

# Pointwise convergence in nilmanifolds along smooth functions of polynomial growth

KONSTANTINOS TSINAS

*University of Crete, Department of Mathematics and Applied Mathematics,  
Voutes University Campus, Heraklion 71003, Greece  
(e-mail: kon.tsinas@gmail.com)*

*(Received 12 April 2022 and accepted in revised form 31 July 2023)*

*Abstract.* We study the equidistribution of orbits of the form  $b_1^{a_1(n)} \cdots b_k^{a_k(n)} \Gamma$  in a nilmanifold  $X$ , where the sequences  $a_i(n)$  arise from smooth functions of polynomial growth belonging to a Hardy field. We show that under certain assumptions on the growth rates of the functions  $a_1, \dots, a_k$ , these orbits are equidistributed on some subnilmanifold of the space  $X$ . As an application of these results and in combination with the Host–Kra structure theorem for measure-preserving systems, as well as some recent seminorm estimates of the author for ergodic averages concerning Hardy field functions, we deduce a norm convergence result for multiple ergodic averages. Our method mainly relies on an equidistribution result of Green and Tao on finite segments of polynomial orbits on a nilmanifold [The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)* **175** (2012), 465–540].

Key words: ergodic averages, equidistribution, nilmanifolds, Hardy fields

2020 Mathematics Subject Classification: 22F30 (Primary); 37A17 (Secondary)

## 1. Introduction and main results

1.1. *History and main goals.* In recent years, there has been an active interest in determining the limiting behaviour of the multiple ergodic averages

$$\frac{1}{N} \sum_{n=1}^N f_1(T^{a_1(n)}x) \cdots f_k(T^{a_k(n)}x) \quad (1)$$

for various sequences  $a_1(n), \dots, a_k(n)$  of integers, where  $T$  is an invertible measure-preserving transformation acting on a probability space  $(X, \mathcal{X}, \mu)$  and  $f_1, \dots, f_k$  are functions in  $L^\infty(\mu)$ . Through the breakthrough work of Furstenberg [8], which delivered a new proof of Szemerédi’s theorem using tools from ergodic theory, it has been apparent that the analysis of the averages in equation (1) has noteworthy applications to number



theory and combinatorics. In particular, we now have substantial generalizations of Szemerédi's theorem, some of which have not been demonstrated with approaches other than the use of ergodic theory.

An integral tool in verifying convergence of the averages in equation (1) is the structure theorem of Host and Kra [11], which in multiple cases reduces the above problem to studying rotations on particular spaces with algebraic structure, which are called nilmanifolds (see [12] for a full presentation of the theory). A nilmanifold is a homogeneous space  $X = G/\Gamma$ , where  $G$  is a nilpotent Lie group and  $\Gamma$  is a discrete cocompact subgroup. The study of nilmanifolds is essential due to its ties to ergodic theory mentioned above, as well as the numerous applications to combinatorics and number theory.

In this article, our central problem is the study of the distribution of orbits in a nilmanifold along sequences that arise from smooth functions with polynomial growth. We suppose that our functions are elements of a Hardy field (for the definition of a Hardy field, we direct the reader to §2). The benefit of working within a Hardy field is that certain 'regularity' properties of the derivatives of a function, which are vital in several parts of our proofs, can be extrapolated from a simple growth condition on the initial function. For instance, a condition like equation (P) below imposes multiple pleasant properties on the derivatives of a function in  $\mathcal{H}$ .

The field of logarithmico-exponential functions is the prototypical example of a Hardy field. It is defined as the collection of functions formed by a finite combination of the operations  $+$ ,  $-$ ,  $\cdot$ ,  $\div$ ,  $\exp$ ,  $\log$  and composition of functions acting on a real variable  $t$  (which takes values on some half-line  $[x, +\infty)$ ) and real constants. The fact that it is a Hardy field was established in [10]. Our results are most interesting for the Hardy field  $\mathcal{LE}$  and one can keep this particular case in mind throughout the article. In addition, we refer the reader to Appendix B for the definition and properties of nilmanifolds, which appear in the subsequent discussion and the main theorems.

Due to its connections to ergodic theory and combinatorics, the investigation of equidistribution properties along Hardy sequences has been carried out several times throughout the literature. First of all, we recall a fundamental result concerning the equidistribution of Hardy sequences, which corresponds to the basic case when the underlying nilmanifold is a finite-dimensional torus. In particular, we restate here [2, Theorem 1.3].

**THEOREM A. (Boshernitzan)** *Let the function  $a \in \mathcal{H}$  have polynomial growth. Then, the sequence  $a(n)$  is equidistributed mod 1 if and only if*

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - p(t)|}{\log t} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Q}[t]. \quad (\text{P})$$

Applying Weyl's equidistribution theorem and the previous result, we can effortlessly show that if the functions  $a_1, \dots, a_k$  have polynomial growth and each non-trivial linear combination of them stays logarithmically away from real multiples of integer polynomials, then the sequence  $(c_1 a_1(n), \dots, c_k a_k(n))$  is equidistributed on  $\mathbb{T}^k$  for all non-zero real numbers  $c_1, \dots, c_k$ . Practically, Theorem A can be used to examine orbits on  $\mathbb{T}^k$  along the sequences  $a_1(n), \dots, a_k(n)$  on  $\mathbb{T}^k$ , answering our problem in the case when the nilmanifold  $X$  is any finite-dimensional torus (the abelian case). Another corollary

of Theorem A is that if  $a(t) \in \mathcal{H}$  stays logarithmically away from real multiples of integer polynomials, then the sequence  $\lfloor a(n) \rfloor a$  is equidistributed on  $\mathbb{T}$  for all irrational  $a \in (0, 1)$ . This phenomenon (namely, that equidistribution properties of  $a(n)$  yield information for the equidistribution properties of  $\lfloor a(n) \rfloor$ ) will be present throughout the article, so the reader can view statements involving  $a(n)$  in place of  $\lfloor a(n) \rfloor$  as being morally the same.

Suppose now that we are given a nilmanifold  $X = G/\Gamma$  (for the definitions of all terms below, see Appendix B) and assume that the group  $G$  is connected and simply connected. We are interested in the behaviour of the sequence

$$v(n) = (b_1^{\lfloor a_1(n) \rfloor} \Gamma, \dots, b_k^{\lfloor a_k(n) \rfloor} \Gamma), \quad (2)$$

where  $b_1, \dots, b_k$  are elements of the group  $G$  and  $a_1, \dots, a_k$  are Hardy field functions. Notice that this is a sequence on the product nilmanifold  $X^k$ . The most fundamental equidistribution result is due to Leibman, who showed that if the functions  $a_1, \dots, a_k$  are integer polynomials, then we have equidistribution on a ‘subspace’ of  $X$  (called a subnilmanifold), as long as we restrict the values of  $n$  to appropriate arithmetic progressions.

More specifically, we present the following theorem [13, Theorem B], an application of which (on the nilmanifold  $X^k$ ) implies the claim in the previous paragraph.

**THEOREM B. (Leibman)** *Let  $X = G/\Gamma$  be a nilmanifold and  $x \in X$ . Consider the sequence*

$$g(n) = b_1^{p_1(n)} \dots b_k^{p_k(n)} \quad (3)$$

*in  $G$ , where  $b_1, \dots, b_k \in G$  and  $p_1, \dots, p_k$  are polynomials with integer coefficients. Then, there exists  $Q \in \mathbb{N}$ , a closed, connected and rational subgroup  $H$  of  $G$  and points  $x_0, \dots, x_{Q-1} \in X$ , such that for every  $r \in \{0, \dots, Q-1\}$ , the sequence  $g(Qn+r)x$  is equidistributed on the subnilmanifold  $Hx_r$ .*

A noteworthy corollary of the previous theorem is that if  $F : X \rightarrow \mathbb{C}$  is a continuous function, then the averages

$$\frac{1}{N} \sum_{n=1}^N F(g(n)x)$$

converge pointwise for all  $x \in X$ . This can be used in conjunction with the Host–Kra structure theory (see Theorem E in §2) to infer that the averages in equation (1) converge in norm, when the sequences  $a_1(n), \dots, a_k(n)$  are integer polynomial sequences. In addition, we deduce (as a corollary of [13, Theorem C] in the same paper) that if  $G$  is connected, the equidistribution of the sequence  $g(n)\Gamma$  is controlled by the projection of  $g(n)\Gamma$  on the ‘abelianization’  $G/[G, G]\Gamma$  of  $G/\Gamma$ , which is a finite-dimensional torus called the horizontal torus of  $X$ .

A major improvement of the above theorem was established by Green and Tao in [9], who characterized the behaviour of polynomial orbits on nilmanifolds in quantitative language. (While their theorem was established under the stronger hypothesis that the

underlying Lie group  $G$  is connected and simply connected, one can typically reduce to this case in many applications.) This theorem has notable applications in number theory and will be undoubtedly vital in this paper. Like Leibman’s theorem in [13] that we mentioned above briefly, this theorem highlights the relation of the equidistribution properties of a polynomial sequence (see Definition B.3) on a nilmanifold with its projection to the horizontal torus. Since there are many technical terms that are required to state this theorem, we have presented its statement in Appendix B along with a sample corollary when the nilmanifold is a torus, as well as all of the required background on the quantitative equidistribution theory on nilmanifolds.

Now, let us consider the more general case when the sequences  $a_1(n), \dots, a_k(n)$  appearing in equation (2) are not just integer polynomials, but functions that belong to a Hardy field  $\mathcal{H}$ . In the case  $k = 1$ , Frantzikinakis established [4] that if the function  $a(t)$  satisfies

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - cp(t)|}{\log t} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Z}[t],$$

then the sequence  $b^{\lfloor a_1(n) \rfloor} x$  is equidistributed on the orbit  $Y = \overline{\{b^n x : n \in \mathbb{N}\}}$  of  $b$  for any  $b \in G$  and  $x \in X$ . In the case of general  $k$ , he also established the following theorem in the same paper.

**THEOREM C.** (Frantzikinakis [4, Theorem 1.3]) *Let  $a_1, \dots, a_k$  be functions of polynomial growth that belong to a Hardy field  $\mathcal{H}$ , such that they have pairwise distinct growth rates and satisfy*

$$t^{k_i} \log t < a_i(t) < t^{k_i+1} \tag{4}$$

for some  $k_i \in \mathbb{N}$ . Then, for any nilmanifold  $X = G/\Gamma$  and  $b_1, \dots, b_k \in G$ , the sequence

$$\left( b_1^{\lfloor a_1(n) \rfloor} x_1, \dots, b_k^{\lfloor a_k(n) \rfloor} x_k \right)_{n \in \mathbb{N}} \tag{5}$$

is equidistributed on  $\overline{(b_1^n x_1)}_{n \in \mathbb{N}} \times \dots \times \overline{(b_k^n x_k)}_{n \in \mathbb{N}}$  for all  $x_1, \dots, x_k \in X$ .

In the same paper, Frantzikinakis conjectured that if the linear combinations of the functions  $a_1, \dots, a_k$  stay logarithmically away from real multiples of integer polynomials, then the sequence in equation (5) is equidistributed on  $\overline{(b_1^n x_1)}_{n \in \mathbb{N}} \times \dots \times \overline{(b_k^n x_k)}_{n \in \mathbb{N}}$ . More specifically, we have the following conjecture.

*Conjecture 1.* [4] Let  $a_1, \dots, a_k$  be functions in a Hardy field  $\mathcal{H}$  with polynomial growth and such that every non-trivial linear combination  $a(t)$  of them satisfies

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - p(t)|}{\log t} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Z}[t].$$

Then, for any nilmanifold  $X = G/\Gamma$ ,  $b_i \in G$  and  $x_i \in X$ , the sequence

$$\left( b_1^{\lfloor a_1(n) \rfloor} x_1, \dots, b_k^{\lfloor a_k(n) \rfloor} x_k \right)_{n \in \mathbb{N}}$$

is equidistributed on  $\overline{(b_1^n x_1)}_{n \in \mathbb{N}} \times \dots \times \overline{(b_k^n x_k)}_{n \in \mathbb{N}}$ .

Recently, Richter established the following equidistribution theorem. We present here a special case of that result, where we assume that the underlying Lie group  $G$  is connected and simply connected so that the elements  $b^s$  are defined for any  $b \in G$  and  $s \in \mathbb{R}$  (see also the first paragraph of §B.2 for a more thorough explanation). We also define

$$\nabla = \text{span}\{a_1, \dots, a_k\} = \{c_1 a_1^{(n_1)}(t) + \dots + c_k a_k^{(n_k)}(t) : c_i \in \mathbb{R}, n_i \in \mathbb{N} \cup \{0\}\}.$$

**THEOREM D.** (Richter [17, Theorem B]) *Let  $X = G/\Gamma$  be a nilmanifold with  $G$  connected and simply connected, and let  $a_1, \dots, a_k$  be functions in a Hardy field  $\mathcal{H}$ , such that for any function  $a \in \nabla = \text{span}\{a_1, \dots, a_k\}$ , we have that*

$$|a(t) - p(t)| \ll 1 \quad \text{or} \quad |a(t) - p(t)| > \log t$$

for any polynomial  $p(t) \in \mathbb{R}[t]$ . Consider any commuting elements  $b_1, \dots, b_k \in G$  and define the sequence

$$v(n) = b_1^{a_1(n)} \dots b_k^{a_k(n)}.$$

Then, there exists a closed, connected and rational subgroup  $H$  of  $G$ , and points  $x_0, \dots, x_{Q-1}$  in  $X$ , such that the sequence  $v(Qn + r)\Gamma$  is equidistributed on the subnilmanifold  $Hx_r$  of  $X$  for all  $r \in \{0, \dots, Q - 1\}$ .

The hypothesis that  $b_1, \dots, b_k$  are commuting is harmless in problems regarding the convergence of ergodic averages or in applications to combinatorics. Furthermore, while in this setting we have the sequences  $a_i(n)$  instead of  $\lfloor a_i(n) \rfloor$  in the exponents, the statement above actually implies an equidistribution theorem for the sequences  $\lfloor a_i(n) \rfloor$ . We remark that the results in [17] are generalized to equidistribution results with respect to (weaker) averaging schemes other than Cesàro averages. Under those averaging schemes, the assumptions on the functions  $a_1, \dots, a_k$  can be weakened significantly (however, our results deal only with Cesàro averages). In the follow-up paper [1], Bergelson, Moreira and Richter employed the above equidistribution results to obtain convergence results for multiple ergodic averages and combinatorial applications for Hardy field sequences.

**1.2. Main results.** To state our results, we will assume that we have a fixed Hardy field  $\mathcal{H}$ , and the only extra hypothesis we require is that it includes the polynomial functions (this is a very mild restriction). Removing this restriction may be possible, though this would certainly complicate our arguments or the notation in the proofs. Unless noted otherwise, our theorems below apply to any such Hardy field. An exception is made only for Theorem 1.3 (we shall reiterate these assumptions in the main theorems).

For a given set of functions  $a_1, \dots, a_k$  in our Hardy field  $\mathcal{H}$ , we use the notation

$$\mathcal{L}(a_1, \dots, a_k) = \{c_1 a_1 + \dots + c_k a_k : (c_1, \dots, c_k) \in \mathbb{R}^k \setminus \{0\}\} \tag{6}$$

to refer to the collection of functions in  $\mathcal{H}$  that are non-trivial linear combinations of the functions  $a_1(t), \dots, a_k(t)$ . The nilmanifolds  $\overline{(b^{\mathbb{R}x})}$  and  $\overline{(b^{\mathbb{N}x})}$  are defined in §B.2.

**THEOREM 1.1.** *Let  $\mathcal{H}$  be a Hardy field containing the polynomial functions. Let  $a_1, \dots, a_k$  be functions in  $\mathcal{H}$  that have polynomial growth. Assume that there exists*

an  $\varepsilon > 0$  (the value of  $\varepsilon$  depends only on the initial collection  $\{a_1, \dots, a_k\}$ ), such that every function  $a \in \mathcal{L}(a_1, \dots, a_k)$  satisfies

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - p(t)|}{t^\varepsilon} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Q}[t]. \tag{7}$$

(Equivalently, we could require that  $p(t) \in \mathbb{Z}[t]$ , because this is a condition on all the linear combinations of the functions  $a_1, \dots, a_k$ .) Then, we have the following.

- (i) For any collection of nilmanifolds  $X_i = G_i / \Gamma_i$ , elements  $b_i \in G_i$  and  $x_i \in X_i$ , the sequence

$$(b_1^{\lfloor a_1(n) \rfloor} x_1, \dots, b_k^{\lfloor a_k(n) \rfloor} x_k)$$

is equidistributed on the nilmanifold  $\overline{(b_1^{\mathbb{N}} x_1)} \times \dots \times \overline{(b_k^{\mathbb{N}} x_k)}$ .

- (ii) For any collection of nilmanifolds  $X_i = G_i / \Gamma_i$  such that the groups  $G_i$  are connected, simply connected, elements  $b_i \in G_i$  and  $x_i \in X_i$ , the sequence

$$(b_1^{a_1(n)} x_1, \dots, b_k^{a_k(n)} x_k)$$

is equidistributed on the nilmanifold  $\overline{(b_1^{\mathbb{R}} x_1)} \times \dots \times \overline{(b_k^{\mathbb{R}} x_k)}$ .

*Remark 1.*

- (a) The connectedness assumptions imposed on part (ii) of the previous theorem ensure that all elements of the form  $b^s$  where  $b \in G$  and  $s \in \mathbb{R}$  are well defined (see also Appendix B for the definition of the element  $b^s$  for non-integer  $s$ ).
- (b) In regards to part (ii) of the previous theorem, we establish the more general statement that if  $b_1, \dots, b_k$  commute, the sequence  $b_1^{a_1(n)} \dots b_k^{a_k(n)} \Gamma$  is equidistributed on the nilmanifold  $\overline{b_1^{\mathbb{R}} \dots b_k^{\mathbb{R}} \Gamma}$ . The fact that this is indeed a more general statement can be seen by passing to the product nilmanifold  $X_1 \times \dots \times X_k$ . A similar assertion holds for Theorem 1.2 below and we provide more details on this deduction after Proposition 4.1.

Observe that, in contrast to Theorem A, we have the term  $t^\varepsilon$  in the denominator, which is just out of reach of the conjectured optimal term  $\log t$ . As an example, using Theorem 1.1, we can prove that for any elements  $b_1, b_2 \in G$ , the sequence  $(b_1^{n \log n} \Gamma, b_2^{n^{3/2}} \Gamma)$  is equidistributed on the nilmanifold  $\overline{(b_1^{\mathbb{R}} \Gamma, b_2^{\mathbb{R}} \Gamma)}$ , assuming that  $G$  satisfies the appropriate connectedness assumptions since we want these elements to be well defined.

If we have functions that are not linearly independent, then the above theorem fails, as can be seen by noting that the sequence  $(n^{3/2}, n^{1/2}, n^{3/2} + n^{1/2})$  is not equidistributed on  $\mathbb{T}^3$ . However, we can relax the linear independence condition in Theorem 1.1 and still obtain a convergence result.

**THEOREM 1.2.** *Let  $\mathcal{H}$  be a Hardy field containing the polynomial functions. Let  $a_1, \dots, a_k$  be functions in  $\mathcal{H}$  that have polynomial growth. Assume that there exists  $\varepsilon > 0$ , such that every function  $a \in \mathcal{L}(a_1, \dots, a_k)$  satisfies either*

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - p(t)|}{t^\varepsilon} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Q}[t] \tag{8}$$

or

$$\text{the limit } \lim_{t \rightarrow +\infty} a(t) \text{ is a real number.} \tag{9}$$

Then, we have the following.

- (i) For any collection of nilmanifolds  $X_i = G_i/\Gamma_i$ , elements  $b_i \in G_i$ ,  $x_i \in X_i$  and continuous functions  $f_1, \dots, f_k$  with complex values, the averages

$$\frac{1}{N} \sum_{i=1}^N f_1(b_1^{\lfloor a_1(n) \rfloor} x_1) \cdots f_k(b_k^{\lfloor a_k(n) \rfloor} x_k)$$

converge.

- (ii) For any collection of nilmanifolds  $X_i = G_i/\Gamma_i$  such that the groups  $G_i$  are connected, simply connected, elements  $b_i \in G_i$ ,  $x_i \in X_i$  and continuous functions  $f_1, \dots, f_k$  with complex values, the averages

$$\frac{1}{N} \sum_{i=1}^N f_1(b_1^{a_1(n)} x_1) \cdots f_k(b_k^{a_k(n)} x_k)$$

converge.

