SINGLE AND MULTIPLE RECURRENCE ALONG NON-POLYNOMIAL SEQUENCES

VITALY BERGELSON, JOEL MOREIRA, AND FLORIAN K. RICHTER

ABSTRACT. We establish new recurrence and multiple recurrence results for a rather large family $\mathcal{F}$ of non-polynomial functions which contains tempered functions and (non-polynomial) functions from a Hardy field with polynomial growth.

In particular, we show that, somewhat surprisingly (and in the contrast to the multiple recurrence along polynomials), the sets of return times along functions from $\mathcal{F}$ are *thick*, i.e., contain arbitrarily long intervals. A major component of our paper is a new result about equidistribution of sparse sequences on nilmanifolds, whose proof borrows ideas from the work of Green and Tao [26].

Among other things, we show that for any $f \in \mathcal{F}$, any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$, and any $\varepsilon > 0$, the sets of returns
$$\big\{n \in \mathbb{N} : \mu(A \cap T^{-\lfloor f(n) \rfloor} A) > \mu^2(A) - \varepsilon\big\}$$
$$\big\{n \in \mathbb{N} : \mu\big(A \cap T^{-\lfloor f(n) \rfloor} A \cap T^{-\lfloor f(n+1) \rfloor} A \cap \cdots \cap T^{-\lfloor f(n+k) \rfloor} A\big) > 0\big\}$$
are thick.

Our recurrence theorems imply, via Furstenberg's correspondence principle, some new combinatorial results. For example, we show that given a set $E \subset \mathbb{N}$ with positive upper density, for every $k \in \mathbb{N}$ there are $a, n \in \mathbb{N}$ such that
$$\big\{a, a + \lfloor f(n) \rfloor, \cdots, a + \lfloor f(n+k) \rfloor\big\} \subset E.$$

When $f(n) = n^c$, with $c > 0$ non-integer, this result provides a positive answer to a question posed by Frantzikinakis [16, Problem 23].

TABLE OF CONTENTS

1. INTRODUCTION

Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving system. The classical Poincaré recurrence theorem states that for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in \mathbb{N} = \{1, 2, \ldots\}$ such that $\mu(A \cap T^{-n}A) > 0$. Over the years it was revealed that the set of return times

$$R_A := \{n \in \mathbb{N} : \mu(A \cap T^{-n}A) > 0\}$$

has quite intricate combinatorial and number-theoretical properties. For example, R_A contains a perfect square [21, Proposition 1.3]. This fact, in turn, implies a theorem of Sárközy [44] which says that for any set $E \subset \mathbb{N}$ with positive *upper Banach density*[1] there exist $n \in \mathbb{N}$ and $a, b \in E$ such that $a - b = n^2$ (see [21, Theorem 1.2]).

A more general version of this result, proved by Furstenberg in [22, 23], asserts that for any $g \in \mathbb{Q}[x]$ with $g(\mathbb{Z}) \subset \mathbb{Z}$ and $g(0) = 0$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-g(n)}A) > 0,$$

which implies that the *upper density*[2] of the set $\{n \in \mathbb{N} : \mu(A \cap T^{-g(n)}A) > 0\}$ is positive. One can actually show that for any $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{-g(kn)}A) > \mu(A)^2 - \varepsilon,$$

which gives the following polynomial version of the classical Khintchine recurrence theorem [32]. (This polynomial Khintchine recurrence theorem also follows from a stronger result obtained in [4].)

Theorem 1.1 (Polynomial Khintchine recurrence theorem). *For any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$, any $g \in \mathbb{Q}[x]$ with $g(\mathbb{Z}) \subset \mathbb{Z}$ and $g(0) = 0$, and any $\varepsilon > 0$, the set of optimal return times $R_{\varepsilon,A} = \left\{ n \in \mathbb{N} : \mu(A \cap T^{-g(n)}A) > \mu^2(A) - \varepsilon \right\}$ is syndetic, i.e., there exists $l \in \mathbb{N}$ such that $R_{\varepsilon,A}$ has non-empty intersection with every interval of length bigger or equal to l.*

In [5, Theorem 7.1] a mean ergodic theorem along *tempered sequences*[3] was proved, which implies that for any tempered function f, any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-\lfloor f(n) \rfloor}A) \geqslant \mu(A)^2,$$

where $\lfloor x \rfloor$ denotes the integer part of a real number x (see [5, Corollary 7.2]). This gives a large class of sequences $f(n)$ for which the set of optimal return times

$$(1.1) \qquad R_{\varepsilon,A} := \left\{ n \in \mathbb{N} : \mu(A \cap T^{-\lfloor f(n) \rfloor}A) > \mu^2(A) - \varepsilon \right\}$$

has positive upper density. Examples include $f(n) = bn^c \log^r(n)$, where $b \in \mathbb{R} \backslash \{0\}$, $c > 0$ with $c \notin \mathbb{N}$ and $r \geqslant 0$, and $f(n) = n^c(\cos(\log^r(n)) + 2)$, where $c > 0$ and $0 < r < 1$.

The goal of this paper is to establish new results about the set of optimal returns $R_{\varepsilon,A}$ defined in (1.1) and about the set of multiple returns

$$(1.2) \quad R_A^{(k)} := \left\{ n \in \mathbb{N} : \mu(A \cap T^{-\lfloor f(n) \rfloor}A \cap T^{-\lfloor f(n+1) \rfloor}A \cap \ldots \cap T^{-\lfloor f(n+k) \rfloor}A) > 0 \right\}, \quad k \in \mathbb{N},$$

---

[1] The *upper Banach density* of a set $E \subset \mathbb{N}$ is defined as $d^*(E) := \limsup_{N-M \to \infty} |E \cap [M, N]|/(N-M)$.

[2] The *upper density* of a set $E \subset \mathbb{N}$ is defined as $\overline{d}(E) := \limsup_{N \to \infty} |E \cap [1, N]|/N$.

[3] A real-valued function f defined on a half-line $[a, \infty)$ is called a *tempered* function if there exist $\ell \in \mathbb{N}$ such that f is ℓ times continuously differentiable, $f^{(\ell)}(x)$ tends monotonically to zero as $x \to \infty$, and $\lim_{x \to \infty} x|f^{(\ell)}(x)| = \infty$. (This notion was introduced in the context of the theory of uniform distribution by Cigler in [13].)

for a rather large class of sequences $f(n)$. As is shown in Lemma A.1, this class includes:
- all sequences $f(n)$, where $f\colon [1, \infty) \to \mathbb{R}$ is a tempered function;
- all sequences $f(n)$, where $f\colon [1, \infty) \to \mathbb{R}$ is a function from a *Hardy field*[4] with the property that for some $\ell \in \mathbb{N} \cup \{0\}$ one has $\lim_{x \to \infty} f^{(\ell)}(x) = \pm\infty$ and $\lim_{x \to \infty} f^{(\ell+1)}(x) = 0$.

Before formulating our results, let us define the class of functions that we will be dealing with.

Definition 1.2. Given a function $f\colon \mathbb{N} \to \mathbb{R}$ let Δf denote its *first order difference* (or *discrete derivative*), $\Delta f(n) := f(n+1) - f(n)$. Define

$$\mathcal{F}_0 \ := \ \left\{ f\colon \mathbb{N} \to \mathbb{R} : \lim_{n \to \infty} f(n) = 0, \ f(n) \text{ eventually monotone, and } \left| \sum_{n=1}^{\infty} f(n) \right| = \infty \right\};$$

$$\mathcal{F}_{\ell+1} \ := \ \left\{ f\colon \mathbb{N} \to \mathbb{R} : \Delta f \in \mathcal{F}_\ell \right\}, \quad \text{for } \ell \in \mathbb{N} \cup \{0\},$$

and let $\mathcal{F} := \bigcup_{\ell=1}^{\infty} \mathcal{F}_\ell$.

Observe that the function $f(n) = n^\alpha$ belongs to $\mathcal{F}$ if and only if $\alpha > 0$ and is not an integer (in fact, in this case we have $n^\alpha \in \mathcal{F}_{\lfloor \alpha \rfloor + 1}$). While our class $\mathcal{F}$ is vastly more general than the family of functions $\{n \mapsto n^\alpha : \alpha \in (0, \infty) \backslash \mathbb{N}\}$, it may be convenient for the reader to keep this example in mind, as our results are already new and interesting even for this family.

We will show that for any $f \in \mathcal{F}$ the sets $R_{\varepsilon, A}$ and $R_A^{(k)}$ defined in (1.1) and (1.2) possess a somewhat unexpected combinatorial property that stands in contrast to the syndeticity featured in Khintchine's recurrence theorem and in Theorem 1.1. The notion in question is dual to the notion of a syndetic set:

Definition 1.3. $R \subset \mathbb{N}$ is called *thick* if it contains arbitrarily long intervals.

Notice that a set is thick if and only if its complement is not syndetic.

While for sequences f belonging to the class $\mathcal{F}$ the sets $R_{\varepsilon, A}$ are, in general, not syndetic (cf. [18, Remark after Theorem 2.3]), they do possess a combinatorial property that we will presently introduce and which can be viewed as a generalized version of syndeticity.

