

COMPLEXITY OF NILSYSTEMS AND SYSTEMS LACKING NILFACTORS

BERNARD HOST, BRYNA KRA, AND ALEJANDRO MAASS

ABSTRACT. Nilsystems are a natural generalization of rotations and arise in various contexts, including in the study of multiple ergodic averages in ergodic theory, in the structural analysis of topological dynamical systems, and in asymptotics for patterns in certain subsets of the integers. We show, however, that many natural classes in both measure preserving systems and topological dynamical systems contain no higher order nilsystems as factors, meaning that the only nilsystems they contain as factors are rotations. Our main result is that in the topological setting, nilsystems have a particular type of complexity of polynomial growth, where the polynomial (with explicit degree) is an asymptotic both from below and above. We also deduce several ergodic and topological applications of these results.

1. THE ROLE OF NILSYSTEMS

1.1. Nilsystems in various contexts. In studying multiple ergodic averages in a *measure preserving system* (X, \mathcal{B}, μ, T) , certain *factors* called *nilfactors* with algebraic structure occur naturally: these are the factors of the given system that have the structure of a measure theoretic nilsystem (see Sections 1.2 and 2.2 for definitions).

As is often the case, the ergodic notions and results have counterparts in topological dynamics. As in the ergodic setting, we refer to topological factors that are nilsystems as *topological nilfactors*. Topological nilfactors naturally arise in the structural analysis of topological dynamical systems [23]. Nilsystems arise naturally in other contexts such as additive combinatorics; as an example, we cite the recent use of nilsequences (introduced in [3], these are functions evaluated on an

2010 *Mathematics Subject Classification.* 54H20, 37A05, 37B40.

Key words and phrases. Nilsystems, complexity, topological dynamics.

The second author was partially supported by NSF grant 1200971 and the third author by Fondap 15090007 and CMM-Basal grants. The authors thank PICS-CNRS for support of the first Franco-Chilean Conference on Dynamics and Combinatorics, where this work was begun.

orbit in a nilsystem) in the work on patterns in the primes, as described in the program laid out in [19].

The main results of this paper are a detailed analysis of the existence (or not) of nilfactors in the topological setting. Although we mostly work in the topological setting, we start with the more classical measure theoretic point of view, as some ergodic results are used in our topological study.

1.2. Nilsystems in ergodic theory. By a measure preserving system, we mean a Lebesgue probability space (X, \mathcal{B}, μ, T) endowed with a measurable, measure preserving transformation $T: X \rightarrow X$. When (X, \mathcal{B}, μ, T) and (Y, \mathcal{D}, ν, S) are measure preserving systems, a map $f: X \rightarrow Y$ is a *factor map* if it is measurable, carries the measure μ to ν , and satisfies $S \circ f = f \circ T$ μ -a.e.

The simplest factor is the invariant σ -algebra that arises in the von Neumann mean ergodic theorem. Assuming henceforth that the system is ergodic, the relevant factor for the study of the average of $f(T^n x)f(T^{2n} x)$ is the *Kronecker factor* $Z_1(X) = Z_1(X, \mathcal{B}, \mu, T)$. This is the factor spanned by the eigenfunctions of the system, and it is isomorphic to a translation on a compact abelian group endowed with its Haar measure.

For more intricate averages, for example the average of $f(T^n x) \cdot f(T^{2n} x) \cdot \dots \cdot f(T^{sn} x)$, we need a more sophisticated structural analysis of the system. In [20], this is done by introducing a series of factors $Z_s(X) = Z_s(X, \mathcal{B}, \mu, T)$ for $s \geq 1$. The convergence of this average then follows by analyzing the averages separately in the factor $Z_s(X)$ and in its orthogonal complement.

Such structural analysis only becomes useful when the factors have some sort of geometric or algebraic structure, and this is the content of the structure theorem in [20]. More precisely, the factor $Z_s(X)$ is the inverse limit of all s -step nilsystems that are factors of (X, \mathcal{B}, μ, T) . We call a factor of a system that is a nilsystem a *nilfactor* (for the definition of a nilsystem, see Section 2.2). The factors $Z_s(X)$ have since been used to understand other multiple ergodic averages (see, for example, [21, 27, 17]) and to prove new results on multiple recurrence (for example in [4, 17, 15]).

Our motivation in starting this work was to give examples of “simple” and “natural systems” with explicit, nontrivial factors $Z_s(X)$ for some $s > 1$. Since all factors $Z_s(X)$, $s \geq 1$, are trivial for weakly mixing systems, we restrict our attention to non-weakly mixing ones. Of course, the notion of a “natural system” is not precisely defined, but it is clear that a system obtained by building an arbitrary extension of

a given nilsystem is “artificial.” Somewhat surprisingly, we found the task of finding non-artificial nilsystems harder than expected. It turns out that for many natural classes of systems, the factors $Z_s(X)$ coincide with the Kronecker factor $Z_1(X)$; there are spectral obstructions that force this, and this is explained in Corollary 2.2. The interest in such a spectral result is that it sets the context for the main results of this paper, which are in the setting of topological dynamical systems.

In Sections 6.1 and 6.2, we give two applications for measure preserving systems that have no nilfactors other than rotation factors. The first result is on optimal lower bounds for intersections of translates of a set, and the second application is a strengthening of results in [22] and [7] on the convergence of weighted polynomial multiple averages.

1.3. Nilsystems in topological dynamics. We turn now to the counterparts of these notions in topological dynamics. Recall that a topological dynamical system is a compact metric space endowed with a homeomorphism T of X . A factor from the topological system (X, T) to the topological factor (Y, S) is a map $f: X \rightarrow Y$, continuous and onto, such that $S \circ f = f \circ T$.

As in the ergodic setting, we refer to topological factors that are nilsystems as *topological nilfactors*. Topological nilfactors naturally arise in the structural analysis of topological dynamical systems [23]. The main results of this paper are a detailed analysis of the existence (or not) of nilfactors in the topological setting.

We recall some definitions ([23], see also [32]). Let (X, T) be a transitive topological system, meaning that X is a compact metric space endowed with a homeomorphism $T: X \rightarrow X$ such that the orbit $\{T^n x: n \in \mathbb{Z}\}$ of some point $x \in X$ is dense. We can associate to this system an increasing sequence $Z_{\text{top},s}(X)$, $s \geq 1$, of topological factors. The first factor $Z_{\text{top},1}(X)$ is the *maximal equicontinuous factor*, also called the *topological Kronecker factor*, of (X, T) . It is spanned by the continuous eigenfunctions of the system, and is topologically isomorphic to a translation on a compact abelian group. For $s > 1$, $Z_{\text{top},s}(X)$ is the inverse limit of all s -step topological nilfactors of (X, T) . In [23] and [32], it is shown that topological nilfactors can be characterized by dynamically defined “cubic subsets” of $X^{2^{s+1}}$, leading to a topological analog of the ergodic theoretic structure theorem.

Finding natural topological systems with nilsystems as topological nilfactors also turns out to be nontrivial. One already has obstructions that arise from the ergodic setting, by considering any invariant measure on the system. Secondly, since nilsystems have zero entropy, it is

only of interest to consider distal systems with zero entropy. The relevant property seems to be some sort of complexity, and the notion we use is the *topological complexity* inspired in the notion of ε - n spanning sets (see for example [35] for the origin of this definition in the work of Dinaburg and Bowen):

Definition 1.1. Let (X, T) be a topological dynamical system and let d be a distance on X defining its topology. For $\varepsilon > 0$ and $n \geq 1$, an ε - n *spanning set* for (X, T) is a finite subset $\{x_1, \dots, x_m\}$ of X such that for every $x \in X$ there exists $j \in \{1, \dots, m\}$ such that

$$d(T^k x, T^k x_j) < \varepsilon \text{ for every } k \in \{0, \dots, n-1\}.$$

Let $\mathcal{S}_{X,T,d}(\varepsilon, n)$ denote the minimal cardinality of an ε - n spanning set of (X, T) . When there is no possible ambiguity, we omit the system and metric from the notation and write $\mathcal{S}(\varepsilon, n)$. We call the function \mathcal{S} the *topological complexity* of the system (X, T) .

Clearly, this notion depends on the choice of the distance d on X . However, if d and d' are distances on X defining its topology, then for every $\varepsilon > 0$ there exist $\eta_1 > 0$ and $\eta_2 > 0$, both tending to zero with ε , such that

$$\mathcal{S}_{X,T,d}(\eta_1, n) \leq \mathcal{S}_{X,T,d'}(\varepsilon, n) \leq \mathcal{S}_{X,T,d}(\eta_2, n)$$

for every $n \in \mathbb{N}$.

We also could have defined this in terms of ε - n separated sets: an ε - n *separated set* for (X, T) is a finite set $\{x_1, \dots, x_m\}$ of X such that for all $x_i \neq x_j$ there exists $k \in \{0, \dots, n-1\}$ such that $d(T^k x_i, T^k x_j) \geq \varepsilon$. Taking $s_{X,T,d}(\varepsilon, n)$ to be the maximal cardinality of an ε - n separated set in X , we obtain the same behavior as n tends to infinity and ε tends to 0. This remark is used (Section 4.2.2) to give an upper bound for the complexity of nilsystems.

The notion of topological complexity used here is closely related to the notion of the complexity of a cover studied in [5] and our results can be rephrased in this language (see Section 3.1).

We show that every nilsystem (X, T) has polynomial complexity (Theorem 3.2) with an explicit degree, and most importantly, the degree is the same both above and below. This places a constraint on any system having such a system as a factor. The upper bound is related to the well known fact that in a nilsystem, the orbits of two distinct points diverge at a polynomial rate. However, our bound is a global invariant, as opposed to such an infinitesimal characterization, and it describes the long time behavior. A polynomial upper bound was given in [9], but without a clear control on the exponent.

A natural question is what systems have the same complexity as nil systems, as shown in Theorem 3.2:

Question 1. *Characterize the minimal topological dynamical systems (X, T) such that for every $\varepsilon > 0$, there exist constants $c(\varepsilon), c'(\varepsilon) > 0$ such that*

$$c(\varepsilon)n \leq \mathfrak{S}_{X,T,d}(\varepsilon, n) \leq c'(\varepsilon)n$$

for every $n \in \mathbb{N}$ and $c(\varepsilon) \rightarrow \infty$ as $\varepsilon \rightarrow 0$. If, in addition, we assume that (X, T) is distal system, then is it a 2-step nilsystem ?

1.4. Nilsystems in symbolic dynamics. We recall that a *subshift* over the alphabet A is a closed, shift invariant subset X of $A^{\mathbb{Z}}$, where A is a finite set and $A^{\mathbb{Z}}$ is endowed with the natural compact topology and with the *shift transformation*. Endowed with the restriction of the shift, a subshift is a particular type of topological dynamical system.

In a similar spirit to the ergodic and topological questions, we ask how to decide whether a given subshift has any nilfactor other than a rotation. Conversely, given a nilsystem, we can ask what kinds of subshifts admit it as a factor? This leads us to a notion of complexity classically used for subshifts:

Definition 1.2 (see for example [12]). Let (X, T) be a subshift on the finite alphabet A . For every integer $n \geq 1$, let $\mathcal{C}_X(n)$ be the number of words of length n occurring in X . The *combinatorial complexity* of (X, T) is the function $n \mapsto \mathcal{C}_X(n)$. When there is no possible ambiguity, we omit the space from the notation and write $\mathcal{C}(n)$.

We show (Corollary 3.3) that a subshift with very low combinatorial complexity does not admit any nilfactor other than a rotation. More generally, subshifts with low combinatorial complexity do not admit any nilfactor of large order. Some applications of these results are given in Section 5.