The main distinction between Theorems 1.1 and 1.2 is that in the second case, we allow for linear dependencies between the functions  $a_1(t), \dots, a_k(t)$  (for example, we may have the functions  $(t \log t, t^{3/2}, t^{3/2} + t \log t)$ ). We will use this theorem to deduce a convergence result for multiple ergodic averages (Theorem 1.3 below).

Theorems 1.1 and 1.2 extend the equidistribution result of Theorem C from [4], where the functions  $a_1, \dots, a_k$  were assumed to have different growth rates and satisfy the growth condition in equation (4). However, our results are complementary to the results in [17], in the sense that both Theorem 1.1 and Theorem D each cover collections of functions that are not implied by the other one. The main difference between our results and the results in the previous literature (in the case of general  $k$ ) is that prior results did not cover functions in the range  $t^\ell < a(t) \ll t^\ell \log t$ , where  $\ell$  is a positive integer. Our method circumvents this restriction and can handle all families of functions of the form  $\sum_{i=1}^k c_i t^{a_i} (\log t)^{b_i}$ , where  $a_i > 0$  and  $b_i, c_i \in \mathbb{R}$  (assuming, of course, that the linear combinations of the involved functions satisfy either equation (8) or equation (9)). However, our method has a drawback. As we stated, there are cases covered in the results of [17] that do not follow from the arguments present in this paper. These examples concern functions that grow slower than fractional powers  $t^\delta$ , such as the function  $\log^c t$  for  $c > 0$  or the function  $\exp(\sqrt{\log t})$ . An example that is not covered by Theorem 1.2 is the pair of functions  $(\log^2 t, t^{3/2})$ . However, this last pair of functions can be covered by the results in [17]. We shall discuss the techniques and limitations of our proof in depth below (§1.3).

Combining Theorem 1.2 and the results in [18] on characteristic factors, we get a mean convergence result for multiple ergodic averages. Since the seminorm estimates for such averages were established in [18] under particular assumptions on our Hardy field  $\mathcal{H}$ , these

have to be incorporated into our statement. We will not need to use these assumptions anywhere else in this article, however.

**THEOREM 1.3.** *Let  $\mathcal{H}$  be a Hardy field that contains the field  $\mathcal{LE}$  of logarithmico-exponential functions and is closed under composition and compositional inversion of functions (when defined). Furthermore, assume that the functions  $a_1, \dots, a_k \in \mathcal{H}$  are as in Theorem 1.2. Then, for any measure-preserving system  $(X, \mu, T)$  and any functions  $f_1, \dots, f_k \in L^\infty(\mu)$ , the averages*

$$\frac{1}{N} \sum_{n=1}^N T^{\lfloor a_1(n) \rfloor} f_1 \dots T^{\lfloor a_k(n) \rfloor} f_k \tag{10}$$

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

An example of a Hardy field that satisfies the above property is the Hardy field of Pfaffian functions (for the definition, see, for instance, [18, §2]).

It follows from the results in [18] that if the functions  $a_1, \dots, a_k$  are as in Theorem 1.1 (actually, the  $t^\epsilon$  term can be replaced with the optimal term  $\log t$ ), then for any ergodic measure-preserving system  $(X, \mu, T)$  and bounded functions  $f_1, \dots, f_k$ , the averages

$$\frac{1}{N} \sum_{i=1}^N f_1(T^{\lfloor a_1(n) \rfloor} x) \dots f_k(T^{\lfloor a_k(n) \rfloor} x)$$

converge in the  $L^2$ -sense to the product of the integrals  $\int f_1 d\mu \dots \int f_k d\mu$ . The methods used in that article cannot work when there are linear dependencies between the functions  $a_1, \dots, a_k$  (since they rely on the joint ergodicity results from [6]). Therefore, to prove Theorem 1.3, we have to show that the Host–Kra factors are characteristic for these averages, reduce the problem to nilmanifolds using the Host–Kra structure theorem (see Theorem E in §2) and then tackle the problem of mean convergence in nilmanifolds. The first part of the above argument follows from the results in [18] (see Proposition 4.3), while Theorem 1.2 gives the stronger result of pointwise convergence when the system  $(X, \mu, T)$  is a nilsystem. We comment here that the optimal restrictions on the functions  $a_1, \dots, a_k$  in Theorem 1.3 are expected to be that the functions are good for convergence when the system  $(X, \mu, T)$  is a rotation on some torus  $\mathbb{T}^d$ . A refuted conjecture of Frantzikinakis appears in [5, Problem 22], although the statement needs to be changed to the following (personal communication).

*Conjecture 2.* Let  $a_1, \dots, a_k$  be functions in  $\mathcal{LE}$  (or any other Hardy field) such that for all real numbers  $t_1, \dots, t_k \in [0, 1)$ , the averages

$$\frac{1}{N} \sum_{n=1}^N e(t_1 \lfloor a_1(n) \rfloor + \dots + t_k \lfloor a_k(n) \rfloor) \tag{11}$$

converge. Then, for any measure-preserving system  $(X, \mu, T)$  and functions  $f_1, \dots, f_k \in L^\infty(\mu)$ , the averages

$$\frac{1}{N} \sum_{n=1}^N T^{\lfloor a_1(n) \rfloor} f_1 \dots T^{\lfloor a_k(n) \rfloor} f_k \tag{12}$$



converge in  $L^2(\mu)$  and, if  $(X, \mu, T)$  is a nilsystem and the functions  $f_1, \dots, f_k$  are continuous, then those averages converge pointwise everywhere.

*Remark 2.* It can be shown that the above condition on the exponentials of the involved sequences is not sufficient if we replace the Hardy sequences with other, more general, sequences. Indeed, [7, Theorem B] provides an example of a sequence  $a(n)$ , such that for  $a_1(n) = a(n)$  and  $a_2(n) = 2a(n)$ , the averages

$$\frac{1}{N} \sum_{n=1}^N e(t_1 a_1(n) + t_2 a_2(n))$$

converge for any  $t_1, t_2 \in [0, 1)$ , but the ergodic averages

$$\frac{1}{N} \sum_{n=1}^N T^{a_1(n)} f_1 \cdot T^{a_2(n)} f_2$$

do not converge in mean.

1.3. *Short overview of the proof and additional remarks.* The main idea of the proof is that functions in  $\mathcal{H}$  of polynomial growth can be approximated sufficiently well by polynomials in short intervals. The equidistribution properties of polynomial sequences in nilmanifolds, even on small intervals, are well understood from [9]. We use these quantitative results of Green and Tao (see Theorem F in Appendix B) to show that the averages over small intervals are ‘close’ to the integral of a continuous function in our nilmanifold. This approach was used in [4] to show that (following the notation of Theorem 1.1) the sequence  $b^{a(n)}$  is equidistributed for all  $b \in G$  and any function  $a(n)$  satisfying equation (P). In the case that we need to cover here, there are more technical difficulties in the proof, since we have to find polynomial expansions for several functions simultaneously, which also tend to be of increased complexity (for example, choosing the length of the short intervals is fairly straightforward in the case of one function, but not when someone deals with several functions in  $\mathcal{H}$ ). This idea of using a common polynomial expansion was also used recently by the author to establish the corresponding problem of finding characteristic factors for ergodic averages involving Hardy field iterates. This approach is well suited to handle functions in the range  $t^k < a(t) < t^k \log t$ , which were previously not known in the literature. Some additional care needs to be taken to separate polynomial functions and functions that we call ‘strongly non-polynomial’ (see Definition 2.1). This is an elementary argument and is carried out in Lemma A.5 in Appendix A. A similar ‘decomposition’ idea is present in [17, Lemma A.3] (also used in [18]), but we cannot use the exact same decomposition here, because some information on the linear combinations of our functions would be lost.

Our argument differs quite a bit from the methods used in [17], which relied on applications of the van der Corput inequality as a means of complexity reduction and qualitative equidistribution results on nilmanifolds. In simplistic terms, this replaces the issue of studying equidistribution for a function  $a(t) \in \mathcal{H}$  by the problem of studying equidistribution properties for the derivatives  $a', a'', a'''$  and so on. This cannot be used

to cover, for example, functions in the range  $t \prec a(t) \ll t \log t$ , because the derivative  $a'$  must satisfy  $a'(t) \ll \log t$ , which does not have good equidistribution properties even on the 1-dimensional torus  $\mathbb{T}$ . As we mentioned above, we can sidestep this situation, but our argument also has limitations. More precisely, we do not cover functions that grow very slowly (which we call sub-fractional functions). A sample of a ‘slow-growing’ function that we cannot handle is the function  $\log^c t$  for  $c > 1$  (for instance, the pair  $(\log^2 t, t \log t)$  is not covered by Theorem 1.1). The main reason is that when we pass to averages on small intervals, these functions become approximately equal to a constant and our method of using the Taylor expansion breaks down. That explains the existence of the function  $t^\varepsilon$  in equation (8) instead of the term  $\log t$ , which is speculated to be optimal.

In addition, we do not cover the case where some of the functions  $a_i(t)$  are real polynomials, because the reduction to a statement on connected simply connected Lie groups becomes a lot more complicated. For example, consider a nilmanifold  $X = G/\Gamma$  with  $G$  connected and simply connected and elements  $b_1, b_2 \in G$  that commute. It is not clear how to describe sufficiently well the orbit of the sequence  $b_1^{n^{3/2}} b_2^{n^2} \Gamma$  on  $X$ . Alternatively, invoking Leibman’s theorem on polynomial orbits (Theorem B), we can describe the orbit  $\overline{\{b^{n^2} \Gamma\}}$ , while the nilmanifold  $\overline{b_1^{\mathbb{R}} \Gamma}$  (which can be shown to be equal to the closure of the orbit  $b^{n^{3/2}} \Gamma$  by Theorem C) can be expressed in a nice form by Ratner’s theorem (see also Lemma B.1). However, we do not know how to accomplish this for their product  $b_1^{n^{3/2}} b_2^{n^2} \Gamma$ . For example, we expect that this sequence equidistributes on some subnilmanifold  $Y$  (possibly after restricting to an arithmetic progression), but we cannot get any information on the underlying Lie group that defines  $Y$ , which is necessary when applying Theorem F.

A simple argument reduces our problem to the case when the Lie group  $G$  is connected and simply connected. Namely, we will prove Theorem 1.2 under the above connectedness assumptions. We sketch this reduction in Appendix B (at the end of §B.1). Therefore, we make the following convention:

*For the rest of the article up until the Appendix, we make the assumption that all nilpotent Lie groups are always connected and simply connected.* (★)

1.4. *Notational conventions.* Throughout this article, we denote by  $\mathbb{N} = \{1, 2, \dots\}$  the set of natural numbers. We denote  $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ ,  $e(t) = e^{2\pi i t}$ , while  $\|x\|_{\mathbb{T}} = d(x, \mathbb{Z})$  and  $\{x\}$  denote the distance of  $x$  from the nearest integer and the fractional part of  $x$ , respectively. For an element  $\mathbf{x} = (x_1, \dots, x_k)$  in  $\mathbb{R}^k$ , we denote  $|\mathbf{x}| = |x_1| + \dots + |x_k|$ . Lastly, we denote by  $\mathbf{1}_A$  the characteristic function of a set  $A$ .

For two sequences  $a_n, b_n$ , we say that  $b_n$  dominates  $a_n$  and write  $a_n \prec b_n$  or  $a_n = o(b_n)$ , when  $a_n/b_n$  goes to 0, as  $n \rightarrow +\infty$ . In addition, we write  $a_n \ll b_n$  or  $a_n = O(b_n)$ , if there exists a positive constant  $C$  such that  $|a_n| \leq C|b_n|$  for large enough  $n$ . When we want to denote the dependence of this constant on some parameters  $h_1, \dots, h_k$ , we will use the notation  $a_n = O_{h_1, \dots, h_k}(b_n)$ . We use identical notation for asymptotic relations between functions on some real variable  $t$ , where we understand that these hold when we take  $t \rightarrow +\infty$ .

Finally, we use the symbol  $\mathbb{E}$  to denote averages (over a range that will be implicit by the corresponding subscripts each time). Throughout the rest of the article, we use the letters  $p, q$  to denote polynomials and  $\chi$  is used to denote a horizontal character. We will use  $b_1, b_2, \dots, b_k$  or  $u_1, u_2, \dots, u_k, w_1, \dots, w_k$  in the proofs to denote elements of a nilpotent Lie group  $G$ .

2. Background material

2.1. *Measure-preserving systems and Host–Kra structure theorem.* A measure-preserving system is a quadruple  $(X, \mathcal{X}, \mu, T)$ , where  $(X, \mathcal{X}, \mu)$  is a Lebesgue probability space and  $T$  is an invertible measure-preserving map, that is,  $\mu(T^{-1}(A)) = \mu(A)$  for all  $A \in \mathcal{X}$ . It is called ergodic if all the  $T$ -invariant functions are constant. For the purposes of this article, a factor of the system  $(X, \mathcal{X}, \mu, T)$  is a  $T$ -invariant sub- $\sigma$ -algebra of  $\mathcal{X}$ . However, when there is no confusion, we will omit the  $\sigma$ -algebra  $\mathcal{X}$  from the quadruple  $(X, \mathcal{X}, \mu, T)$ .

Let  $(X, \mu, T)$  be a measure-preserving system and let  $f \in L^\infty(\mu)$ . We define the Host–Kra uniformity seminorms inductively as follows:

$$\|f\|_{0,T} = \int f \, d\mu$$

and, for  $s \in \mathbb{N}$ ,

$$\|f\|_{s+1,T}^{2^{s+1}} = \lim_{H \rightarrow \infty} \mathbb{E}_{0 \leq h \leq H} \| \bar{f} \cdot T^h f \|_{s,T}^{2^s}. \tag{13}$$

In the ergodic case, the existence of these limits and the fact that these quantities are indeed seminorms was established in [11]. In the same article, it was shown that these seminorms give rise to a factor  $\mathcal{Z}_{s-1}(X)$  of  $X$  for every  $s \geq 1$ , which is characterized by the following relation:

$$f \perp L^2(\mathcal{Z}_{s-1}(X)) \iff \|f\|_{s,T} = 0.$$

The significance of these factors hinges on the following remarkable structure theorem of Host and Kra [11].

**THEOREM E. (Host–Kra)** *Let  $(X, \mu, T)$  be an ergodic system. Then, the factor  $\mathcal{Z}_s(X)$  is an inverse limit of  $s$ -step nilsystems.*

The last property implies that there exist  $T$ -invariant sub- $\sigma$ -algebras  $\mathcal{Z}_s(n)$ ,  $n \in \mathbb{N}$  that span  $\mathcal{Z}_s$ , such that the factor  $\mathcal{Z}_s(n)$  is isomorphic as a system to an  $s$ -step nilsystem.

2.2. *Background on Hardy fields.* Let  $\mathcal{B}$  denote the set of germs at infinity of real-valued functions defined on a half-line  $[x, +\infty)$ . Then,  $(\mathcal{B}, +, \cdot)$  is a ring, and a sub-field  $\mathcal{H}$  of  $\mathcal{B}$  that is closed under differentiation is called a Hardy field. We will say that  $a(n)$  is a Hardy sequence if for  $n \in \mathbb{N}$  large enough, we have  $a(n) = f(n)$  for a function  $f \in \mathcal{H}$ .

Any two functions  $f, g \in \mathcal{H}$  with  $g$  not identically zero are comparable, that is, the limit

$$\lim_{t \rightarrow \infty} \frac{f(t)}{g(t)}$$

exists and thus it makes sense to compare their growth rates. In addition, every non-constant function in  $\mathcal{H}$  is eventually monotone and, therefore, has a constant sign eventually. In Appendix A, we have collected some lemmas about growth rates of functions in  $\mathcal{H}$ , which will be used frequently throughout the proofs. The proofs of these lemmas can be found in [18], so we shall omit most of them.

We define below some notions that will be used repeatedly throughout the remainder of the paper.

*Definition 2.1.* Let  $a$  be a function in  $\mathcal{H}$ .

- (a) The function  $a$  has polynomial growth if there exists a positive integer  $k$  such that  $f(t) \ll t^k$ . The smallest positive integer  $k$  for which this holds will be called the degree of  $a$ .
- (b) The function  $a$  is called sub-linear if  $a(t) \prec t$ .
- (c) The function  $a$  is called sub-fractional if  $a(t) \prec t^\varepsilon$  for all  $\varepsilon > 0$ .
- (d) The function  $a$  is called strongly non-polynomial if, for any positive integer  $k$ , we have that the functions  $a(t)$  and  $t^k$  have distinct growth rates.

If  $a \in \mathcal{H}$  has polynomial growth, we will also say that the corresponding sequence  $a(n)$  has polynomial growth throughout the article. To understand the definition, consider the functions  $a_1(t) = t^{2/3}$ ,  $a_2(t) = \log^2 t$ ,  $a_3(t) = t + t^{1/2}$  and  $a_4(t) = \exp(t)$ . The first two functions are sub-linear, but the functions  $a_3, a_4$  are not. The function  $a_2(t)$  is the only sub-fractional function among the four functions (it grows slower than all fractional powers), while the strongly non-polynomial functions are  $a_1, a_2$  and  $a_4$  (note that  $a_3$  grows like the polynomial  $p(t) = t$ ). The function  $a_4$  does not have polynomial growth.

*Remark 3.* The definition of strongly non-polynomial presented here is slightly different than the one given in [18]. The definition in that article was that we have the growth relation  $t^k \prec a(t) \prec t^{k+1}$  for  $k \in \mathbb{N}$ , which imposes polynomial growth on our function. In addition, our new definition also allows the inclusion of functions  $a(t)$  such that  $\lim_{t \rightarrow +\infty} |a(t)| = 0$ , while the old one excludes these functions (we do this solely for technical reasons).

### 3. Preparations for the proof

In this section, we will collect some lemmas and make some reductions, which will be useful when we delve into the proof of Theorems 1.1 and 1.2 in the next section. In addition, we provide a specific example, which illustrates the central ideas of the proof of Theorem 1.2 and does not involve a lot of computations.

First of all, we present a lemma, which appears in [4, Lemma 3.3]. We will use this lemma to reduce our problem of studying the long averages over an interval  $[1, N]$  (like those appearing in Theorem 1.2) to averages in short intervals. Its proof is elementary and so we omit it.

LEMMA 3.1. Let  $(a(n))_{n \in \mathbb{N}}$  be a bounded sequence of complex numbers. Assume that

$$\lim_{N \rightarrow +\infty} \mathbb{E}_{N \leq n \leq N+L(N)} a(n) = \alpha$$

for some positive function  $L(t)$  with  $1 \prec L(t) \prec t$ . Then, we also have

$$\lim_{N \rightarrow +\infty} \mathbb{E}_{1 \leq n \leq N} a(n) = \alpha.$$

3.1. *An example of convergence.* Assume  $X = G/\Gamma$  is a nilmanifold with  $G$  connected and simply connected. We will show that the averages

$$\mathbb{E}_{1 \leq n \leq N} f(b_1^{n^{3/2}} x) \cdot g(b_2^{n \log n} x)$$

converge for any  $x \in X$ , where  $b_1, b_2 \in G$ .

Using Lemma 3.1, it suffices to show that the averages

$$\mathbb{E}_{N \leq n \leq N+L(N)} f(b_1^{n^{3/2}} x) \cdot g(b_2^{n \log n} x)$$

converge, for some sub-linear function  $L(t)$ . Passing to the nilmanifold  $X \times X$ , we see that our problem reduces to showing that the averages

$$\mathbb{E}_{N \leq n \leq N+L(N)} F(b_1^{n^{3/2}} b_2^{n \log n} x)$$

converge for any nilmanifold  $X = G/\Gamma$ , commuting elements  $b_1, b_2 \in G$  and function  $F \in C(X)$ . (When we pass to the product  $X \times X$ , we have to study the actions of the elements  $(b_1, e_G)$  and  $(e_G, b_2)$ , which clearly commute.) Due to density, we can actually pick  $F \in \text{Lip}(X)$ . We provide more details for this deduction in the next section (after Proposition 4.1).