Definition 1.4. Let $W \in \mathcal{F}_1 \cup \{\mathrm{Id} : n \mapsto n\}$ be positive. A set $R \subset \mathbb{N}$ is called *W-syndetic* if there exists $l \in \mathbb{N}$ such that for any interval $[M, N] \subset \mathbb{N}$ with $W(N) - W(M) \geqslant l$ one has

$$\sum_{n \in R \cap [M,N]} \Delta W(n) \geqslant 1.$$

Any W-syndetic set R has positive *lower* and *upper W-density*:

$$\underline{d}_W(R) := \liminf_{N \to \infty} \frac{1}{W(N)} \sum_{n \in R \cap [1,N]} \Delta W(n) > 0.$$

$$\overline{d}_W(R) := \limsup_{N \to \infty} \frac{1}{W(N)} \sum_{n \in R \cap [1,N]} \Delta W(n) > 0.$$

Note that when $W = \mathrm{Id}$ we recover from Definition 1.4 the usual notion of syndeticity, and $\underline{d}_W(R)$ and $\overline{d}_W(R)$ become the usual upper and lower density, denoted simply as $\underline{d}(R)$ and

---

[4]A *germ at* ∞ of a given function $f\colon [a, \infty) \to \mathbb{R}$ is any equivalence class of functions $g\colon [b, \infty) \to \mathbb{R}$ under the equivalence relationship $(f \sim g) \Leftrightarrow (\exists c \geqslant \max\{a, b\}$ such that $f(x) = g(x)$ for all $x \in [c, \infty))$. Let G denote the set of all germs at ∞ of real valued functions defined on a half-line $[a, \infty)$. Any subfield of the ring $(G, +, \cdot)$, where $+$ and $\cdot$ denote pointwise addition and multiplication, that is closed under differentiation is called a *Hardy field* [27, 28]. By abuse of language, we say that a function $f\colon [1, \infty) \to \mathbb{R}$ belongs to some Hardy field if its germ belongs to that Hardy field.

$\overline{d}(R)$. Also, for any $W \in \mathcal{F}_1$ the inequality $\underline{d}(R) \leqslant \underline{d}_W(R) \leqslant \overline{d}_W(R) \leqslant \overline{d}(R)$ holds[5], which implies that any W-syndetic set has positive upper density.

Our first main result pertains to the sets $R_{\varepsilon,A}$.

Theorem A. *Let $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$ and define $W := \Delta^{\ell-1} f$. Then for any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ and any $\varepsilon > 0$ the set*

$$R_{\varepsilon,A} = \left\{ n \in \mathbb{N} : \mu(A \cap T^{-\lfloor f(n) \rfloor} A) > \mu^2(A) - \varepsilon \right\}$$

is thick and W-syndetic.

We describe now the results which deal with the sets of multiple recurrence $R_A^{(k)}$ defined in (1.2). Let us first recall the polynomial multiple recurrence theorem obtained in [9] (see also [6, 10, 8]).

Theorem 1.5 (Polynomial multiple recurrence theorem). *For any $k \in \mathbb{N}$, any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$ and any polynomials $g_1, \ldots, g_k \in \mathbb{Q}[x]$ with $g_i(\mathbb{Z}) \subset \mathbb{Z}$ and $g_i(0) = 0$ for $i = 1, \ldots, k$, the set*

$$(1.3) \qquad R = \left\{ n \in \mathbb{N} : \mu(A \cap T^{-g_1(n)} A \cap \ldots \cap T^{-g_k(n)} A) > 0 \right\}$$

is syndetic.

By a classical theorem of algebra (cf. [39] or [41]) any polynomial $g \in \mathbb{Q}[x]$ with $g(\mathbb{Z}) \subset \mathbb{Z}$ and $\deg g = d$ can be represented as an integer linear combination of binomial coefficients, $g(n) = a_d \binom{n}{d} + \ldots + a_1 \binom{n}{1} + a_0 \binom{n}{0}$. Since $\Delta \binom{n}{d} = \binom{n}{d-1}$, it follows that for any $g \in \mathbb{Q}[x]$ with $g(\mathbb{Z}) \subset \mathbb{Z}$ and $\deg g = d$ there exists $p(x) = a_d x^d + \ldots + a_1 x + a_0 \in \mathbb{Z}[x]$ such that $g(n) = p(\Delta) \binom{n}{d}$, where the operator $p(\Delta)$ is defined as $p(\Delta) f := a_d \Delta^d f + \ldots + a_1 \Delta f + a_0 f$. This leads to the following equivalent formulation of Theorem 1.5.

Theorem 1.6. *Let $k, d \in \mathbb{N}$ and $(X, \mathcal{B}, \mu, T)$ an invertible probability measure preserving system. For any $A \in \mathcal{B}$ with $\mu(A) > 0$ and any polynomials $p_1, \ldots, p_k \in \mathbb{Z}[x]$ with $\deg(p_i) < d$ for $i = 1, \ldots, k$, the set*

$$(1.4) \qquad R = \left\{ n \in \mathbb{N} : \mu\left(A \cap T^{-p_1(\Delta)\binom{n}{d}} A \cap \ldots \cap T^{-p_k(\Delta)\binom{n}{d}} A\right) > 0 \right\}$$

is syndetic.

Our second main result is a multiple recurrence result in the spirit of Theorem 1.6.

Theorem B. *Let $f \in \mathcal{F}_\ell$ and define $W := \Delta^{\ell-1} f$. For any $k \in \mathbb{N}$, any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$ and any polynomials $p_1, \ldots, p_k \in \mathbb{Z}[x]$, the set*

$$(1.5) \qquad R = \left\{ n \in \mathbb{N} : \mu\left(A \cap T^{-\lfloor p_1(\Delta) f(n) \rfloor} A \cap \ldots \cap T^{-\lfloor p_k(\Delta) f(n) \rfloor} A\right) > 0 \right\}$$

is thick and W-syndetic.

Remark 1.7. Unlike the return-time set in (1.4), the return-time set in (1.5) is in general not syndetic. Nonetheless, Theorem B shows that the set R in (1.5) is combinatorially rich in a new and different way.

Remark 1.8. Multiple recurrence results dealing with non-polynomial functions from Hardy fields were obtained in [17, 18, 19, 20]. In these papers it was established (among other things) that, for any $k \in \mathbb{N}$, any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$ and for a function f from a Hardy field satisfying certain growth conditions there

---

[5]The middle inequality is trivial, and the left and right inequalities can be established by a method similar to that used in the proof of [34, Lemma 1.7.1]. See also [25, Theorem 3.5] for a uniform version.

exists $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{\lfloor f(n) \rfloor} A \cap T^{2\lfloor f(n) \rfloor} A \cap \cdots \cap T^{k \lfloor f(n) \rfloor} A) > 0.$$

Additionally, it was shown that under the same conditions, given functions $f_1, \ldots, f_k$ from a Hardy field of different growth rates, there exists $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{\lfloor f_1(n) \rfloor} A \cap T^{\lfloor f_2(n) \rfloor} A \cap \cdots \cap T^{\lfloor f_k(n) \rfloor} A) > 0.$$

By way of contrast, Theorem B applies to the functions $p_1(\Delta)f, \ldots, p_k(\Delta)f$ which, in general, do not have different growth rates (in fact they may not even be linearly independent), nor are they all a scalar multiple of a common function. These novel features are established with the help of some new ideas and techniques which are particularly patent in Section 5.

By utilizing the identity

$$f(n+k) = (1+\Delta)^k f(n), \qquad \forall n \in \mathbb{N}, \ \forall k \in \mathbb{N} \cup \{0\},$$

one obtains the following corollary from Theorem B.

Corollary C. *Let $f \in \mathcal{F}_\ell$, let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system and define $W := \Delta^{\ell-1} f$. Then for any $A \in \mathcal{B}$ with $\mu(A) > 0$ and any $k \in \mathbb{N}$ the set*

$$R_A^{(k)} = \left\{ n \in \mathbb{N} : \mu\big(A \cap T^{\lfloor f(n) \rfloor} A \cap T^{\lfloor f(n+1) \rfloor} A \cap \cdots \cap T^{\lfloor f(n+k) \rfloor} A\big) > 0 \right\}$$

is thick and W-syndetic.

The fact that the set $R_A^{(k)}$ is non-empty for the special case when $f(n) = n^\alpha$ answers [16, Problem 23]. On the other hand, we remark that Corollary C does *not* hold when f is a polynomial with integer coefficients. To illustrate this contrast, fix $\alpha \geqslant 1$. Then the following is true if and only if α is not an integer: For every invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, for any $A \in \mathcal{B}$ with $\mu(A) > 0$ and for every $k \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that

$$(1.6) \qquad \mu(A \cap T^{\lfloor n^\alpha \rfloor} A \cap T^{\lfloor (n+1)^\alpha \rfloor} A \cap \cdots \cap T^{\lfloor (n+k)^\alpha \rfloor} A) > 0.$$

Indeed, when $\alpha \notin \mathbb{N}$ the function $f(n) = n^\alpha$ belongs to $\mathcal{F}$ and thus (1.6) follows from Corollary C. When α is an integer, one can take $(X, \mathcal{B}, \mu, T)$ to be the rotation on 2 points and $k = 1$ to obtain a counterexample. In fact, one can obtain a counterexample for any system with a non-trivial Kronecker factor.

Corollary C implies, via Furstenberg's correspondence principle (see [21, Section 1] or [1, Theorem 1.1]) several interesting combinatorial applications. For instance, it follows that for every $f \in \mathcal{F}$ and $k \in \mathbb{N}$, a set with positive Banach density must contain a configuration of the form

$$\{a, \ a + \lfloor f(n) \rfloor, \ a + \lfloor f(n+1) \rfloor, \ \ldots, \ a + \lfloor f(n+k) \rfloor\}$$

for some $a, n \in \mathbb{N}$. A stronger result is Corollary F below.