The classical Morse-Hedlund Theorem states that a subshift X is finite (and thus only consists of periodic sequences) if and only if $\mathcal{C}(n) = n$ for some n . On the other hand, *Sturmian systems* satisfy $\mathcal{C}(n) = n+1$ for every $n \geq 1$, and such systems are codings of irrational rotations on the circle. This leads to the following question:

Question 2. *Find an “optimal” coding of a minimal nilsystem (X, T) . More precisely, define a subshift (Y, S) having (X, T) as a nilfactor and with minimal possible complexity \mathcal{C}_Y .*

1.5. Some questions. The discussion in Section 1.2 leads naturally to other questions. In the topological study of nilsystems (Section 1.3),

notions of complexity play a key role in understanding the existence of nilfactors. We ask if something analogous holds in the ergodic setting:

Question 3. *Does the conclusion of Corollary 2.2 remain valid when the spectral hypothesis is replaced by some hypothesis on the “measure-theoretic complexity” [13] or on the “slow entropy” [24] of the system?*

A related question is:

Question 4. *Compute the “measure-theoretic complexity” or the “slow entropy” of ergodic nilsystems.*

The ergodic results of this paper deal with factors. A natural generalization consists in proving similar results for joinings:

Conjecture. *Let (X, μ, T) be an ergodic system satisfying the spectral hypothesis of Corollary 2.2 and let (Y, ν, S) be an ergodic nilsystem. Then every joining of these two systems is relatively independent with respect to the corresponding joining of their Kronecker factors.*

We conclude by returning to our original motivation:

Question 5. *Find “natural” nontrivial systems in any of the settings considered here (ergodic, topological or symbolic) that have an explicit nilfactor other than a rotation, meaning systems such that $Z_1(X) \neq Z_2(X)$. More generally, describe classes of systems with $Z_s(X) \neq Z_{s+1}(X)$ for some $s \geq 1$.*

Acknowledgments. We thank Dave Morris, Terry Tao, and Jean-Paul Thouvenot for helpful discussions.

2. NOTATION AND BASIC DEFINITIONS

2.1. Measure preserving systems. Throughout the article, we omit the σ -algebra of measure preserving systems from our notation and write (X, μ, T) instead of (X, \mathcal{B}, μ, T) . All subsets of X and functions on X are implicitly assumed to be measurable. If X is given a topological structure, the σ -algebra is assumed to be the Borel σ -algebra. For simplicity we always assume that the transformation T is invertible. When f is a function on X , we write, as usual, Tf instead of $f \circ T$. Also, in a mild abuse of notation, we use T to denote the unitary operator $f \mapsto f \circ T$ of $L^2(\mu)$.

2.2. Nilsystems. Let G be a group. For $a, b \in G$, the commutator (making a conventional choice in the order) of these elements is defined to be

$$[a, b] = aba^{-1}b^{-1}.$$

Throughout, we make use of several standard identities: for $a, b \in G$, we have $[b, a] = [a, b]^{-1}$ and for $a, b, c \in G$,

$$(1) \quad [a, bc] = [a, b] [b, [a, c]] [a, c].$$

If A, B are subsets of G and $a \in A$, we write $[a, B]$ for the group spanned by $\{[a, b] : b \in B\}$ and $[A, B]$ for the group spanned by $\{[a, b] : a \in A, b \in B\}$. The subgroups G_j , $j \geq 1$, of G are defined inductively by

$$G_1 = G ; G_{j+1} = [G, G_j] \text{ for every } j \geq 1.$$

One can check that

$$(2) \quad \text{for all } k, \ell \geq 1, \quad [G_k, G_\ell] \subset G_{k+\ell}.$$

Let $s \geq 1$ be an integer. The group G is *s-step nilpotent* if $G_{s+1} = \{1_G\}$. In particular, G is 1-step nilpotent if and only if it is abelian. If $s \geq 2$ and G is *s-step nilpotent* but not $(s-1)$ -step nilpotent, then we have

$$(3) \quad G_1 \supsetneq G_2 \supsetneq G_3 \supsetneq \cdots \supsetneq G_{s-1} \supsetneq G_s \neq \{1_G\}.$$

If G is a Lie group, then G_0 denotes the connected component of its unit element 1_G .

Let $s \geq 2$ be an integer, G be an *s-step nilpotent* Lie group, and Γ be a discrete and cocompact subgroup of G . The compact manifold $X = G/\Gamma$ is called an *s-step nilmanifold*. The group G acts on X by left translation, and we write this action as $(g, x) \mapsto g \cdot x$. The *Haar measure* μ on X is the unique Borel probability measure μ on X that is invariant under this action.

Let τ be a fixed element of G and let $T : X \rightarrow X$ be the map $x \mapsto \tau \cdot x$. Then (X, T) is called a *topological s-step nilsystem* and (X, μ, T) a *measure theoretical s-step nilsystem*, or just an *s-step nilsystem*. The basic properties of nilsystems were established in [1] and [30], and a more modern presentation is found in [25]. In particular, we have the equivalences: (X, T) is transitive if and only if it is minimal if and only if it is uniquely ergodic if and only if (X, μ, T) is ergodic.

If (X, T) is minimal then, writing G' for the subgroup of G spanned by G_0 and τ and setting $\Gamma' = \Gamma \cap G'$, we have that $G = G'\Gamma$. Thus $(X, T) \cong (X', T')$ where $X' = G'/\Gamma'$ and T' is the translation by τ on X' . Therefore, without loss of generality we can restrict to the case that G is spanned by G_0 and τ .

We can also assume that G_0 is simply connected (see for example [28] or [1] for the case that $G = G_0$ and [26] for the general case).

2.3. The spectrum of a nilsystem. Improving a result of Green [1], Stepin proved (see Starkov [33] for history and comments):

Theorem ([34]). *Let $(X = G/\Gamma, \mu, T)$ be an ergodic nilsystem with connected, simply connected group G . Then $L^2(\mu)$ can be written as the orthogonal sum $L^2(\mu) = \mathcal{H} \oplus \mathcal{H}'$ of two T -invariant subspaces. The space \mathcal{H} consists of functions $f \in L^2(\mu)$ that factorize through $G/G_2\Gamma$ and the restriction of T to this space has discrete spectrum. The restriction of T to \mathcal{H}' has Lebesgue spectrum of infinite multiplicity.*

Green, Stepin, and Starkov only consider the case of a connected, simply connected group. But, in view of our applications, we need a similar result without any assumption that the group be connected. While the extension of Stepin's result to connected, but not simply connected groups, is standard, the generalization for non-connected groups is harder. Nilsystems arising from non-connected groups can be quite different than those arising from connected ones, and there does not seem to be a direct method of deducing the general case from the particular one. Adapting the existing proofs requires many changes, and so instead of modifying existing proofs we give a different one.

For any $s > 2$, any s -step ergodic nilsystem that is not a rotation admits a 2-step nilfactor that is not a rotation. Thus, we only need such a spectral result for 2-step nilsystems:

Proposition 2.1. *Let $(X = G/\Gamma, \mu, T)$ be an ergodic 2-step nilsystem that is not a rotation. Then $L^2(\mu)$ can be written as the orthogonal sum $L^2(\mu) = \mathcal{H} \oplus \mathcal{H}'$ of two closed T -invariant subspaces such that the restriction of T to \mathcal{H} has discrete spectrum and its restriction to \mathcal{H}' has Lebesgue spectrum of infinite multiplicity.*

The proof is elementary, but lengthy, and so we postpone it to Appendix A.

Corollary 2.2. *Let (X, μ, T) be an ergodic system and assume that its spectrum does not admit a Lebesgue component with infinite multiplicity. Then this system does not admit any nilsystem as a factor, other than a rotation factor.*

This result applies, in particular, to

- Weakly mixing systems.
- Systems with singular maximal spectral type.
- Systems with finite spectral multiplicity. This class includes:
 - Systems of finite rank, for example substitution dynamical systems, linearly recurrent systems, Bratteli-Vershik systems of finite topological rank, and interval exchange

transformations (for definitions and references, see for example [29, 31, 8, 6, 18]).

- Systems of local rank one or of funny rank one (see [10, 11], where the definitions are attributed to J.-P. Thouvenot).

Since nilsystems have zero entropy, the result also applies to

- Systems whose Pinsker factor belongs to one of the preceding types.

3. COMPLEXITY OF TOPOLOGICAL NILSYSTEMS

3.1. Complexity and commutator dimension. Before stating the theorem, we define:

Definition 3.1. If $(X = G/\Gamma, T)$ is an s -step nilsystem for some $s \geq 1$ and if $\tau \in G$ is the element defining T , the *total commutator dimension* p of X is defined to be

$$(4) \quad p = \sum_{\ell=1}^{s-1} \dim(\text{range}(\text{Ad}_\tau - \text{id})^\ell).$$

Our main result is:

Theorem 3.2. *Let $(X = G/\Gamma, T)$ be a minimal s -step nilsystem for some $s \geq 2$ and assume that (X, T) is not an $(s - 1)$ -step nilsystem. Let d_X be a distance on X defining its topology. Then for every $\varepsilon > 0$ that is sufficiently small, there exist positive constants $C(\varepsilon)$ and $C'(\varepsilon)$ such that the topological complexity $\mathcal{S}(\varepsilon, n)$ of (X, T) for the distance d_X satisfies*

$$(5) \quad C(\varepsilon)n^p \leq \mathcal{S}(\varepsilon, n) \leq C'(\varepsilon)n^p \text{ for every } n \geq 1,$$

where p is the total commutator dimension of X . Moreover, $p \geq s - 1$ and $C(\varepsilon) \rightarrow +\infty$ when $\varepsilon \rightarrow 0$.

Furthermore, for a suitably chosen distance on X , one can take $C(\varepsilon) = C\varepsilon^{-d}$ and $C'(\varepsilon) = C'\varepsilon^{-d}$, where C and C' are positive constants and d is the dimension of X .

The result can be translated into the language of complexity of covers studied in [5]. We start by reviewing the definition. Let \mathcal{U} be an open cover of X , and for every integer $n \in \mathbb{N}$, write

$$\mathcal{U}_0^{n-1} := \bigvee_{j=0}^{n-1} T^{-j}\mathcal{U}.$$

Define $\mathcal{N}(\mathcal{U}, n)$ to be the minimal cardinality of a subcover of \mathcal{U}_0^{n-1} ; the *complexity function* of \mathcal{U} is the map $n \mapsto \mathcal{N}(\mathcal{U}, n)$. We have that

$\mathcal{S}(\varepsilon, n) \leq c(\varepsilon)$ for every $\varepsilon > 0$ is equivalent to $\mathcal{N}(\mathcal{U}, n) \leq C'(\mathcal{U})$ for every open cover \mathcal{U} of X , and these conditions are equivalent to the system being a rotation. The upper bound in (5) can be rephrased as saying that for every open cover \mathcal{U} of X , there exists a constant $C'(\mathcal{U})$ with $\mathcal{N}(\mathcal{U}, n) \leq C'(\mathcal{U})n^p$ for every $n \in \mathbb{N}$. The lower bound means that there exists an open cover \mathcal{U} and a constant $C(\mathcal{U})$ such that $\mathcal{N}(\mathcal{U}, n) \geq C(\mathcal{U})n^p$ for every $n \in \mathbb{N}$.

Corollary 3.3. *Let $s \geq 1$ and let $(X = G/\Gamma, T)$ be a minimal nilsystem that is not an s -step nilsystem. Then*

$$\liminf_{n \rightarrow +\infty} \frac{1}{n^s} \mathcal{S}(\varepsilon, n) \rightarrow +\infty \text{ as } \varepsilon \rightarrow 0.$$

Applications of these results are given in Section 5.

3.2. Conventions and notation. In the sequel, $s \geq 2$ is an integer and $(X = G/\Gamma, T)$ is a minimal s -step nilsystem that is not an $(s - 1)$ -step nilsystem. We let τ denote the element of G defining the transformation T .

As explained in Section 2.2, the assumption of minimality allows us to assume that G is spanned by the connected component G_0 of the unit element 1_G and τ . We make further assumptions on the choice of the presentation G/Γ of the nilsystem X and then prove Theorem 3.2 under these additional assumptions. This clearly implies the result in the general case.

Let μ denote the Haar measure of X and let λ denote the Haar measure of G . We normalize λ such that the measure of any (Borel) fundamental domain of the projection $\pi: G \rightarrow X$ is equal to 1.

Here, and again in Section 4.1, we impose conditions on the distance d_X on X defining the topology on X . Again, the conclusions of Theorem 3.2 remain valid for a general distance defining the topology.