Let  $X'$  denote the subnilmanifold  $\overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$  of  $X$ . By Lemma B.1, this set is indeed a subnilmanifold of  $X$  and has a representation as  $H/\Delta$ , with  $H$  connected, simply connected and containing all elements  $b_1^s$  and  $b_2^s$  for any  $s \in \mathbb{R}$ . In this example, we will also assume that  $X' = \overline{b_1^{\mathbb{Z}} b_2^{\mathbb{Z}} \Gamma}$ . In the main proof, we will use Lemma B.2 to reduce the general case of the theorem to this one.

Using the Taylor expansion around the point  $N$ , we can write

$$(N + h)^{3/2} = N^{3/2} + \frac{3}{2}hN^{1/2} + \frac{3h^2}{8N^{1/2}} - \frac{h^3}{16N^{3/2}} + \frac{3h^4}{128\xi_h^{5/2}} \quad \text{for some } \xi_h \in [N, N + h]$$

for every  $0 \leq h \leq L(N)$ . If we choose  $L(t)$  to satisfy

$$t^{1/2} \prec L(t) \prec t^{5/8},$$

then the last term in the above expansion is smaller than  $o_N(1)$ , while the second to last term is unbounded. Similarly, we can write

$$(N + h) \log(N + h) = N \log N + h(\log N + 1) + \frac{h^2}{2N} - \frac{h^3}{6\psi_h^2} \quad \text{for some } \psi_h \in [N, N + h].$$

If we choose again  $L(t)$  to satisfy

$$t^{1/2} \prec L(t) \prec t^{2/3},$$

we can show that the last term is  $o_N(1)$ , while the  $h^2$  term is unbounded. For instance, we can choose  $L(t) = t^{3/5}$  and both growth conditions that we imposed will be satisfied.

Since the function  $F$  is continuous, we can disregard the highest order terms in the above expansion since they are both  $o_N(1)$ . Our problem reduces to showing that the averages

$$\mathbb{E}_{0 \leq h \leq L(N)} F(b_1^{N^{3/2} + (3/2)hN^{1/2} + (3h^2/8N^{1/2}) - (h^3/16N^{3/2})} b_2^{N \log N + h(\log N + 1) + h^2/2N} x)$$

converge. For the sake of simplicity, we will show that the averages

$$\mathbb{E}_{0 \leq h \leq L(N)} F(b_1^{h^3/N^{3/2}} b_2^{h^2/N} x)$$

converge, since both of these statements follow from the same arguments. For convenience, we will also assume that  $x = \Gamma$ .

Let  $\delta > 0$ . We consider the finite sequence

$$(v(h)\Gamma)_{0 \leq h \leq L(N)} = (b_1^{h^3/N^{3/2}} b_2^{h^2/N} \Gamma)_{0 \leq h \leq L(N)}$$

and we show that if  $N$  is large enough, then it is  $\delta$ -equidistributed on the subnilmanifold  $X' = \overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$  of  $X$ . It is apparent that  $v(n)\Gamma$  is a polynomial sequence in  $X'$ . We consider the horizontal torus  $Z$  of  $X'$ , which is isomorphic to some  $\mathbb{T}^d$  ( $d \in \mathbb{N}$ ) and we also let  $\pi$  denote the projection map from  $X'$  to  $Z$ . If the given sequence is not  $\delta$ -equidistributed (for a fixed value of  $N$ ), we can invoke Theorem F to find a positive constant  $M = M(X', \delta)$  and a non-trivial horizontal character  $\chi_N$  of modulus at most  $M$  and such that

$$\|\chi_N(\pi(v(h)\Gamma))\|_{C^\infty[L(N)]} \leq M.$$

Suppose  $\chi_N$  descends to the character

$$(t_1, \dots, t_d) \rightarrow e(k_{1,N}t_1 + \dots + k_{d,N}t_d)$$

on  $\mathbb{T}^d$ , where  $k_{1,N}, \dots, k_{d,N}$  are integers. The fact that the modulus is bounded by  $M$  implies that

$$|k_{1,N}| + \dots + |k_{d,N}| \leq M.$$

Let us also write  $\pi(b_1\Gamma) = (x_1, \dots, x_d)$  and  $\pi(b_2\Gamma) = (y_1, \dots, y_d)$ . Then, the last inequality implies that

$$\left\| e\left(\frac{h^3}{N^{3/2}} \sum_{i=1}^d k_{i,N}x_i + \frac{h^2}{N} \sum_{i=1}^d k_{i,N}y_i\right)\right\|_{C^\infty[L(N)]} \leq M. \tag{14}$$

Assume there are infinitely many  $N$  for which this holds. Since there are only finitely many possible choices for the numbers  $k_{1,N}, \dots, k_{d,N}$  above, we conclude that there exists a character  $\chi$  such that  $\|\chi(\pi(a(h)\Gamma))\|_{C^\infty[L(N)]} \leq M$  holds for infinitely many  $N \in \mathbb{N}$ . Then, we rewrite equation (14) ( $k_i$  are some integers independent of  $N$ ) as

$$\left\| e\left(\frac{h^3}{N^{3/2}} \sum_{i=1}^d k_i x_i + \frac{h^2}{N} \sum_{i=1}^d k_i y_i\right)\right\|_{C^\infty[L(N)]} \leq M,$$

and this inequality holds for infinitely many  $N$ .

The definition of the  $C^\infty[L(N)]$  norms implies that we have the relations

$$L(N)^3 \left\| \frac{\sum_{i=1}^d k_i x_i}{N^{3/2}} \right\|_{\mathbb{T}} \leq M$$

and

$$L(N)^2 \left\| \frac{\sum_{i=1}^d k_i y_i}{N} \right\|_{\mathbb{T}} \leq M$$

for infinitely many  $N$ . Due to our choice of the function  $L(N)$ , these relations fail for  $N$  sufficiently large unless

$$\sum_{i=1}^d k_i x_i \in \mathbb{Z} \quad \text{and} \quad \sum_{i=1}^d k_i y_i \in \mathbb{Z}.$$

This implies that  $\chi \circ \pi(b_1\Gamma) = \chi \circ \pi(b_2\Gamma) = 0$  and, consequently, we must also have  $\chi \circ \pi(b_1^m b_2^n \Gamma) = 0$  for any  $m, n \in \mathbb{Z}$ . Since elements of this form are dense in  $\overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$  by our initial hypothesis, we get that  $\chi$  must be the trivial character, which is a contradiction.

In conclusion, we have established that the sequence  $(v(h)\Gamma)_{0 \leq h \leq L(N)}$  is  $\delta$ -equidistributed for large enough  $N$  on  $X' = \overline{b_1^{\mathbb{R}} b_2^{\mathbb{R}} \Gamma}$ . The result now follows by sending  $\delta \rightarrow 0$ . We also notice that the limit of the averages is  $\int_{X'} F \, dm_{X'}$ .

*Remark 4.* We describe briefly here why we have to use the  $t^\varepsilon$  term in equation (8) instead of the conjectured optimal term  $\log t$ . Assuming we had the functions  $\log^2 t$  and  $t \log t$  in this example, then for any choice of the sub-linear function  $L(t)$  that would give a good polynomial approximation for the function  $t \log t$ , we would have

$$\max_{0 \leq h \leq L(N)} |\log^2(N+h) - \log^2 N| = o_N(1),$$

which suggests that the sequence  $\log^2 n$  is essentially constant in the small intervals  $[N, N + L(N)]$ . If we proceed exactly as in the above argument, the best we can actually show is that

$$\left| \mathbb{E}_{N \leq n \leq N+L(N)} F(b_1^{\log^2 n} b_2^{n \log n} \Gamma) - \int_{Y_2} F(b_1^{\log^2 N} y) \, dm_{Y_2}(y) \right| \leq \delta \|F(b_1^{\log^2 N} \cdot)\|_{\text{Lip}(Y_2)}$$

for large enough  $N$ , where  $Y_2 = \overline{b_2^{\mathbb{R}} \Gamma}$  and  $F(b_1^{\log^2 N} \cdot)$  denotes the function  $y \rightarrow F(b_1^{\log^2 N} y)$  defined on the nilmanifold  $Y_2$ . However, the Lipschitz norm above is of the order  $\log^2 N \|F\|_{\text{Lip}(X)}$ , which diverges as  $N \rightarrow +\infty$ , so this bound cannot be useful for any purposes.

Another approach would be to use the fact that the parameter  $M$  in Theorem F is of the form  $\delta^{-O(1)}$ , namely we have bounds that are polynomial in  $\delta$ . Thus, one could allow the parameter  $\delta$  to vary with  $N$ . For instance, establishing a bound of the form  $(\log N)^{-(2+\varepsilon)} \|F(b_1^{\log^2 N} \cdot)\|_{\text{Lip}(Y_2)}$  in place of the term  $\delta \|F(b_1^{\log^2 N} \cdot)\|_{\text{Lip}(Y_2)}$  (namely, showing that our sequence is  $(\log N)^{-(2+\varepsilon)}$ -equidistributed) on the right-hand side of the above equation leads to a solution to the more general problem. (It would actually suffice to obtain this statement for almost all  $N \in \mathbb{N}$  in the sense of natural density.) However,

any bound of this type is incorrect in general. Indeed, assume that the horizontal torus of  $b_2^{\mathbb{R}}\Gamma$  was  $\mathbb{T}^2$  and also let  $(b_{2,1}, b_{2,2}) \in \mathbb{T}^2$  denote the image of the element  $b_2\Gamma$  under the projection map. Following the same approximations as the ones in the example, we would like to show that the finite polynomial sequence  $b_2^{h^2/N}\Gamma$ , where  $0 \leq h \leq L(N)$ , is  $(\log N)^{-(2+\epsilon)}$ -equidistributed for almost all  $N \in \mathbb{N}$  and for some suitable sub-linear function  $L(t)$  satisfying only  $L(t) \succ t^{1/2}$ . Then, an application of Theorem F implies that if this assertion does not hold, then there exists a positive constant  $C$  and a horizontal character  $\chi$  of modulus at most  $\log^C N$ , such that

$$\|\chi(b_2^{h^2/N}\Gamma)\|_{C^\infty(L(N))} \leq \log^C N.$$

Equivalently, there exist integers  $k_1, k_2$  with  $|k_1| + |k_2| \leq \log^C N$  such that

$$L^2(N) \left\| \frac{k_1 b_{2,1} + k_2 b_{2,2}}{N} \right\|_{\mathbb{T}} \leq \log^C N.$$

Thus, we would get a contradiction if we showed that

$$\min_{|k_1|, |k_2| \leq \log^C N} |k_1 b_{2,1} + k_2 b_{2,2}| \geq \frac{N \log^C N}{L^2(N)}$$

holds for  $N$  in a set of density 1. However, we note that bounds like the above depend on the diophantine properties of the numbers  $b_{2,1}, b_{2,2}$ . Indeed, let us suppose that  $\alpha = b_{2,1}/b_{2,2} \leq 1$ . If we divide by  $b_{2,2}$ , the last inequality can be rewritten as

$$\min_{|k_1|, |k_2| \leq \log^C N} |k_1 \alpha + k_2| \geq \frac{N \log^C N}{|b_{2,2}| L^2(N)}.$$

For a fixed choice of  $k_1$ , the absolute value is minimized by picking  $k_2$  to be the nearest integer to  $-k_1 \alpha$ . Thus, we would need to show that

$$\min_{|k_1| \leq \log^C N} \|k_1 \alpha\|_{\mathbb{T}} \geq \frac{N \log^C N}{|b_{2,2}| L^2(N)}$$

and we can find  $b_{2,1}, b_{2,2} \in (0, 1)$  for which this inequality fails for all  $N$  in a set of positive upper density. A simpler example that avoids the complicated function on the right-hand side of the last equation is to show that we can find  $\alpha \in (0, 1)$  for which the inequality  $\min_{|k| \leq N} \|k\alpha\|_{\mathbb{T}} \geq 2^{-n}$  fails for all  $N \in \mathbb{N}$  in a set of upper density 1. Indeed, we can construct an  $\alpha \in (0, 1)$  such that  $\liminf_{n \rightarrow +\infty} 2^{2^n} \|n\alpha\|_{\mathbb{T}} = 0$ . Thus, there is a sequence  $q_n$  such that  $\|q_n \alpha\|_{\mathbb{T}} \leq 2^{-2^{q_n}}$  which implies that  $\min_{|k| \leq N} \|k\alpha\|_{\mathbb{T}} \leq 2^{-2^{q_n}} \leq 2^{-N}$  for every  $N$  with  $q_n \leq N \leq 2^{q_n}$ . Thus, the set of  $N$  for which the above inequality fails has upper density 1.

**3.2. Removing the integer parts.** In this part, we will establish a lemma that practically implies that part (a) of Theorem 1.2 follows from part (b) of the same theorem. The fact that part (a) of Theorem 1.1 follows from part (b) of the same theorem is precisely the statement of [4, Lemma 5.1], which is proven using very similar arguments to the proof of Lemma 3.2 below. If a collection of sequences of real numbers has the property that the averages



$$\mathbb{E}_{1 \leq n \leq N} f_1(b_1^{a_1(n)} x_1) \cdots f_k(b_k^{a_k(n)} x_k) \tag{15}$$

converge for all nilmanifolds  $X_i = G_i / \Gamma_i$ , elements  $b_i \in G_i$ , points  $x_i \in X_i$  and continuous functions  $f_i$  defined on  $X_i$ , we will say that this collection is pointwise good for nilsystems. The notation  $b_i^{a_i(n)}$  makes sense here due to the connectedness assumptions we have imposed on the Lie groups  $G_i$ .

LEMMA 3.2. *Let  $a_1(n), \dots, a_k(n)$  be sequences of real numbers that satisfy the following.*

- (a) *The collection  $a_1(n), \dots, a_k(n)$  is pointwise good for nilsystems.*
- (b) *For every  $1 \leq i \leq k$ , we have that the sequence  $(a_i(n)\mathbb{Z})_{n \in \mathbb{N}}$  satisfies one of the following:*
  - (1) *it is equidistributed on  $\mathbb{T}$ ;*
  - (2) *it converges to some  $c = c(i) \in \mathbb{T}$  different from 0;*
  - (3) *it converges to 0 and the sequence  $\{a_i(n)\} - \frac{1}{2}$  has a constant sign eventually.*

*Then, the sequences  $[a_1(n)], \dots, [a_k(n)]$  are pointwise good for nilsystems.*

Remark 5. The number  $\frac{1}{2}$  in the third condition is arbitrary since we could have used any number  $\alpha \in (0, 1)$ . We primarily use this condition in the following manner: suppose we have a function  $f(t)$ , which converges monotonically to some  $k \in \mathbb{Z}$  as  $t \rightarrow +\infty$ . Then, we clearly have  $\|f(t)\|_{\mathbb{T}} \rightarrow 0$  and we also observe that the sequence  $\{f(n)\}$  does not oscillate between intervals of the form  $[0, \varepsilon]$  and  $[1 - \varepsilon, 1)$  (due to the monotonicity assumption). Thus, the sequence  $\{f(n)\} - \frac{1}{2}$  will indeed have a constant sign (positive if  $f$  increases to  $k$  and negative otherwise).

Proof. Let  $X_i = G_i / \Gamma_i$  be nilmanifolds with  $G_i$  connected and simply connected and  $b_i \in G_i$ . Let  $f_1, \dots, f_k$  be continuous functions defined on  $X_1, \dots, X_k$ , respectively. Under the hypotheses of the lemma on the sequences  $a_1(n), \dots, a_k(n)$ , we want to show that the averages

$$\mathbb{E}_{1 \leq n \leq N} f_1(b_1^{[a_1(n)]} x_1) \cdots f_k(b_k^{[a_k(n)]} x_k) \tag{16}$$

converge for any choice of the  $x_i \in X_i$ .

Fix some  $i \in \{1, 2, \dots, k\}$ . If the sequence  $a_i(n)$  satisfies the second condition, namely that  $a_i(n)\mathbb{Z}$  converges to  $c\mathbb{Z}$  ( $c \neq 0$ ), then for  $n$  sufficiently large, we have

$$[a_i(n)] = a_i(n) - \{c\} + o_n(1).$$

This implies that  $b_i^{[a_i(n)]} = b_i^{-\{c\}} b_i^{a_i(n) + o_n(1)}$ . Since the function  $f_i$  is continuous, we can disregard the contribution of the  $o_n(1)$  term, while the  $b_i^{-\{c\}}$  term can be absorbed by the  $x_i$ . Therefore, we notice that in this case, we can remove the integer part for the sequence  $a_i(n)$ . An entirely similar argument demonstrates that the same holds if  $a_i(n)$  satisfies the third condition.

To complete the proof, we will consider below the case that each of the sequences  $a_i(n)\mathbb{Z}$  is equidistributed on  $\mathbb{T}$  for convenience (namely, they all satisfy the first condition). Since we can easily remove the integer parts for those sequences that satisfy the second

or third condition as we did above, the argument below easily adapts to the general setting with some changes in notation.

We rewrite the averages in equation (16) as

$$\mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k f_i(b_i^{-\{a_i(n)\}} b_i^{a_i(n)} x_i) = \mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k \tilde{f}_i(a_i(n)\mathbb{Z}, b_i^{a_i(n)} x_i),$$

where  $\tilde{f}_i : \mathbb{T} \times X_i \rightarrow \mathbb{C}$  is the function defined by the relation

$$\tilde{f}_i(s\mathbb{Z}, gx) = f_i(b_i^{-\{s\}} gx).$$

Let  $v_i(n)$  be the sequence  $(a_i(n)\mathbb{Z}, b_i^{a_i(n)} x_i)$ . By our hypothesis, for any continuous functions  $f'_i$  on  $\tilde{X}_i = \mathbb{T} \times X_i$ , the averages of  $\prod_{i=1}^k f'_i(v_i(n))$  converge. However, note that the functions  $\tilde{f}_i$  that we are dealing with may have discontinuities when  $s$  becomes close to an integer. Our goal is to approximate each  $\tilde{f}_i$  by a continuous function and then use the above observation.

Let  $\varepsilon > 0$ . For every  $1 \leq i \leq k$ , we define a continuous function  $f_{i,\varepsilon}$  that agrees everywhere with  $\tilde{f}_i$  on  $[\varepsilon, 1 - \varepsilon] \times X_i$  and such that  $f_{i,\varepsilon}$  is bounded uniformly by  $2\|\tilde{f}_i\|_\infty$ . Observe that

$$\begin{aligned} \left| \mathbb{E}_{1 \leq n \leq N} \tilde{f}_i(v_i(n)) - \mathbb{E}_{1 \leq n \leq N} f_{i,\varepsilon}(v_i(n)) \right| &= \frac{1}{N} \left| \sum_{\substack{1 \leq n \leq N \\ a_i(n) \notin [\varepsilon, 1-\varepsilon]}} \left( \tilde{f}_i(v_i(n)) - f_{i,\varepsilon}(v_i(n)) \right) \right| \\ &\ll \varepsilon \|\tilde{f}_i\|_\infty + o_N(1), \end{aligned} \tag{17}$$

where the last bound follows from the triangle inequality and the fact that  $a_i(n)$  is equidistributed (mod 1), which indicates that the set  $\{n \in \mathbb{N} : a_i(n) \notin [\varepsilon, 1 - \varepsilon]\}$  has asymptotic density  $2\varepsilon$ .

Combining equation (17) with a simple telescoping argument, we deduce that

$$\limsup_{N \rightarrow +\infty} \left| \mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k \tilde{f}_i(v_i(n)) - \mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k f_{i,\varepsilon}(v_i(n)) \right| \ll k\varepsilon \prod_{i=1}^k \|\tilde{f}_i\|_\infty.$$

Since the averages  $\mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k f_{i,\varepsilon}(v_i(n))$  converge as  $N \rightarrow \infty$  by our hypothesis (the functions involved here are continuous), we infer that the averages

$$\mathbb{E}_{1 \leq n \leq N} \prod_{i=1}^k \tilde{f}_i(v_i(n))$$

form a Cauchy sequence and, therefore, converge. The conclusion follows. □

Using the previous lemma, we can establish that the first part of Theorem 1.2 follows from the second part. We postpone this until the next section, where we also prove the second part of Theorem 1.2.

4. *Proofs of main theorems*

The main tool we are going to use in our proof is the quantitative Green–Tao theorem on polynomial orbits (Theorem F). A technical obstruction in our proof is that among the functions  $a_1, \dots, a_k$  in the statement of Theorem 1.2, we must separate the polynomial functions from the strongly non-polynomial ones. We will accomplish this using an elementary lemma (Lemma A.5) which is proven in the Appendix. We restate Theorem 1.2 here.