Remark 1.9. We would like to note that Corollary C implies Szemerédi's theorem [47] which states that every set $A \subset \mathbb{N}$ with positive upper Banach density contains arbitrarily long arithmetic progressions. To verify this claim, let $A \subset \mathbb{N}$ have positive upper Banach density and let $k \in \mathbb{N}$. Take any function $f \in \mathcal{F}_2$ and let $W = \Delta f$. Since $W \in \mathcal{F}_1$, we have $\Delta W(n) \to 0$ as $n \to \infty$, and therefore $|W(n + 10k^2) - W(n)| \to 0$ as $n \to \infty$. Using Corollary C find $a, n \in \mathbb{N}$ with n large enough so that $|W(n + 10k^2) - W(n)| < 1/(5k)$ and such that

$$\{a + \lfloor f(n) \rfloor, a + \lfloor f(n+1) \rfloor, \cdots, a + \lfloor f(n + 10k^2) \rfloor\} \subset A.$$

Using the pigeonhole principle, find $t \in \{1, 2, \ldots, 5k\}$ such that $\|tW(n)\|_{\mathbb{T}} < 1/(10k)$, where $\|x\|_{\mathbb{T}}$ denotes the distance from the real number x to the closest integer. Observe that the set

$\{a + f(n), a + f(n+t), \cdots, a + f(n+2kt)\}$ is, up to a small error, the arithmetic progression $\{a + f(n), a + f(n) + tW(n), a + f(n) + 2tW(n), \cdots, a + f(n) + 2ktW(n)\}$ whose common difference $tW(n)$ is almost an integer. Therefore either the first k or the last k terms of the set

$$\{a + \lfloor f(n) \rfloor, a + \lfloor f(n+t) \rfloor, \cdots, a + \lfloor f(n+2kt) \rfloor\}$$

form an arithmetic progression of length k.

We will derive Theorem B from a more general result dealing with weighted multiple ergodic averages along sequences from $\mathcal{F}$. To state this theorem we need to introduce some notation first.

Definition 1.10. Let $a \colon \mathbb{N} \to \mathbb{C}$ be bounded, let $W \in \mathcal{F}_1$ and let $M < N \in \mathbb{N}$. Define

$$\mathbb{E}^W_{n \in [M,N]} a(n) := \frac{1}{W(N) - W(M)} \sum_{n=M}^{N} \Delta W(n) a(n).$$

We denote by

$$W\text{-}\lim_{n \to \infty} a(n) := \lim_{N \to \infty} \mathbb{E}^W_{n \in [1,N]} a(n)$$

the *(simple) Riesz mean*[6] of a with respect to W, whenever this limit exists, and

$$UW\text{-}\lim_{n \to \infty} a(n) := \lim_{W(N) - W(M) \to \infty} \mathbb{E}^W_{n \in [M,N]} a(n)$$

the *uniform Riesz mean* of a with respect to W, whenever this limit exists.

Remark 1.11. When $W = \mathrm{Id}$ the uniform Riesz mean with respect to W is the same as the more well-known *Uniform Cesàro mean*, which is a natural form of convergence often used in ergodic theory and is also closely related to the classical notion of *well-distribution* introduced by Hlawka [29] and Peterson [40] in the 1950s (cf. Definition 2.1 below).

Theorem D. *Let $k \in \mathbb{N}$, $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$, $W := \Delta^{\ell-1} f$ and $p_1, \ldots, p_k \in \mathbb{Z}[x]$. Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system.*
  (1) For any $h_1, \ldots, h_k \in L^\infty(X, \mathcal{B}, \mu)$ the limit

$$UW\text{-}\lim_{n \to \infty} T^{\lfloor p_1(\Delta) f(n) \rfloor} h_1 T^{\lfloor p_2(\Delta) f(n) \rfloor} h_2 \cdot \ldots \cdot T^{\lfloor p_k(\Delta) f(n) \rfloor} h_k$$

  exists in $L^2(X, \mathcal{B}, \mu)$.
  (2) For any $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$(1.7) \qquad UW\text{-}\lim_{n \to \infty} \mu\big(A \cap T^{-\lfloor p_1(\Delta) f(n) \rfloor} A \cap T^{-\lfloor p_2(\Delta) f(n) \rfloor} A \cap \cdots \cap T^{-\lfloor p_k(\Delta) f(n) \rfloor} A\big) > 0.$$

Remark 1.12. If f is a tempered function (see Footnote 3) then we conjecture that all uniform Riesz means with respect to W appearing in Theorem D can actually be replaced with conventional Cesàro averages, but this is not a consequence of our results. On the other hand, there are functions in $\mathcal{F}$, which even grow faster than linear, for which the results in Theorem D do not hold when Riesz means are replaced by Cesàro averages, even for the special case $k = 1$ and $p_1(\Delta) = \mathrm{id}$. Examples of such functions f can be found in [11, p. 65].

As a corollary of Theorem D we get the following strengthening of Theorem B, which quantifies the appearance frequency of long intervals in the set R defined in (1.5).

Corollary E. *Take $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$, $k \in \mathbb{N}$, $p_1, \ldots, p_k \in \mathbb{Z}[x]$, $(X, \mathcal{B}, \mu, T)$ an invertible probability measure preserving systems and $A \in \mathcal{B}$ with $\mu(A) > 0$. Define $W := \Delta^{\ell-1} f$. Then for any*

---

[6]We follow the terminology of Kuipers and Niederreiter [34, Example 7.3 on page 63], but use a different notation.

$L \in \mathbb{N}$ the set of m such that

$$[m, m+L] \subset \left\{ n \in \mathbb{N} : \mu\left(A \cap T^{-\lfloor p_1(\Delta)f(n)\rfloor}A \cap \ldots \cap T^{-\lfloor p_k(\Delta)f(n)\rfloor}A\right) > 0 \right\}$$

is W-syndetic.

Finally we present a combinatorial application of Corollary E, which can de derived using Furstenberg's correspondence principle.

Corollary F. *Take* $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$, $k \in \mathbb{N}$, $p_1, \ldots, p_k \in \mathbb{Z}[x]$, *and let* $E \subset \mathbb{N}$ *have positive upper Banach density. Then* E *contains configurations of the form*

$$(1.8) \qquad \left\{ a, \ a + \lfloor p_1(\Delta)f(n)\rfloor, \ a + \lfloor p_2(\Delta)f(n)\rfloor, \ \ldots, \ a + \lfloor p_k(\Delta)f(n)\rfloor \right\}$$

with $a, n \in \mathbb{N}$. *Moreover, if for each* $n \in \mathbb{N}$ *one denotes by* $A(n)$ *the set of* $a \in \mathbb{N}$ *for which the configuration in* (1.8) *is contained in* E, *then for any* $L \in \mathbb{N}$ *the set of* m *such that*

$$[m, m+L] \subset \left\{ n \in \mathbb{N} : d^*\big((A(n)\big) > 0 \right\}$$

is W-*syndetic, where* $W := \Delta^{\ell-1}f$.

It is natural to ask whether Theorem A can be improved in the spirit of Corollary F. Unfortunately, our proof of Theorem A (see Section 3) does not seem to allow for such an extension.

Organization of the paper. To establish (multiple) recurrence it is often helpful to start by analyzing translations on tori, where recurrence results can be obtained using the theory of uniform distribution. In Section 2 we study well-distribution of orbits along sequences from $\mathcal{F}$ in compact abelian groups. To that end we employ Riesz means, which are important for two reasons. On the one hand, they allow us to treat functions, such as $f(n) = \log n$, which are *not* uniformly distributed mod 1. On the other hand, even for functions which are uniformly distributed mod 1, such as $f(n) = \sqrt{n}$, by using Riesz means we are able to obtain a variant of well-distribution (see Definition 2.1 and Theorems 2.12 and 1.13), which is not discernible by Cesàro averages and is crucially needed in our proof of Theorem D.

In Section 3 we prove Theorem A using equidistribution results from Section 2; an outline of the proof is presented at the beginning of the section.

Sections 4 and 5 are devoted to the proof of Theorem D. The method of the proof bears some similarities to the method used by Frantzikinakis in [18], but has several differences which are partly due to the fact that we are using Riesz means. The first step of the proof of Theorem D, executed in Section 4.2, is to show that the nilfactor is "characteristic" for the pertinent expressions. We adapt a PET induction argument introduced in [2] together with results from [30] to fit the framework of Riesz means. The rest of Section 4 is dedicated to reducing Theorem D to the following equidistribution result on nilmanifolds, which is of independent interest.

Theorem 1.13. *Let* G *be a connected and simply connected nilpotent Lie group, let* $\Gamma \subset G$ *be a co-compact discrete subgroup and let* $X = G/\Gamma$. *Let* $k, \ell \in \mathbb{N}$ *with* $k \leqslant \ell$, *let* $f \in \mathcal{F}_\ell$, *let* $W = \Delta^{\ell-1}f$, *let* $b_0, \ldots, b_{k-1} \in G$ *be commuting elements, let*

$$Y = \overline{\left\{ b_0^{t_0} \cdots b_{k-1}^{t_{k-1}} \Gamma : t_0, \ldots, t_{k-1} \in \mathbb{R} \right\}} \subset X,$$

and let μ_Y *be the normalized Haar measure on* Y. *Then for every continuous function* $H \in C(X)$ *we have*

$$\underset{n\to\infty}{UW\text{-}\lim} \, H\left(b_0^{f(n)} b_1^{\Delta f(n)} \cdots b_{k-1}^{\Delta^{k-1}f(n)} \Gamma \right) = \int_Y H(y) \, \mathrm{d}\mu_Y(y).$$

The proof of Theorem 1.13 is presented in Section 5 and is based on a new induction procedure, loosely based on the scheme used by Green and Tao in [26]. Some of the technical results

which are used in the proof of Theorem 1.13 are collected in Appendix A .

2. WELL-DISTRIBUTION WITH RESPECT TO RIESZ MEANS

In this section we establish some facts about well-distribution of sequences in compact abelian groups with respect to Riesz means. These results will be utilized in the sequel.

2.1. WEYL'S CRITERION AND VAN DER CORPUT'S LEMMA FOR RIESZ MEANS

Definition 2.1 (cf. [14, 45])**.** Let G be a compact abelian group, let H be a closed subgroup of G and let μ_H denote the (normalized) Haar measure on H. Let $W \in \mathcal{F}_1$ and let $x \colon \mathbb{N} \to G$. We say that $x(n)$ is μ_H-*well distributed* (μ_H-*w.d.*) *with respect to* W-$\lim$ if for every continuous function $F \in C(G)$,

$$UW\text{-}\lim F\big(x(n)\big) = \lim_{W(N)-W(M)\to\infty} \frac{1}{W(N)-W(M)} \sum_{n=M}^{N} \Delta W(n) F\big(x(n)\big) = \int_G F \, \mathsf{d}\mu_H.$$

If $H = G$ we simply say that $x(n)$ is *well distributed* (*w.d.*) *with respect to* W-$\lim$.