Throughout the proof, we often fix some $\varepsilon > 0$ and assume that ε is sufficiently small. This means that it is smaller than some constant depending only on the nilsystem $(X = G/\Gamma, T)$ and the distance d_X defined on it, and not on any other parameter such as the integer n .

Finally, we choose a bounded Borel fundamental domain D of the projection $\pi: G \rightarrow X$.

3.3. Some preliminaries.

3.3.1. Choosing a distance on X . First, we choose a distance d_G on the group G that defines its topology. For the moment, we only assume that this distance is invariant under right translations, meaning that

for all $g, g', h \in G$,

$$d_G(gh, g'h) = d_G(g, g').$$

The nilmanifold X is endowed with the quotient distance, meaning that for $x, y \in X$,

$$(6) \quad d_X(x, y) = \inf \{d_G(g, h) : \pi(g) = x \text{ and } \pi(h) = y\}.$$

In other words, for all $g, h \in G$, we have

$$(7) \quad d_X(\pi(g), \pi(h)) = \inf_{\alpha, \beta \in \Gamma} d_G(g\alpha, h\beta) = \inf_{\gamma \in \Gamma} d_G(g, h\gamma).$$

Since the inverse image under π of every point of X is discrete, the infimums in these last two formulas are attained.

We recall that for $2 \leq j \leq s$, G_j and $G_j\Gamma$ are closed subgroups of G and that $\Gamma \cap G_j$ is a cocompact subgroup of G_j . In particular, we deduce that there exists $\delta_0 > 0$ such that for $2 \leq j \leq s$,

$$(8) \quad \text{if } \gamma \in \Gamma \text{ is within a distance } \delta_0 \text{ of } G_j, \text{ then } \gamma \in G_j \cap \Gamma.$$

In particular, we deduce that

$$(9) \quad \text{if } d_G(g, g') < \delta_0 \text{ and } \pi(g) = \pi(g'), \text{ then } g = g'.$$

3.3.2. Commutators. The following lemma is used to prove that the exponent p of Theorem 3.2 is $\geq s - 1$.

Lemma 3.4. *For $1 \leq \ell \leq s - 1$, we have that $[\tau, G_\ell] \not\subset G_{\ell+2}$.*

Proof. Let $G' = G/G_{\ell+2}$, $\pi: G \rightarrow G'$ be the quotient homomorphism, $\Gamma' = \pi(\Gamma)$ and $\tau' = \pi(\tau)$. Let T' be the translation by τ' on $X' = G'/\Gamma'$. Then for every $j \geq 1$, we have that $\pi(G_j) = G'_j$. Therefore $G'_{\ell+2} = \{1_{G'}\}$ and (X', T') is an $(\ell + 1)$ -step nilsystem. But $G_{\ell+1} \neq G_{\ell+2}$ and so $G'_{\ell+1} \neq \{1_{G'}\}$. Thus (X', T') is not a ℓ -step nilsystem. Moreover, since (X', T') is a factor of (X, T) , it is minimal.

We remark that $\pi([\tau, G_\ell]) = [\tau', G'_\ell]$. Therefore, substituting X' for X , we are reduced to showing that τ does not commute with G_ℓ when X is a $(\ell + 1)$ -step, but not ℓ -step, nilsystem.

Assume that τ commutes with G_ℓ . Consider the commutator map $G_\ell \times G_0 \rightarrow G_{\ell+1}$. Since G is $(\ell + 1)$ -step nilpotent, by (1) this map is multiplicative in each coordinate separately. But by (2), $[G_\ell, G_2]$ is trivial and so this commutator map induces a continuous map $G_\ell \times G_0/G_2 \rightarrow G_{\ell+1}$. Finally, G_ℓ and G_0/G_2 are abelian and so this continuous map is bilinear.

On the other hand, by minimality the subgroup of G spanned by Γ and τ is dense in G . This and the hypothesis imply that $[G_\ell \cap \Gamma, G] = [G_\ell \cap \Gamma, \Gamma] \subset \Gamma$. Therefore, for $\gamma \in G_\ell \cap \Gamma$, the map $g \mapsto [\gamma, g]$

continuously maps the connected group G_0 to the discrete group Γ , and so this map is trivial. We conclude that $G_\ell \cap \Gamma$ commutes with G_0 .

Therefore, the commutator map $G_\ell \times G_0 \rightarrow G_{\ell+1}$ induces a bilinear continuous map

$$(10) \quad \psi: \frac{G_\ell}{G_\ell \cap \Gamma} \times \frac{G_0}{G_2} \rightarrow G_{\ell+1}.$$

Let $\chi: G_{\ell+1} \rightarrow \mathbb{T}$ be a character of the abelian group $G_{\ell+1}$. Then $\chi \circ \psi: G_\ell/(G_\ell \cap \Gamma) \times G_0/G_2 \rightarrow \mathbb{T}$ induces a continuous group homomorphism from group G_0/G_2 to the dual group of the compact abelian group $G_\ell/(G_\ell \cap \Gamma)$. Since this dual group is discrete and G_0/G_2 is connected, this group homomorphism is trivial. It follows that $\chi \circ \psi$ is the trivial map.

As this holds for every character χ of $G_{\ell+1}$, ψ is the trivial map. Combining this with definition (10), it follows that the commutator map $G_\ell \times G_0 \rightarrow G_{\ell+1}$ is trivial, and so G_ℓ commutes with G_0 .

By assumption, G is spanned by G_0 and τ and G_ℓ commutes with τ . It follows that G_ℓ is included in the center of G , that $G_{\ell+1}$ is trivial, and thus that G is ℓ -step nilpotent, a contradiction. \square

3.3.3. *Some linear algebra.* We make use of the following estimate from linear algebra. The proof is postponed to Appendix B.

Proposition 3.5. *Let \mathbb{R}^d be endowed with the Euclidean norm $\|\cdot\|$ and let the Lebesgue measure of a Borel subset K of \mathbb{R}^d be written $|K|$. Let A be a $d \times d$ matrix and assume that it is unipotent, meaning that $(\text{id} - A)^d = 0$. For every integer $n \geq 2$, let*

$$\mathcal{W}_n = \{ \xi \in \mathbb{R}^d : \|A^k \xi\| \leq 1 \text{ for } 1 \leq k < n \}.$$

If

$$p = \sum_{k=1}^{d-1} \dim(\text{range}(\text{id} - A)^k),$$

there exist positive constants C and C' (depending on d and on A) such that

$$(11) \quad Cn^{-p} \leq |\mathcal{W}_n| \leq C'n^{-p}$$

for every n .

3.4. Reduction to a local problem. Throughout this section, $n > 1$ is an integer. We assume that ε is given and is sufficiently small, x_0 is a point in X , and $h \in D$ is chosen such that $\pi(h) = x_0$.

For the moment, we view x_0 and h as fixed, but it is important that the bounds do not depend on $x_0 \in X$, and thus also not on $h \in D$.

We study the set $V \subset X$ defined by

$$V = \{x \in X : d_X(T^k x_0, T^k x) < \varepsilon \text{ for } 0 \leq k < n\}.$$

Since the infimum in the definition of d_X (see (7)) is attained, for every $x \in V$ there exists $g \in G$ with

$$(12) \quad x = g \cdot x_0 \text{ and } d_G(g, 1_G) = d_X(x, x_0).$$

Furthermore, for sufficiently small ε , it follows from (9) that these conditions completely define g . Let W denote the set of elements g associated in this way to points of V . Again, the choice of small ε implies that W is included in the connected component G_0 of 1_G in G .

Convention. In the sequel, we often view x , and so g , fixed. When needed, we emphasize the dependence of g on x satisfying (12) by writing $g(x)$ instead of g .

By hypothesis and (7), for $0 \leq k < n$, there exists $\gamma_k \in \Gamma$ with

$$(13) \quad d_G(\tau^k h \gamma_k, \tau^k g h) = d_X(\tau^k \cdot x_0, \tau^k g \cdot x_0) = d_X(T^k x_0, T^k x) < \varepsilon$$

and, by (9) again, this γ_k is unique. Clearly, $\gamma_0 = 1_G$.

In the remainder of this section, we show:

Lemma 3.6. *Let x_0 , x , g , and γ_k be as above. Then $\gamma_k = 1_G$ for $0 \leq k < n$.*

3.4.1. *Initial computations.* From the characterization (13) of γ_k , it follows that for $0 \leq k < n$,

$$\begin{aligned} \varepsilon &> d_G([\tau^k h, \gamma_k] \gamma_k \tau^k h, \tau^k g h) \\ &= d_G([\tau^k h, \gamma_k] \gamma_k \tau^k, \tau^k g) \text{ by right invariance} \\ &= d_G([\tau^k h, \gamma_k] \gamma_k \tau^k, g[g^{-1}, \tau^k] \tau^k) \\ &= d_G([\tau^k h, \gamma_k] \gamma_k, g[g^{-1}, \tau^k]) \text{ by right invariance (again)}. \end{aligned}$$

Since $d_G(g, 1_G) < \varepsilon$, we have that

$$(14) \quad d_G([\tau^k h, \gamma_k] \gamma_k, [g^{-1}, \tau^k]) < 2\varepsilon \text{ for } 0 \leq k < n.$$

We proceed by induction on the degree s of the s -step nilsystem X .

3.4.2. *The case $s = 2$.* Since G_2 is included to the center of G , it follows from (1) that the map $(x, y) \mapsto [x, y]$ from $G \times G$ to G_2 is bilinear and thus $[g^{-1}, \tau^k] = [g, \tau]^{-k} = [\tau, g]^k$ for every k . Moreover, $[\tau^k h, \gamma_k]$ belongs to the center of G , and Equation (14) can be rewritten as

$$(15) \quad d_G(\gamma_k, [\tau^k h, \gamma_k]^{-1} [\tau, g]^k) < 2\varepsilon \text{ for } 0 \leq k < n.$$

Therefore, for $0 \leq k < n$, we have that γ_k is within distance 2ε of G_2 . By Remark (8), we have that

$$\gamma_k \in G_2 \cap \Gamma,$$

and in particular γ_k belongs to the center of G . Then (15) implies that

$$(16) \quad d_G(\gamma_k, [\tau, g]^k) < 2\varepsilon \text{ for } 0 \leq k < n.$$

Since $d_G(g, 1_G) < \varepsilon$, it follows that $d_G([\tau, g], 1_G) < C\varepsilon$ for some $C > 0$. Thus $d_G(\gamma_1, 1_G) < (C + 2)\varepsilon$. Since Γ is discrete and ε is small,

$$\gamma_0 = \gamma_1 = 1_G.$$

On the other hand, by (16) again,

$$d_G(\gamma_k \gamma_{k+1}^{-2} \gamma_{k+2}, 1_G) < 8\varepsilon \text{ for } 0 \leq k < n - 2.$$

Since Γ is discrete and ε is small,

$$\gamma_k \gamma_{k+1}^{-2} \gamma_{k+2} = 1_G \text{ for } 0 \leq k < n - 2.$$

We deduce that $\gamma_k = 1_G$ for $0 \leq k < n$, as announced.

3.4.3. The general case. Assume that X is an s -step nilsystem for some $s > 2$ and that the result has been proven for an $(s - 1)$ -step nilsystem. Maintaining the same notation and conventions as above, we show that $\gamma_k = 1$ for every k .

Taking the quotient by G_s , the induction hypothesis implies that

$$\gamma_k \in G_s \text{ for } 1 \leq k < n$$

and in particular γ_k belongs to the center of G . For every $k \in \mathbb{Z}$, we write

$$(17) \quad g(k) := [g^{-1}, \tau^k].$$

By the estimate in Equation (14), we have that

$$(18) \quad d_G(\gamma_k, g(k)) < 2\varepsilon \text{ for } 0 \leq k < n.$$

A new difficulty arises here that does not come up for $s = 2$, as in general $g(k)$ does not belong to the center of G . We begin by showing that the map $k \mapsto g(k)$ is a polynomial map from \mathbb{Z} to G_2 (see [25]). The computations for this are fairly explicit and we include them.