**THEOREM 1.2.** *Let  $\mathcal{H}$  be a Hardy field that contains the polynomial functions. Let  $a_1, \dots, a_k$  be functions in  $\mathcal{H}$  that have polynomial growth. Assume that there exists  $\varepsilon > 0$ , such that every function  $a \in \mathcal{L}(a_1, \dots, a_k)$  satisfies either*

$$\lim_{t \rightarrow +\infty} \frac{|a(t) - p(t)|}{t^\varepsilon} = +\infty \quad \text{for any polynomial } p(t) \in \mathbb{Q}[t],$$

or

$$\text{the limit } \lim_{t \rightarrow +\infty} a(t) \text{ is a real number.}$$

Then, we have the following.

- (i) *For any collection of nilmanifolds  $X_i = G_i/\Gamma_i$ , elements  $b_i \in G_i$ ,  $x_i \in X_i$  and continuous functions  $f_1, \dots, f_k$  with complex values, the averages*

$$\frac{1}{N} \sum_{i=1}^N f_1(b_1^{\lfloor a_1(n) \rfloor} x_1) \cdots f_k(b_k^{\lfloor a_k(n) \rfloor} x_k)$$

*converge.*

- (ii) *For any collection of nilmanifolds  $X_i = G_i/\Gamma_i$  such that the groups  $G_i$  are connected, simply connected, elements  $b_i \in G_i$ ,  $x_i \in X_i$  and continuous functions  $f_1, \dots, f_k$  with complex values, the averages*

$$\frac{1}{N} \sum_{i=1}^N f_1(b_1^{a_1(n)} x_1) \cdots f_k(b_k^{a_k(n)} x_k)$$

*converge.*

First of all, we show that the first part follows from the second part. This is accomplished by using Lemma 3.2. We remark again that in part (i), there are no connectedness assumptions made on the groups  $G_i$ . Nonetheless, the convention equation (\*) in §1 allows us to consider only the case that the Lie groups  $G_i$  are connected and simply connected. We implicitly work under this assumption in the proof below.

*Proof of part (i) of Theorem 1.2, assuming part (ii).* We will have to confirm that the conditions of Lemma 3.2 are satisfied. Let  $a_1, \dots, a_k \in \mathcal{H}$  be as in the statement of Theorem 1.2. Condition (a) of Lemma 3.2 is satisfied by our hypothesis. Now, we verify the second condition.

Fix some  $i \in \{1, 2, \dots, k\}$ . We consider three cases.

- (i) Assume that the function  $a_i(t)$  is such that  $|a_i(t) - q(t)| \succ t^\varepsilon$  for all polynomials  $q(t)$  with rational coefficients. Then, the sequence  $a_i(n)\mathbb{Z}$  is equidistributed on  $\mathbb{T}$  (satisfying condition (1)), due to Theorem A.

- (ii) Assume that the function  $a_i(t)$  is such that  $\lim_{t \rightarrow +\infty} a_i(t) = c \notin \mathbb{Z}$ . Then, the sequence  $a_i(n)$  satisfies condition (2) of Lemma 3.2.
- (iii) Assume that neither of the above conditions is true. Since  $a_i(t)$  must satisfy equation (9), we deduce that  $a_i(t)$  converges to some integer  $c$ . However, since  $a_i(t)$  converges to  $c$  monotonically (functions in  $\mathcal{H}$  are eventually monotone), we deduce that condition (3) of Lemma 3.2 is satisfied and we are done. □

Now we switch our attention to the proof of part (ii). First, we will apply Lemma A.5 from Appendix A to replace the original functions  $a_1, \dots, a_k$  with a collection of functions that are more manageable. This will enable us to separate the polynomial functions from strongly non-polynomial ones. In addition, among the strongly non-polynomial functions, we have to isolate those that are sub-fractional, because they behave differently when we try to employ the Taylor expansion. This whole process will reduce Proposition 4.1 below to Lemma 4.2, which we will then proceed to establish.

Following all these reductions, we use the Taylor expansion to substitute the strongly non-polynomial functions with polynomials in some small intervals. Now, this reduces the original problem to a quantitative equidistribution problem of finite polynomial sequences in a nilmanifold, although the coefficients of the polynomials vary depending on the underlying short interval. Finally, in step 3, we use the quantitative equidistribution results to show that averages of Lipschitz functions in the nilmanifold over these ‘variable’ polynomial sequences are very close to an integral over a subnilmanifold, which ultimately allows us to evaluate the limit of the initial averages.

We make one final reduction: let  $a_1, \dots, a_k \in \mathcal{H}$  be functions as in the statement of Theorem 1.2. Passing to the product nilmanifold, we infer that our problem follows from the following statement.

**PROPOSITION 4.1.** *Let  $X = G/\Gamma$  be a nilmanifold,  $b_1, \dots, b_k \in G$  are commuting elements and  $a_1, \dots, a_k \in \mathcal{H}$  have polynomial growth. Assume that there exists  $\varepsilon > 0$ , such that every function  $a \in \mathcal{L}(a_1, \dots, a_k)$  satisfies either equation (8) or equation (9). Then, for any  $x \in X$  and continuous function  $F : X \rightarrow \mathbb{C}$ , we have that the averages*

$$\mathbb{E}_{1 \leq n \leq N} F(b_1^{a_1(n)} \dots b_k^{a_k(n)} x) \tag{18}$$

converge.

*Proof that Proposition 4.1 implies Theorem 1.2.* We want to show that the averages

$$\frac{1}{N} \sum_{i=1}^N f_1(b_1^{a_1(n)} x_1) \dots f_k(b_k^{a_k(n)} x_k)$$

converge for all  $x_i \in X_i$ , where the nilmanifolds  $X_i = G_i/\Gamma_i$ , the elements  $b_i$  and the functions  $a_i \in \mathcal{H}$  are as in the statement of part (ii) of Theorem 1.2. We define the continuous function  $F$  on the product nilmanifold  $X_1 \times \dots \times X_k$  by the relation

$$F(y_1, \dots, y_k) = f_1(y_1) \dots f_k(y_k).$$

We also denote by  $\tilde{b}_i$  the element on  $G_1 \times \dots \times G_k$ , whose  $i$ th coordinate is equal to  $b_i$ , while all of its other coordinates are equal to the respective identity element. Observe that

the elements  $\tilde{b}_1, \dots, \tilde{b}_k$  are pairwise commuting. Finally, let us also denote by  $x$  the point  $(x_1, \dots, x_k)$  on the product  $X_1 \times \dots \times X_k$ . Then, a simple computation implies that our initial average is equal to

$$\mathbb{E}_{1 \leq n \leq N} F(b_1^{a_1(n)} \dots b_k^{a_k(n)} x)$$

and the claim now follows. □

Now, we will reduce Proposition 4.1 to the following lemma.

LEMMA 4.2. *Let  $G/\Gamma$  be a nilmanifold and suppose that  $u_1, \dots, u_s$  are elements in  $G$ , such that*

$$\overline{u_1^{\mathbb{R}} \dots u_s^{\mathbb{R}} \Gamma} = \overline{u_1^{\mathbb{Z}} \dots u_s^{\mathbb{Z}} \Gamma}. \tag{19}$$

*In addition, assume that the nilmanifold  $X' = \overline{u_1^{\mathbb{R}} \dots u_s^{\mathbb{R}} \Gamma}$  can be represented as  $G'/\Gamma'$ , where  $G'$  is connected, simply connected and contains all elements  $u_1, \dots, u_s$ . Let  $s_0, s$  be positive integers and define the sequence  $v(n)$*

$$\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_i(n)+x_i(n)}, \tag{20}$$

where:

- (a)  $p_i, \tilde{p}_j$  are polynomials with real coefficients, such that every non-trivial linear combination of the polynomials  $\tilde{p}_{s_0+1}, \dots, \tilde{p}_s$  is not an integer polynomial;
- (b) the functions  $x_i$  are all strongly non-polynomial, the functions  $x_1, \dots, x_{s_0}$  are not sub-fractional and have pairwise distinct growth rates and the functions  $x_{s_0+1}, \dots, x_s$  are sub-fractional.

*Then, for any Lipschitz function  $F$  on  $X'$  with Lipschitz norm at most 1, the averages*

$$\mathbb{E}_{1 \leq n \leq N} F\left(\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_i(n)+x_i(n)} \Gamma'\right)$$

*converge to the integral  $\int_{X'} F dm_{X'}$ .*

While the statement may seem relatively convoluted at first, the sequence  $v(n)$  above has a convenient form, so that the Taylor approximation can be used directly.

First of all, we prove that Lemma 4.2 implies Proposition 4.1. We will rely on Lemma A.5 to make the required reductions on the Hardy field functions in the iterates and we will also use Lemma B.2 to get the equality equation (19), where  $u_1, \dots, u_s$  will be some appropriate elements of the Lie group  $G$  (they will be products of powers of the elements  $b_i$  in Proposition 4.1).

*Proof that Lemma 4.2 implies Proposition 4.1.* Applying Lemma A.5, we can find a basis  $f_1, \dots, f_s$  for the set  $\mathcal{L}(a_1, \dots, a_k)$  of non-trivial linear combinations. The collection of functions  $f_1, \dots, f_s$  can be written in the form  $(g_1, \dots, g_m, h_1, \dots, h_\ell)$  where  $g_i, h_i$  are as in Lemma A.5. We will not use this specific property until a little further below, so as to avoid cumbersome notation. Note that the fact that  $f_1, \dots, f_s$  form a basis indicates

that the assumptions on the linear combinations of the  $a_1, \dots, a_k$  in the statement of Proposition 4.1 are now transferred to the functions  $f_1, \dots, f_s$ .

If we write

$$a_i(t) = \sum_{j=1}^s c_{i,j} f_j(t) \tag{21}$$

for some real numbers  $c_{i,j}$ , then we can rewrite the average in equation (18) as

$$\mathbb{E}_{1 \leq n \leq N} F(u_1^{f_1(n)} \dots u_s^{f_s(n)} x) \tag{22}$$

for some commuting elements  $u_1, \dots, u_s \in G$  (here, the fact that the elements  $b_1, \dots, b_k$  commute is required). We denote

$$v(n) = u_1^{f_1(n)} \dots u_s^{f_s(n)},$$

which is a sequence in  $G$ . We want to establish that the averages of the sequence  $F(v(n)x)$  converge for all  $x \in X$  and any continuous function  $F$ . If one of the functions  $f_1, \dots, f_m$  is such that the limit  $\lim_{t \rightarrow +\infty} f_i(t)$  is a real number (which can be the case when a linear combination of the original functions satisfies equation (9)), we can invoke the continuity of  $F$  to eliminate the corresponding term  $u_i^{f_i(n)}$  in the product and replace it by a constant. Hence, we may assume that all of the functions  $f_1(t), \dots, f_s(t)$  go to  $\pm\infty$ , as  $t \rightarrow +\infty$ .

Now we use the particular structure of the functions  $f_1, \dots, f_s$ . The statement of Lemma A.5 implies that the collection of functions  $f_1, \dots, f_s$  has the form  $(g_1, \dots, g_m, h_1, \dots, h_\ell)$  (clearly,  $m + \ell = s$ ) such that the functions  $g_i$  can be written in the form  $p_i(t) + x_i(t)$ , where the functions  $x_1(t), \dots, x_m(t)$  are strongly non-polynomial and have pairwise distinct (and non-trivial) growth rates, while the functions  $h_i$  can be written in the form  $\tilde{p}_i(t) + y_i(t)$ , where  $y_i(t)$  converges to 0. Here,  $p_i$  and  $\tilde{p}_i$  are polynomials with real coefficients.

We may rearrange the functions  $f_i$  so that  $f_i = g_i$  for all  $1 \leq i \leq m$  and  $f_j = h_{j-m}$  for each  $m + 1 \leq j \leq s$ . Rewrite the sequence  $v(n)$  as

$$v(n) = \prod_{i=1}^m u_i^{g_i(n)} \cdot \prod_{i=1}^{\ell} u_{m+i}^{h_i(n)} = \prod_{i=1}^m u_i^{p_i(n)+x_i(n)} \cdot \prod_{i=1}^{\ell} w_i^{\tilde{p}_i(n)+y_i(n)},$$

where we use the notation  $w_i$  for the element  $u_{i+m}$  in the last equality. Without loss of generality, assume that

$$x_1(t) \succ x_2(t) \succ \dots \succ x_m(t) \succ 1.$$

First, we need to distinguish between the sub-fractional functions and the ‘fast’ growing functions among the functions  $x_i(t)$  (this will be important later when we use the polynomial expansion). Thus, let  $0 \leq s_0 \leq m$  be a natural number such that  $x_{s_0}(t) \gg t^\varepsilon$  for some  $\varepsilon > 0$ , while  $x_{s_0+1}$  is a sub-fractional function. This also implies that all the functions  $x_i$  for  $i$  satisfying  $s_0 + 1 \leq i \leq m$  are sub-fractional since we have arranged the functions so that their growth rates are in descending order.

Once again, we rewrite the sequence  $v(n)$  in the form

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^m u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\tilde{p}_i(n)+y_i(n)}.$$

Because the function  $F$  is continuous, we can discard the functions  $y_1, \dots, y_{\ell}$ , since they all converge to zero. The hypotheses in equations (8) and (9) on the linear combinations of the remaining functions in the exponents continue to hold. Indeed, this can be seen by noting that equations (8) and (9) still hold when replacing one of the functions (say  $a_1$ ) by a function of the form  $a_1(t) + e(t)$ , with  $e(t) \rightarrow 0$ . Consequently, we can redefine  $v(n)$  to be the sequence

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^m u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\tilde{p}_i(n)}.$$

We will now reduce our problem to the case that the polynomials  $\tilde{p}_1(t), \dots, \tilde{p}_{\ell}(t)$  are linearly independent. Due to our hypothesis (namely equations (8),(9)), every non-trivial linear combination of the functions  $\tilde{p}_1(t), \dots, \tilde{p}_{\ell}(t)$  must satisfy either equation (8) or equation (9). Thus, every linear combination of the polynomials  $\tilde{p}_1(t), \dots, \tilde{p}_{\ell}(t)$  is not a polynomial with integer coefficients unless it is the zero polynomial. If the second case is true, there exist  $c_1, \dots, c_{\ell-1} \in \mathbb{R}$  such that

$$\tilde{p}_{\ell} = c_1 \tilde{p}_1 + \dots + c_{\ell-1} \tilde{p}_{\ell-1}.$$

Then, we have

$$\prod_{i=1}^{\ell} w_i^{\tilde{p}_i(n)} = \prod_{i=1}^{\ell-1} (w_i w_{\ell}^{c_i})^{\tilde{p}_i(n)}.$$

If the polynomials  $\tilde{p}_1, \dots, \tilde{p}_{\ell-1}$  are linearly independent, then we are done. Otherwise, we proceed similarly to eliminate  $\tilde{p}_{\ell-1}$ . After a finite number of steps, we will reach a collection of linearly independent polynomials.

In view of the above, we are allowed to assume that  $\tilde{p}_1, \dots, \tilde{p}_{\ell}$  are linearly independent. Now, we show that we can reduce to the case that the polynomials  $p_{s_0+1}, \dots, p_m, \tilde{p}_1, \dots, \tilde{p}_{\ell}$ . Indeed, the linear independence assumption on the polynomials  $\tilde{p}_1, \dots, \tilde{p}_{\ell}$  implies that the polynomials  $p_{s_0+1}, \dots, p_m, \tilde{p}_1, \dots, \tilde{p}_{\ell}$  are linearly independent. To see how this works, observe that if there are real numbers  $c_i, d_j$  such that

$$\sum_{i=1}^{m-s_0} c_i p_{s_0+i} + \sum_{i=1}^{\ell} d_i p'_i = 0,$$

then the function

$$\sum_{i=1}^{m-s_0} c_i (p_{s_0+i} + x_{s_0+i}) + \sum_{i=1}^{\ell} d_i \tilde{p}_i = \sum_{i=1}^{m-s_0} c_i x_{s_0+i}$$

is a sub-fractional function that does not converge to 0, since the functions  $x_{s_0+i}$  are sub-fractional and have pairwise distinct growth rates. This contradicts our hypothesis (specifically equation (8)) and our claim follows.

In conclusion, we see that the sequence  $v(n)$  can be written in the form

$$\prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^m u_i^{p_i(n)+x_i(n)} \prod_{i=1}^{\ell} w_i^{\tilde{p}_i(n)}, \tag{23}$$

where the functions  $x_i$  are strongly non-polynomial with distinct growth rates, the functions  $x_1, \dots, x_{s_0}$  are not sub-fractional, the functions  $x_{s_0+1}, \dots, x_s$  are sub-fractional and every non-trivial linear combination of the polynomials  $p_{s_0+1}, \dots, p_m, \tilde{p}_1, \dots, \tilde{p}_\ell$  is not an integer polynomial. We also recall that we have arranged the functions  $x_i$  to be in decreasing order with respect to their growth rates.

We can combine the last two factors of this product into one factor to simplify our problem a bit more. More specifically, we can rewrite the sequence  $v(n)$  in the form (we make some mild modifications in our notation here)

$$v(n) = \prod_{i=1}^{s_0} u_i^{p_i(n)+x_i(n)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_i(n)+x_i(n)}, \tag{24}$$

where  $s = m + l$ ,  $p_i, \tilde{p}_j$  are real polynomials, the functions  $x_i$  are strongly non-polynomial with distinct growth rates,  $x_1, \dots, x_{s_0}$  are not sub-fractional,  $x_{s_0+1}, \dots, x_s$  are sub-fractional and every non-trivial linear combination of the polynomials  $\tilde{p}_i$  is not an integer polynomial. Namely, our functions satisfy hypotheses (a) and (b) of Lemma 4.2.

To establish our assertion, it suffices to show that the sequence  $v(n)x$  (where  $v(n)$  is as in equation (24)) is equidistributed on the nilmanifold  $X' = \overline{u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} x}$  for any  $x \in X$ . We will prove this in the case  $x = \Gamma$  since the general case follows from this using the change of base point trick that we discuss in Appendix B (see §B.1.2). In addition, we can invoke Lemma B.2 to find a real number  $s_0$ , such that  $X' = \overline{(u_1^{s_0})^{\mathbb{Z}} \cdots (u_s^{s_0})^{\mathbb{Z}} \Gamma}$ . Replacing the functions  $p_i(t) + x_i(t)$  ( $1 \leq i \leq s_0$ ) by the functions  $(p_i(t) + x_i(t))/s_0$  and  $\tilde{p}_i(t) + x_i(t)$  ( $s_0 + 1 \leq i \leq s$ ) by  $(\tilde{p}_i(t) + x_i(t))/s_0$  (the assumptions on the linear combinations of the functions remain unaffected), we can reduce our problem to the case that  $X' = \overline{u_1^{\mathbb{Z}} \cdots u_s^{\mathbb{Z}} \Gamma}$ .

We want to show that for any continuous function  $F$  from  $X' = G'/\Gamma'$  ( $G'$  is connected, simply connected and  $\Gamma'$  is a uniform subgroup), the averages

$$\mathbb{E}_{1 \leq n \leq N} F(v(n)\Gamma')$$

converge to the integral  $\int_{X'} F \, dm_{X'}$ . Since Lipschitz functions are dense in the space  $C(X')$ , we may assume that  $F$  is Lipschitz continuous. In addition, we may assume after rescaling that  $\|F\|_{\text{Lip}(X')} \leq 1$ . Now, our claim follows immediately from Lemma 4.2.  $\square$

In the following part, we will prove Lemma 4.2. We divide the proof into two steps. During step 1, we will approximate the functions  $x_1, \dots, x_s$  by polynomials in a suitable short interval. Our goal is to reach an average over a short interval of the form  $[N, N + L(N)]$  of a sequence of the form  $F(g(n)x)$ , where  $F$  is Lipschitz and  $g(n)$  is a polynomial sequence on the nilmanifold  $X'$  (the polynomial sequence will vary with the parameter  $N$ ). This will be ensured by Proposition A.4. In step 2, we will use Theorem F to deduce that these averages are close to the integral of  $F$  for large values of  $N$ .