We remark that well-distribution with respect to W-$\lim$ is related to W-syndetic sets in the same way as the conventional notion of well-distribution connects to syndetic sets.

Here are two results which illustrate Definition 2.1. They both follow directly from Lemma 2.13 below.

Example 2.2. Let $W(n) = \log n$. Then the sequence $x(n) = (\log n \bmod 1, \log(n+1) \bmod 1) \in \mathbb{T}^2$ is μ_H-w.d. with respect to W-$\lim$ where $H = \{(x,x) : x \in \mathbb{T}\}$ is the diagonal on $\mathbb{T}^2$. Observe that $x(n)$ is not μ_H-u.d. (a sequence $y(n)$ is μ_H-u.d. if for any $f \in C(\mathbb{T}^2)$ we have $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(y(n)) = \int_H f \, \mathsf{d}\mu_H$).

Example 2.3. Let $H := \{(x, y, 2y - x) : x, y \in \mathbb{T}\} \subset \mathbb{T}^3$ and consider the sequence $x(n) = \big(n^{3/2} \bmod 1, (n+1)^{3/2} \bmod 1, (n+2)^{3/2} \bmod 1\big) \in \mathbb{T}^3$. Since $(n+2)^{3/2} = 2(n+1)^{3/2} - n^{3/2} + \mathrm{o}_{n\to\infty}(n)$, the sequence $x(n)$ is μ_H-u.d. with respect to Cesàro means. But $x(n)$ is not μ_H-w.d. because if $F \colon H \to \mathbb{C}$ denotes the function $F(x, y, 2y - x) = e^{2\pi i(x-y)}$ for all $x, y \in \mathbb{T}$ then it is not hard to find $M, N \in \mathbb{N}$ with $N - M$ arbitrarily large such that $\frac{1}{N-M} \sum_{n=M}^{N} F(x(n))$ is not equal to $\int_H F \, \mathsf{d}\mu_H = 0$. However, if $W(n) = \sqrt{n}$ then $x(n)$ is μ_H-w.d. with respect to W-$\lim$.

We remark that in both examples above the sequence $x(n)$ is not contained in the subgroup H.

We will make use of the following version of Weyl's criterion.

Proposition 2.4 (Weyl criterion)**.** *Let G be a compact abelian group, let H be a closed subgroup of G, let $\widehat{G}$ denote the dual group of G and let $\Gamma := \{\chi \in \widehat{G} : \chi(y) = 1 \text{ for all } y \in H\}$. Let $W \in \mathcal{F}_1$ and let $x \colon \mathbb{N} \to G$. Then x is μ_H-w.d. with respect W-$\lim$ if and only if for every $\chi \in \widehat{G}$,*

$$UW\text{-}\lim_{n\to\infty} \chi(x(n)) = \begin{cases} 1, & \text{if } \chi \in \Gamma, \\ 0, & \text{if } \chi \notin \Gamma. \end{cases}$$

Proof. Every continuous function on G can be uniformly approximated by linear combinations of characters. So the proof follows from the fact that for each character χ we have $\int \chi \, \mathsf{d}\mu_H = 1$ if $\chi \in \Gamma$ and $\int \chi \, \mathsf{d}\mu_H = 0$ if $\chi \notin \Gamma$. $\square$

We also need a version of van der Corput's lemma. The proof is similar to that of [3, Theorem 2.2].

Proposition 2.5 (van der Corput). *Let $\mathscr{H}$ be a Hilbert space, let $W \in \mathcal{F}_1$, and let $(N_t)_{t\in\mathbb{N}}$ and $(M_t)_{t\in\mathbb{N}}$ be sequences of positive integers with $W(N_t) - W(M_t) \to \infty$ as $t \to \infty$. Let $f_1, f_2, \dots \colon \mathbb{N} \to \mathscr{H}$ satisfy $\sup_{t,n\in\mathbb{N}} \|f_t(n)\| < \infty$ and assume that for all but finitely many $d \in \mathbb{N}$ we have*

$$\lim_{t\to\infty} \mathbb{E}^W_{n\in[M_t,N_t]} \langle f_t(n+d), f_t(n) \rangle = 0.$$

Then

$$\lim_{t\to\infty} \mathbb{E}^W_{n\in[M_t,N_t]} f_t(n) = 0 \qquad \text{in norm.}$$

Proof. For each $d \in \mathbb{N}$ we have $\mathbb{E}^W_{n\in[M_t,N_t]} f_t(n) = \mathbb{E}^W_{n\in[M_t,N_t]} f_t(n+d) + o_{d;t\to\infty}(1)$.[7] Averaging over d and then using Jensen's inequality, we obtain

$$
\begin{aligned}
\left\| \mathbb{E}^W_{n\in[M_t,N_t]} f_t(n) \right\|^2
&= \left\| \mathbb{E}_{d\in[1,D]} \mathbb{E}^W_{n\in[M_t,N_t]} f_t(n+d) \right\|^2 + o_{D;t\to\infty}(1) \\
&\leqslant \mathbb{E}^W_{n\in[M_t,N_t]} \left\| \mathbb{E}_{d\in[1,D]} f_t(n+d) \right\|^2 + o_{D;t\to\infty}(1) \\
&= \mathbb{E}^W_{n\in[M_t,N_t]} \mathbb{E}_{d,d'\in[1,D]} \langle f_t(n+d), f_t(n+d') \rangle + o_{D;t\to\infty}(1) \\
&= \mathbb{E}_{d,d'\in[1,D]} \mathbb{E}^W_{n\in[M_t,N_t]} \langle f_t(n+d-d'), f_t(n) \rangle + o_{D;t\to\infty}(1)
\end{aligned}
$$

When $d = d'$ the inner average equals $\|f_t(n)\|^2$ which is bounded by a constant. When $d \neq d'$ the assumption implies that the inner average is $o_{D;t\to\infty}(1)$. Therefore $\|\mathbb{E}^W_{n\in[M_t,N_t]} f_t(n)\|^2 \leqslant \frac{1}{D} + o_{D;t\to\infty}(1)$. Letting $t \to \infty$ and then $D \to \infty$ we obtain the desired conclusion. $\square$

Combining Proposition 2.5 with Proposition 2.4 we deduce the following corollary.

Corollary 2.6. *Let G be a compact abelian group, let $u\colon \mathbb{N} \to G$ and let $W \in \mathcal{F}_1$. If for every $h \in \mathbb{N}$, the sequence $n \mapsto u(n+h) - u(n)$ is w.d. with respect to W-$\lim$, then the sequence $n \mapsto u(n)$ is w.d. with respect to W-$\lim$ as well.*

2.2. Well-distribution of sequences in $\mathcal{F}$ with respect to Riesz means

The following result is very similar to [34, Theorem 1.7.16], but we don't require any smoothness and, additionally, obtain well-distribution.

Lemma 2.7. *Let $\pi\colon \mathbb{R} \to \mathbb{T} := \mathbb{R}/\mathbb{Z}$ be the canonical quotient map and let $W \in \mathcal{F}_1$. Then the sequence $\pi(W(n))$ is w.d. with respect to W-$\lim$.*

Proof. When convenient, we identify $\mathbb{T}$ with the interval $[0,1)$ in the standard way. Since finite linear combinations of indicator functions $1_{[0,x)}$ of intervals $[0,x) \subset \mathbb{T}$ with $x \in [0,1]$ are uniformly dense in $C(\mathbb{T})$, it suffices to show that

$$U\underset{n\to\infty}{W\text{-}\lim}\, 1_{[0,x)}\big(\pi(W(n))\big) = x \qquad \text{for every } x \in [0,1].$$

Define $A_{M,N} := \mathbb{E}^W_{n\in[M,N]} 1_{[0,x)}(\pi(W(n)))$. Our goal is to show that $\lim_{W(N)-W(M)\to\infty} A_{M,N} = x$.

For each $m \in \mathbb{Z}$, let $S(m)$ be the smallest integer such that $W(S(m)) > m$. Given $M, N \in \mathbb{N}$, let w_M and w_N be the largest integers for which $S(w_M) \leqslant M$ and $S(w_N) \leqslant N$ and let $s_M = S(w_M)$ and $s_N = S(w_N)$. It suffices to verify the following claims:
CLAIM 1: $A_{s_M,s_N} = x + o_{W(N)-W(M)\to\infty}(1)$;
CLAIM 2: $|A_{M,N} - A_{s_M,s_N}| = o_{W(N)-W(M)\to\infty}(1)$.

---

[7]Here and throughout the paper we write $o_{d;t\to\infty}(1)$ to denote a quantity that tends to 0 as $t \to \infty$ and d is kept fixed.

Proof of Claim 1: Fix $\varepsilon > 0$ and choose m_0 such that $\Delta W(n) < \varepsilon$ for all $n > S(m_0)$. Let $c(m)$ be the smallest integer in the interval $[S(m), S(m+1)]$ such that $W(c(m)) \geqslant m + x$. A simple calculation shows that

$$A_{s_M, s_N} = \frac{1}{w_M - w_N} \sum_{m=s_M}^{s_N} \big[W(c(m)) - W(S(m)) \big] = x + \mathrm{O}\left(\varepsilon + \frac{m_0}{w_N - w_M} \right).$$

This finishes the proof of Claim 1.

Proof of Claim 2: Observe that $w_M \leqslant W(M) < w_M + 1$ and $w_N \leqslant W(N) < w_N + 1$ and hence (assuming $W(N) - W(M) > 2$) we have $s_M \leqslant M \leqslant s_N \leqslant N$. Therefore

$$
\begin{aligned}
|A_{M,N} - A_{s_M, s_N}| &= \left| \mathbb{E}_{n \in [M,N]}^W 1_{[0,x)}\big(\pi(W(n)) \big) - \mathbb{E}_{n \in [s_M, s_N]}^W 1_{[0,x)}\big(\pi(W(n)) \big) \right| \\
&\leqslant \left| \frac{1}{w_N - w_M} - \frac{1}{W(N) - W(M)} \right| (w_N - W(M)) \\
&\qquad + \frac{W(M) - w_M}{w_N - w_M} + \frac{W(N) - w_N}{W(N) - W(M)} \\
&\leqslant \frac{4}{W(N) - W(M) - 2}.
\end{aligned}
$$

$\square$

The following corollary follows immediately from Lemma 2.7 and the definition of W-lim.

