbf{y}) + u(T^{\underline{n}_{t_j}(\widehat{\mathbf{x}})}\mathbf{y}) - u(\mathbf{y}) + \underline{n}_{t_j}(\widehat{\mathbf{x}}) \cdot \int_{\mathbb{T}^d} \Psi \, d\text{Leb}_d. \end{aligned}$$

Let  $\bar{u}$  and  $\bar{\Psi}$  be the maximum of  $|u|$  and of  $\Psi$  over  $\mathbb{T}^d$ . Since, by definition,  $t_j - \bar{\Psi} \leq S_{\underline{n}_{t_j}(\widehat{\mathbf{x}})}(\Psi)(\mathbf{y}) = S_{n_{t_j}(\mathbf{y})}(\Psi)(\mathbf{y}) \leq t_j$ , from the previous equation it follows that for all  $\widehat{\mathbf{x}} \notin \widehat{X}(t_j, C)$ ,

$$t_j - \bar{\Psi} - C - 2\bar{u} \leq \underline{n}_{t_j}(\widehat{\mathbf{x}}) \cdot \int_{\mathbb{T}^d} \Psi \, d\text{Leb}_d \leq t_j + C + 2\bar{u}.$$

In particular, there exists a constant  $K$  such that for all  $t_j$  there are at most  $K$  possible values of  $\underline{n}_{t_j}(\widehat{\mathbf{x}})$ . Therefore, there exists a sequence  $n_j = \underline{n}_{t_j}(\mathbf{x}_j)$  such that  $\text{Leb}(|S_{n_j}(\Psi^\perp)(\widehat{\mathbf{x}}, x_d)| \leq C) \geq \delta/K$ , so that  $\limsup_{n \rightarrow \infty} \text{Leb}(|S_n(\Psi^\perp)| \leq C) \geq \delta/K > 0$ , in contradiction with Theorem 2.4.  $\square$

*Remark 3.* Straightforward computations show that  $\partial_d(S_n(\Psi)) = S_n(\partial_d\Psi) = S_n(\partial_d\Psi^\perp)$  and  $\partial_d^2(S_n(\Psi)) = S_n(\partial_d^2\Psi) = S_n(\partial_d^2\Psi^\perp)$  for all  $n \geq 1$ . Indeed,  $\partial_d\Psi = \partial_d\Psi^\perp$ , since  $\psi = \int \Psi dx_d$  does not depend on  $x_d$ ; moreover, as a map in the  $x_d$ -coordinate,  $T^i$  is a translation for all  $i \geq 1$ , hence  $\partial_d(\Psi \circ T^i) = \partial_d\Psi \circ T^i$ .

**7.3. The Mixing Criterion.** Let  $Q = \prod_{i=1}^d [w_i, w'_i] \times [0, h]$  be a given cube. Choose  $\delta_0 \in (0, 1)$  such that  $(1 - \delta_0)(1 - D'\delta_0^{d'} - m\delta_0) \geq 1 - \delta$ , where  $D', d'$  are given by Lemma 5.3 w.r.t.  $\|\cdot\|$ , with  $\left\| \sum_{|j| \leq m} \alpha_j e(jz) \right\| = \max_j |\alpha_j|$ . Let  $\varepsilon_0, N_0, C_0$  be chosen appropriately as in [1, §4.5]; let  $\chi$  be a continuous function such that

$$\chi(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \prod_{i=1}^{d-1} [w_i, w'_i] \times [w_d + \varepsilon_0(w'_d - w_d), w'_d - \varepsilon_0(w'_d - w_d)], \\ 0 & \text{if } \mathbf{x} \notin \prod_{i=1}^{d-1} [w_i, w'_i] \times [w_d + \varepsilon_0/2(w'_d - w_d), w'_d - \varepsilon_0/2(w'_d - w_d)]. \end{cases} \quad (7.3)$$

Finally, let  $t_0 > 0$  be such that for all  $t \geq t_0$  we have  $\text{Leb}(\mathbb{T}^{d-1} \setminus \widehat{X}(t, C_0)) \leq \delta_0$ . Set  $\widehat{X}(t, C_0) = \widehat{X}(t)$ .