All the reductions above allow us to write  $v(n)$  in a form that will be appropriate for the application of the quantitative equidistribution theorem (after we perform the Taylor expansion). When we apply the Taylor expansion in the first step, the functions  $x_{s_0+1}, \dots, x_s$  will become approximately constant and thus the desired equidistribution will be mainly ‘affected’ by the polynomials  $\tilde{p}_{s_0+1}, \dots, \tilde{p}_s$ . However, the functions  $x_1, \dots, x_{s_0}$  will play a meaningful role in the equidistribution of our sequence. In particular, the presence of the functions  $x_1, \dots, x_{s_0}$  will imply ‘closeness’ of our averages to the integral of the Lipschitz function  $F$ , unless the projections of the elements  $u_1, \dots, u_{s_0}$  on the horizontal torus are zero. In this second case, condition (a) on the polynomials completes the proof. Lastly, the ‘linear independence’ condition of the polynomials  $\tilde{p}_{s_0+1}, \dots, \tilde{p}_s$  guarantees that the projection of the sequence  $v(n)$  on  $X'$  will be equidistributed on the entire nilmanifold  $u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} \Gamma$ , since otherwise, we would need to pass to some subnilmanifold to guarantee equidistribution (and to an appropriate arithmetic progression).

*Proof of Lemma 4.2. Step 1. Approximating by polynomials.* Let  $L(t)$  be a sub-linear function with  $\lim_{t \rightarrow +\infty} L(t) = +\infty$  that we will determine later. It suffices to show that the sequence of the averages

$$\mathbb{E}_{N \leq n \leq N+L(N)} F(v(n)\Gamma') \tag{25}$$

converges to  $\int_{X'} F \, dm_{X'}$ , since the conclusion would follow from Lemma 3.1. Reordering if necessary, we assume again that

$$x_1(t) \succ \cdots \succ x_{s_0}(t).$$

Let  $r$  be a very large natural number compared to the degrees of the polynomials  $p_i, \tilde{p}_j$  and the degrees of the functions  $x_i(t)$ . If  $r$  is sufficiently large, we have that  $x_i^{(r)}(t) = o_i(1)$  for all  $i \in \{1, \dots, s_0\}$ . Assuming again that  $r$  is sufficiently large, then for any function  $L(t)$  that satisfies

$$(x_i^{(r)}(t))^{-1/r} < L(t) < t^{1-\varepsilon'}$$

for some  $\varepsilon' > 0$  and all  $i \in \{1, \dots, s_0\}$ , we have that for each  $i \in \{1, \dots, s_0\}$ , there is a unique natural number  $k_i \geq r$  so that the sub-class  $S(x_i, k_i)$  contains the function  $L(t)$  (this follows from Proposition A.3). The fact that the function  $L(t)$  belongs to  $S(x_i, k_i)$  indicates that we have the relations

$$(x_i^{(k_i)}(t))^{-1/k_i} < L(t) < (x_i^{(k_i+1)}(t))^{-1/(k_i+1)}. \tag{26}$$

We can guarantee that the numbers  $k_i$  are also very large compared to the degrees of the polynomials  $p_j, \tilde{p}_j$  by enlarging the number  $r$  in the beginning. (For example, assuming that  $k_i$  is at least 10 times as large as the maximal degree appearing among the polynomials  $p_i, \tilde{p}_j$  and 10 times as large as the number  $s$  of all existing polynomials would suffice for our arguments.)

We use the Taylor expansion for the functions  $x_1(t), \dots, x_{s_0}(t)$  to write

$$x_i(N+h) = x_i(N) + \cdots + \frac{x_i^{k_i}(N)h^k}{k_i!} + o_N(1) = q_{i,N}(h) + o_N(1) \tag{27}$$

for  $0 \leq h \leq L(N)$  (for the explanation of the  $o_N(1)$  term, see the discussion after Proposition A.2). If, however, we have  $i > s_0$  (namely, in the case where the function  $x_i$  is sub-fractional), then

$$\max_{0 \leq h \leq L(N)} |x_i(N + h) - x_i(N)| = o_N(1). \tag{28}$$

In addition, we denote  $p_{i,N}(h) = p_i(N + h)$  and similarly  $\tilde{p}_{i,N}(h) = \tilde{p}_i(N + h)$  for every admissible value of  $i$ . Thus, we rewrite the expression in equation (25) as

$$\mathbb{E}_{0 \leq h \leq L(N)} F(w_N \prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_{i,N}(h)} \Gamma'), \tag{29}$$

where we discarded the  $o_N(1)$  terms, because  $F$  is continuous. Here,  $w_N = \prod_{i=s_0+1}^s u_i^{x_i(N)}$  but the explicit form of this term will not concern us, since we will only require that the element  $w_N$  belongs to the underlying group  $G'$  defining the nilmanifold  $X' = \overline{u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} \Gamma}$ .

In conclusion, we have reduced our problem to showing that given the nilmanifold  $X' = \overline{u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} \Gamma}$  (which is also equal to  $\overline{u_1^{\mathbb{Z}} \cdots u_s^{\mathbb{Z}} \Gamma}$ ), the averages in equation (29) converge. Here, the polynomials  $q_{i,N}$  are defined in equation (27) (they are essentially the Taylor polynomials of the Hardy field functions  $x_i$ ), while the polynomials  $p_{i,N}, \tilde{p}_{j,N}$  were defined by the relations  $p_{i,N} = p_i(N + h)$  and  $\tilde{p}_{j,N} = \tilde{p}_j(N + h)$ , where the  $p_i, \tilde{p}_j$  are polynomials with real coefficients. We also recall that the polynomials  $\tilde{p}_i$  are such that every non-trivial linear combination of them is not an integer polynomial. Under all these assumptions, we will show that the polynomial sequence (restricted to the range  $0 \leq h \leq L(N)$ ) inside the function  $F$  is  $\delta$ -equidistributed for  $N$  sufficiently large in the following step. We remark that the growth conditions in equation (26) imposed on the function  $L(t)$  will also play a crucial role in this.

*Step 2. Using the quantitative equidistribution theorem.* Let  $Z \cong \mathbb{T}^d$  be the horizontal torus of the nilmanifold  $X' = \overline{u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} \Gamma}$  and let  $\pi : X' \rightarrow Z$  denote the projection map. Let  $\delta > 0$  be sufficiently small (in the sense that Theorem F is applicable). We assert that the finite polynomial sequence

$$\left( \prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_{i,N}(h)} \Gamma' \right)_{0 \leq h \leq L(N)} \tag{30}$$

is  $\delta$ -equidistributed on the nilmanifold  $X'$  for  $N$  sufficiently large. If the claim does not hold for a natural number  $N$ , then by Theorem F, there exists a real number  $M > 0$  and a non-trivial horizontal character  $\chi_N$  of modulus  $\leq M$  such that

$$\left\| \chi_N \circ \pi \left( \prod_{i=1}^{s_0} u_i^{q_{i,N}(h)+p_{i,N}(h)} \prod_{i=s_0+1}^s u_i^{\tilde{p}_{i,N}(h)} \Gamma' \right) \right\|_{C^\infty[L(N)]} \leq M. \tag{31}$$

(The constant  $M$  depends only on  $\delta$ , the nilmanifold  $X'$  as well as the degrees of the polynomials  $q_i, p_i$ , which are all fixed in our arguments. The central property we need is that it is independent of the variable  $N$ .) Thus, if our prior assertion fails, then the above relation would hold for infinitely many  $N \in \mathbb{N}$ .

Our first goal is to eliminate the dependence of the characters  $\chi_N$  on the variable  $N$ . Note that the function  $\chi_N \circ \pi$  is a character on  $\mathbb{T}^d$  of modulus  $\leq M$  and, thus, has the form

$$(t_1, \dots, t_d) \rightarrow e(k_{1,N}t_1 + \dots + k_{d,N}t_d)$$

for  $k_{i,N} \in \mathbb{Z}$  with  $|k_{1,N}| + \dots + |k_{d,N}| \leq M$ . We also write  $\pi(u_i) = (u_{i,1}, \dots, u_{i,d})$  for the projections of the elements  $u_i$  on the horizontal torus. Then, a straightforward computation allows us to rewrite equation (31) as

$$\left\| e \left( \sum_{i=1}^{s_0} (q_{i,N}(h) + p_{i,N}(h))(k_{1,N}u_{i,1} + \dots + k_{d,N}u_{i,d}) + \sum_{i=s_0+1}^s (\tilde{p}_{i,N}(h))(k_{1,N}u_{i,1} + \dots + k_{d,N}u_{i,d}) \right) \right\|_{C^\infty[L(N)]} \leq M. \tag{32}$$

Since there are only finitely many choices for the numbers  $k_{1,N}, \dots, k_{d,N}$ , we have that if our claim fails, there are  $k_1, \dots, k_d \in \mathbb{Z}$ , so that the inequality

$$\left\| e \left( \sum_{i=1}^{s_0} (q_{i,N}(h) + p_{i,N}(h))(k_1u_{i,1} + \dots + k_du_{i,d}) + \sum_{i=s_0+1}^s (\tilde{p}_{i,N}(h))(k_1u_{i,1} + \dots + k_du_{i,d}) \right) \right\|_{C^\infty[L(N)]} \leq M \tag{33}$$

holds for infinitely many  $N \in \mathbb{N}$ . We will also denote the horizontal character corresponding to the  $d$ -tuple  $(k_1, \dots, k_d)$  by  $\chi$ . Thus, we have eliminated the dependence of the character  $\chi$  on  $N$ .

Denote  $\tilde{u}_i = k_1u_{i,1} + \dots + k_du_{i,d}$ . We will show that the above hypotheses imply that all the numbers  $\tilde{u}_i$  equal 0. Thus, suppose that this is not valid and we will reach a contradiction. We consider two cases.

*Case 1.* First, suppose that all of the numbers  $\tilde{u}_i$  with  $1 \leq i \leq s_0$  are zero, which implies that the first summand in equation (33) vanishes. Naturally, equation (33) is simplified to

$$\left\| e \left( \sum_{i=s_0+1}^s \tilde{p}_{i,N}(h)\tilde{u}_i \right) \right\|_{C^\infty[L(N)]} \leq M. \tag{34}$$

We recall here that we had defined  $\tilde{p}_{i,N}(h) = \tilde{p}_i(N + h)$ . Let  $Q(t) = \sum_{i=s_0+1}^s \tilde{u}_i \tilde{p}_i(t)$ . This is a linear combination of the polynomials  $\tilde{p}_i(t)$ . However, this linear combination is not a polynomial in  $\mathbb{Q}[t]$  due to our assumptions on the polynomials  $\tilde{p}_i(n)$ , unless, of course, all the coefficients  $\tilde{u}_i$  (for  $s_0 + 1 \leq i \leq s$ ) in this combination are zero, which we have supposed to not be the case. Thus,  $Q(t)$  has at least one irrational coefficient (except the constant term) and is equidistributed on  $\mathbb{T}$ . The relation in equation (34) implies that  $\|e(Q(N + h))\|_{C^\infty[L(N)]} \leq M$  for infinitely many  $N$ . It is not difficult to see by calculating the coefficients in  $Q(N + h)$  that this fails for  $N$  large enough.

*Case 2.* Suppose now that at least one of the numbers  $\tilde{u}_i$  with  $1 \leq i \leq s_0$  is non-zero. Furthermore, assume  $l$  is a positive integer that is larger than the degrees of the polynomials

$p_{i,N}(h), \tilde{p}_{j,N}(h)$  (for all admissible values of the indices  $i, j$ ) as well as the degrees of the functions  $x_i$ , but  $l$  is also smaller than all the numbers  $k_i$ . Recall that we have picked  $k_i$  to be very large in relation to the degrees of the polynomials  $p_i, \tilde{p}_j$  and degrees of the functions  $x_i$  in the beginning, thus we can find ‘many’ such numbers  $l$ . The fact that  $l$  is larger than the degrees of the functions  $x_i$  combined with Proposition A.1 implies that  $x_i^{(l)}(t) \rightarrow 0$ , as  $t \rightarrow +\infty$ .

For a number  $l$  as above, the coefficient of the term  $h^l$  in the polynomial appearing in equation (33) is equal to

$$\frac{1}{l!} \sum_{i=1}^{s_0} x_i^{(l)}(N) \tilde{u}_i$$

and, thus, it does not depend on the polynomials  $p_i, \tilde{p}_j$ . Using the definition of the smoothness norms, equation (33) implies that

$$L(N)^l \left\| \frac{1}{l!} \sum_{i=1}^{s_0} x_i^{(l)}(N) \tilde{u}_i \right\|_{\mathbb{T}} \leq M$$

for infinitely many  $N \in \mathbb{N}$ . The last inequality becomes

$$L(N)^l \left| \sum_{i=1}^{s_0} x_i^{(l)}(N) \tilde{u}_i \right| \leq l! M$$

for large enough  $N$ , because all functions  $x_i^{(l)}(t)$  go to 0. However, the Hardy field function inside the absolute value above has the same growth rate as the function  $x_1^{(l)}(t)$ , since the functions  $x_1, \dots, x_{s_0}$  are strongly non-polynomial and have distinct growth rates (recall that  $x_1$  has the largest growth rate among the  $x_i$ ), unless, of course,  $\tilde{u}_1 = 0$ . If the latter does not hold, we get

$$\left| x_1^{(l)}(N) \tilde{u}_1 \right| \leq \frac{C}{L(N)^l}$$

for infinitely many  $N$  and some constant  $C$ , which contradicts equation (26). Thus, we eventually deduce that  $\tilde{u}_1 = k_1 u_{1,1} + \dots + k_d u_{1,d} = 0$ . Repeating the same argument, we get inductively that  $\tilde{u}_i = k_1 u_{i,1} + \dots + k_d u_{i,d} = 0$  for all  $1 \leq i \leq s_0$ , which is a contradiction.

To summarize, we have shown that if the sequence in equation (30) is not  $\delta$ -equidistributed for all large enough  $N$ , then all the numbers  $\tilde{u}_i = k_1 u_{i,1} + \dots + k_d u_{i,d}$  are zero. Equivalently, we have  $\chi \circ \pi(u_i) = 0$  for all  $1 \leq i \leq s$ . This implies that the character  $\chi$  is the trivial character on  $X'$ . Indeed, the character  $\chi$  annihilates all elements  $u_1^{n_1} \dots u_s^{n_s} \Gamma$ , where  $n_1, \dots, n_s \in \mathbb{Z}$  and by density of those elements on  $X'$  (recall our assumption that  $X'$  is also equal to the nilmanifold  $u_1^{\mathbb{Z}} \dots u_s^{\mathbb{Z}} \Gamma$ ),  $\chi$  is zero everywhere. This is a contradiction (the horizontal characters appearing when we applied Theorem F are assumed to be non-trivial).

In conclusion, we have that the finite polynomial sequence in equation (30) is  $\delta$ -equidistributed for  $N$  sufficiently large. Thus, we conclude that the averages in equation (25) are  $\delta \|F(w_N \cdot)\|_{\text{Lip}(X')} = \delta \|F\|_{\text{Lip}(X')}$  close to the quantity  $\int_{X'} F(w_N x) dm_{X'}(x)$ .

The action of  $w_N$  on  $X'$  preserves the Haar measure of  $X'$ , so we get that the last integral is equal to  $\int_{X'} F(x) dm_{X'}(x)$ . Taking  $\delta \rightarrow 0$ , we finish the proof.  $\square$

*Proof of Theorem 1.1.* As we explained in the previous section (before the statement of Lemma 3.2), the first part follows from the second part (see also [4, Lemma 5.1]) and, in turn, this second part follows using similar arguments as in the proof of Theorem 1.2. We only highlight the main differences here. All the disparities appear in the part where we reduce Proposition 4.1 to Lemma 4.2.

(a) In equation (22), all the functions  $f_1, \dots, f_s$  satisfy equation (8) (there are no functions among the  $f_i$  that satisfy  $\lim_{t \rightarrow +\infty} |f_i(t)| < \infty$ ). We also have  $k = s$ .

(b) We do not have to make the reduction to the case where the polynomials  $\tilde{p}_1, \dots, \tilde{p}_\ell$  are linearly independent. There cannot be a non-trivial linear combination of them that is zero, because that would violate equation (8).

(c) The limit of the averages is again  $\int_{X'} F(x) dm_{X'}(x)$ , where  $X' = \overline{u_1^{\mathbb{R}} \cdots u_s^{\mathbb{R}} \Gamma}$  by Lemma 4.2. We would like to show that the limit is equal to  $\int_{X''} F dm''_{X'}$ , where  $X''$  is the nilmanifold  $\overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}} \Gamma}$ . Recall that each  $u_i$  is equal to  $b_1^{c_{i,1}} \cdots b_k^{c_{i,k}}$  (by equation (21)) and the numbers  $c_{i,j}$  form an invertible  $k \times k$  matrix (due to the linear independence assumption on the original functions  $a_1, \dots, a_k$ ). Thus, we can also write  $b_i = \prod_{j=1}^k u_j^{c'_{i,j}}$  for some numbers  $c'_{i,j}$  (here, we also use that the elements  $b_i$  are pairwise commuting). Combining the above, we have that  $\overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}}} = \overline{u_1^{\mathbb{R}} \cdots u_k^{\mathbb{R}}}$  and thus the closures of their projections on  $G/\Gamma$  define the same subnilmanifold.  $\square$

4.1. *Proof of Theorem 1.3.* Finally, we provide a proof of Theorem 1.3. We use the following proposition from [18]. Although it will not be used in the proof, we have to assume below that the Hardy field  $\mathcal{H}$  that we work with is closed under composition and compositional inversion of functions, since the following proposition was proven under this assumption.

PROPOSITION 4.3. [18, Proposition 3.1] *Let  $\mathcal{H}$  be a Hardy field that contains the field  $\mathcal{LE}$  of logarithmico-exponential functions and is closed under composition and compositional inversion of functions (when defined). Assume that the functions  $a_1, \dots, a_k \in \mathcal{H}$  have polynomial growth and suppose that the following two conditions hold:*

- (i) *the functions  $a_1, \dots, a_k$  dominate the logarithmic function  $\log t$ ;*
- (ii) *the pairwise differences  $a_i - a_j$  dominate the logarithmic function  $\log t$  for any  $i \neq j$ .*

*Then, there exists a positive integer  $s$ , such that for any measure-preserving system  $(X, \mu, T)$ , functions  $f_1 \in L^\infty(\mu)$  and  $f_{2,N}, \dots, f_{k,N} \in L^\infty(\mu)$ , all bounded by 1, with  $f_1 \perp Z_s(X)$ , the expression*

$$\sup_{|c_n| \leq 1} \left\| \mathbb{E}_{1 \leq n \leq N} c_n T^{\lfloor a_1(n) \rfloor} f_1 \cdot T^{\lfloor a_2(n) \rfloor} f_{2,N} \cdots T^{\lfloor a_k(n) \rfloor} f_{k,N} \right\|_{L^2(\mu)} \tag{35}$$

*converges to 0, as  $N \rightarrow +\infty$ .*

*Proof of Theorem 1.3.* Using a standard ergodic decomposition argument, we may assume that the system  $(X, \mu, T)$  is ergodic. We can also rescale the functions  $f_i \in L^\infty(\mu)$  so that they are 1-bounded. Our first objective is to apply Proposition 4.3 to reduce the problem to the case where the system  $X$  is a nilsystem. If the functions  $a_1, \dots, a_k$  are such that the conditions of Proposition 4.3 are satisfied, then this can be done instantly. If this does not hold, we have to perform a series of reductions to be able to apply Proposition 4.3. We do this in two steps.

(a) First, assume there exists one function among the  $a_i$  (say  $a_1$  for simplicity), which has growth rate smaller than or equal to  $\log t$ . Then, using equations (8) and (9), we deduce that  $a_1$  converges monotonically to some real number  $c$  and the integer part of  $a_1(n)$  becomes a constant. Thus, the asymptotic behaviour of the averages in equation (10) is the same if we substitute the term  $T^{\lfloor a_1(n) \rfloor} f_1$  with the term  $T^{\lfloor c \rfloor} f_1$ . Consequently, we only need to show that the averages

$$\mathbb{E}_{1 \leq n \leq N} T^{\lfloor a_2(n) \rfloor} f_2 \dots T^{\lfloor a_k(n) \rfloor} f_k$$

converge in norm. Repeating the same argument, we eliminate all functions  $a_i$  that grow slower than  $\log t$ .