Corollary 2.8. *Let $W \in \mathcal{F}_1$ and $\alpha \in \mathbb{R} \backslash \{0\}$. Then the sequence $\alpha W(n) \bmod 1$ is w.d. in $[0,1)$ with respect to W-lim.*

Here is another corollary which will be used later.

Corollary 2.9. *Let $W \in \mathcal{F}_1$ and let $f : \mathbb{N} \to \mathbb{R}$ be such that $f(n) \to 0$ as $n \to \infty$. Then the set $A := \{ n \in \mathbb{N} : \lfloor W(n) + f(n) \rfloor \neq \lfloor W(n) \rfloor \}$ has zero Banach W-density[8].*

Proof. For each $\varepsilon > 0$ let $A_\varepsilon = \{ n \in \mathbb{N} : \| W(n) \|_{\mathbb{T}} < \varepsilon \}$.[9] It follows from Lemma 2.7 that $d_W^*(A_\varepsilon) = 2\varepsilon$. Since A_ε contains A, up to possibly finitely many exceptions, it follows that $d_W^*(A) < 2\varepsilon$, and hence taking $\varepsilon \to 0$ we obtain $d_W^*(A) = 0$ as desired. $\square$

The next lemma shows that the property of being w.d. is stable under small perturbations.

Lemma 2.10. *Let G be a compact abelian group, let $W \in \mathcal{F}_1$ and let $u \colon \mathbb{N} \to G$ be a sequence that is w.d. with respect to W-lim. Then for every $g \colon \mathbb{N} \to G$ such that $\lim_{n \to \infty} g(n) = 0$, the sequence $u(n) + g(n)$ is w.d. with respect to W-lim.*

Proof. Let $\chi \colon G \to \mathbb{C}$ be a non-trivial character. We need to show that $UW\text{-}\lim \chi\big(u(n) + g(n) \big) = 0$. Since χ is a continuous homomorphism it satisfies $\chi\big(u(n) + g(n) \big) = \chi(u(n)) \chi(g(n))$ and $\chi(g(n)) = 1 + o_{n \to \infty}(1)$, so $\chi\big(u(n) + g(n) \big) - \chi(u(n)) = o_{n \to \infty}(1)$. This implies that $UW\text{-}\lim \chi\big(u(n) + g(n) \big) = UW\text{-}\lim \chi\big(u(n) \big) = 0$. $\square$

The following classical identity will be used often throughout this paper.

Lemma 2.11. *For every function $f \colon \mathbb{N} \to \mathbb{C}$ and $h \in \mathbb{N}$ we have*

$$(2.1) \qquad f(n + h) = \big(\Delta + 1 \big)^h f(n) = \sum_{i=0}^{h} \binom{h}{i} \Delta^i f(n).$$

---

[8]Given a set $E \subset \mathbb{N}$ its *Banach W-density* is defined as $d_W^*(E) := UW\text{-}\lim_{n \to \infty} 1_E(n)$ whenever this limit exists.

[9]Here, and elsewhere in the paper, we use the notation $\| x \|_{\mathbb{T}}$ to denote the distance from $x \in \mathbb{R}$ to the closest point in $\mathbb{Z}$.

Theorem 2.12. *Let $\pi \colon \mathbb{R} \to \mathbb{T}$ be the canonical quotient map, let $\ell \geqslant 0$ be an integer, let $f \in \mathcal{F}_{\ell+1}$ and let $W = \Delta^\ell f$. Then for any coefficients $c_0, \ldots, c_\ell \in \mathbb{R}$, not all 0, and any $g \colon \mathbb{N} \to \mathbb{R}$ which satisfies $\lim g(n) = 0$, the function $F := \pi \circ \left(c_0 f + c_1 \Delta f + \cdots + c_\ell \Delta^\ell f + g \right)$ is w.d. with respect to W-$\lim$.*

Proof. We prove this by induction on ℓ. For $\ell = 0$ the conclusion follows directly from Corollary 2.8 and Lemma 2.10. Assume now that $\ell \geqslant 1$ and the result has already been established for all smaller ℓ.

Let F be as in the statement of the theorem. If $c_0 = 0$ the result follows by induction, hence, let us assume $c_0 \neq 0$. We will use Corollary 2.6 to show that F is w.d. with respect to the Riesz mean W-$\lim$. Hence it suffices to show that for any $h \in \mathbb{N}$ the sequence $n \mapsto F(n+h) - F(n)$ is w.d. with respect to W-$\lim$. In view of Lemma 2.11 we have

$$F(n+h) = \sum_{j=0}^\ell c_j \Delta^j f(n+h) + g(n) = \sum_{j=0}^\ell c_j \sum_{i=0}^h \binom{h}{i} \Delta^{j+i} f(n) + g(n)$$

so $F(n+h) - F(n) = b_1 \Delta f + \cdots + b_k \Delta^\ell f + o(1)$, where

$$b_k = \sum_{i=0}^k c_{k-i} \binom{h}{i} - c_k = \sum_{i=1}^k c_{k-i} \binom{h}{i}.$$

In particular $b_1 = h c_0 \neq 0$. Since $\Delta f \in \mathcal{F}_\ell$ and $\Delta^{\ell-1}(\Delta f) = W$ we deduce from the induction hypothesis that indeed $F(n+h) - F(n)$ is w.d. with respect to W-$\lim$. $\qquad \square$

In view of the Weyl criterion, we can extrapolate the above result to well-distribution of sequences in higher-dimensional tori.

Lemma 2.13. *Let $\ell \geqslant 0$, let $f \in \mathcal{F}_{\ell+1}$ and let $W = \Delta^\ell f$. Let $d \in \mathbb{N}$, let $\pi \colon \mathbb{R}^d \to \mathbb{T}^d$ be the canonical quotient map and, for each $i = 0, \ldots, \ell$, let $\alpha_i \in \mathbb{R}^d$. Define the sequence $F \colon \mathbb{N} \to \mathbb{R}^{(\ell+1)d}$ by*

$$F(n) = \left(\alpha_i \Delta^i f(n) \right)_{i=0,\ldots,\ell},$$

and consider the subgroup

$$H := \bigotimes_{i=0}^\ell \overline{\{ \pi(t \alpha_i) : t \in \mathbb{R} \}} \subset \mathbb{T}^{(\ell+1)d}$$

with corresponding normalized Haar measure μ_H. Then the sequence $\pi\big(F(n)\big)$ is μ_H-w.d. with respect to W-$\lim$.

Proof. Let Γ be the set of characters of $\mathbb{T}^{(\ell+1)d}$ that become trivial when restricted to H. Given a character χ of $\mathbb{T}^{(\ell+1)d}$, we can find $\tau_0, \ldots, \tau_\ell \in \mathbb{Z}^d$ such that $\chi\big(\pi(x_0, \ldots, x_\ell)\big) = e\big(\langle x_0, \tau_0 \rangle + \cdots + \langle x_\ell, \tau_\ell \rangle \big)$. Then $\chi \in \Gamma$ if and only if $\langle \alpha_i, \tau_i \rangle = 0$ for all $i = 0, \ldots, \ell$. In view of Theorem 2.12 we have

$$U\!\!\underset{n\to\infty}{W\text{-}\lim}\, \chi\Big(\pi\big(F(n)\big) \Big) = U\!\!\underset{n\to\infty}{W\text{-}\lim}\, e\left(\sum_{i=0}^\ell \langle \alpha_i, \tau_i \rangle \Delta^i f(n) \right) = 1_\Gamma(\chi),$$

where $e(\theta) := e^{2\pi i \theta}$, and hence the Weyl criterion (Proposition 2.4) implies that the sequence $\pi\big(F(n)\big)$ is μ_H-w.d. with respect to W-$\lim$. $\qquad \square$

Since functions $f \in \mathcal{F}$ are real valued, we consider the associated sequences $g(n) := \lfloor f(n) \rfloor$. Given a real number $x \in \mathbb{R}$ we denote by $\{x\} := x - \lfloor x \rfloor$ the fractional part of x. Since Δ is a linear operator we have that $\Delta^i g(n) = \Delta^i f(n) - \Delta^i \{f(n)\}$.

Lemma 2.14. *Let $f \in \mathcal{F}$, let $g(n) = \lfloor f(n) \rceil$ and let $h \in \mathbb{N}$. Then*

$$(2.2) \qquad \Delta^h g(n) = \Delta^h f(n) - \sum_{t=0}^{h} \binom{h}{t} (-1)^{h-t} \left\{ \sum_{s=0}^{t} \binom{t}{s} \Delta^s f(n) \right\}$$

Proof. The following well known identity, which is similar to (2.1), can be derived with the help of Newton's binomial formula:

$$(2.3) \qquad \Delta^h \tilde{f}(n) = \sum_{i=0}^{h} (-1)^{h-i} \binom{h}{i} \tilde{f}(n+i).$$

Equation (2.2) follows directly by applying (2.3) to the function $\tilde{f}(n) = \{f(n)\}$ and then using Lemma 2.11. $\qquad \square$

We record for future reference the following inequality, which follows from Lemma 2.14:

$$(2.4) \qquad \forall\, h, n \in \mathbb{N} \qquad \left| \Delta^h g(n) - \Delta^h f(n) \right| < 2^h.$$

Theorem 2.15. *Let $\ell, d \in \mathbb{N}$, let $f \in \mathcal{F}_{\ell+1}$, let $W := \Delta^\ell f$, let $g(n) := \lfloor f(n) \rceil$ and let $\pi \colon \mathbb{R}^{(d+1)(\ell+1)} \to \mathbb{T}^{(d+1)(\ell+1)}$ be the canonical projection. Then, for all $\alpha \in \mathbb{R}^d$, the sequence*

$$G(n) := \pi\Big(\big(\alpha \Delta^i g(n), \Delta^i f(n) \big)_{i=0,\dots,\ell} \Big)$$

is $\mu_{(K \times \mathbb{T})^{\ell+1}}$-w.d. with respect to W-$\lim$, where $K := \overline{\{ n\alpha \bmod \mathbb{Z}^d : n \in \mathbb{N} \}} \subset \mathbb{T}^d$.