We recall (5.4),

$$S_n(\Psi^\perp)(\widehat{\mathbf{x}}, x_d) = \sum_{0 < |l| \leq m} c_{l,n}(\widehat{\mathbf{x}}) e(lx_d),$$

and denote  $c'_{l,n}(\widehat{\mathbf{x}}) = 2\pi i l c_{l,n}(\widehat{\mathbf{x}})$  so that we can write

$$S_n(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d) = S_n(\partial_d\Psi^\perp)(\widehat{\mathbf{x}}, x_d) = \sum_{0 < |l| \leq m} c'_{l,n}(\widehat{\mathbf{x}}) e(lx_d). \quad (7.4)$$

Let

$$\mathcal{P}_0(t, \widehat{\mathbf{x}}) = \{(\widehat{\mathbf{x}}, x_d) \in \{\widehat{\mathbf{x}}\} \times \mathbb{T} : |S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d)| \geq \delta_0 \left\| S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi) \right\|\},$$

which is a union of intervals in the  $x_d$ -coordinate, since, for fixed  $n$ ,  $S_n(\Psi^\perp)(\widehat{\mathbf{x}}, \cdot)$  is a polynomial in  $x_d$  of degree  $m$ . Let  $\mathcal{P}_1(t, \widehat{\mathbf{x}})$  be the partial partition obtained by discarding from  $\mathcal{P}_0(t, \widehat{\mathbf{x}})$  all intervals of length less than  $\delta_0$ . By Lemma 5.3, we have  $\text{Leb}(\mathcal{P}_0(t, \widehat{\mathbf{x}})) \geq 1 - D'\delta_0^{d'}$ . Again, since  $S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d)$  is a trigonometric polynomial of degree  $m$ , there are at most  $2m$  points in each level set; therefore  $\mathcal{P}_1(t, \widehat{\mathbf{x}})$  is obtained from  $\mathcal{P}_0(t, \widehat{\mathbf{x}})$  by removing at most  $m$  intervals of length smaller than  $\delta_0$ . The size of the partial partition  $\mathcal{P}_1(t, \widehat{\mathbf{x}})$  satisfies

$$\text{Leb}(\mathcal{P}_1(t, \widehat{\mathbf{x}})) \geq 1 - D'\delta_0^{d'} - m\delta_0,$$

thus, by Fubini's Theorem,

$$\text{Leb}_{d-1} \left( \bigcup_{\widehat{\mathbf{x}} \in \widehat{X}(t)} \mathcal{P}_1(t, \widehat{\mathbf{x}}) \right) \geq (1 - \delta_0)(1 - D'\delta_0^{d'} - m\delta_0) \geq 1 - \delta, \quad (7.5)$$

by the choice of  $\delta_0$ .

The following lemma ensures that on each element of the partition the stretch is large enough. For all  $I \in \mathcal{P}_1(t, \widehat{\mathbf{x}})$ , denote by  $\underline{n}_t(I) = \min_{x_d} n_t(\widehat{\mathbf{x}}, x_d)$ ,  $\overline{n}_t(I) = \max_{x_d} n_t(\widehat{\mathbf{x}}, x_d)$ , and  $\Delta n_t(I) = \overline{n}_t(I) - \underline{n}_t(I) + 1$ .

**Lemma 7.3.** *For all  $I \in \mathcal{P}_1(t, \widehat{\mathbf{x}})$  we have that*

$$|S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d)| \geq \frac{\pi\delta_0}{m} C_0, \quad \text{for all } (\widehat{\mathbf{x}}, x_d) \in I; \quad (7.6)$$

$$|S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d)| \geq \frac{\delta_0}{2m} |S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d\Psi)(\widehat{\mathbf{x}}, x'_d)|, \quad \text{for all } (\widehat{\mathbf{x}}, x_d), (\widehat{\mathbf{x}}, x'_d) \in I. \quad (7.7)$$

*Proof.* From (7.4), for all  $(\widehat{\mathbf{x}}, x_d) \in \mathbb{T}^d$  and  $n \geq 1$  we have that

$$|S_n(\partial_d\Psi)(\widehat{\mathbf{x}}, x_d)| \leq (2m) \max_{0 < |l| \leq m} |c'_{l,n}(\widehat{\mathbf{x}})|;$$



hence, from the definition of  $\mathcal{P}_1(t, \widehat{\mathbf{x}}) \subset \mathcal{P}_0(t, \widehat{\mathbf{x}})$ ,

$$\min_{x_d \in \mathbb{T}} |S_n(\partial_d \Psi)(\widehat{\mathbf{x}}, x_d)| \geq \frac{\delta_0}{2m} \max_{x_d \in \mathbb{T}} |S_n(\partial_d \Psi)(\widehat{\mathbf{x}}, x_d)|.$$

This proves (7.6). Moreover, by definition of  $\widehat{X}(t)$ , there exists  $\mathbf{x} = (\widehat{\mathbf{x}}, x_d)$  for which  $|S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d \Psi)(\mathbf{x})| \geq C_0$ . Thus,  $\max_{0 < |l| \leq m} |c_{l,n}(\widehat{\mathbf{x}})| \geq C_0/(2m)$ , so that  $\max_{0 < |l| \leq m} |c'_{l,n}(\widehat{\mathbf{x}})| \geq 2\pi \max_{0 < |l| \leq m} |c_{l,n}(\widehat{\mathbf{x}})| \geq \pi C_0/m$ . We conclude (7.7) from the definition of  $\mathcal{P}_0(t, \widehat{\mathbf{x}})$ .  $\square$

From the previous estimates, it is possible to deduce the following properties; for the proof we refer to [1, Lemmas 11,12].