(b) Due to the reduction in the previous step, we have a sub-collection of the original functions, so that all functions in this new set dominate  $\log t$ . We will denote this collection by  $a_1, \dots, a_k$  again, and our task is to show that the averages

$$\mathbb{E}_{1 \leq n \leq N} T^{\lfloor a_1(n) \rfloor} f_1 \dots T^{\lfloor a_k(n) \rfloor} f_k$$

converge in mean (for all systems). Our next objective is to eliminate pairs of functions, whose difference grows slower than  $\log t$  so that we can ultimately apply Proposition 4.3.

Assume that two of the functions (say  $a_1, a_2$ ) are such that their difference is dominated by  $\log t$ . We observe that the function  $a_1(t)$  goes to  $\pm\infty$  as  $t \rightarrow +\infty$ , since it dominates  $\log t$ . In that case, the function  $a_1(t)$  satisfies equation (8) and by Theorem A, the sequence  $a_1(n)$  is equidistributed (mod 1). Observe that since  $a_1 - a_2$  must satisfy equation (9), we must have  $a_2(t) = a_1(t) + c + x(t)$ , where the function  $x(t) \in \mathcal{H}$  converges to 0 monotonically and  $c$  is a real number. Thus, for  $t \in \mathbb{R}$  sufficiently large, we have

$$\lfloor a_2(t) \rfloor = \lfloor a_1(t) + x(t) + c \rfloor = \lfloor a_1(t) \rfloor + \lfloor c \rfloor + \varepsilon(t), \tag{36}$$

where  $\varepsilon(t) \in \{0, \pm 1, \pm 2\}$  and the value of  $\varepsilon(t)$  depends on whether the inequalities

$$\{a_1(t) + c\} + \{x(t)\} \leq 1$$

and

$$\{a_1(t)\} + \{c\} \leq 1$$

hold or not, as well as whether  $x(t)$  is eventually positive or negative.

Define  $A_z = \{t \in \mathbb{R}, \varepsilon(t) = z\}$  for  $z \in \{0, \pm 1, \pm 2\}$ . Then, we see that our multiple averages are equal to the sum

$$\sum_{z \in \{0, \pm 1, \pm 2\}} \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{A_z}(n) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor + z} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \dots T^{\lfloor a_k(n) \rfloor} f_k.$$

For a fixed  $z$ , we want to show that the corresponding average converges. For  $n \in \mathbb{N}$  large enough, we will approximate the sequence  $\mathbf{1}_{A_z}(n)$  by sequences of the form  $F(\{a_1(n)\})$ , where  $F$  is a continuous function.

We establish this for  $z = 0$  (the other cases follow similarly). Assume that  $x(t)$  decreases to 0 (the other case) is similar, which means that  $x(t)$  is eventually positive and also  $\{x(t)\} = x(t)$  for  $t$  sufficiently large. In addition, we can also assume that  $c$  is positive. Observe that for  $t \in A_0$ , we have

$$\lfloor a_2(t) \rfloor = \lfloor a_1(t) \rfloor + \lfloor c \rfloor$$

by the definition of  $A_0$ . This is equivalent to the inequalities

$$\begin{aligned} \{a_1(t) + c\} + \{x(t)\} &\leq 1, \\ \{a_1(t)\} + \{c\} &\leq 1, \end{aligned}$$

which can be condensed into

$$\{a_1(t)\} + \{x(t)\} \leq 1 - \{c\}, \tag{37}$$

since we assumed for simplicity that  $x(t)$  is eventually positive. To summarize, we have shown that

$$n \in A_0 \iff \{a_1(n)\} + \{x(n)\} \leq 1 - \{c\}. \tag{38}$$

Let  $\varepsilon > 0$  be a small number. Since we have that the function  $x(t)$  decreases to 0, we have that  $\{x(t)\} < \varepsilon$  for  $t$  large enough. Consider the set

$$A_\varepsilon = \{n \in \mathbb{N} : \{a_1(n)\} \leq 1 - \{c\} - \varepsilon\}.$$

Then, for sufficiently large values of  $n$ , we observe that if  $n \in A_\varepsilon$ , then the inequality

$$\{a_1(t)\} + \{x(t)\} \leq 1 - \{c\}$$

holds as well. Namely,  $A_\varepsilon \subseteq A_0$ . Let us denote  $B_\varepsilon = [0, 1 - c - \varepsilon]$  for convenience and observe that

$$\mathbf{1}_{A_\varepsilon}(n) = \mathbf{1}_{B_\varepsilon}(\{a_1(n)\}).$$

Now we approximate the function  $\mathbf{1}_{B_\varepsilon}$  by a continuous function in the uniform norm, where  $\mathbf{1}_{B_\varepsilon}$  is considered a function on the torus  $\mathbb{T}$  in the natural way. We can define a continuous function on  $\mathbb{T}$ , such that  $F_\varepsilon$  agrees with  $\mathbf{1}_{B_\varepsilon}$  on the set

$$[\varepsilon, 1 - \{c\} - 2\varepsilon] \cup [1 - \{c\}, 1 - \varepsilon]$$

and such that  $\|F_\varepsilon - \mathbf{1}_{B_\varepsilon}\|_\infty \leq 2$ . (In the case that  $c$  is an integer, we make natural modifications to this set. For example, one could define the function  $F_\varepsilon$  so that it agrees with  $\mathbf{1}_{B_\varepsilon}$  on  $[\varepsilon, 1 - 2\varepsilon]$ . Basically, we only require the function  $F_\varepsilon$  to agree with  $\mathbf{1}_{B_\varepsilon}$  on a set of measure  $1 - O(\varepsilon)$  for our argument to work.) We suppose that  $\varepsilon$  is small enough so that these intervals are well defined. Observe that  $\mathbf{1}_{B_\varepsilon}$  is equal to 1 on the first interval of this union and equal to 0 on the second interval.

Observe that

$$\begin{aligned} A_0 \setminus A_\varepsilon &= \{n \in \mathbb{N} : 1 - \{c\} - \varepsilon < \{a_1(n)\} \leq 1 - \{c\} - \{x(n)\}\} \\ &\subseteq \{n \in \mathbb{N} : \{a_1(n)\} \in [1 - \{c\} - \varepsilon, 1 - \{c\}]\}. \end{aligned}$$

Since the function  $a_1(t)$  is equidistributed modulo 1, we conclude that the set  $A_0 \setminus A_\varepsilon$  has upper density at most  $\varepsilon$ . Therefore, we have

$$\begin{aligned} & \left\| \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{A_0}(n) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right. \\ & \quad \left. - \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{B_\varepsilon}(\{a_1(n)\}) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right\|_{L^2(\mu)} \\ & \leq \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{A_0 \setminus A_\varepsilon}(n) \|\mathbf{1}_{A_\varepsilon} - \mathbf{1}_{A_\varepsilon \setminus A_0}\|_\infty \leq 2\varepsilon + o_N(1), \end{aligned} \tag{39}$$

where we used the fact that  $\mathbf{1}_{A_\varepsilon}(n) = \mathbf{1}_{B_\varepsilon}(\{a_1(n)\})$  for all  $n \in \mathbb{N}$ , the trivial bound for the values of  $n \in A_0 \setminus A_\varepsilon$  and the fact that the set  $A_0 \setminus A_\varepsilon$  has upper density at most  $\varepsilon$ .

We do a similar comparison for the averages weighted by  $F_\varepsilon(\{a_1(n)\})$  and  $\mathbf{1}_{B_\varepsilon}(\{a_1(n)\})$ . To be more specific, we reiterate that the functions  $\mathbf{1}_{B_\varepsilon}$  and  $F_\varepsilon$  agree on the set

$$[\varepsilon, 1 - \{c\} - 2\varepsilon] \cup [1 - \{c\}, 1 - \varepsilon].$$

Accordingly, we have  $\mathbf{1}_{B_\varepsilon}(\{a_1(n)\}) = F_\varepsilon(\{a_1(n)\})$ , unless

$$\{a_1(n)\} \in [0, \varepsilon] \cup (1 - \{c\} - 2\varepsilon, 1 - \{c\}) \cup (1 - \varepsilon, 1).$$

Let  $C_\varepsilon$  denote the set of  $n \in \mathbb{N}$  for which  $\{a_1(n)\}$  belongs to this union. This union has measure  $4\varepsilon$ , which implies that the upper density of  $C_\varepsilon$  is at most  $4\varepsilon$  (since  $a_1(n)$  is equidistributed modulo 1). Hence, we infer that

$$\begin{aligned} & \left\| \mathbb{E}_{1 \leq n \leq N} F_\varepsilon(\{a_1(n)\}) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right. \\ & \quad \left. - \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{B_\varepsilon}(\{a_1(n)\}) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right\|_{L^2(\mu)} \\ & \leq \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{C_\varepsilon}(n) \|F_\varepsilon - \mathbf{1}_{B_\varepsilon}\|_\infty \leq 8\varepsilon + o_N(1), \end{aligned} \tag{40}$$

where we used the fact that  $\mathbf{1}_{B_\varepsilon}(\{a_1(n)\}) = F_\varepsilon(\{a_1(n)\})$  for all  $n$  on the complement of  $C_\varepsilon$ , the trivial bound for the values of  $n \in C_\varepsilon$  and the fact that  $C_\varepsilon$  has upper density at most  $4\varepsilon$ .

Combining equations (39) and (40), we deduce that

$$\begin{aligned} & \left\| \mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{A_0}(n) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right. \\ & \quad \left. - \mathbb{E}_{1 \leq n \leq N} F_\varepsilon(\{a_1(n)\}) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdot T^{\lfloor a_3(n) \rfloor} f_3 \cdots T^{\lfloor a_k(n) \rfloor} f_k \right\|_{L^2(\mu)} \\ & \leq 10\varepsilon + o_N(1). \end{aligned} \tag{41}$$

Taking  $\varepsilon \rightarrow 0$ , we deduce that it is sufficient to verify that the averages

$$\mathbb{E}_{1 \leq n \leq N} F(\{a_1(n)\}) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdots T^{\lfloor a_k(n) \rfloor} f_k$$

converge for any continuous function  $F$  on  $\mathbb{T}$ . This would imply that the averages

$$\mathbb{E}_{1 \leq n \leq N} \mathbf{1}_{A_0}(n) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdots T^{\lfloor a_k(n) \rfloor} f_k$$

converge in norm.



After approximating  $F$  by trigonometric polynomials (in the uniform norm), it suffices to show that the averages

$$\mathbb{E}_{1 \leq n \leq N} e(l_1 a_1(n)) T^{\lfloor a_1(n) \rfloor} (f_1 \cdot T^{\lfloor c \rfloor} f_2) \cdots T^{\lfloor a_k(n) \rfloor} f_k$$

converge in norm for any  $l_1 \in \mathbb{Z}$ . Note that the function  $a_2(t)$  has vanished and its role has been replaced by the sequence  $e(l_1 a_1(n))$ .

We repeat this process until we eliminate all pairs of functions, whose difference grows slower than  $\log t$ , where at each step, our averages are multiplied by a sequence of the form  $e(l_i a_i(n))$  ( $l_i \in \mathbb{Z}$ ). After finitely many iterations, our problem eventually reduces to the following: let  $a_1, \dots, a_k$  satisfy equation (8) or equation (9) and let  $b_1, \dots, b_m$  be a subset of  $\{a_1, \dots, a_k\}$ , so that the functions  $b_1, \dots, b_m$  satisfy the hypotheses of Proposition 4.3. Then, for any integers  $l_1, \dots, l_k$ , the averages

$$\mathbb{E}_{1 \leq n \leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n)) T^{\lfloor b_1(n) \rfloor} f_1 \cdots T^{\lfloor b_m(n) \rfloor} f_m$$

converge in  $L^2(\mu)$  for all functions  $f_1, \dots, f_m \in L^\infty(\mu)$ .

Now we can apply Proposition 4.3 and use a standard telescopic argument to show that the limiting behaviour of the above averages does not change if we replace the functions  $f_i$  by their projections to the factor  $Z_s(X)$  (the number  $s$  is the one given by Proposition 4.3). However, by Theorem E, the factors  $Z_s(X)$  are inverse limits of  $s$ -step nilsystems. Thus, by another standard limiting argument, we may reduce to the case that the space  $X$  is a nilmanifold and  $\mu$  is its Haar measure, while the transformation  $T$  is the action (by left multiplication) of an element  $g$  on  $X$ . Finally, we can approximate the functions  $f_i$  by continuous functions and reduce our problem to the following.

If  $X = G/\Gamma$  is a nilmanifold with  $g \in G$  and the functions  $a_1, \dots, a_k, b_1, \dots, b_m \in \mathcal{H}$  are as above, then for any continuous functions  $f_1, \dots, f_m$ , the averages

$$\mathbb{E}_{1 \leq n \leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n)) f_1(g^{\lfloor b_1(n) \rfloor} x) \cdots f_m(g^{\lfloor b_m(n) \rfloor} x)$$

converge in mean.

We show that these averages converge pointwise for every  $x \in X$ . We recall that the functions  $b_1, \dots, b_m$  belong to the set  $\{a_1, \dots, a_k\}$  (this is the only thing that we will need to use for the rest of the proof).

First of all, it suffices to show that the averages

$$\mathbb{E}_{1 \leq n \leq N} e(l_1 a_1(n) + \cdots + l_k a_k(n)) f_1(g^{b_1(n)} x) \cdots f_m(g^{b_m(n)} x)$$

converge pointwise, where  $X = G/\Gamma$  is such that  $G$  is connected, simply connected nilpotent Lie group (basically, we can remove the integer parts appearing in the iterates). This follows by standard modifications in the proof of Lemma 3.2 (the fact that we have the coefficients  $e(l_1 a_1(n) + \cdots + l_k a_k(n))$  in the final expression does not affect the argument), so we omit the details.

Now, observe that we can write the above averages in the form

$$\mathbb{E}_{1 \leq n \leq N} F_0(g_0^{l_1 a_1(n) + \cdots + l_k a_k(n)} \tilde{x}) F_1(\tilde{g}^{b_1(n)} \tilde{x}) \cdots F_m(\tilde{g}^{b_m(n)} \tilde{x}),$$

where  $g_0 = (1_{\mathbb{T}}, e_G)$  and  $\tilde{g} = (1_{\mathbb{T}}, g)$  act on the product nilmanifold  $\mathbb{T} \times X$ , the point  $\tilde{x}$  is just  $(\mathbb{Z}, x)$  and the functions  $F_i$  are defined by

$$F_0(y\mathbb{Z}, a\Gamma) = e(y) \quad \text{and} \quad F_i(y\mathbb{Z}, a\Gamma) = f_i(a\Gamma) \quad \text{for } i \geq 1.$$

These are continuous functions on  $\mathbb{T} \times X$ . The functions  $l_1 a_1(t) + \dots + l_k a_k(t), b_1(t), \dots, b_m(t)$  satisfy the hypotheses of Theorem 1.2 (since the functions  $a_1, \dots, a_k$  do) and the result follows.  $\square$

*Acknowledgements.* I would like to thank my PhD advisor Nikos Frantzikinakis for many helpful discussions. I would also like to thank the anonymous referee for pointing out corrections in the previous versions of the paper and for several additional valuable suggestions that improved the overall presentation of the article. The author was supported by the Research Grant ELIDEK HFRI-FM17-1684 and ELIDEK-Fellowship number 5367 (3rd Call for HFRI Ph.D. Fellowships) during the preparation of this article.

A. *Appendix. Hardy field functions in short intervals*

A.1. *Growth rates of Hardy field functions.* All of the results presented below were proven in [18] and thus we omit their proofs. We refer the reader to the example in §3, where we establish Theorem 1.2 in the case where we have two simple functions. In that example, we do not need any special lemmas to show that we can find a common Taylor expansion, because we can perform the calculations by hand. However, in the proofs of Theorems 1.1 and 1.2, we need to show that we can always do the same common polynomial expansion for general functions.

The first two propositions are some elementary facts concerning the growth rates of derivatives of functions in a Hardy field.

PROPOSITION A.1. [18, Proposition A.1] *Let  $f \in \mathcal{H}$  have polynomial growth. Then, for any natural number  $k$ , we have*

$$f^{(k)}(t) \ll \frac{f(t)}{t^k}.$$

*In addition, if  $t^\delta < f(t)$  for some  $\delta > 0$ , we have*

$$f'(t) \sim \frac{f(t)}{t}.$$

The above proposition establishes that if we have a function in  $\mathcal{H}$  that has polynomial growth, then its derivatives of large enough order will be functions that converge to 0. The next lemma implies that a particular growth relation holds between consecutive derivatives (of large enough order).

PROPOSITION A.2. [18, Proposition A.2] *Let  $f \in \mathcal{H}$  be strongly non-polynomial with  $f(t) \succ \log t$ . Then, for  $k$  sufficiently large, we have*

$$1 < |f^{(k)}(t)|^{-1/k} < |f^{(k+1)}(t)|^{-1/(k+1)} < t.$$

Let us demonstrate how this proposition is used to get a polynomial expansion in short intervals for a single function. Let  $a \in \mathcal{H}$  be a strongly non-polynomial function that satisfies the growth condition  $a(t) \succ \log t$ . Let  $k$  be a positive integer that is large enough so that we can apply the two preceding propositions. We argue that we can find a function  $L(t)$  (not necessarily in  $\mathcal{H}$ ) such that

$$|a^{(k)}(t)|^{-1/k} \prec L(t) \prec |a^{(k+1)}(t)|^{-1/(k+1)}. \tag{A.1}$$

For instance, the geometric mean of the functions  $|a^{(k)}(t)|^{-1/k}$  and  $|a^{(k+1)}(t)|^{-1/(k+1)}$  is a suitable choice for our purposes.

We will examine the function  $a$  in intervals of the form  $[N, N + L(N)]$  and approximate it by a polynomial, which will vary with  $N$ . Observe that if  $0 \leq h \leq L(N)$ , then we have

$$a(N + h) = a(N) + \dots + \frac{h^k a^{(k)}(N)}{k!} + \frac{h^{k+1} a^{(k+1)}(\xi_{h,N})}{(k + 1)!}$$

for some  $\xi_{h,N} \in [N, N + h]$ . Using the largeness of  $k$ , Proposition A.2 implies that  $|a^{(k+1)}(t)| \rightarrow 0$  monotonically (the monotonicity follows from the fact that the function  $a^{(k+1)}(t)$  belongs to  $\mathcal{H}$ ). Then, for  $N$  sufficiently large,

$$\left| \frac{h^{k+1} a^{(k+1)}(\xi_{h,N})}{(k + 1)!} \right| \leq \left| \frac{L(N)^{k+1} a^{(k+1)}(N)}{(k + 1)!} \right| \prec 1,$$

because of equation (A.1). Furthermore, we have that

$$\left| \frac{L(N)^k a^{(k)}(N + L(N))}{k!} \right| \rightarrow +\infty.$$

Indeed, since  $L(t)$  is a sub-linear function by Proposition A.2, we infer that the two functions  $a^{(k)}(t + L(t))$  and  $a^{(k)}(t)$  have the same growth rate and thus we only need to prove that

$$\left| \frac{L(N)^k a^{(k)}(N)}{k!} \right| \rightarrow +\infty. \tag{A.2}$$

This follows similarly as above. To summarize, we have

$$a(N + h) = a(N) + \dots + \frac{h^k a^{(k)}(N)}{k!} + o_N(1) \quad \text{for } 0 \leq h \leq L(N). \tag{A.3}$$

Therefore, functions that satisfy equation (A.1) have the following distinctive property: the sequence  $a(n)$ , when restricted to the intervals  $[N, N + L(N)]$  as above, is asymptotically equal to a polynomial sequence (that depends on  $N$ ) of degree exactly  $k$ . This motivates us to study the properties of functions that satisfy equation (A.1). The main goal is to accomplish the same for several functions  $a_1, \dots, a_m$  in a Hardy field  $\mathcal{H}$ . This is relatively straightforward to do by hand in explicit examples, like the one in §3. In the more abstract setting, if we manage to show that we can find a function  $L(t)$ , so that equation (A.1) is satisfied for all functions  $a_1, \dots, a_m$  (the integer  $k$  is allowed to be different for each function), then we will establish that a polynomial expansion like the one

in equation (A.3) holds for all the functions  $a_1, \dots, a_m$  simultaneously. We will introduce some notions shortly that will assist us in this endeavour.