First, since $d_G(g, 1_G) < \varepsilon$, there exists a constant $C > 0$ such that $d_G([g^{-1}, \tau^k], 1_G) < C\varepsilon$ for $0 \leq k \leq s - 1$ and, by definition (17) of $g(k)$ and (18), $d_G(\gamma_k, 1_G) < (C + 2)\varepsilon$ for $0 \leq k \leq s - 1$. Since Γ is discrete and ε is small, we deduce that

$$(19) \quad \gamma_k = 1_G \text{ for } 0 \leq k \leq s - 1.$$

On the other hand, recall that the *difference operator* D maps a sequence $(h(k): k \geq 0)$ with values in G to the sequence $(Dh(k): k \geq 0)$ given by

$$Dh(k) = h(k)^{-1}h(k+1) \text{ for every } k \geq 0.$$

Applying this definition to the sequence $(g(k): k \geq 0)$ defined by (17), it is easy to check by induction that for every $j \geq 1$,

$$D^j g(k) = \tau^k v_j \tau^{-k},$$

where

$$v_1 = [g^{-1}, \tau] \text{ and } v_{j+1} = [v_j^{-1}, \tau].$$

Therefore, for every $j \geq 1$ we have $v_j \in G_{j+1}$ and $D^j g(k) \in G_{j+1}$ for every k . (In the vocabulary of Leibman [25], this means that the sequence $(g(k): k \geq 0)$ belongs to the class $\mathcal{P}_{(0,1,2,\dots)}G$.) In particular $v_s = 1_G$ and

$$D^s g(k) = 1_G \text{ for every } k \geq 0.$$

By the definition of the difference operator and by induction, we have that the sequence $(g(k): k \geq 0)$ satisfies a recurrence relation of the form: for every $k \geq 0$,

$$1_G = D^s g(k) = g(k+m_1)^{\eta_1} g(k+m_2)^{\eta_2} \dots g(k+m_{2^s})^{\eta_{2^s}},$$

where

$$(20) \quad 0 \leq m_j < s \text{ and } \eta_j = \pm 1 \text{ for } 1 \leq j < 2^s; \quad m_{2^s} = s \text{ and } \eta_{2^s} = 1.$$

Since γ_ℓ belongs to the center of G for every ℓ ,

$$\begin{aligned} & d_G \left(1_G, \gamma_{k+m_1}^{\eta_1} \dots \gamma_{k+m_{2^s}}^{\eta_{2^s}} \right) \\ &= d_G \left((g(k+m_1)^{\eta_1} \dots g(k+m_{2^s})^{\eta_{2^s}}) (\gamma_{k+m_1}^{\eta_1} \dots \gamma_{k+m_{2^s}}^{\eta_{2^s}})^{-1}, 1_G \right) \\ &= d_G \left((g(k+m_1) \gamma_{k+m_1}^{-1})^{\eta_1} \dots (g(k+m_{2^s}) \gamma_{k+m_{2^s}}^{-1})^{\eta_{2^s}}, 1_G \right) \\ &\leq \sum_{j=1}^{2^s} d_G (g(k+m_j) \gamma_{k+m_j}^{-1}, 1_G) < 2^{s+1} \varepsilon \quad \text{by (18)}. \end{aligned}$$

Since Γ is discrete and ε is small, we deduce that

$$(21) \quad \gamma_{k+m_1}^{\eta_1} \dots \gamma_{k+m_{2^s}}^{\eta_{2^s}} = 1_G \text{ for every } k \geq 0.$$

In other words, the s^{th} iterated difference $(D^s \gamma_k)$ of the sequence $(\gamma_k: k \geq 0)$ is trivial, and this sequence is a polynomial sequence in the abelian group $G_s \cap \Gamma$.

Recalling that $m_{2^s} = s$ and $\eta_{2^s} = 1$, combining (19), (21) and (20) and using induction, we have that $\gamma_k = 1_G$ for $0 \leq k < n$ (see also Proposition 3.1 in [25]).

This concludes the proof of Lemma 3.6. \square

3.4.4. A summary.

Corollary 3.7. *Let $x_0 \in X$ and $h \in D$ with $\pi(h) = x_0$. Define*

$$V = \{x \in X : d_X(T^k x_0, T^k x) < \varepsilon \text{ for } 0 \leq k < n\}.$$

Then, for every $x \in V$, there exists a unique $g = g(x) \in G$ satisfying

$$x = g \cdot x_0 \text{ and } d_G(g, 1) = d_X(x, x_0).$$

We have

$$(22) \quad d_G(\tau^k g \tau^{-k}, 1_G) < \varepsilon \text{ for } 0 \leq k < n.$$

The last formula follows from (13), Lemma 3.6 and invariance of the distance under right translations.

4. PROOF OF THEOREM 3.2: COMPUTING THE COMPLEXITY

4.1. Working in the Lie algebra. As discussed in Section 2.2, we can assume that the connected component G_0 of the identity 1_G is simply connected.

As we consider neighborhoods W of 1_G such that $d_G(g, 1_G) < \varepsilon$ for every $g \in W$, we have that $W \subset G_0$. Thus it suffices to give a description of G_0 .

Recall that the exponential map is a homeomorphism from the Lie algebra \mathcal{G} onto G_0 . For every $j \geq 1$, we denote the Lie algebra of G_j by \mathcal{G}_j and the exponential map takes \mathcal{G}_j onto G_j .

Let \mathcal{G} be endowed with some Euclidean norm $\|\cdot\|$ and some orthonormal basis for this norm. We use this basis to identify \mathcal{G} and \mathbb{R}^d .

Let the distance d_G on G be the right invariant Riemannian distance such that the associated norm on the Lie algebra \mathcal{G} is $\|\cdot\|$. We have

$$(23) \quad d_G(\exp(\xi), 1_G) = \|\xi\| \text{ for every } \xi \in \mathcal{G}.$$

As before, we denote the Lebesgue measure of a subset \mathcal{L} of \mathcal{G} by $|\mathcal{L}|$. In general, the exponential map does not take the Lebesgue measure of \mathcal{G} to the Haar measure λ of G . But, because all elements g under consideration belong to the compact subset

$$K := \{g \in G : d_G(g, 1_G) \leq 1\}$$

of G , the density (with respect to λ) of the image of Lebesgue measure under the exponential map is bounded above and below by some positive constants C and C' :

$$(24) \quad \text{for every } L \subset K, \\ C |\{\xi \in \mathcal{G} : \exp(\xi) \in L\}| \leq \lambda(L) \leq C' |\{\xi \in \mathcal{G} : \exp(\xi) \in L\}|.$$

We recall that the linear map $\text{Ad}_\tau: \mathcal{G} \rightarrow \mathcal{G}$ is defined to be the differential of the map $g \mapsto \tau g \tau^{-1}$ from G to itself, evaluated at the point 1_G .

We have that

$$(25) \quad \tau \exp(\xi) \tau^{-1} = \exp(\text{Ad}_\tau \xi) \text{ for every } \xi \in \mathcal{G}.$$

Notation. We write $\Phi_\tau: G \rightarrow G$ for the map $g \mapsto \tau g \tau^{-1} g^{-1}$.

The differential of Φ_τ evaluated at 1_G is $\text{Ad}_\tau - \text{id}: \mathcal{G} \rightarrow \mathcal{G}$. Since G is s -step nilpotent, the s^{th} iterate of Φ_τ is the constant map 1_G . Thus

$$(\text{Ad}_\tau - \text{id})^s = 0,$$

meaning that the map $\text{Ad}_\tau - \text{id}$ is nilpotent. More precisely, for every $k \geq 1$, we have that $\Phi_\tau(G_k) \subset G_{k+1}$ and thus $(\text{Ad}_\tau - \text{id})\mathcal{G}_k \subset \mathcal{G}_{k+1}$.

Lemma 4.1. *The total commutator dimension p (Definition 3.1) of X satisfies*

$$p \geq \dim(\text{range}(\text{Ad}_\tau - \text{id})) \geq s - 1.$$

Proof. For $1 \leq \ell \leq s - 1$, the restriction of the map Φ_τ to G_ℓ maps this group to $G_{\ell+1}$. Composing it with the factor map $G_{\ell+1} \rightarrow G_{\ell+1}/G_{\ell+2}$ we obtain a map $\Psi_\ell: G_\ell \rightarrow G_{\ell+1}/G_{\ell+2}$. By (1), this map is a group homomorphism. By Lemma 3.4, this homomorphism is not trivial.

Now we take differentials at the unit element 1_G of G . The differential of Φ_τ at 1_G is $\text{Ad}_\tau - \text{id}$, the differential at 1_G of the quotient map $G_{\ell+1} \rightarrow G_{\ell+1}/G_{\ell+2}$ is the quotient map $p_\ell: \mathcal{G}_{\ell+1} \rightarrow \mathcal{G}_{\ell+1}/\mathcal{G}_{\ell+2}$, and thus the differential of Ψ_ℓ at this point is $p_\ell \circ (\text{Ad}_\tau - \text{id})$. Since the group homomorphism Ψ_ℓ is not trivial, its differential at 1_G is not zero and thus $p_\ell \circ (\text{Ad}_\tau - \text{id})$ is not trivial.

We conclude that $(\text{Ad}_\tau - \text{id})\mathcal{G}_\ell \not\subset \mathcal{G}_{\ell+2}$. Since $(\text{Ad}_\tau - \text{id})\mathcal{G}_{\ell+1} \subset \mathcal{G}_{\ell+2}$, it follows that $(\text{Ad}_\tau - \text{id})\mathcal{G}_\ell \neq (\text{Ad}_\tau - \text{id})\mathcal{G}_{\ell+1}$. Thus

$$\dim((\text{Ad}_\tau - \text{id})\mathcal{G}_\ell) \geq \dim((\text{Ad}_\tau - \text{id})\mathcal{G}_{\ell+1}) + 1.$$

Summing this inequality for $1 \leq \ell \leq s - 1$, we conclude that

$$\dim((\text{Ad}_\tau - \text{id})\mathcal{G}) \geq s - 1.$$

□

4.2. Bounding the complexity.

Notation. Set

$$\mathcal{K} := \{\xi \in \mathcal{G} : \exp(\xi) \in K\},$$

and note that \mathcal{K} is a compact subset of \mathcal{G} . Set

$$\begin{aligned} \mathcal{W}_n &:= \{\xi \in \mathcal{G} : \|\text{Ad}_\tau^k \xi\| \leq 1 \text{ for } 0 \leq k < n\}; \\ W_{\varepsilon,n} &:= \{\exp(\xi) : \xi \in \varepsilon \mathcal{W}_n\} \\ &= \{g \in G : d_G(\tau^k g \tau^{-k}, 1_G) < \varepsilon \text{ for } 0 \leq k < n\} \text{ by (25) and (23)}. \end{aligned}$$

(In the last line, we can write $g \in G$ instead of G_0 , as ε small implies that all points of G at a distance $< \varepsilon$ of 1_G belong to G_0 .)

Since the matrix of Ad_τ is unipotent, by definition (4) of p and Proposition 3.5, we have that $C\varepsilon^d n^{-p} \leq |\varepsilon \mathcal{W}_n| \leq C'\varepsilon^d n^{-p}$. Thus, since $W_{\varepsilon,n} \subset K$, by (24) it follows that:

Corollary 4.2. *For every sufficiently small $\varepsilon > 0$ and every $n \in \mathbb{N}$,*

$$(26) \quad C\varepsilon^d n^{-p} \leq \lambda(W_{\varepsilon,n}) \leq C'\varepsilon^d n^{-p}.$$

4.2.1. *Lower bound for the complexity.* We use this description to prove the lower bound of Theorem 3.2.

Assume that $\varepsilon > 0$ is sufficiently small and that $\{x_1, \dots, x_N\}$ is an ε - n spanning set for X . Let

$$V_j = \{x \in X : d_X(T^k x_j, T^k x) < \varepsilon \text{ for } 0 \leq k < n\}.$$

By hypothesis, the union of the sets V_j cover X .