**Lemma 7.4** ([1, Lemmas 11,12]). *For all  $I \in \mathcal{P}_1(t, \widehat{\mathbf{x}})$  and for all  $\underline{n}_t(I) \leq n \leq \bar{n}_t(I)$ , we have*

$$\frac{1}{2} |S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d \Psi)(\widehat{\mathbf{x}}, x_d)| \leq |S_n(\partial_d \Psi)(\widehat{\mathbf{x}}, x_d)| \leq \frac{3}{2} |S_{\underline{n}_t(\widehat{\mathbf{x}})}(\partial_d \Psi)(\widehat{\mathbf{x}}, x_d)| \quad \text{for all } (\widehat{\mathbf{x}}, x_d) \in I.$$

Moreover, the function  $x_d \mapsto n_t(\widehat{\mathbf{x}}, x_d)$  is monotone and  $\Delta n_t(I) \geq \pi \delta_0^2 C_0 / (2m \min \Psi)$ .

Let us subdivide each interval  $I \in \mathcal{P}_1(t, \widehat{\mathbf{x}})$  into  $\Delta n_t(I)$  subintervals on which  $x_d \mapsto n_t(\widehat{\mathbf{x}}, x_d)$  is locally constant and let us group them into  $N_t(I) + 1$  groups, the first  $N_t(I)$  of which made by  $N_t(I)$  consecutive intervals, where  $N_t(I) = \lfloor \sqrt{\Delta n_t(I)} \rfloor$ . Denote by  $\mathcal{P}_m(t, \widehat{\mathbf{x}})$  the partition into intervals  $J$  obtained in this way. The estimate on the total measure (7.5) still holds. Moreover, each  $J \in \mathcal{P}_m(t, \widehat{\mathbf{x}})$  satisfies the following properties, which can be proved using the estimates on the stretch and on the size of the intervals, see [1, Lemma 13].

**Lemma 7.5** ([1, Lemma 13]). *For each  $J \in \mathcal{P}_m(t, \widehat{\mathbf{x}})$ , for all  $(\widehat{\mathbf{x}}, x_d), (\widehat{\mathbf{x}}, x'_d) \in J$  and all  $\underline{n}_t(J) \leq n \leq \bar{n}_t(J)$  we have*

$$\left| \frac{\Delta n_t(J)}{\Delta S_{\underline{n}_t(J)}(J)} - 1 \right| \leq \varepsilon_0; \quad (7.8)$$

$$\frac{1}{\Delta n_t(J)} \sum_{n=\underline{n}_t(J)}^{\bar{n}_t(J)} \chi \circ T^n(\widehat{\mathbf{x}}, x_d) \geq (1 - \varepsilon_0)^2 \prod_{i=1}^d (w'_i - w_i); \quad (7.9)$$

$$\text{Leb}_1(J) \leq \frac{w'_d - w_d}{2} \varepsilon_0; \quad (7.10)$$

$$\left| \frac{\Delta S_{\underline{n}_t(J)}(J)}{\Delta S_n(J)} - 1 \right| \leq \varepsilon_0. \quad (7.11)$$

Moreover, denoting  $J_n^h = \{(\widehat{\mathbf{x}}, x_d) \in J : t - h < S_n(\Psi)(\widehat{\mathbf{x}}, x_d) \leq t\}$ , we have

$$\left| \frac{\Delta S_n(J) \text{Leb}_1(J_n^h)}{\text{Leb}_1(J)h} - 1 \right| \leq \varepsilon_0. \quad (7.12)$$