A.2. *The sub-classes  $S(a, k)$ .* Let  $a \in \mathcal{H}$  be a strongly non-polynomial function such that  $a(t) \gg t^\delta$ , for some  $\delta > 0$  (namely, we exclude sub-fractional functions). For  $k \in \mathbb{N}$  sufficiently large (we only require that  $a^{(k)}(t) \rightarrow 0$ ), we define the subclass  $S(a, k)$  of  $\mathcal{H}$  as

$$S(a, k) = \{g : g(t) \prec t \text{ and } |a^{(k)}(t)|^{-1/k} \leq g(t) \prec |a^{(k+1)}(t)|^{-1/(k+1)}\},$$

where the notation  $g(t) \leq f(t)$  signifies that the limit  $\lim_{t \rightarrow \infty} |f(t)/g(t)|$  is non-zero. The purpose of the classes  $S(a, k)$  is to characterize the growth relation in equation (A.1). We will use the following lemma.

LEMMA A.3. [18, Lemma A.3] *Let  $a \in \mathcal{H}$  be a strongly non-polynomial function with  $a(t) \gg t^\delta$ , for some  $\delta > 0$ .*

- (i) *The class  $S(a, k)$  is non-empty for  $k$  sufficiently large.*
- (ii) *For any  $0 < c < 1$  sufficiently close to 1, there exists  $k_0 \in \mathbb{N}$  (depending on  $c$ ), such that the function  $t \rightarrow t^c$  belongs to  $S(a, k_0)$ .*
- (iii) *The class  $S(a, k)$  does not contain all functions of the form  $t \rightarrow t^c$  for  $c$  sufficiently close to 1.*

A naive way to think of the sub-classes is like a sequence of disjoint intervals on a line (with no gaps between consecutive intervals). Property (ii) in the above lemma implies that each function of the form  $t^c$  for  $c$  close to 1 belongs to a unique  $S(a, k)$ . We can demonstrate that this actually holds if the fractional power  $t^c$  is replaced by any function  $g$  satisfying a growth condition of the form  $t^{c_1} \prec g(t) \prec t^{c_2}$ , where  $c_1$  must be sufficiently close to 1.

PROPOSITION A.4. *Let  $a_1, \dots, a_k$  be strongly non-polynomial functions in  $\mathcal{H}$  of polynomial growth, such that all the functions  $a_i$  dominate some fractional power  $t^\delta$  for some  $\delta > 0$ . There exists  $0 < C < 1$  depending only on the functions  $a_1, \dots, a_k$ , such that if the function  $L(t)$  satisfies*

$$t^C \prec L(t) \prec t^{1-\varepsilon}$$

*for some  $\varepsilon > 0$ , then there exist positive integers  $k_i$  (that depend on  $L(t)$ ), such that  $L(t) \in S(a_i, k_i)$  for every  $i \in \{1, \dots, k\}$ . In addition, for any positive real number  $M$ , there exists a constant  $A = A(M, a_1, \dots, a_k) \in (0, 1)$ , such that if*

$$t^A \prec L(t) \prec t^{1-\varepsilon}$$

*for some  $\varepsilon > 0$ , then we have  $k_i > M$  for every  $i \in \{1, \dots, k\}$ .*

*Proof.* It is apparent that we only need to establish the assertion in the case  $k = 1$  (namely, when we have only one function). Therefore, we fix a strongly non-polynomial function  $a$  that is not sub-fractional and recall that by Lemma A.3, there exists a constant  $C < 1$  depending only on  $a(t)$ , such that every function of the form  $t^c$  with  $c > C$  belongs to the class  $S(a, n_c)$  for some natural number  $n_c$ . Now, assume that the function  $L(t)$  satisfies

$$t^C \prec L(t) \prec t^{c_1} \tag{A.4}$$

for some  $C < c_1 < 1$ . Then, because both  $t^C$  and  $t^{c_1}$  belong to the sub-classes  $S(a, n_C)$  and  $S(a, n_{c_1})$ , respectively, for some  $n_C, n_{c_1} \in \mathbb{N}$ , we get that  $L(t)$  belongs to  $S(a, n_3)$  for some integer  $n_3$  that satisfies  $n_C \leq n_3 \leq n_{c_1}$ .

Now we establish the second part. Let  $M$  be a fixed real number and consider a fractional power  $t^{c_2}$  with  $C < c_2 < 1$ , so that  $t^{c_2}$  belongs to  $S(a, n_{c_2})$  for some  $c_2 > M$ . Such a fractional power exists, which is evident by combining the second and third statements of Lemma A.3. Thus, if  $L(t)$  satisfies

$$t^{c_2} \prec L(t) \prec t^{1-\varepsilon}$$

for some  $\varepsilon > 0$ , we have that  $L(t) \in S(a, k')$  (by the first part) for a positive integer  $k'$  with  $k' \geq n_{c_2} > M$ . The claim follows.  $\square$

The first part of Proposition A.4 implies that if we are given functions  $a_1, \dots, a_k$  that satisfy the hypotheses, then we can find a sub-linear function  $L(t)$ , such that  $L(t) \in S(a_i, k_i)$ . This asserts that the function  $a_i$  will be approximated by a polynomial of degree  $k_i$  in short intervals of the form  $[N, N + L(N)]$  for every  $i \in \{1, \dots, k\}$ . Furthermore, the second part establishes that we can make the degrees  $k_i$  of the Taylor polynomials arbitrarily large, as long as we take the function  $L(t)$  to grow ‘sufficiently fast’ (faster than some appropriate power  $t^C$  with  $C < 1$ ).

The sub-classes  $S(a, k)$  were defined for functions that are not sub-fractional. The above argument does not extend to these latter functions. As an example, let us fix a number  $\delta$  with  $0 < \delta < 1$  and a sub-fractional function  $a \in \mathcal{H}$ . If we consider the function  $L(t) = t^\delta$  and try to repeat the same approximations to obtain an analogue of equation (A.3), we run into an issue. Clearly, it is easy to see that

$$\max_{0 \leq h \leq L(N)} |a(N + h) - a(N)| = o_N(1),$$

using the mean value theorem. Thus, the sequence  $a(n)$ , when it is restricted to the interval  $[N, N + L(N)]$ , is  $o_N(1)$  close to the value  $a(N)$ , which signifies that it is approximately equal to a constant on this interval (or equivalently, all polynomial expansions we get are of degree 0). This could be circumvented if we considered sub-linear functions  $L(t)$  that grow faster than all the powers  $t^\delta, 0 < \delta < 1$ , such as the function  $t / \log t$ . If we do this however, the growth condition in equation (A.1) can never hold for functions that are not sub-fractional (in simple terms, there can be no polynomial approximation of finite degree). (Concerning the problem of finding characteristic factors for ergodic averages involving Hardy field iterates, there was a workaround for this issue in [18] using a double-averaging trick. Unfortunately, the same argument breaks down in the setting of pointwise convergence on nilmanifolds. See also Remark 4.) We omit the specific details of this deduction.

**A.3. Decomposing Hardy field functions.** We consider a Hardy field  $\mathcal{H}$  that contains the polynomials and let  $a$  be a function in  $\mathcal{H}$ . We partition  $\mathcal{H}$  into equivalence classes by the relation  $f \sim g$ , which is equivalent to saying that the limit of  $f(t)/g(t)$  as  $t \rightarrow +\infty$  is a non-zero real number. In simple terms,  $f, g$  are in the same equivalence class if and only if they have the same growth rate. We put the zero function in its own equivalence class.

We will define the strongly non-polynomial growth rate of a function  $a \in \mathcal{H}$  as follows.

(i) If  $a$  is a strongly non-polynomial function (recall the definition in §2), we define it to be the equivalence class of  $a$ .

(ii) If  $a$  is not strongly non-polynomial, then it can be written in the form  $p(t) + x(t)$ , where  $p(t)$  is a polynomial and  $x(t)$  is a strongly non-polynomial function (or the zero function) with  $x(t) < p(t)$ . Observe that  $x(t)$  is a function in  $\mathcal{H}$ , since our Hardy field contains the polynomials. We define the strongly non-polynomial growth rate of  $a$  as the equivalence class of the function  $x \in \mathcal{H}$ .

The strongly non-polynomial growth rate is defined for any function  $a \in \mathcal{H}$ . It is well defined, in the following sense: consider a function  $a \in \mathcal{H}$  like in case (ii) above, which has two different representations as  $p_1(t) + x_1(t)$  and  $p_2(t) + x_2(t)$ , where  $p_1, p_2$  are polynomials,  $x_1, x_2$  are strongly non-polynomial, and  $x_1(t) < p_1(t)$  and  $x_2(t) < p_2(t)$ . Then, we must have  $x_1(t) \sim x_2(t)$ . An example where such distinct representations may exist is the function  $a(t) = t^2 + t + t^{3/2}$ . We can choose  $p_1(t) = t^2, x_1(t) = t^{3/2} + t$  and  $p_2(t) = t^2 + t, x_2(t) = t^{3/2}$ . While  $x_1 \neq x_2$ , these two functions have the same growth rate.

A simple observation is that if a function  $a \in \mathcal{H}$  is written in the form  $p(t) + x(t)$ , where  $p$  is polynomial and  $x$  is strongly non-polynomial, then the functions  $a$  and  $x$  have the same strongly non-polynomial growth rate (one could alternatively use this remark to present another equivalent definition of the strongly non-polynomial growth rate).

Finally, we also say that  $a \in \mathcal{H}$  has trivial growth rate if  $\lim_{t \rightarrow +\infty} a(t) = 0$ . Recall that we also included these functions when we defined the strongly non-polynomial functions. We will now prove the following lemma.

LEMMA A.5. *Let  $\mathcal{H}$  be a Hardy field that contains the polynomials and let  $a_1, \dots, a_k \in \mathcal{H}$  be arbitrary functions. Then, the set  $\mathcal{L}(a_1, \dots, a_k)$  of non-trivial linear combinations has a basis  $(g_1, \dots, g_m, h_1, \dots, h_\ell)$ , where  $m, \ell$  are non-negative integers, such that the functions  $h_1, \dots, h_\ell$  have the form  $p_i(t) + o_i(1)$ , where  $p_i$  is a real polynomial for every  $1 \leq i \leq \ell$  and  $g_1, \dots, g_m$  have distinct and non-trivial strongly non-polynomial growth rates.*

*Proof.* We can restrict our attention to the case that the functions  $a_1, \dots, a_k$  are linearly independent (otherwise, we pass to a maximal subset of these functions whose elements are linearly independent). We induct on  $k$ . For  $k = 1$ , we have nothing to prove. Assume the claim holds for all integers smaller than  $k$ . All functions considered below are implicitly assumed to belong to  $\mathcal{H}$ .

We may write each of the functions  $a_1, \dots, a_k$  in the form  $p_i(t) + x_i(t)$ , where  $p_i$  are real polynomials and  $x_i(t)$  are strongly non-polynomial functions (either one of the functions  $p_i, x_i$  may also be identically zero). After reordering, we may assume that

$$x_1(t) \gg x_2(t) \gg \dots \gg x_k(t).$$

Now, we define the number  $l \in \{0, 1, \dots, k\}$  to be the smallest natural number for which all functions  $x_{l+1}(t), x_{l+2}(t)$  and so on have limit zero (as  $t \rightarrow +\infty$ ). If none of the  $x_i$  have limits going to 0, then we just set  $l = k$ .

We consider two cases.

(i) If the functions  $x_1, \dots, x_l$  have distinct growth rates, then we are done. In this case, the functions  $g_j$  appearing in the statement are the functions  $p_i(t) + x_i(t)$  for  $1 \leq i \leq l$ , while the role of the functions  $h_j$  is performed by the functions  $p_i(t) + x_i(t)$  for  $i > l$  (observe that for  $i > l$ , we have that  $x_i(t)$  have trivial growth rate due to the definition of  $l$ ). The strongly non-polynomial growth rates of the former set of functions are equal to the growth rates of the functions  $x_1, \dots, x_l$ , which are pairwise distinct.

(ii) Assume now two of the functions among  $x_1, \dots, x_l$  have the same growth rate. In particular, let  $k_0$  be the smallest integer such that  $x_{k_0} \sim x_{k_0+1}$  (obviously  $k_0 < l$ ) and let  $r \geq 1$  be the largest integer such that

$$x_{k_0} \sim x_{k_0+1} \sim \dots \sim x_{k_0+r}.$$

For  $k_0 + 1 \leq i \leq k_0 + r$ , we can write  $x_i(t) = x_{k_0}(t) + y_i(t)$ , where  $y_i(t) \prec x_i(t)$ . Using this, we can write  $a_{k_0}(t) = p_{k_0}(t) + x_{k_0}(t)$  and

$$a_i(t) = (p_{k_0}(t) + x_{k_0}(t)) + (p_i(t) - p_{k_0}(t) + y_i(t)) \quad \text{for } k_0 + 1 \leq i \leq k_0 + r.$$

Now we apply the induction hypothesis on the collection of functions

$$\{p_{k_0+1}(t) - p_{k_0}(t) + y_{k_0+1}(t), \dots, p_{k_0+r}(t) - p_{k_0}(t) + y_{k_0+r}(t), \\ p_{k_0+r+1}(t) + x_{k_0+r+1}(t), \dots, p_k(t) + x_k(t)\}.$$

This gives a basis  $(g_1, \dots, g_m, u_1, \dots, u_\ell)$  for this set of functions, with the properties outlined in the statement. We add the functions  $p_1(t) + x_1(t), \dots, p_{k_0}(t) + x_{k_0}(t)$  to the functions  $g_1, \dots, g_m$  and add the functions  $p_i(t) + x_i(t)$ ,  $l < i \leq k$ , to the collection  $u_1, \dots, u_\ell$ . (Recall that  $x_i(t)$  goes to 0 for  $l < i \leq k$ .) In this way, we construct a basis for the original collection  $a_1, \dots, a_k$  with the asserted properties (if the functions that we have constructed are not linearly independent, then we can just pass to a subset of these functions that will form a basis). Indeed, we only have to check that the functions

$$p_1(t) + x_1(t), \dots, p_{k_0}(t) + x_{k_0}(t), g_1(t), \dots, g_m(t)$$

have distinct strongly non-polynomial growth rates. This follows by noting that the strongly non-polynomial growth rates of the functions  $g_1, \dots, g_m$  cannot be larger than the growth rates of the functions  $y_i$ , which all grow strictly slower than  $x_{k_0}$ . Thus, the function  $p_{k_0}(t) + x_{k_0}(t)$  has a bigger strongly non-polynomial growth rate than all of the functions  $g_1, \dots, g_m$ . Furthermore, the strongly non-polynomial growth rate of the function  $p_i(t) + x_i(t)$  ( $1 \leq i \leq k_0$ ) is the same as  $x_i(t)$ , and these are all pairwise distinct by the definition of  $k_0$ . The claim follows. □

*Remark A.6.*

- (i) Note that we do not require that the functions  $a_1, \dots, a_k$  have polynomial growth in the above lemma.
- (ii) A very simple example that illustrates the above decomposition is the following: assume that we have the functions  $a_1(t) = t^2 + t^{3/2}$ ,  $a_2(t) = t^{3/2}$ ,  $a_3(t) = 2t^{3/2} + t^2$  and  $a_4(t) = t^{3/2} + t \log t + t^3$ . These four functions are clearly linearly dependent. The above lemma provides the basis  $(g_1, g_2, h_1)$ , where  $g_1(t) = t^{3/2}$ ,

$g_2(t) = t \log t + t^3$  and  $h_1(t) = t^2$ . The main property (which will be important in the proof of Theorem 1.2) is that the functions  $g_1, g_2$  have distinct strongly non-polynomial growth rates ( $t^{3/2}, t \log t$ , respectively), even though  $g_2$  grows like  $t^3$  (i.e. a polynomial).

B. *Appendix. Nilmanifolds and quantitative equidistribution theory*

B.1. *Background on nilmanifolds.* A large portion of the material concerning nilmanifolds (excluding the quantitative equidistribution results) can be found in [12, Part 3], where there is a focus on the ergodic theoretic point of view. For a more general presentation of the theory of nilpotent Lie groups, see also [3].

Let  $G$  be a topological group. A subgroup  $H$  of a topological group  $G$  is called discrete if there is a cover of  $H$  by open sets of  $G$ , such that each of these open sets contains exactly one element of  $H$ . It is called co-compact if the quotient topology makes  $G/H$  a compact space. We call a subgroup with both of the above properties uniform and we will use the letters  $\Gamma$  or  $\Delta$  to denote such subgroups.

Let  $G$  be a  $k$ -step nilpotent Lie group and  $\Gamma$  be a uniform subgroup. The space  $X = G/\Gamma$  is called a  $k$ -step nilmanifold.

Let  $b$  be any element in  $G$ . Then,  $b$  acts on  $G$  by left multiplication. Let  $m_X$  be the image of the Haar measure of  $G$  on  $X$  under the natural projection map. Then,  $m_X$  is invariant under the action of the element  $b$  (and therefore the action of  $G$ ). If we set  $T(g\Gamma) = (bg)\Gamma$ , then the transformation  $T$  is called a nilrotation, and  $(X, m_X, T)$  is called a nilsystem. If the transformation  $T$  is ergodic, we say that  $b$  acts ergodically on the nilmanifold  $X$ . It can be proven that  $b$  acts ergodically on  $X$  if and only if the sequence  $(b^n x)_{n \in \mathbb{N}}$  is dense on  $X$  for all  $x \in X$  (see, for instance, [12, Ch. 11]).

Let  $x_n$  be a sequence of elements on  $X = G/\Gamma$ . We say that  $x_n$  is equidistributed on  $X = G/\Gamma$  if and only if for every continuous function  $F : X \rightarrow \mathbb{C}$ , we have

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{1 \leq n \leq N} F(x_n) = \int_X F dm_X,$$

where  $m_X$  is the (normalized) Haar measure of  $X$ .

A rational subgroup  $H$  is a subgroup of  $G$  such that  $H \cdot e_X$  is a closed subset of  $X = G/\Gamma$ , where  $e_X$  is the identity element of  $X$ . Equivalently,  $H\Gamma$  is a closed subset of the space  $G$ . This also implies that  $H$  must be closed in  $G$  (see [12, Ch. 10, Lemma 14]). A subnilmanifold of  $X$  is a set  $Y \subset X$  of the form  $H \cdot x$ , where  $x$  is an element of  $X$  and  $H$  is a rational subgroup of  $G$ .

B.1.1. *Horizontal torus and characters.* Assume  $X = G/\Gamma$  is a  $k$ -step nilmanifold with  $G$  connected and simply connected and consider the subgroup  $G_2 = [G, G]$ . The nilmanifold  $Z = G/(G_2\Gamma)$  is called the horizontal torus of  $X$ . We observe that  $Z$  is a connected, compact Abelian Lie group, and thus isomorphic to some torus  $\mathbb{T}^d$ . For a  $b \in G$ , it can be shown that the nilrotation induced by  $b$  is ergodic if and only if the induced action of  $b$  on  $Z$  is ergodic [15, Theorem 3] (see also the theorem in [13, §2.17]).

A horizontal character  $\chi$  is a continuous group morphism  $\chi : G \rightarrow \mathbb{C}$ , such that  $\chi(g\gamma) = \chi(g)$  for all  $\gamma \in \Gamma$ . We observe that  $\chi$  also annihilates  $G_2$  and therefore descends



to the horizontal torus  $Z$ . Thus, under the natural projection map  $\pi$ ,  $\chi$  becomes a character on some torus  $\mathbb{T}^d$ . We will often use the notation  $\chi \circ \pi$  when working in the horizontal torus, while we reserve the letter  $\chi$  to denote the same character in the original group  $G$ .

**B.1.2. Change of base point.** For every  $b \in G$ , we have that the sequence  $b^n \Gamma$  is equidistributed in the set  $\overline{\{b^n \Gamma : n \in \mathbb{Z}\}}$ . Therefore, if  $g$  is any other element in  $G$ , we have that the sequence  $b^n g \Gamma$  is equidistributed in the nilmanifold  $g\{(g^{-1}bg)^n \Gamma, n \in \mathbb{N}\}$ . This follows by noting that  $b^n g = g(g^{-1}bg)^n$ . An analogous relation holds for the elements of the set  $(b^s g)_{s \in \mathbb{R}}$ , which we define below. This trick, which is called the change of base point trick, can be used when we want to show that some sequence  $v(n)x$  is equidistributed (on some specific nilmanifold depending on  $x$ ) to change the base point  $x$  to  $\Gamma$ .