For $1 \leq j \leq N$, choose $h_j \in D$ such that $\pi(h_j) = x_j$. We are in the setting of Corollary 3.7. For every $j \geq 1$ and every $x \in V_j$, there exists a unique $g_j(x) \in G$ with $x = g_j(x) \cdot x_j = \pi(g_j(x)h_j)$ and $d_G(g_j(x), 1_G) = d_X(x_j, x) < \varepsilon$. By (22), we have that $d_G(\tau^k g_j(x) \tau^{-k}, 1_G) < \varepsilon$ for $0 \leq k < n$. In other words, by definition of the set $W_{\varepsilon,n}$, we have that $g_j(x) \in W_{\varepsilon,n}$. It follows that

$$\pi\left(\bigcup_{j=1}^N W_{\varepsilon,n} h_j\right) = X.$$

Therefore, by choice of the normalization of λ ,

$$1 \leq \lambda\left(\bigcup_{j=1}^N W_{\varepsilon,n} h_j\right) \leq N\lambda(W_{\varepsilon,n}) \leq NC'\varepsilon^d n^{-p},$$

where the inequality follows from Corollary 4.2. We conclude that $N > C\varepsilon^{-d} n^p$ for some constant C . \square

4.2.2. *Upper bound of the complexity.* Recall that D is a Borel fundamental domain of the projection $\pi: G \rightarrow X$. Let D_1 be a compact subset of G containing all points at a distance at most 1 from D . We make use of the definition of complexity using ε - n separate sets (Section 1.3) to show:

Lemma 4.3. *Assume that $\varepsilon < 1$. There exists a subset $\{h_1, \dots, h_N\}$ of D with*

$$(27) \quad N \leq \lambda(D_1)\lambda(W_{\varepsilon/2,n})^{-1}$$

such that

$$(28) \quad D \subset \bigcup_{j=1}^N W_{\varepsilon,n}h_j.$$

Proof. Let N be the maximal cardinality of a subset $\{h_1, \dots, h_N\}$ of D such that the sets $W_{\varepsilon/2,n}h_j$ are disjoint.

Since every element $g \in W_{\varepsilon/2,n}$ satisfies $d_G(g, 1_G) < \varepsilon/2 < 1$, all the subsets $W_{\varepsilon,n}h_j$ are contained in D_1 . Therefore,

$$\lambda(D_1) \geq \sum_{j=1}^N \lambda(W_{\varepsilon/2,n}h_j) = N\lambda(W_{\varepsilon,n}).$$

We claim that the set $\{h_1, \dots, h_N\}$ satisfies (28). Assume instead that this does not hold and that $h \in D$ does not belong to this union. It follows immediately from the definition that $W_{\varepsilon/2,n}$ is symmetric and that $W_{\varepsilon/2,n} \cdot W_{\varepsilon/2,n} \subset W_{\varepsilon,n}$. Therefore, for $1 \leq j \leq n$, since $h \notin W_{\varepsilon,n}h_j$, we have that $W_{\varepsilon/2,n}h_j \cap W_{\varepsilon/2,n}h = \emptyset$.

Setting $h_{N+1} = h$, we have that the set $\{h_1, \dots, h_N, h_{N+1}\}$ satisfies the imposed condition, contradicting the maximality of N . \square

We now show that the upper bound of Theorem 3.2 holds, thereby completing the proof. Let $\varepsilon > 0$ be sufficiently small. Let N and $\{h_1, \dots, h_N\}$ be defined as in the conclusion of Lemma 4.3.

Let $x_j = \pi(h_j)$ for $1 \leq j \leq N$. We claim that $\{x_1, \dots, x_N\}$ is an ε - n spanning set for X .

Let $x \in X$ and $h \in D$ be such that $\pi(h) = x$. There exists j with $1 \leq j \leq N$ such that $h \in W_{\varepsilon,n}h_j$, meaning that there exists $g \in W_{\varepsilon,n}$ with $h = gh_j$. For $0 \leq k < n$, we have

$$\begin{aligned} d_X(T^k x, T^k x_j) &= d_X(\pi(\tau^k gh_j), \pi(\tau^k h_j)) \leq d_G(\tau^k gh_j, \tau^k h_j) \\ &= d_G(\tau^k g, \tau^k) = d_G(1_G, \tau^k g \tau^{-k}) < \varepsilon, \end{aligned}$$

since $g \in W_{\varepsilon,n}$. This proves the claim.

On the other hand, by (27) and Corollary 4.2, we have that $N \leq C\varepsilon^{-d}n^p$ for some $C > 0$, concluding the proof of Theorem 3.2.

5. SOME TOPOLOGICAL APPLICATIONS

As in the ergodic setting, we find classes of systems such that all factors $Z_{\text{top},s}(X)$, $s \geq 1$, of (X, T) are equal to $Z_{\text{top},1}(X)$.

It follows directly from Corollary 2.2 that transitive systems of finite topological rank satisfy this property; in particular, this is the case for minimal substitution dynamical systems and minimal interval exchange transformations. Namely, let (X, T) be such a system with nilfactor (Y, S) . Since (Y, S) is transitive, it is uniquely ergodic (see [1] and [30]), and its invariant measure is the Haar measure ν of Y . Let μ be an invariant ergodic measure on (X, T) . Then (X, μ, T) is a system of measure theoretical finite rank, and the topological factor map $X \rightarrow Y$ is also a measure-theoretic factor map. Thus by Corollary 2.2, (Y, ν, S) is measure theoretically isomorphic to a rotation. By the rigidity properties of nilsystems (see for example [23, Appendix A]), (Y, S) is topologically isomorphic to this rotation.

Corollary 3.3 can also be used to find other such classes:

Proposition 5.1. *Let (X, T) be a transitive subshift and assume that*

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \mathcal{C}_X(n) < +\infty.$$

Then (X, T) does not admit any topological nilfactor other than rotations. Therefore, for every $s \geq 1$, the topological factor $Z_{\text{top},s}(X)$ of X is equal to its topological Kronecker factor $Z_{\text{top},1}(X)$.

More generally, if for some $s \geq 1$ we have

$$(29) \quad \liminf_{n \rightarrow +\infty} \frac{1}{n^s} \mathcal{C}_X(n) < +\infty,$$

then (X, T) does not admit any nilsystem as a topological factor that is not an s -step nilsystem. Therefore, for every $t \geq s$, the topological factor $Z_{\text{top},t}(X)$ of X is equal to $Z_{\text{top},s}(X)$.

Proof. The last statement follows immediately from the fact that $Z_{\text{top},s}(X)$ is the inverse limit of all s -step topological nilfactors of (X, T) .

Assuming (29), it suffices to show that there is no topological factor map $\phi: (X, T) \rightarrow (Y, S)$, where (Y, S) is a minimal $(s+1)$ -step nilsystem that is not an s -step nilsystem. Indeed, any minimal nilsystem that is not an s -step nilsystem admits an $(s+1)$ -step nilsystem as a factor that is not an s -step nilsystem.

Assume instead that such a factor map $\phi: (X, T) \rightarrow (Y, S)$ exists and let d_Y be a distance on Y defining its topology. Recall that (X, T) is a transitive subshift on the finite alphabet A . We write $x \in X$ as $x = (x_n: n \in \mathbb{Z})$, and if $I \subset \mathbb{Z}$ is a finite interval, then we write x_I for the finite sequence $(x_n: n \in I)$.

Let $\varepsilon > 0$. Since ϕ is continuous, there exists an integer $L = L(\varepsilon) > 0$ such that $d_Y(\phi(x), \phi(y)) < \varepsilon$ whenever $x, y \in X$ satisfy $x_{[-L,L]} =$

$y_{[-L,L]}$. Therefore, if $x, y \in X$ satisfy $x_{[-L,n+L]} = y_{[-L,n+L]}$ for some $n \geq 1$, then $d_Y(T^k\phi(x), T^k\phi(y)) < \varepsilon$ for every $k \in \{0, \dots, n\}$.

Fix $n \geq 1$ and set $m = \mathcal{C}_X(n + 2L + 1)$.

By definition, there exist m elements $x^{(1)}, \dots, x^{(m)}$ of X such that for every $x \in X$, there exists $j \in \{1, \dots, m\}$ with $x_{[-L,n+L]} = x_{[-L,n+L]}^{(j)}$. By definition of L and since ϕ is onto, the set $\{\phi(x^{(j)}): 1 \leq j \leq m\}$ is an ε - n spanning set of (Y, S) . Thus

$$\mathcal{S}_{Y,S,d_Y}(\varepsilon, n) \leq \mathcal{C}_X(n + 2L + 1).$$

Since this holds for every $n \geq 1$, it follows from the hypothesis that

$$\liminf_{n \rightarrow +\infty} \frac{1}{n^s} \mathcal{S}_{Y,S,d_Y}(\varepsilon, n) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n^s} \mathcal{C}_X(n) < +\infty.$$

But this contradicts the statement of Corollary 3.3. \square

Remark. Under the hypothesis of “linear complexity,” that is, that there exists a constant $c > 0$ such that $\mathcal{C}_X(n) \leq cn$ for every $n \in \mathbb{N}$, the first statement of Proposition 5.1 can also be deduced from Corollary 2.2, by the method discussed at the beginning of this section; in this case, the system (X, T) has topological finite rank [12].

6. SOME ERGODIC APPLICATIONS

6.1. First application: lower bounds for multiple recurrence.

Theorem 6.1. *Assume that (X, μ, T) is an ergodic system satisfying $Z_s(X) = Z_1(X)$ for all $s > 1$, for example a system satisfying one of the properties listed after Corollary 2.2. Let p_1, \dots, p_k be integer polynomials satisfying $p_i(0) = 0$ for $1 \leq i \leq k$. Then for every $A \subset X$ and every $\varepsilon > 0$, the set*

$$(30) \quad \{n \in \mathbb{N}: \mu(A \cap T^{-p_1(n)}A \cap T^{-p_2(n)}A \cap \dots \cap T^{-p_k(n)}A) > \mu(A)^{k+1} - \varepsilon\}$$

is syndetic.

Remark. Here and in Theorems 6.2 and 6.3, the hypothesis can be replaced by $Z_2(X) = Z_1(X)$; it is known that this condition implies that $Z_s(X) = Z_1(X)$ for all $s > 1$ (this follows, for example, from the inclusions (3) in Section 2.2).

In [3] it is showed that the conclusion of Theorem 6.1 does not hold for non-ergodic systems, even in the simple case of $k = 2$, $p_1(n) = n$ and $p_2(n) = 2n$. The conclusion also fails for general ergodic systems, for example for $k \geq 4$ and $p_j(n) = jn$ for $1 \leq j \leq k$. In both of these cases, the set of integers defined by (30) may be empty.

On the other hand, the conclusion of Theorem 6.1 holds for weakly mixing systems [2]. Similar lower bounds for some particular choices of polynomials are found in [3, 16, 14].

For convenience, we begin the proof with the case of linear exponents and then explain how the method extends to the polynomial case. We first show:

Theorem 6.2. *Assume that (X, μ, T) is an ergodic system with $Z_s(X) = Z_1(X)$ for all $s > 1$, for example a system satisfying one of the properties listed after Corollary 2.2. Then for any integer $k \geq 1$, any set $A \subset X$, and any $\varepsilon > 0$, the set*

$$(31) \quad \{n \in \mathbb{N} : \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > \mu(A)^{k+1} - \varepsilon\}$$

is syndetic.

Proof. Fix $A \subset X$, an integer $k \geq 1$, and $\varepsilon > 0$.

Set $g = \mathbb{E}(\mathbf{1}_A \mid Z_1(X))$ and notice that $0 \leq g \leq 1$. Recall that the Kronecker factor $(Z_1(X), \nu, T)$ of (X, μ, T) is a compact abelian group, endowed with its Haar measure ν , Borel σ -algebra \mathcal{Z}_1 , and the transformation T is translation by some element $\alpha \in Z_1$. For simplicity, in this proof we write Z_1 instead of $Z_1(X)$. For each $t \in Z_1$, let g_t be the function given by $g_t(x) = g(x+t)$. Then there exists a neighborhood U of 0 in Z_1 such that $\|g_t - g\|_{L^1(\nu)} < \varepsilon/k^2$ for every $t \in U$. Therefore, for $t \in U$, we have that $\|g_{jt} - g\|_{L^1(\nu)} < j\varepsilon/k^2$ for every integer $1 \leq j \leq k$ and

$$\int g \cdot g_t \cdot \dots \cdot g_{kt} d\nu > \int g^{k+1} d\nu - \sum_{j=1}^k \frac{j\varepsilon}{k^2} \geq \mu(A)^{k+1} - \varepsilon.$$

Let $\Lambda := \{n \in \mathbb{N} : n\alpha \in U\}$. Then for every $n \in \Lambda$,

$$\int g \cdot T^n g \cdot \dots \cdot T^{kn} g d\nu > \mu(A)^{k+1} - \varepsilon.$$

Furthermore, Λ is a syndetic set, and thus there exists an integer $L > 0$ such that every interval of length L in \mathbb{N} contains at least one element of Λ .