It remains to prove (7.2) of the Mixing Criterion. By definition,  $J_n^h$  is the set of points in  $J$  that after time  $t$  undergo exactly  $n$  iterations of  $T$  (recall that  $h < \min \Psi$ ) and are mapped inside  $\mathbb{T}^d \times [0, h]$ . In particular, for different values of  $n$ , they are all disjoint. If, for  $J = \{\widehat{\mathbf{x}}\} \times (x'_d, x''_d) \in \mathcal{P}_m(t, \widehat{\mathbf{x}})$ , we have that  $\chi(T^n(\widehat{\mathbf{x}}, x'_d)) > 0$ , by the estimate (7.10) on the size of  $J$  and the definition of  $\chi$  (7.3), it follows that  $T^n(\widehat{\mathbf{x}}, x_d) \in \prod_i [w_i, w'_i]$  for all  $(\widehat{\mathbf{x}}, x_d) \in J_n^h$  and thus  $T_t^\Psi(\widehat{\mathbf{x}}, x_d) \in Q$ . We deduce that

$$\text{Leb}_1(J \cap T_{-t}^\Psi(Q)) \geq \sum_{n=\underline{n}_t(J)}^{\bar{n}_t(J)} \chi \circ T^n(\widehat{\mathbf{x}}, x'_d) \text{Leb}_1(J_n^h).$$

Using (7.8), (7.9), (7.11) and (7.12), we conclude

$$\begin{aligned}
& \sum_{n=\underline{n}_t(J)}^{\bar{n}_t(J)} \chi \circ T^n(\widehat{\mathbf{x}}, x'_d) \text{Leb}_1(J^h) \\
&= \frac{1}{\Delta n_t(J)} \sum_{n=\underline{n}_t(J)}^{\bar{n}_t(J)} \chi \circ T^n(\widehat{\mathbf{x}}, x_d) \frac{\Delta n_t(J)}{\Delta S_{\underline{n}_t(J)}(J)} \frac{\Delta S_{\underline{n}_t(J)}(J)}{\Delta S_n(J)} \frac{\Delta S_n(J) \text{Leb}_1(J^h)}{\text{Leb}_1(J)h} h \text{Leb}_1(J) \\
&\geq (1 - \varepsilon_0)^5 h \text{Leb}(J) \prod_{i=1}^d w'_i - w_i = (1 - \varepsilon_0)^5 \text{Leb}(J) \text{Leb}(Q).
\end{aligned}$$

#### REFERENCES

- [1] A. Avila, G. Forni, and C. Ulcigrai. Mixing for time-changes of heisenberg nilflows. *Journal of Differential Geometry*, 89(3):369–410, 2011.
- [2] A.D.R. Choudary and C. Niculescu. *Real Analysis on Intervals*. Springer India, 2014.
- [3] L. Corwin and F.P. Greenleaf. *Representations of Nilpotent Lie Groups and Their Applications: Volume 1, Part 1, Basic Theory and Examples*. Cambridge University Press, 2004.
- [4] M. Einsiedler and T. Ward. *Ergodic theory*, volume 259. Springer-Verlag London, 2011.
- [5] B.R. Fayad. Analytic mixing reparametrizations of irrational flows. *Ergodic Theory and Dynamical Systems*, 22(2):437–468, 2002.
- [6] L. Flaminio and G. Forni. Equidistribution of nilflows and applications to theta sums. *Ergodic Theory and Dynamical Systems*, 26(2):409–433, 2006.
- [7] G. Forni and C. Ulcigrai. Time-changes of horocycle flows. *Journal of Modern Dynamics*, 6(2):251–273, 2012.
- [8] H. Furstenberg. Strict ergodicity and transformation of the torus. *American Journal of Mathematics*, 83(4):573–601, 1961.
- [9] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, 1981.
- [10] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer-Verlag Berlin Heidelberg, 2004.
- [11] A. Katok. *Combinatorial constructions in Ergodic Theory and Dynamics*. Number 30. American Mathematical Soc., 2003.
- [12] B. Marcus. Ergodic properties of horocycle flows for surfaces of negative curvature. *Annals of Mathematics*, 105(1):81–105, 1977.
- [13] D. Ravotti. Quantitative mixing for locally hamiltonian flows with saddle loops on compact surfaces. *Annales Henri Poincaré*, 18(12):3815–3861, 2017.
- [14] L.D. Simonelli. Absolutely continuous spectrum for parabolic flows/maps. *Discrete and Continuous Dynamical Systems*, 38(1):263–292, 2018.
- [15] Ya.G. Sinai and K.M. Khanin. Mixing for some classes of special flows over rotations of the circle. *Functional Analysis and Its Applications*, 26(3):155–169, 1992.
- [16] R. Tiedra de Aldecoa. Spectral analysis of time-changes of horocycle flows. *Journal of Modern Dynamics*, 6(2):275–285, 2012.
- [17] C. Ulcigrai. Mixing of asymmetric logarithmic suspension flows over interval exchange transformations. *Ergodic Theory and Dynamical Systems*, 27(3):991–1035, 2007.

SCHOOL OF MATHEMATICS, UNIVERSITY OF BRISTOL, UNIVERSITY WALK, BS8 1TW BRISTOL, UK  
*E-mail address:* `davide.ravotti@bristol.ac.uk`