**B.1.3. Reduction to connected–simply connected Lie groups.** Let  $G$  be a  $k$ -step nilpotent Lie group and let  $\Gamma$  be a uniform subgroup of  $G$ . Then, the space  $X = G/\Gamma$  is called a  $k$ -step nilmanifold. The space  $X$  may have several representations of the form  $G/\Gamma$  (with possible variance in the degree of nilpotency). Let  $G^\circ$  be the connected component of  $e_G$  in  $G$ . If we assume that  $G/G^\circ$  is finitely generated (without loss of generality, we can assume that in this article, because our results deal with the action of  $G$  on finitely many elements of  $X$ ), then by passing to the universal cover  $\tilde{G}$  of  $G$ , it can be shown that  $X$  has a representation  $\tilde{G}/\tilde{\Gamma}$  where now the underlying group  $\tilde{G}$  is simply connected. In addition, we can argue as in [13, §1.11] to deduce that  $X$  can be embedded as a subnilmanifold in some nilmanifold  $G'/\Gamma'$ , where  $G'$  is a connected and simply connected nilpotent Lie group and every translation on  $X$  has a representation in  $X' = G'/\Gamma'$ . This means that for any  $x \in X$ ,  $b_1, \dots, b_k \in G$  and continuous function  $F : X \rightarrow \mathbb{C}$ , we can find  $x' \in X'$ ,  $b'_1, \dots, b'_k \in G'$  and  $F' : X' \rightarrow \mathbb{C}$ , such that  $F(b_1^{n_1} \dots b_k^{n_k} x) = F'((b'_1)^{n_1} \dots (b'_k)^{n_k} x')$  for all  $n_1, \dots, n_k \in \mathbb{Z}$ .

**B.2. Nilorbits and Ratner’s theorem.** Let  $G$  be a connected and simply connected Lie group. It is well known that the exponential map  $\exp$  from the Lie algebra of  $G$  to  $G$  is a diffeomorphism. In particular, it is a bijection between  $G$  and its Lie algebra  $\mathfrak{g}$ . For  $b \in G$  and  $t \in \mathbb{R}$ , we can then define the element  $b^t$  as the unique element of  $G$  satisfying  $b^t = \exp(tX)$ , where  $\exp(X) = b$ . As a corollary of Ratner’s theorem [16], we get the following lemma.

**LEMMA B.1.** *Let  $G/\Gamma$  be a nilmanifold with  $G$  connected and simply connected. For any elements  $b_1, \dots, b_k \in \Gamma$ , we have that the set*

$$\overline{b_1^{\mathbb{R}} \dots b_k^{\mathbb{R}} \Gamma} = \overline{\{b_1^{t_1} \dots b_k^{t_k} \Gamma : t_1, \dots, t_k \in \mathbb{R}\}}$$

*is a subnilmanifold of  $X$  with a representation  $H/\Delta$  for some closed, connected and rational subgroup  $H$  of  $G$  that contains the elements  $b_1^s, \dots, b_k^s$  for all  $s \in \mathbb{R}$  and  $\Delta$  is a uniform subgroup of  $H$ .*

We call the set  $\overline{\{b^t \Gamma : t \in \mathbb{R}\}}$  the nilorbit of the element  $b$ . We will analogously denote by  $\overline{b^{\mathbb{Z}} \Gamma}$  the set  $\overline{\{b^n \Gamma : n \in \mathbb{Z}\}}$  and  $\overline{b^{\mathbb{N}} \Gamma} = \overline{\{b^n \Gamma : n \in \mathbb{N}\}}$ .

We establish the following lemma, which will be necessary for our proofs.

LEMMA B.2. *Let  $X = G/\Gamma$  be a nilmanifold and let  $b_1, \dots, b_k \in \Gamma$  be any pairwise commuting elements. Then, there exists a real number  $t$  such that*

$$\overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}} \Gamma} = \overline{\{b_1^{n_1 t} \cdots b_k^{n_k t} \Gamma : n_1, \dots, n_k \in \mathbb{Z}\}}.$$

*Proof.* We want to find some  $t \in \mathbb{R}$  so that the sequence

$$\phi_t(n_1, \dots, n_k) = b_1^{n_1 t} \cdots b_k^{n_k t}$$

is equidistributed on the nilmanifold  $Y = \overline{b_1^{\mathbb{R}} \cdots b_k^{\mathbb{R}} \Gamma}$ . By Lemma B.1,  $Y$  has a representation as  $H/\Delta$ , where  $H$  is connected, simply connected and rational. Observe that  $\phi_t$  naturally induces a  $\mathbb{Z}^k$  action on  $Y$  by  $(\phi_t(n_1, \dots, n_k), h\Delta) \rightarrow b_1^{n_1 t} \cdots b_k^{n_k t} h\Delta$ . It is sufficient to show that this  $\mathbb{Z}^k$ -action is ergodic on  $Y$ , since this implies that  $Y = \{\phi_t(\mathbf{n})y, \mathbf{n} \in \mathbb{Z}^k\}$  for all  $y \in Y$ . However, using the results in [13] (specifically, Theorem 2.17), the above action is ergodic if and only if it is ergodic on the horizontal torus  $Z$  of  $Y$ , which is homeomorphic to some torus  $\mathbb{T}^d$ . Equivalently, if we denote by  $(b_{i,1}, \dots, b_{i,d})$  the projection of the point  $b_i \Gamma$  on  $Z$ , then we need to check whether the sequence

$$(t(n_1 b_{i,1} + \cdots + n_k b_{k,1}), \dots, t(n_1 b_{1,d} + \cdots + n_k b_{k,d}))$$

is dense on  $\mathbb{T}^d$ . It suffices to choose  $t$  so that  $1/t$  is rationally independent of any integer combination of the coordinates  $b_{i,j}$ . This completes the proof. □

B.3. *Polynomial sequences on nilmanifolds.* We provide the general definition of polynomial sequences with respect to some filtration.

*Definition B.3.* A filtration  $G_\bullet$  of degree  $d$  on a nilpotent Lie group  $G$  is a sequence of closed connected subgroups

$$G = G^{(0)} = G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(d)} \supseteq G^{(d+1)} = e_G,$$

such that  $[G^{(i)}, G^{(j)}] \subseteq G^{(i+j)}$  for all  $i, j \geq 0$ . The filtration is called rational if all groups  $G^{(i)}$  appearing in the above sequence are rational subgroups of  $G$ . A polynomial sequence on  $G$  with respect to the above filtration is a sequence  $g(n)$  such that, for all positive integers  $h_1, \dots, h_k$ , we have that the sequence  $\partial_{h_1} \cdots \partial_{h_k} g$  takes values in  $G^{(k)}$ , for all  $k \in \mathbb{N}$ , where  $\partial_h$  denotes the ‘differencing operator’ that maps the sequence  $(g(n))_{n \in \mathbb{N}}$  to the sequence  $(g(n+h)(g(n))^{-1})_{n \in \mathbb{N}}$ .

An example of a filtration is the lower central series of the group  $G$ . For the purposes of this article, we will only need to consider polynomial sequences of the form

$$v(n) = b_1^{p_1(n)} \cdots b_k^{p_k(n)}, \tag{B.5}$$

where  $b_i \in G$  for all  $1 \leq i \leq k$  and  $p_i$  are real polynomials. Note that the terms  $b_i^{p_i(n)}$  are well defined, due to our connectedness assumptions. To see that this is indeed a polynomial sequence with our initial definition, we construct a specific filtration on  $G$ . We assume that  $G$  is  $k$ -step nilpotent and we also denote the maximum degree among the polynomials  $p_i$  as  $d$ . We consider the filtration (of degree  $dk$ )  $G_\bullet = (G^{(i)})_{0 \leq i \leq dk}$ , where

$G^{(i)} = G_{\lfloor i/d \rfloor + 1}$  and  $G_j$  are the commutator subgroups of  $G$ . This is a rational filtration because all commutator subgroups of  $G$  are rational (see [12, Ch. 10, Proposition 22] for the proof). Then, the sequence  $v(n)$  in equation (B.5) is a polynomial sequence with respect to this filtration. We direct the reader to the discussion after [9, Corollary 6.8], where these last observations were made originally. We will also call the projected sequence  $v(n)\Gamma$  on  $X = G/\Gamma$  a polynomial sequence on  $X$ .

**B.4. Quantitative equidistribution.** Assume that  $p(t)$  is a polynomial. Then,  $p(n)$  can be expressed uniquely in the form

$$p(n) = \sum_{i=0}^d a_i n^i$$

for some real numbers  $a_i$  and  $d \in \mathbb{N}$ . For  $N \in \mathbb{N}$ , we define the smoothness norm

$$\|e(p(n))\|_{C^\infty[N]} = \max_{1 \leq i \leq d} (N^i \|a_i\|_{\mathbb{R}/\mathbb{Z}}). \tag{B.6}$$

(The definition of the smoothness norms is a bit different in [9]. There, the authors write the polynomials in the form  $p(n) = \sum_{i=0}^d a_i \binom{n}{i}$  and define the smoothness norm using the same definition as equation (B.6) (the coefficients  $a_i$  are different). However, these definitions give two equivalent norms and, thus, all theorems can be stated for both norms, up to changes in the absolute constants.)

A filtration on a Lie group  $G$  gives rise to a basis on its Lie algebra  $\mathfrak{B}$ , which is called a Mal'cev basis [14]. Mal'cev bases play an essential role in the theory of quantitative equidistribution on nilmanifolds. First, we give the following definition.

*Definition B.4.* Let  $X = G/\Gamma$  be a  $k$ -step nilmanifold with a rational filtration  $G_\bullet = (G^{(i)})_{i \geq 0}$ . Define  $m = \dim(G)$  and  $m_i = \dim(G^{(i)})$ . A basis  $(\xi_1, \dots, \xi_m)$  of the associated Lie algebra  $\mathfrak{g}$  over  $\mathbb{R}$  is called a Mal'cev basis adapted to  $G_\bullet$ , if the following conditions are met.

- (i) For each  $0 \leq j \leq m - 1$ ,  $\mathfrak{h}_j = \text{span}(\xi_{h+1}, \dots, \xi_m)$  is a Lie algebra ideal on  $\mathfrak{g}$  and thus  $H_j = \exp(\mathfrak{h}_j)$  is a normal Lie subgroup of  $G$ .
- (ii) For every  $0 \leq i \leq k$ , we have  $G^{(i)} = H_{m-m_i}$ .
- (iii) Each  $b \in G$  can be uniquely written in the form  $\exp(t_1 \xi_1) \cdots \exp(t_m \xi_m)$  for  $t_i \in \mathbb{R}$ .
- (iv) The subgroup  $\Gamma$  consists precisely of those elements which, when written in the above form, have all  $t_i \in \mathbb{Z}$ .

Suppose that the element  $b$  is written in the form  $\exp(t_1 \xi_1) \cdots \exp(t_m \xi_m)$ . The map  $\psi : G \rightarrow \mathbb{R}^m$  defined by  $\psi(b) = (t_1, \dots, t_m)$  is a diffeomorphism from  $G$  to  $\mathbb{R}^m$ . The numbers  $(t_1, \dots, t_m)$  are called the coordinates of  $g$  with respect to the associated Mal'cev basis. If we consider the Euclidean metric on  $\mathbb{R}^m$ , we can construct a Riemannian metric  $d_G$  on  $G$ , whose value at the origin is equal to the Euclidean metric of  $\mathbb{R}^m$  at the origin (of  $\mathbb{R}^m$ ) composed with the inverse map  $\psi^{-1}$ . This metric is invariant under right translations and induces a metric  $d_X$  on  $X = G/\Gamma$  defined by the relation:

$$d_X(g\Gamma, h\Gamma) = \inf\{d_G(b, b'), bg^{-1} \in \Gamma, b'h^{-1} \in \Gamma\}.$$

The metric used in [9] is slightly different than the one we consider here, but as the authors remark, these metrics are equivalent and all theorems hold as well by changing the absolute constants.

The sequence  $(g(n)\Gamma)_{1 \leq n \leq N}$  is said to be  $\delta$ -equidistributed on the nilmanifold  $X = G/\Gamma$  if and only if for any Lipschitz function  $F : X \rightarrow \mathbb{C}$ , we have that

$$\left| \frac{1}{N} \sum_{n=1}^N F(g(n)\Gamma) - \int_X F d\mu_X \right| \leq \delta \|F\|_{\text{Lip}(X)},$$

where

$$\|F\|_{\text{Lip}(X)} = \|F\|_\infty + \sup_{x,y \in X, x \neq y} \frac{|F(x) - F(y)|}{d_X(x, y)}.$$

We now fix a  $k$ -step nilmanifold  $X = G/\Gamma$ , as well as a positive integer  $d$ . We equip it with the rational filtration  $G_\bullet$  of degree  $dk$  that we defined above (after Definition B.3), as well as a Mal'cev basis adapted to this filtration and the corresponding coordinate map  $\psi : G \rightarrow \mathbb{R}^m$  ( $m$  is the dimension of  $G$ ). Observe that under this filtration, we have that  $G^{(d+1)} = G_2$  and property (ii) in Definition B.4 implies that  $G_2 = H_{m-m_{d+1}}$ . Thus, the Mal'cev basis induces an isometric identification of the horizontal torus  $Z = G/G_2\Gamma$  with the torus  $\mathbb{T}^{m-m_{d+1}}$  equipped with the standard metric.

Let  $\pi : X \rightarrow Z$  denote the projection map and let  $\chi$  be a horizontal character on  $G$ . Consider an element  $b \in G$  with coordinates  $(t_1, \dots, t_m)$ . Then, by properties (iii) and (iv) in Definition B.4, we have that there is some  $\vec{\ell} = (\ell_1, \dots, \ell_{m-m_{d+1}}) \in \mathbb{Z}^{m-m_{d+1}}$  such that

$$\chi \circ \pi(b) = \ell_1 t_1 + \dots + \ell_{m-m_{d+1}} t_{m-m_{d+1}}.$$

Thus, we get a character on the torus  $\mathbb{T}^{m-m_{d+1}}$  (written here with additive notation). We can then define the modulus  $\|\chi\|$  of the character  $\chi$  to be equal to

$$\|\vec{\ell}\| = |\ell_1| + \dots + |\ell_{m-m_{d+1}}|. \tag{B.7}$$

If  $v(n)$  is the polynomial sequence in equation (B.5) (recall that it is a polynomial sequence with respect to the filtration  $G_\bullet$ ), then the sequence  $\chi \circ \pi(v(n)\Gamma)$  is a polynomial sequence on the horizontal torus  $Z \cong \mathbb{T}^{m-m_{d+1}}$ . Indeed, if we denote  $\psi(b_i) = (t_{i,1}, \dots, t_{i,m})$ , then a simple calculation shows that

$$\begin{aligned} \chi(\pi(v(n)\Gamma)) &= \chi(\pi(b_1^{p_1(n)} \dots b_k^{p_k(n)})) \\ &= p_1(n)(\ell_1 t_{1,1} + \dots + \ell_{m-m_{d+1}} t_{1,m-m_{d+1}}) \\ &\quad + \dots + p_k(n)(\ell_1 t_{k,1} + \dots + \ell_{m-m_{d+1}} t_{k,m-m_{d+1}}), \end{aligned}$$

which makes the fact that  $\chi(\pi(v(n)\Gamma))$  is a polynomial sequence more evident.

The primary tool that we shall use is the following theorem of Green and Tao which describes the orbits of polynomial sequences in finite intervals. We present it in the case of our filtration  $G_\bullet$ , although the statement holds for any rational filtration. Some quantitative information (specifically relating to the concepts of quantitative rationality of Mal'cev bases) has been suppressed, since in our applications, the nilmanifold will be fixed and the above condition on the Mal'cev bases is guaranteed if we take  $\delta$  small enough.

**THEOREM F.** [9, Theorem 2.9] *Let  $d$  be a non-negative integer,  $X = G/\Gamma$  be a nilmanifold with  $G$  connected and simply connected, and we equip the nilmanifold  $X$  with the Mal'cev basis adapted to the  $dk$  filtration  $G_\bullet$  as above. Assume  $\delta$  is a sufficiently small (depending only on  $X, d$ ) parameter. Then, there exist a positive constant  $C = C(X, d)$  with the following property: for every  $N \in \mathbb{N}$ , if  $(v(n))_{n \in \mathbb{N}}$  is a polynomial sequence with respect to  $G_\bullet$  such that the finite sequence  $(v(n)\Gamma)_{1 \leq n \leq N}$  is not  $\delta$ -equidistributed, then for some non-trivial horizontal character  $\chi$  (that depends on  $N$  and the sequence  $v(n)$ ) of modulus  $\|\chi\| \leq \delta^{-C}$ , we have*

$$\|\chi(\pi(v(n)\Gamma))\|_{C^\infty(N)} \leq \delta^{-C},$$

where  $\pi$  denotes the projection map from  $X$  to its horizontal torus.

To get a sense of how this theorem works, let us consider an application on a polynomial sequence on  $\mathbb{T}$ . Let  $d$  be a positive integer and  $\delta > 0$  a small real number. Then, there exists a constant  $C$  that depends only on  $d$ , such that for any polynomial

$$p(t) = a_d x^d + \dots + a_1 t + a_0$$

of degree  $d$ , we have either that

$$\left| \mathbb{E}_{1 \leq n \leq N} e(p(n)) \right| < \delta$$

or there exists an integer  $q$  with  $|q| \leq \delta^{-C}$ , such that

$$N^k \|qa_k\|_{\mathbb{T}} \leq \delta^{-C}$$

for every  $1 \leq k \leq d$ . Thus, either the exponential sums of the polynomial sequence  $p(n)$  are small or the non-constant coefficients  $a_k$  satisfy a ‘major-arc’ condition (they are ‘close’ to a rational with denominator bounded by  $\delta^{-C}$ ). Observe that the constant  $C$  does not depend on the length of the interval  $N$ .

## REFERENCES

- [1] V. Bergelson, J. Moreira and F. Richter. Multiple ergodic averages along functions from a Hardy field: convergence, recurrence and combinatorial applications. *Preprint*, 2020, [arXiv:2006.03558](https://arxiv.org/abs/2006.03558).
- [2] M. Boshernitzan. Uniform distribution and Hardy fields. *J. Anal. Math.* **62** (1994), 225–240.
- [3] L. Corwin and F. P. Greenleaf. *Representations of Nilpotent Lie Groups and their Applications. Part 1: Basic Theory and Examples (Cambridge Studies in Advanced Mathematics, 18)*. Cambridge University Press, Cambridge, 1990.
- [4] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Anal. Math.* **109** (2009), 353–395.
- [5] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.
- [6] N. Frantzikinakis. Joint ergodicity of sequences. *Adv. Math.* **417** (2023), 63pp.
- [7] N. Frantzikinakis, E. Lesigne and M. Wierdl. Sets of  $k$ -recurrence but not  $(k+1)$ -recurrence. *Ann. Inst. Fourier (Grenoble)* **56**(4) (2006), 839–849.
- [8] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.* **71** (1977), 204–256.
- [9] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)* **175** (2012), 465–540.
- [10] G. H. Hardy. Properties of logarithmico-exponential functions. *Proc. Lond. Math. Soc. (2)* **10** (1912), 54–90.

- [11] B. Host and B. Kra. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math. (2)* **161**(2) (2005), 397–488.
- [12] B. Host and B. Kra. *Nilpotent Structures in Ergodic Theory (Mathematical Surveys and Monographs, 236)*. American Mathematical Society, Providence, RI, 2018.
- [13] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Ergod. Th. & Dynam. Sys.* **25**(1) (2005), 201–213.
- [14] A. Mal'cev. On a class of homogeneous spaces. *Izvestiya Akad. Nauk SSSR, Ser Mat.* **13** (1949), 9–32.
- [15] W. Parry. Dynamical systems on nilmanifolds. *Bull. Lond. Math. Soc.* **2** (1970), 37–40.
- [16] M. Ratner. Raghunatan's topological conjecture and distribution of unipotent flows. *Duke Math. J.* **61**(1) (1991), 235–280.
- [17] F. K. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field. *J. Anal. Math.* **149** (2023), 421–483.
- [18] K. Tsinas. Joint ergodicity of Hardy field sequences. *Trans. Amer. Math. Soc.* **376** (2023), 3191–3263.