On the other hand, by hypothesis, $g = \mathbb{E}(\mathbf{1}_A \mid Z_s(X))$ since $Z_s(X) = Z_1$ for all $s \geq 1$. Thus by [3, Corollary 4.6], the difference

$$a_n := \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) - \int g \cdot T^n g \cdot \dots \cdot T^{kn} g d\nu$$

converges to 0 in uniform density, meaning that

$$\lim_{N \rightarrow +\infty} \sup_{M \in \mathbb{N}} \frac{1}{N} \sum_{n=M}^{M+N-1} |a_n| = 0.$$

In particular, the set

$$\Lambda' := \left\{ n \in \mathbb{N} : \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > \int g \cdot T^n g \cdot \dots \cdot T^{kn} g \, d\nu - \varepsilon \right\}$$

has lower Banach density one, meaning that

$$\lim_{N \rightarrow \infty} \inf_{M \in \mathbb{N}} \frac{1}{N} |\Lambda' \cap [M, M + N]| = 1.$$

Thus there exists an integer $N > 0$ such that every interval of length N in \mathbb{N} contains L consecutive elements of Λ' . By definition of L , any interval of length N in \mathbb{N} contains some $n \in \Lambda' \cap \Lambda$ and this integer n satisfies (31). \square

Now we prove Theorem 6.1, that is, the extension of Theorem 6.2 for polynomial iterates. As the proof is similar but notationally more cumbersome, we only include an outline of the steps.

Proof of Theorem 6.1. Assume that (X, μ, T) is an ergodic system with $Z_s(X) = Z_1(X)$ for all $s > 1$ and that p_1, \dots, p_k are integer polynomials satisfying $p_i(0) = 0$ for $1 \leq i \leq k$. In this proof, we write Z_1, Z_s, \dots , instead of $Z_1(X), Z_s(X), \dots$.

By [21] and [26], there exists an integer $s \geq 1$ such that for all functions $f_0, f_1, \dots, f_k \in L^\infty(\mu)$, the averages over $[M_i, N_i)$ of

$$\begin{aligned} & \int f_0 \cdot T^{p_1(n)} f_1 \cdot \dots \cdot T^{p_k(n)} f_k \, d\mu \\ & - \int \mathbb{E}(f_0 | Z_s) \cdot T^{p_1(n)} \mathbb{E}(f_1 | Z_s) \cdot \dots \cdot T^{p_k(n)} \mathbb{E}(f_k | Z_s) \, d\mu \end{aligned}$$

converge to zero for all sequences (M_i) and (N_i) of integers such that $N_i - M_i \rightarrow +\infty$. Proceeding as in the proof of the deduction of Corollary 4.5 from Theorem 4.4 of [3], we deduce that

$$\begin{aligned} & \int f_0 \cdot T^{p_1(n)} f_1 \cdot \dots \cdot T^{p_k(n)} f_k \, d\mu \\ & - \int \mathbb{E}(f_0 | Z_{s+1}) \cdot T^{p_1(n)} \mathbb{E}(f_1 | Z_{s+1}) \cdot \dots \cdot T^{p_k(n)} \mathbb{E}(f_k | Z_{s+1}) \, d\mu \end{aligned}$$

converges to zero in uniform density. By hypothesis and Corollary 2.2, we have that $Z_{s+1} = Z_1$. Applying this with $f_0 = f_1 = \dots = f_k = \mathbf{1}_A$, and writing $g = \mathbb{E}(\mathbf{1}_A | Z_1)$, we conclude that

$$\mu(A \cap T^{-p_1(n)}A \cap \dots \cap T^{-p_k(n)}A) - \int g \cdot T^{p_1(n)} g \cdot \dots \cdot T^{p_k(n)} g \, d\mu$$

converges to zero in uniform density.

We continue as in the proof of Theorem 6.2. Let ν denote the Haar measure of Z_1 and let α be the element of Z_1 defining its transformation. For t and $z \in Z_1$, write $g_t(z) = g(z+t)$; choose an open neighborhood U of 0 in Z_1 such that $\|g_t - g\|_{L^1(\nu)} < \varepsilon/k^2$ for every $t \in U$. We now use a standard equidistribution method. Let H be the closed subgroup of \mathbb{T}^k spanned by $(p_1(n)t, \dots, p_k(n)t)$ for $n \in \mathbb{Z}$ and $t \in \mathbb{T}$. We have that H is equal to the set of $(t_1, \dots, t_k) \in \mathbb{T}^k$ such that $a_1 t_1 + \dots + a_k t_k = 0$ for every choice of $(a_1, \dots, a_k) \in \mathbb{Z}^k$ such that $a_1 p_1 + \dots + a_k p_k$ is identically zero. By Weyl's Theorem [36], the sequence $(p_1(n)\alpha, \dots, p_k(n)\alpha)$ is well distributed in H , meaning that for every continuous function ϕ on \mathbb{T}^k ,

$$\frac{1}{N} \sum_{n=M}^{N+M-1} \phi((p_1(n)\alpha, \dots, p_k(n)\alpha)) \longrightarrow \int \phi dm_H, \text{ uniformly in } M,$$

where m_H denotes the Haar measure on H . We deduce that for the open set U ,

$$(p_1(n)\alpha, \dots, p_k(n)\alpha) \in U \times \dots \times U$$

for a syndetic set of n . We conclude as in the proof of Theorem 6.2. \square

6.2. Second application: weighted multiple averages. The second application is a strengthening of results in [22] and [7] on the convergence of weighted polynomial multiple averages. Recall that the Kronecker factor of an ergodic system is naturally endowed with a topology, making it a compact abelian group.

Theorem 6.3. *Let (X, T) be a uniquely ergodic topological dynamical system with invariant measure μ . Assume that (X, μ, T) satisfies $Z_s(X) = Z_1(X)$ for all $s > 1$, for example a system satisfying one of the properties listed after Corollary 2.2, and that the projection π_1 of X onto its Kronecker factor is continuous. Then for any Riemann integrable function ϕ on X , any $x \in X$, any system (Y, ν, S) , any $k \geq 1$, any functions $f_1, \dots, f_k \in L^\infty(\nu)$, and any integer polynomials p_1, \dots, p_k , the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} \phi(T^n x) \cdot S^{p_1(n)} f_1 \cdot S^{p_2(n)} f_2 \cdot \dots \cdot S^{p_k(n)} f_k$$

converge in $L^2(\nu)$ as $N \rightarrow +\infty$.

In particular, this result applies for substitution dynamical systems [31], and more generally for many linearly recurrent systems and systems of finite topological rank [6]. In [22], it was proved in the case of linear polynomials for particular sequences, including, for example the Thue-Morse sequence.

Proof. Assume that (X, T) is a uniquely ergodic system satisfying the hypotheses of Theorem 6.3. Since for every $s \geq 1$ we have that $Z_s(X) = Z_1(X)$, by hypothesis the projection $\pi_s: X \rightarrow Z_s(X)$ is continuous. Theorem 6.3 follows immediately by combining two results in the literature. The first is a weighted ergodic average for nilsequences:

Theorem ([22], Theorem 2.19 and Proposition 7.1). *Let (X, T) be a uniquely ergodic system with invariant measure μ and let $m \geq 1$ be an integer. Assume that the factor map $\pi_s: X \rightarrow Z_s(X)$ is continuous. Then for any Riemann integrable function ϕ on X , $x \in X$, and s -step nilsequence $\underline{b} = (b_n: n \in \mathbb{Z})$, the limit*

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} \phi(T^n x) b_n$$

exists.

The second is a weighted ergodic theorem for multiple convergence along integer polynomials proved by Chu that generalized the linear case in [22]:

Theorem ([7], Theorem 1.3). *For any $k, d \in \mathbb{N}$, there exists an integer $s \geq 1$ with the following property: for any bounded sequence $\underline{a} = (a_n: n \in \mathbb{Z})$, if the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} a_n b_n$$

converge as $N \rightarrow +\infty$ for every s -step nilsequence $\underline{b} = (b_n: n \in \mathbb{Z})$, then for every system (Y, ν, S) , all $f_1, \dots, f_k \in L^\infty(\nu)$, and all integer polynomials p_1, \dots, p_k of degree $\leq d$, the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} a_n S^{p_1(n)} f_1 \cdot S^{p_2(n)} f_2 \cdot \dots \cdot S^{p_k(n)} f_k$$

converge in $L^2(\nu)$.

□

APPENDIX A. PROOF OF PROPOSITION 2.1

For convenience, we repeat the statement of Proposition 2.1:

Proposition. *Let $(X = G/\Gamma, \mu, T)$ be an ergodic 2-step nilsystem that is not a rotation. Then $L^2(\mu)$ can be written as the orthogonal sum $L^2(\mu) = \mathcal{H} \oplus \mathcal{H}'$ of two closed T -invariant subspaces such that the restriction of T to \mathcal{H} has discrete spectrum and its restriction to \mathcal{H}' has Lebesgue spectrum of infinite multiplicity.*

A.1. Reductions and notation. Before proving the proposition, we make some usual reductions that present the nilmanifold in a standard way. Note that these reductions are not the same as those of Section 2.2.

Let $(X = G/\Gamma, \mu, T)$ be an ergodic 2-step nilsystem that is not a rotation. As in Section 2.2, we write $\tau \in G$ for the element defining the transformation T and G_0 for the connected component of the identity. By minimality we can assume that the subgroup spanned by G_0 and τ is dense in G . This implies that G_2 is connected and thus included in G_0 .

Let Γ' be the largest normal subgroup of G contained in Γ . By substituting G/Γ' for G and Γ/Γ' for Γ , we reduce to the case that Γ does not contain any normal subgroup of G . Thus the action of G on X is faithful and it follows that Γ is abelian and that G_2 is compact. Since this Lie group is abelian and connected, it is a finite dimensional torus. Furthermore, it follows that each of subgroups spanned by $[\tau, \Gamma]$, by $[\tau, G]$ and by $[\tau, G_0]$ is dense in G_2 .

Let $q: G \rightarrow G/G_2$ and $\pi: G \rightarrow X$ denote the quotient maps. We recall (see [30] and [26]) that the Kronecker factor of X is the compact abelian Lie Group $Z_1 = G/G_2\Gamma$, endowed with its Lebesgue measure m_{Z_1} and translation by $\alpha = p \circ \pi(\tau)$, where $p: X \rightarrow Z_1$ denotes the factor map.

We remark that for every $a \in G$, the map $g \mapsto [a, g]$ is a group homomorphism from G to G_2 (see Section 2.2) and that the kernel of this homomorphism contains G_2 .

A.2. Lebesgue spectrum. For $\chi \in \widehat{G_2}$, set

$$\mathcal{H}_\chi = \{f \in L^2(\mu) : f(u \cdot x) = \chi(u)f(x) \text{ for every } u \in G_2 \text{ and } \mu\text{-a.e. } x\}.$$

Each space \mathcal{H}_χ is invariant under T and $L^2(\mu)$ is the orthogonal sum of these spaces. We have that \mathcal{H}_1 is the space of functions that factorize through Z_1 , and each space \mathcal{H}_χ is invariant under multiplication by functions belonging to $L^\infty(\mu) \cap \mathcal{H}_1$.

We show:

Lemma A.1. *Let χ be a nontrivial character of G_2 . Then the spectral measure associated to any function in \mathcal{H}_χ is absolutely continuous with respect to the Lebesgue measure $m_{\mathbb{T}}$ of \mathbb{T} .*

Proof. Let χ be a nontrivial character of G_2 .