Proof. Let Γ be the set of characters of $\mathbb{T}^{(d+1)(\ell+1)}$ which become trivial when restricted to $(K \times \mathbb{T}^d)^{\ell+1}$. Let χ be a character of $\mathbb{T}^{(d+1)(\ell+1)}$. For each $i = 0, \dots, \ell$ there exist $\tau_i \in \mathbb{Z}^d$ and $h_i \in \mathbb{Z}$ such that

$$\chi\Big(\pi\big((x_i, y_i)_{i=0,\dots,\ell} \big) \Big) = e\left(\sum_{i=0}^{\ell} \langle x_i, \tau_i \rangle + h_i y_i \right).$$

Then $\chi \in \Gamma$ if and only if $\langle \alpha, \tau_i \rangle \in \mathbb{Z}$ and $h_i = 0$ for every $i = 0, \dots, \ell$. In particular, if $\chi \in \Gamma$ then $\chi\big(G(n) \big) = 1$ for every $n \in \mathbb{N}$. In view of the Weyl criterion (Proposition 2.4) it thus suffices to show that given $(\tau_i, h_i) \in \mathbb{Z}^{d+1}$, $i = 0, \dots, \ell$, if there exists $i \in \{0, \dots, \ell\}$ such that either $\alpha_i := \langle \tau_i, \alpha \rangle \notin \mathbb{Z}$ or $h_i \neq 0$, then one has

$$(2.5) \qquad UW\text{-}\lim_{n \to \infty} e\left(\sum_{i=0}^{\ell} \alpha_i \Delta^i g(n) + h_i \Delta^i f(n) \right) = 0.$$

Using Lemma 2.14 we can rewrite (2.5) as

$$(2.6) \qquad UW\text{-}\lim_{n \to \infty} \psi\big(F(n) \big) = 0$$

where $F \colon \mathbb{N} \to \mathbb{R}^{2(\ell+1)}$ is the sequence

$$F(n) := \big(\Delta^i f(n), \ \alpha_i \Delta^i f(n) \big)_{i=0,\dots,\ell}$$

and $\psi \colon \mathbb{R}^{2(\ell+1)} \to \mathbb{C}$ is the map

$$
\begin{aligned}
\psi(x_0, \dots, x_\ell, y_0, \dots, y_\ell) \ &:= \ e\left(\sum_{i=0}^{\ell} y_i + h_i x_i - \alpha_i \sum_{t=0}^{i} \binom{i}{t} (-1)^{i-t} \left\{ \sum_{s=0}^{t} \binom{t}{s} x_s \right\} \right) \\
&= \ e\left(\sum_{i=0}^{\ell} y_i + h_i x_i \right) e\left(- \sum_{t=0}^{\ell} \left\{ \sum_{s=0}^{t} \binom{t}{s} x_s \right\} \sum_{i=t}^{\ell} \alpha_i \binom{i}{t} (-1)^{i-t} \right).
\end{aligned}
$$

In view of Lemma 2.13, the sequence $F(n) \bmod \mathbb{Z}^{2(\ell+1)}$ is μ_H-w.d. with respect to $W\text{-}\lim$, where

$$H = \bigotimes_{i=0}^{\ell} H_i \subset \mathbb{T}^{2(\ell+1)}$$

and $H_i := \overline{\{(t, \alpha_i t) : t \in \mathbb{R}\}} \subset \mathbb{T}^2$.

Since ψ is Riemann integrable and periodic modulo $\mathbb{Z}^{2(\ell+1)}$, (2.6) will follow if we show that the integral $\int_H \psi \, d\mu_H$ of ψ over H (with respect to the Haar measure μ_H on H) equals 0. For convenience, let $\psi_{\ell+1} \equiv 1$ and for each $i = 0, \dots, \ell$, define, recursively

$$\psi_i(x_0, \dots, x_{i-1}) :=$$

$$\int_{H_i} e\left(y_i + h_i x_i - \left\{ \sum_{s=0}^{i} \binom{i}{s} x_s \right\} \sum_{j=i}^{\ell} \alpha_j \binom{j}{i} (-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \dots, x_i) \, d\mu_{H_i}(x_i, y_i).$$

Notice that $\psi_0 = \int_H \psi \, d\mu_H$ (and in particular ψ_0 is a constant).

To show that $\psi_0 = 0$ we distinguish two cases: the case where some α_j is irrational, and the case where all α_i are rational for $i \in \{0, \dots, \ell\}$.

For the first case, let $j \in \{0, \dots, \ell\}$ be such that α_j is irrational. In this case $H_j = \mathbb{T}^2$ and hence $d\mu_{H_j}(x_j, y_j) = dx_j \, dy_j$. Therefore, for any $x_0, \dots, x_{j-1} \in \mathbb{T}$,

$$\psi_j(x_0, \dots, x_{j-1})$$

$$= \int_{\mathbb{T}} e(y_j) \, dy_j \cdot \int_{\mathbb{T}} e\left(h_i x_i - \left\{ \sum_{s=0}^{i} \binom{i}{s} x_s \right\} \sum_{j=i}^{\ell} \alpha_j \binom{j}{i} (-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \dots, x_i) \, dx_j$$

$$= 0.$$

Therefore, for every $i \leqslant j$ we have $\psi_i \equiv 0$ and in particular $\psi_0 = 0$.

Next we treat the case when all α_i are rational. Write $\alpha_i = a_i/b_i$ with $a_i \in \mathbb{Z}$ and $b_i \in \mathbb{N}$ coprime. Then H_i can be parameterized as $H_i = \{(x, \alpha_i x) : x \in [0, b_i)\}$ and, more generally, as $H_i = \{(x, \alpha_i x) : x \in [c, c + b_i]\}$ for any $c \in \mathbb{R}$. Let $c = c(x_0, \dots, x_{i-1}) = -\sum_{s=0}^{i-1} \binom{i}{s} x_s$ and note that

$$\psi_i(x_0, \dots, x_{i-1})$$

$$= \int_c^{c+b_i} e\left((\alpha_i + h_i) x_i - \left\{ \sum_{s=0}^{i} \binom{i}{s} x_s \right\} \sum_{j=i}^{\ell} \alpha_j \binom{j}{i} (-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \dots, x_i) \, dx_i$$

$$= \sum_{t=0}^{b_i-1} \int_{c+t}^{c+t+1} e\left((\alpha_i + h_i) x_i - \{x_i - c\} \sum_{j=i}^{\ell} \alpha_j \binom{j}{i} (-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \dots, x_i) \, dx_i$$

$$= \sum_{t=0}^{b_i-1} \int_{c+t}^{c+t+1} e\left((\alpha_i + h_i) x_i - (x_i - c - t) \sum_{j=i}^{\ell} \alpha_j \binom{j}{i} (-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \dots, x_i) \, dx_i.$$

Let i be the largest index for which either $\alpha_i \notin \mathbb{Z}$ or $h_i \neq 0$. In order to finish the proof of Theorem 2.15 we need to following lemma:

Lemma 2.16. *For every $r = i+1, \dots, \ell$*

$$\psi_r(x_0, \dots, x_{r-1}) = e\left(-\sum_{s=0}^{r-1} x_s \sum_{j=r}^{\ell} a_j \binom{j}{s} \sum_{k=r}^{j} \binom{j-s}{k-s} (-1)^{j-k} \right).$$

Proof of Lemma 2.16. We will use backward induction on r. For $r = \ell + 1$ the result is trivial. Assume now that we have established this for $r + 1$.

For each $r = i + 1, \ldots, \ell$ we have that $\alpha_r = a_r$, $b_r = 1$ and $h_r = 0$. Therefore

$$\psi_r = \int_c^{c+1} e\left(a_r x_r - (x_r - c) \sum_{j=r}^{\ell} a_j \binom{j}{r}(-1)^{j-r} \right.$$
$$\left. - \sum_{s=0}^{r} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{s} \sum_{k=r+1}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right) \, \mathrm{d}x_r$$

Observe that the coefficient of x_r inside the exponential is 0, so the integrand is in fact a constant function and hence

$$\psi_r = e\left(c \sum_{j=r}^{\ell} a_j \binom{j}{r}(-1)^{j-r} - \sum_{s=0}^{r-1} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{s} \sum_{k=r+1}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right)$$

$$= e\left(-\sum_{s=0}^{r-1} \binom{r}{s} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{r}(-1)^{j-r} - \sum_{s=0}^{r-1} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{s} \sum_{k=r+1}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right)$$

Using the identity $\binom{r}{s}\binom{j}{r} = \binom{j}{s}\binom{j-s}{r-s}$ we conclude that

$$\psi_r = e\left(-\sum_{s=0}^{r-1} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{s}\binom{j-s}{r-s}(-1)^{j-r} - \sum_{s=0}^{r-1} x_s \sum_{j=r+1}^{\ell} a_j \binom{j}{s} \sum_{k=r+1}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right)$$

$$= e\left(-\sum_{s=0}^{r-1} x_s \sum_{j=r}^{\ell} a_j \binom{j}{s} \sum_{k=r}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right)$$

as desired. $\qquad\square$

Continuation of the proof of Theorem 2.15: Lemma 2.16 gives us a rather explicit (albeit cumbersome) expression for ψ_i. Recall that i was chosen as the largest index so that either $\alpha_i \notin \mathbb{Z}$ or $h_i \neq 0$. We further divide into two cases: $h_i = 0$ (in which case $\alpha_i \notin \mathbb{Z}$) and $h_i \neq 0$.