The group homomorphism $g \mapsto \chi[g, \tau]$ factorizes through G/G_2 and so there exists a character $\tilde{\chi}$ of G/G_2 satisfying

$$(32) \quad \chi[g, \tau] = \tilde{\chi} \circ q(g) \text{ for every } g \in G.$$

Let $\gamma \in \Gamma$. Since γ commutes with Γ and with G_2 , the map $g \mapsto [\gamma, g]$ factorizes through the Kronecker factor Z_1 . Thus there exists a group homomorphism $\widehat{\gamma}: Z_1 \rightarrow G_2$ such that

$$(33) \quad \widehat{\gamma} \circ p \circ \pi(g) = [\gamma, g] \text{ for every } g \in G.$$

Let K be a compact subset of G such that the restriction to K of the projection $\pi: G \rightarrow X$ is onto. For every $n \in \mathbb{Z}$, we choose γ_n and g_n with

$$(34) \quad \gamma_n \in \Gamma, \quad g_n \in K \text{ and } \tau^n = g_n \gamma_n.$$

The family L of characters $\widetilde{\chi}_{m,n}$ of G/G_2 defined by $\widetilde{\chi}_{m,n} \circ q(g) = \chi[g, g_n g_m^{-1}]$ is a bounded subset of $\widehat{G/G_2}$. But the closed subgroup of $\widehat{G/G_2}$ spanned by $\widetilde{\chi}$ is not compact; if not, the restriction of $\widetilde{\chi}$ to some open subgroup H of G/G_2 would be trivial and so the restriction of $\widetilde{\chi} \circ q$ to G_0 would also be trivial, implying that $\chi([g, \tau]) = 1$ for every $g \in G_0$ and χ would be trivial (again, the subgroup spanned by $[\tau, G_0]$ is dense in G_2), a contradiction. Thus this group is discrete and in particular the set

$$\Lambda := \{k \in \mathbb{Z}: \widetilde{\chi}^k \in L\}$$

is finite.

For every $n \in \mathbb{Z}$, write

$$(35) \quad \theta_n = \chi \circ \widehat{\gamma}_n.$$

We claim that for every $\theta \in \widehat{Z}_1$,

$$(36) \quad |\{n \in \mathbb{Z}: \theta_n = \theta\}| \leq |\Lambda|.$$

To check this, let m and n be integers with $\theta_m = \theta_n$. By definitions (35) and (33), for every $g \in G$ we have $\chi([\gamma_m, g]) = \chi([\gamma_n, g])$ and, by the choice (34) of γ_n and g_n , we have $\chi(\tau^{m-n}, g) = \chi([g_n g_m^{-1}, g]) = \widetilde{\chi}_{m,n} \circ q(g)$. By (32), we conclude that $\widetilde{\chi}^{m-n} = \widetilde{\chi}_{m,n} \in \Lambda$ and the claim follows.

Let $f \in \mathcal{H}_\chi$ be a function belonging to the space $\mathcal{C}^k(X)$ of k -times differentiable functions on X , for some k to be defined later. Note that for every $n \in \mathbb{Z}$, the function $x \mapsto f(g_n \cdot x) \cdot \overline{f(x)}$ belongs to the space \mathcal{H}_1 and can be written as $h_n \circ p$ for some function h_n on Z_1 . Since all the elements g_n belong to the compact subset K of G and the action of G by translation on $\mathcal{C}^k(X)$ is continuous (with respect to the usual topology of $\mathcal{C}^k(X)$), all functions $x \mapsto f(g_n \cdot x) \cdot \overline{f(x)}$ belong to some compact subset of $\mathcal{C}^k(X)$. It follows that all the functions h_n belong to some compact subset of $\mathcal{C}^k(Z_1)$. Taking k to be sufficiently large, this

implies that

$$(37) \quad \sum_{\theta \in \widehat{Z}_1} \sup_{n \in \mathbb{Z}} |\widehat{h}_n(\theta)| < +\infty.$$

The Fourier-Stieljes transform of the spectral measure σ_f of f is given by:

$$\begin{aligned} \widehat{\sigma}_f(n) &:= \int f(T^n x) \cdot \overline{f(x)} d\mu(x) = \int f(g_n \gamma_n \cdot x) \cdot \overline{f(x)} d\mu(x) \text{ by (34)} \\ &= \int f(g_n \widehat{\gamma}_n \circ p(x) \cdot x) \cdot \overline{f(x)} d\mu(x) \text{ by (33)} \\ &= \int \chi \circ \widehat{\gamma}_n \circ p(x) \cdot f(g_n \cdot x) \cdot \overline{f(x)} d\mu(x) \text{ by the definition of } \mathcal{H}_\chi \\ &= \int_{Z_1} \theta_n(z) h_n(z) dm_{Z_1}(z) \text{ by (35) and the definition of } h_n. \end{aligned}$$

We deduce:

$$\begin{aligned} \sum_{n \in \mathbb{Z}} |\widehat{\sigma}_f(n)| &= \sum_{n \in \mathbb{Z}} |\widehat{h}_n(\theta_n^{-1})| \\ &\leq \sum_{\theta \in \widehat{Z}_1} |\{n \in \mathbb{Z} : \theta_n = \theta\}| \sup_{n \in \mathbb{Z}} |\widehat{h}_n(\theta)| \\ &\leq |\Lambda| \sum_{\theta \in \widehat{Z}_1} \sup_{n \in \mathbb{Z}} |\widehat{h}_n(\theta)| < +\infty \text{ by (36) and (37)}. \end{aligned}$$

Therefore, the spectral measure of f is absolutely continuous with respect to the Lebesgue measure $m_{\mathbb{T}}$ of \mathbb{T} . By density, this property extends to every function in \mathcal{H}_χ . \square

We use this to complete the proof of Proposition 2.1.

A.3. End of the proof of Proposition 2.1. For $\chi = 1$, note that a function belonging to \mathcal{H}_1 has discrete spectral measure.

By ergodicity, $\{\alpha^n : n \in \mathbb{Z}\}$ is dense in Z_1 . The group of eigenvalues of (X, T) is $E := \{\theta(\alpha) : \theta \in \widehat{Z}\}$. Since Z_1 admits a (nontrivial) torus as an open subgroup, E is dense in \mathbb{T} .

Let χ be a nontrivial character of G_2 . For every $\theta \in \widehat{Z}_1$, the space \mathcal{H}_χ is invariant under multiplication by the function $\theta \circ p$, as this function belongs to $L^\infty(\mu) \cap \mathcal{H}_1$. For $f \in \mathcal{H}_\chi$, the spectral measure of $f \cdot \theta \circ p$ is equal to the image of σ_f under translation by $\theta(\alpha)$. Therefore, the maximal spectral type σ_χ of the restriction of T to \mathcal{H}_χ is quasi-invariant under translation by λ for every $\lambda \in E$, meaning that if $A \subset \mathbb{T}$ satisfies $\sigma_\chi(A) = 0$, then $\sigma_\chi(A + \lambda) = 0$. But σ_χ is absolutely continuous

with respect to Lebesgue measure and so σ_χ is equivalent to Lebesgue measure. It follows that there exists a function $f_\chi \in \mathcal{H}_\chi$ such that $\sigma_{f_\chi} = m_{\mathbb{T}}$.

On the other hand, the invariant spaces \mathcal{H}_χ , for $\chi \neq 1 \in \widehat{G}_2$, are mutually orthogonal, completing the proof. \square

APPENDIX B. PROOF OF PROPOSITION 3.5

For convenience, we repeat the statement of Proposition 3.5:

Proposition. *Let \mathbb{R}^d be endowed with the Euclidean norm $\|\cdot\|$ and let the Lebesgue measure of a Borel subset K of \mathbb{R}^d be written $|K|$. Let A be a $d \times d$ matrix and assume that it is unipotent, meaning that $(\text{id} - A)^d = 0$. For every integer $n \geq 2$, let*

$$\mathcal{W}_n = \{ \xi \in \mathbb{R}^d : \|A^k \xi\| \leq 1 \text{ for } 1 \leq k < n \}.$$

If

$$p = \sum_{k=1}^{d-1} \dim(\text{range}(\text{id} - A)^k),$$

there exist positive constants C and C' (depending on d and on A) such that

$$Cn^{-p} \leq |\mathcal{W}_n| \leq C'n^{-p}$$

for every n .

Notation. Let J_r denote the $r \times r$ upper triangular elementary Jordan matrix whose entries are given by

$$J_{r,i,j} = \begin{cases} 1 & \text{for } 1 \leq i \leq r \text{ and } j = i; \\ 1 & \text{for } 1 \leq i \leq r - 1 \text{ and } j = i + 1; \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the matrix J_r has 1's on the diagonal and on the superdiagonal, and 0's elsewhere.

We begin with a lemma:

Lemma B.1. *There exists a constant $C = C(r)$ such that*

$$(38) \quad \text{if } |x_j| \leq \frac{1}{n^{j-1}} \text{ for some } n \geq 2 \text{ and } 1 \leq j \leq r,$$

$$\text{then } \left| \sum_{j=1}^r (J_r^k)_{i,j} x_j \right| \leq C \text{ for } 1 \leq i \leq r \text{ and } 0 \leq k < n.$$

On the other hand, there exists a constant $C' = C'(r) > 0$ such that

(39)

$$\text{if } \left| \sum_{j=1}^r (J_r^k)_{i,j} x_j \right| \leq 1 \text{ for } 1 \leq i \leq r \text{ and } 0 \leq k < n \text{ for some } n \geq 2,$$

$$\text{then } |x_j| \leq \frac{C'}{n^{j-1}} \text{ for } 1 \leq j \leq r.$$

Proof. For $k \geq 1$,

$$(40) \quad (J_r^k)_{i,j} = \binom{k}{j-i},$$

where we make use of the convention that $\binom{k}{p} = 0$ if $p < 0$ or $p > k$.

To prove the first statement, assume that x_1, \dots, x_r satisfy the hypothesis of (38). Then for $1 \leq i \leq r$,

$$\left| \sum_{j=1}^r (J_r^k)_{i,j} x_j \right| \leq \sum_{j=i}^r \binom{k}{j-i} k^{-j+1} \leq C_1(r) \sum_{j=i}^r k^{j-i} k^{-j+1} \leq C_2(r),$$

completing the proof of (38).

The proof of (39) requires more work. Assume that x_1, \dots, x_r satisfy the hypothesis of (39).

Taking $k = 0$, we have that

$$(41) \quad |x_j| \leq 1 \text{ for } 1 \leq j \leq r.$$

Thus, without loss of generality, we can restrict ourselves to the case that n is sufficiently large, and assume that $n \geq 2^r + 1$. Define the integer $q \geq 1$ by

$$(42) \quad 2^r q \leq n - 1 < 2^r(q + 1).$$

In the sequel, we only make use of hypothesis (39) with $i = 1$ and $k = 2^m q$ with $0 \leq m \leq r$. Formula (40) for the coefficients of the matrix J_r^k gives

$$\left| \sum_{j=1}^r \binom{2^m q}{j-1} x_j \right| \leq 1 \text{ for } 0 \leq m \leq r.$$

Since $\binom{2^m q}{0} = 1$ and $|x_1| \leq 1$, by (41) we have that

$$(43) \quad \left| \sum_{j=2}^r \binom{2^m q}{j-1} x_j \right| \leq 2 \text{ for } 0 \leq m \leq r.$$

Define

$$(44) \quad p_{2,j}(k) = \frac{1}{k} \binom{k}{j-1} \text{ for } 2 \leq j \leq r.$$

Then $p_{2,j}$ is a polynomial of degree exactly $j-2$ in the variable k . Formula (43) implies that

$$(45) \quad \left| \sum_{j=2}^r p_{2,j}(2^m q) x_j \right| \leq \frac{2}{2^m q} \text{ for } 0 \leq m \leq r.$$

We continue by induction and assume that for some ℓ with $2 \leq \ell < r$, we have

$$(46) \quad \left| \sum_{j=\ell}^r p_{\ell,j}(2^m q) x_j \right| \leq \frac{C}{(2^m q)^{\ell-1}} \text{ for } \ell-2 \leq m \leq r,$$

where $p_{\ell,j}(k)$ is a polynomial of degree exactly $j-\ell$ for $\ell \leq j \leq r$.