If $h_i = 0$ then a quick computation shows that the integrand

$$e\left((\alpha_i + h_i)x_i - (x_i - c - t) \sum_{j=i}^{\ell} \alpha_j \binom{j}{i}(-1)^{j-i} \right) \cdot \psi_{i+1}(x_0, \ldots, x_i)$$

does not depend on x_i, and hence we have

$$\psi_i = \sum_{t=0}^{b_i-1} e\left((c+t) \sum_{j=i}^{\ell} \alpha_j \binom{j}{i}(-1)^{j-i} - \sum_{s=0}^{i-1} x_s \sum_{j=i+1}^{\ell} a_j \binom{j}{s} \sum_{k=i+1}^{j} \binom{j-s}{k-s}(-1)^{j-k} \right)$$

Notice that after factoring out of the sum the terms which do not depend on t we end up with a simple geometric sum. Since every α_r for $r > i$ is an integer (unless of course $i = \ell$ in which case there is no $r > i$), we conclude

$$\psi_i = \text{constant} \cdot \sum_{t=0}^{b_i-1} e\left(t \sum_{j=i}^{\ell} \alpha_j \binom{j}{i}(-1)^{j-i} \right)$$

$$= \text{constant} \cdot \sum_{t=0}^{b_i-1} e\left(t\alpha_i \right)$$

$$= \text{ constant} \cdot \sum_{t=0}^{b_i-1} e\left(\frac{a_i}{b_i}\right)^t$$

$$= 0.$$

Finally we address the case when $h_i \neq 0$. In this case, for each $t = 0, \ldots, b_i - 1$, the integrand

$$e\left((\alpha_i + h_i)x_i - (x_i - c - t)\sum_{j=i}^{\ell}\alpha_j\binom{j}{i}(-1)^{j-i}\right) \cdot \psi_{i+1}(x_0, \ldots, x_i)$$

is a constant multiple of $e(h_i x_i)$, so when integrated over an interval of length 1, it vanishes. $\quad\square$

The following corollary of Theorem 2.15 will be used in the proof of Theorem 3.10 below.

Corollary 2.17. *Let $\ell \in \mathbb{N}$, let $f \in \mathcal{F}_{\ell+1}$ and let $W = \Delta^\ell f$. Then for every $\alpha \in \mathbb{R}\backslash\mathbb{Z}$ the sequence $(\alpha\lfloor f(n)\rfloor)$ is μ_H-w.d. with respect to W-$\lim$, where μ_H is the normalized Haar measure of the closed subgroup of $\mathbb{T}$ defined by $H := \overline{\{n\alpha : n \in \mathbb{N}\}}$. In particular, for every $\alpha \in \mathbb{R}\backslash\mathbb{Z}$ we have*

$$\underset{n\to\infty}{UW\text{-}\lim}\, e\big(\alpha\lfloor f(n)\rfloor\big) = 0.$$

3. OPTIMAL SINGLE RECURRENCE ALONG SEQUENCES IN $\mathcal{F}$

In this section we give a proof of Theorem A. Here is the statement again, for the convenience of the reader.

Theorem A. *Let $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$ and define $W := \Delta^{\ell-1}f$. Then for any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ and any $\varepsilon > 0$ the set*

$$R_{\varepsilon,A} = \big\{n \in \mathbb{N} : \mu(A \cap T^{-\lfloor f(n)\rfloor}A) > \mu^2(A) - \varepsilon\big\}$$

is thick and W-syndetic.

Outline of the proof: In the proof of Theorem A we will utilize the Jacobs-de Leeuw-Glicksberg decomposition:

Theorem 3.1 (cf. [33, §2.4] or [3, Theorem 2.3]). *Let $\mathcal{H}$ be a Hilbert space and $U : \mathcal{H} \to \mathcal{H}$ a unitary operator. Then any $h \in \mathcal{H}$ can be written as $h = h_{wm} + h_c$, where $h_c \perp h_{wm}$ and*
- *h_{wm} is a weakly mixing element, that is, for every $h' \in \mathcal{H}$ we have*

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}|\langle U^n h_{wm}, h'\rangle| = 0.$$

- *h_c is a compact element, i.e., the closure of $\{U^n h_c : n \in \mathbb{N}\}$ is a compact subset of $\mathcal{H}$.*

In light of Theorem 3.1 we can write $1_A = h_{wm} + h_c$, where $h_{wm} \in L^2(X, \mathcal{B}, \mu)$ is weakly mixing and $h_c \in L^2(X, \mathcal{B}, \mu)$ is compact.

As for the component h_{wm}, it will be shown in the next subsection that for any $\ell \in \mathbb{N}$, any $h' \in L^2(X, \mathcal{B}, \mu)$ and any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\frac{1}{N}\sum_{n=1}^{N}\left|\langle T^{p(n)}h_{wm}, h'\rangle\right| < \varepsilon$ uniformly over all polynomials $p \in \mathbb{Z}[x]$ of degree ℓ and having leading coefficient belonging to a set of lower Banach density equal to 1.

In Subsection 3.2 we use the equidistribution results obtained in Section 2 to show that for any $\varepsilon > 0$ there is a thick set $E \subset \mathbb{N}$ with the property that for any $n \in E$ the function $T^{\lfloor f(n)\rfloor}h_c$ is ε-close to h_c in L^2-norm and for all $n \in E$ the Taylor expansion of $k \mapsto f(n+k)$, truncated at an appropriate level, corresponds to a polynomial whose leading coefficient belongs to the set of lower Banach density equal to 1.

Finally, in Subsection 3.3, these results are "glued together" to yield a proof of Theorem A.

3.1. The weakly mixing component

The following remark will be utilized in the proof of Theorem 3.3 below.

Remark 3.2. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $h \in L^2(X, \mathcal{B}, \mu)$. It is not hard to see from the definition that h is weakly mixing if and only if for every $\varepsilon > 0$ and every $h' \in L^2(X)$, there exists a set $D \subset \mathbb{N}$ with $d_*(D) = 1$ such that $|\langle T^d h, h' \rangle| < \varepsilon$ for every $d \in D$. Here, d_* denotes the lower Banach density, defined by the formula

$$d_*(D) = \lim_{N \to \infty} \inf_{M \in \mathbb{N}} \frac{\left| D \cap \{M, M+1, \ldots, M+N\} \right|}{N}.$$

Theorem 3.3. Let $\ell \in \mathbb{N} \cup \{0\}$, let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $h \in L^2(X, \mathcal{B}, \mu)$ be a weakly mixing function with $\|h\|_2 \leqslant 1$. Then for any $\varepsilon > 0$, any $h' \in L^2(X, \mathcal{B}, \mu)$ with $\|h'\|_2 \leqslant 1$ and sufficiently large $N \in \mathbb{N}$ there exists $D \subset \mathbb{N}$ with $d_*(D) = 1$ such that for every polynomial $p \in \mathbb{Z}[x]$ of degree ℓ with $\Delta^\ell p \in D$, one has

$$\frac{1}{N} \sum_{n=1}^{N} \left| \langle T^{p(n)} h, h' \rangle \right| \;\leqslant\; \varepsilon.$$

(Observe that $\Delta^\ell p$ is a constant function whenever p is a polynomial of degree ℓ and the (only) value it takes is $\ell!$ times the leading coefficient of p.)

Proof. For $\ell = 0$ the conclusion follows directly from Remark 3.2. We now proceed by induction on ℓ, assuming that $\ell > 0$ and that the result has been established for $\ell - 1$. Observe that

$$\left| \langle T^{p(n)} h, h' \rangle \right|^2 \;=\; \int_X T^{p(n)} h \cdot h' \, \mathsf{d}\mu \int_X T^{p(n)} \bar{h} \cdot \bar{h}' \, \mathsf{d}\mu$$

$$=\; \int_{X \times X} (T \times T)^{p(n)} (h \otimes \bar{h}) \cdot (h' \otimes \bar{h}') \, \mathsf{d}(\mu \otimes \mu).$$

It follows easily from the definition that if h is a weak mixing function then so is $h \otimes \bar{h}$. Therefore, renaming $h \otimes \bar{h}$ as h and $T \times T$ as T, it suffices to prove that

$$\left\| \frac{1}{N} \sum_{n=1}^{N} T^{p(n)} h \right\| < \varepsilon.$$

To this end we will employ a version of the van der Corput inequality (cf. [34, Lemma 1.3.1]), which states that for every Hilbert space vectors $u_1, u_2, \ldots$ with norm bounded by 1, one has

$$(3.1) \qquad \left\| \frac{1}{N} \sum_{n=1}^{N} u_n \right\| \leqslant \frac{2}{\sqrt{N}} + \sqrt{ \frac{1}{\sqrt{N}} + \frac{2}{\sqrt{N}} \sum_{1 \leqslant m \leqslant \sqrt{N}} \left| \frac{1}{N-m} \sum_{n=1}^{N-m} \langle u_{n+m}, u_n \rangle \right| }.$$

Therefore, it suffices to show that for every $\delta > 0$ and large enough $N \in \mathbb{N}$ there exists a set $D \subset \mathbb{N}$ with $d_*(D) = 1$ such that for every polynomial $p \in \mathbb{Z}[x]$ of degree ℓ with $\Delta^\ell p$ in D and all $m \in \{1, \ldots, \lfloor \sqrt{N} \rfloor\}$, one has

$$(3.2) \qquad \left| \frac{1}{N} \sum_{n=1}^{N} \left\langle T^{p(n+m)} h, T^{p(n)} h \right\rangle \right| < \delta.$$

Observe that

$$\left\langle T^{p(n+m)} h, T^{p(n)} h \right\rangle \;=\; \int_X T^{p(n+m)} h \cdot T^{p(n)} \bar{h} \, \mathsf{d}\mu$$

$$=\; \int_X T^{p(n+m)-p(n)} h \cdot \bar{h} \, \mathsf{d}\mu$$

and $p_m \colon n \mapsto p(n+m) - p(n)$ is a polynomial of degree $\ell - 1$ with $\Delta^{\ell-1} p_m = m \Delta^\ell p$.