The same formula applied with $m-1$ substituted for m leads to

$$(47) \quad \left| \sum_{j=\ell}^r p_{\ell,j}(2^{m-1} q) x_j \right| \leq \frac{2^{\ell-1} C}{(2^m q)^{\ell-1}} \text{ for } \ell-1 \leq m \leq r.$$

For $\ell+1 \leq j \leq r$, define

$$p_{\ell+1,j}(k) = \frac{1}{k} (p_{\ell,j}(k) - p_{\ell,j}(k/2)).$$

Then $p_{\ell+1,j}(k)$ is a polynomial of degree $\leq j-\ell-1$ in the variable k . In fact, this polynomial has degree exactly $j-\ell-1$, as the coefficients of maximal degree of $p_{\ell,j}(k)$ and $p_{\ell,j}(k/2)$ are not the same.

Taking the difference between the formulas (46) and (47), the constant term $p_{\ell,\ell}(2^m q) - p_{\ell,\ell}(2^{m-1} q)$ vanishes and we have that

$$(48) \quad \left| \sum_{j=\ell+1}^r p_{\ell+1,j}(2^m q) x_j \right| \leq \frac{C}{(2^m q)^\ell} \text{ for } \ell-1 \leq m \leq r.$$

By induction, Inequality (47) is proven for $2 \leq \ell < r$. The polynomial $p_{\ell,\ell}$ is a nonzero constant and we have that

$$(49) \quad |x_r| \leq \frac{C}{(2^r q)^{r-1}}.$$

By backwards induction, we now show that

$$(50) \quad |x_\ell| \leq \frac{C}{(2^r q)^{\ell-1}} \text{ for } 1 \leq \ell \leq r.$$

For $\ell = r$, this is exactly (49). Assume that $1 \leq \ell < r$ and that this bound holds for $\ell + 1, \ell + 2, \dots, r$. By (46) applied with $m = r$,

$$|x_\ell| \leq \frac{C}{(2^r q)^{\ell-1}} + \sum_{j=\ell+1}^r |p_{\ell,j}(2^r q)| \frac{C}{(2^r q)^{j-1}}.$$

Since $p_{\ell,j}$ is a polynomial of degree $j - \ell$, we have $|p_{\ell,j}(k)| \leq Ck^{j-\ell}$ for some $C > 0$ and (50) follows.

We conclude the proof by using (42) to conclude that $2^r q > (n - 1)q/(q + 1) \geq n/4$. \square

Corollary B.2. *Let \mathbb{R}^r be endowed with the supremum norm $\|\cdot\|_\infty$ and $|K|$ denote the Lebesgue measure of a Borel subset K of \mathbb{R}^r . Then for every $n \geq 2$, we have*

$$Cn^{-r(r-1)/2} \leq \left| \left\{ x \in \mathbb{R}^r : \|J_r^k x\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\} \right| \leq C'n^{-r(r-1)/2}.$$

Using this, we complete the proof of Proposition 3.5. Since the matrix A is unipotent, there exists a $d \times d$ invertible matrix Φ such that the matrix $B := \Phi A \Phi^{-1}$ is in Jordan form. Thus all the entries of B are all equal to 0 other than $m \geq 1$ diagonal square blocks, each of which is an elementary Jordan $r_j \times r_j$ matrix J_{r_j} for $1 \leq j \leq m$.

We have

$$|\mathcal{W}_n| = |\det(\Phi)|^{-1} |\Phi(\mathcal{W}_n)|$$

for every n . Moreover,

$$\Phi(\mathcal{W}_n) = \left\{ \eta \in \mathbb{R}^d : \|\Phi B^k \eta\| \leq 1 \text{ for } 0 \leq k < n \right\}$$

and thus there exist positive constants c and c' with

$$\begin{aligned} c \left\{ \eta \in \mathbb{R}^d : \|B^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\} &\subseteq \Phi(\mathcal{W}_n) \\ &\subseteq c' \left\{ \eta \in \mathbb{R}^d : \|B^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\}. \end{aligned}$$

Therefore

$$\begin{aligned} c^d \left| \left\{ \eta \in \mathbb{R}^d : \|B^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\} \right| &\leq |\Phi(\mathcal{W}_n)| \\ &\leq c'^d \left| \left\{ \eta \in \mathbb{R}^d : \|B^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\} \right|. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left\{ \eta \in \mathbb{R}^d : \|B^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\} \\ = \prod_{j=1}^m \left\{ \eta \in \mathbb{R}^{r_j} : \|J_{r_j}^k \eta\|_\infty \leq 1 \text{ for } 0 \leq k < n \right\}. \end{aligned}$$

Combining these remarks with Corollary B.2, we have that

$$Cn^{-p} \leq |\mathcal{W}_n| \leq C'n^{-p},$$

where

$$\begin{aligned} p &= \sum_{j=1}^m \frac{r_j(r_j - 1)}{2} = \sum_{j=1}^m \sum_{k=1}^{r_j-1} k \\ &= \sum_{j=1}^m \sum_{k=1}^{d-1} k (\dim(\ker(\text{id} - J_{r_j})^{k+1}) - \dim(\ker(\text{id} - J_{r_j})^k)) \\ &= \sum_{k=1}^{d-1} k (\dim(\ker(\text{id} - A)^{k+1}) - \dim(\ker(\text{id} - A)^k)) \\ &= \sum_{k=1}^{d-1} \dim(\text{range}(\text{id} - A^k)). \quad \square \end{aligned}$$

REFERENCES

- [1] L. AUSLANDER, L. GREEN & F. HAHN. *Flows on homogeneous spaces*. Ann. Math. Studies **53**, Princeton Univ. Press, 1963.
- [2] V. BERGELSON. Weakly mixing PET. *Erg. Th. & Dyn. Sys.*, **7** (1987), 337–349.
- [3] V. BERGELSON, B. HOST & B. KRA, with an appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Invent. Math.*, **160** (2005), 261–303.
- [4] V. BERGELSON, A. LEIBMAN, A & E. LESIGNE, Intersective polynomials and the polynomial Szemerédi theorem. *Adv. Math.*, **219** (2008), no. 1, 369–388.
- [5] F. BLANCHARD, B. HOST & A. MAASS. Topological complexity. *Erg. Th. & Dyn. Sys.*, **20** (2000), no. 3, 641–662.
- [6] X. BRESSAUD, F. DURAND & A. MAASS. On the eigenvalues of finite rank Bratteli-Vershik dynamical systems. *Erg. Th. & Dyn. Sys.*, **30** (2010), no. 3, 639–664.
- [7] Q. CHU. Convergence of weighted polynomial multiple ergodic averages. *Proc. Amer. Math. Soc.*, **137** (2009), no. 4, 1363–1369.
- [8] F. DURAND. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Erg. Th. & Dyn. Sys.*, **20** (2000), no. 4, 1061–1078.
- [9] P. DONG, S. DONOSO, A. MAASS, S. SHAO & X. YE. Infinite-step nil-systems, independence and complexity. To appear in *Erg. Th. & Dyn. Sys.*, [arXiv:1105.3584](https://arxiv.org/abs/1105.3584)
- [10] S. FERENCZI. Systèmes localement de rang un. *Ann. Inst. H. Poincaré Probab. Statist.* **20** (1984), no. 1, 35–51.
- [11] FERENCZI, S. Systèmes de rang un gauche. *Ann. Inst. H. Poincaré Probab. Statist.* **21** (1985), no. 2, 177–186.
- [12] S. FERENCZI. Rank and symbolic complexity. *Erg. Th. & Dyn. Sys.*, **16** (1996), no. 4, 663–682.

- [13] S. FERENCZI. Measure-theoretic complexity of ergodic systems. *Israel J. of Math.*, **100** (1997), no. 4, 189–207.
- [14] N. FRANTZIKINAKIS. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.*, **360** (2008), no. 10, 5435–5475.
- [15] N. FRANTZIKINAKIS. Multiple recurrence and convergence for Hardy field sequences of polynomial growth. *J. d'Analyse Math.*, **112** (2010), 79–135
- [16] N. FRANTZIKINAKIS & B. KRA. Ergodic averages for independent polynomials and applications. *J. Lond. Math. Soc.*, **74** (2006), 131–142.
- [17] N. FRANTZIKINAKIS & M. WIERDL. A Hardy field extension of Szemerédi's theorem. *Adv. Math.*, **222** (2009), 1–43
- [18] R. GJERDE & Ø. JOHANSEN. Bratteli-Vershik models for Cantor minimal systems associated to interval exchange transformations. *Math. Scand.*, **90** (2002), no. 1, 87–100.
- [19] B. GREEN & T. TAO. Linear equations in primes. *Annals of Math.*, **171** (2010), no. 3, 1753–1850.
- [20] B. HOST & B. KRA. Nonconventional ergodic averages and nilmanifolds. *Annals of Maths.*, **161** (2005), 397–488.
- [21] B. HOST & B. KRA. Convergence of polynomial ergodic averages. *Israel J. of Math.*, **149** (2005), 1–20.
- [22] B. HOST & B. KRA. Uniformity seminorms on ℓ^∞ and applications. *J. Anal. Math.*, **108** (2009), 219–276.
- [23] B. HOST, B. KRA & A. MAASS. Nilsequences and a topological structure theorem. *Adv. Math.*, **224** (2010), no. 1, 103–129.
- [24] A. KATOK & J.-P. THOUVENOT. Slow entropy type invariants and smooth realization of commuting measure-preserving transformations. *Ann. Inst. H. Poincaré Probab. Statist.*, **33** (1997), no. 3, 323–338.
- [25] A. LEIBMAN. Polynomial sequences in groups. *J. Algebra* **201** (1998), no. 1, 189–206.
- [26] A. LEIBMAN. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Erg. Th. & Dyn. Sys.*, **25** (2005), no. 1, 201–213.
- [27] A. LEIBMAN. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.*, **146** (2005), 303–316.
- [28] A. MALCEV. *On a class of homogeneous spaces*. Amer. Math. Soc. Transl. **39** (1951).
- [29] D. ORNSTEIN, D. RUDOLPH & B. WEISS. *Equivalence of measure preserving transformations*. Mem. Amer. Math. Soc., **37** (1982), no. 262.
- [30] W. PARRY. Dynamical systems on nilmanifolds. *Bull. London Math. Soc.*, **2** (1970), 37–40.
- [31] M. QUEFFÉLEC. *Substitution dynamical systems — spectral analysis*. Second edition. Lecture Notes in Mathematics, Vol. 1294. Springer-Verlag, Berlin, 2010.
- [32] S. SHAO & X. YE. Regionally proximal relation of order d is an equivalence one for minimal systems and a combinatorial consequence. [arXiv:1007.0189](https://arxiv.org/abs/1007.0189).
- [33] STARKOV. *Dynamical systems on homogeneous spaces*. Translated from the 1999 Russian original by the author. Translations of Mathematical Monographs, 190. American Mathematical Society, Providence, RI, 2000.

- [34] A.M. STEPIN. Flows on solvable manifolds. *Upekhi Mat. Nauk* **24** (1969), no. 2, 241–242 (Russian).
- [35] P. WALTERS *An Introduction to Ergodic Theory*. Graduate Texts in Mathematics, **79**. Springer-Verlag, New York-Berlin, 1982.
- [36] H. WEYL. Über die Gleichverteilung von Zahlen mod Eins. *Math. Ann.*, **77** (1916), 313–352.

BH: LABORATOIRE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES, UNIVERSITÉ PARIS-EST MARNE LA VALLÉE & CNRS UMR 8050, 5 BD. DESCARTES, CHAMPS SUR MARNE, 77454 MARNE LA VALLÉE CEDEX 2, FRANCE

E-mail address: `bernard.host@univ-mlv.fr`

BK: DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, 2033 SHERIDAN ROAD EVANSTON, IL 60208-2730, USA

E-mail address: `kra@math.northwestern.edu`

AM: DEPARTAMENTO DE INGENIERÍA MATEMÁTICA, UNIVERSIDAD DE CHILE & CENTRO DE MODELAMIENTO MATEMÁTICO UMI 2071 UCHILE-CNRS, CASILLA 170/3 CORREO 3, SANTIAGO, CHILE.

E-mail address: `amaass@dim.uchile.cl`