By the induction hypothesis there exists a set $D_0 \subset \mathbb{N}$ with $d_*(D_0) = 1$ such that for any polynomial $q \in \mathbb{Z}[x]$ of degree $\ell - 1$ with $\Delta^{\ell-1} q$ in D_0,

$$\frac{1}{N} \sum_{n=1}^{N} |\langle T^{q(n)} h, \bar{h} \rangle| \ \leqslant \ \delta.$$

Note now that the set $D_0/m := \{a \in \mathbb{N} : am \in D_0\}$ satisfies $d_*(D_0/m) = 1$ for any $m \in \mathbb{N}$. Moreover, the collection of sets with lower Banach density 1 has the finite intersection property. It follows that the set

$$D := \bigcap_{m=1}^{\lfloor \sqrt{N} \rfloor} D_0/m$$

has $d_*(D) = 1$.

Putting everything together, if $p \in \mathbb{Z}[x]$ is a polynomial of degree ℓ with $\Delta^\ell p$ in D, then the polynomial $p_m(n) = p(n+m) - p(n)$ has degree $\ell - 1$ and $\Delta^{\ell-1} p_m$ is in D_0 for every $m \in \{1, \ldots, \lfloor \sqrt{N} \rfloor\}$. This implies that (3.2) holds, which concludes the proof. $\qquad\square$

3.2. THE COMPACT COMPONENT

To deal with the compact component in the proof of Theorem A, we will use the equidistribution results developed in Section 2. We start with the following observation.

Lemma 3.4. *Let $\ell \geqslant 0$, let $f \in \mathcal{F}_{\ell+1}$, let $g(n) = \lfloor f(n) \rceil$ and let $W = \Delta^\ell f$. For any $\alpha \in \mathbb{R}^d$ and any $\delta > 0$ the set*[10]

$$D := \left\{ n \in \mathbb{N} : \left\| \Delta^i g(n) \alpha \right\|_{\mathbb{T}^d} \leqslant \delta; \ \{\Delta^i f(n)\} < \delta \text{ for all } i \in \{0, 1, \ldots, \ell\} \right\}$$

satisfies $d_W(D) > 0$.

Proof. It follows from Theorem 2.15 that the sequence

$$n \mapsto \Big(g(n)\alpha, f(n), \Delta g(n)\alpha, \Delta f(n), \ldots, \Delta^\ell g(n)\alpha, \Delta^\ell f(n) \Big)$$

is μ_H-w.d. with respect to $W\text{-}\lim$ for some closed subgroup H of $\mathbb{T}^{(d+1)(\ell+1)}$. Let $F_\delta \colon (\mathbb{T}^d \times \mathbb{T})^{(\ell+1)} \to \{0,1\}$ denote the function

$$F_\delta(x_0, y_0, \ldots, x_\ell, y_\ell) = \begin{cases} 1, & \text{if } \|x_i\|_{\mathbb{T}^d} \leqslant \delta \text{ and } \{y_i\} < \delta \text{ for all } 0 \leqslant i \leqslant \ell.; \\ 0, & \text{otherwise.} \end{cases}$$

Notice that while F_δ is not continuous, the sets of discontinuity for F_δ as δ changes are disjoint, and hence for all but at most countably many values of δ, the function F_δ is almost everywhere continuous with respect to the Haar measure μ_H of H. Replacing if necessary F_δ with $F_{\delta'}$ for some $\delta' < \delta$, and noticing that $F_{\delta'} \leqslant F_\delta$, we can then assume that F_δ is μ_H-a.e. continuous. We now use well-distribution to conclude that

$$\begin{aligned} d_W(D) &= \lim_{N \to \infty} \frac{1}{W(N)} \sum_{n=1}^{N} \Delta W(n) F\Big(g(n)\alpha, f(n), \ldots, \Delta^\ell g(n)\alpha, \Delta^\ell f(n) \Big) \\ &= \int_H F \, \mathsf{d}\mu_H > 0. \end{aligned}$$

$\qquad\square$

---

[10] Here, and elsewhere in the paper we use the notation $\|x\|_{\mathbb{T}^d}$ to denote the distance from the vector $x \in \mathbb{R}^d$ to the closest point in the lattice $\mathbb{Z}^d$.

Lemma 3.5. *Let $f \in \mathcal{F}_{\ell+1}$, let $g(n) = \lfloor f(n) \rfloor$, let $W = \Delta^\ell f$ and let $D \subset \mathbb{N}$ be such that $d_W(D) > 0$. Define $B := \Delta^\ell g(D) = \{\Delta^\ell g(n) : n \in D\}$. Then $\bar{d}(B) > 0$.*

Proof. Let $s \leqslant N$ be natural numbers. Denote by $(\Delta^\ell g)^{-1}(s)$ the set $\{n : \Delta^\ell g(n) = s\}$. Let $M = M(N) = \max\{m : \Delta^\ell g(m) \leqslant N\}$ and let

$$P(s, N) := \frac{1}{W(M)} \sum_{n \in (\Delta^\ell g)^{-1}(s)} \Delta W(n).$$

From (2.4) it follows that for large enough N we have $W(M)/N \geqslant 1/2$. Moreover, (2.4) also implies that if $n \in (\Delta^\ell g)^{-1}(s)$ then $W(n) \in (s - 2^\ell, s + 2^\ell)$. Let $W^{-1}(a) := \min\{x \in \mathbb{N} : W(x) \geqslant a\}$. We have

$$\sum_{m \in (\Delta g)^{-1}(s)} \Delta W(m) \leqslant \sum_{m = W^{-1}(s - 2^\ell)}^{W^{-1}(s + 2^\ell) - 1} \Delta W(m) \leqslant 2^{\ell+1}$$

and hence, for large enough N, we have

$$\frac{d_W(D)}{2} \;\leqslant\; \frac{1}{W(M)} \sum_{m=1}^{M} 1_D(m) \Delta W(m) = \frac{1}{W(M)} \sum_{s \in B \cap [N]} \sum_{m \in (\Delta g)^{-1}(s)} 1_D(m) \Delta W(m)$$

$$\leqslant \sum_{s \in B \cap [N]} P(s, N) \leqslant \frac{2^{\ell+1} |B \cap [N]|}{W(M)} \leqslant \frac{2^{\ell+2} |B \cap [N]|}{N} \leqslant 2^{\ell+3} \bar{d}(B).$$

$\square$

Lemma 3.6. *For every $\ell, N \in \mathbb{N}$ there exists $\delta > 0$ such that if $f \in \mathcal{F}_{\ell+1}$ and $a \in \mathbb{N}$ satisfy*

$$\forall i = 0, \ldots, \ell \qquad \{\Delta^i f(a)\} < \delta \qquad and \qquad \Delta^{\ell+1} f(a) < \delta$$

then $\Delta^\ell g$ is constant in the interval $[a, a+N]$, where $g(m) = \lfloor f(m) \rfloor$.

Proof. Since $\Delta^i g$ is always an integer it suffices to show that $\left| \Delta^\ell g(m+1) - \Delta^\ell g(m) \right| < 1$ for every $m \in [a, a+N)$. Then we use Lemma 2.14 and the observation that $\{x + y\} \leqslant \{x\} + \{y\}$ to deduce that

$$\left| \Delta^\ell g(m+1) - \Delta^\ell g(m) \right|$$

$$\leqslant \Delta^{\ell+1} f(m) + \sum_{t=0}^{\ell} \binom{\ell}{t} \sum_{s=0}^{t} \binom{t}{s} \left(2\{\Delta^s f(m)\} + \{\Delta^{s+1} f(m)\} \right)$$

$$\leqslant \Delta^{\ell+1} f(m) + 3^{\ell+1} \max_{i \geqslant 0} \{\Delta^i f(m)\}.$$

Finally notice that in view of Lemma 2.11 we have

$$\{\Delta^i f(m)\} = \left\{ \sum_{j=0}^{m-a} \binom{m-a}{j} \Delta^{i+j} f(a) \right\} \leqslant 2^N \delta$$

so we just need to choose δ so that $(1 + 3^{\ell+1} 2^N)\delta < 1$. $\square$

We can now prove the main theorem of this subsection.

Theorem 3.7. *Let $f \in \mathcal{F}_{\ell+1}$, let $g(n) = \lfloor f(n) \rfloor$, let $N \in \mathbb{N}$, $\varepsilon > 0$ and let $F \subset \mathbb{T}$ be a finite set. Then there exists a set $B \subset \mathbb{N}$ with positive upper Banach density and for every $s \in B$ there exists $a \in \mathbb{N}$ such that for every $n \in \{0, 1, \ldots, N\}$*
(1) $\|g(a+n)x\|_{\mathbb{T}} < \varepsilon$ for all $x \in F$,

(2) $\Delta^\ell g(a+n) = s$.

Proof. Let $N \in \mathbb{N}$, $\varepsilon > 0$ and $F = \{x_1, \ldots, x_d\} \subset \mathbb{T}$ be given and define $\alpha := (x_1, \ldots, x_d)$. Let δ_0 be given by Lemma 3.6, set $\delta = \min(\delta_0, \frac{\varepsilon}{(\ell+1)N^\ell})$ and let $D \subset \mathbb{N}$ be defined as

$$(3.3) \qquad D := \left\{ a \in \mathbb{N} : \left\| \Delta^i g(a)\alpha \right\|_{\mathbb{T}^d} \leqslant \delta; \ \{\Delta^i f(a)\} < \delta \text{ for all } i \in \{0, 1, \ldots, \ell\} \right\}.$$

According to Lemma 3.4 the set D has positive d_W density (where $W = \Delta^k f$). By chopping off a finite subset of D if necessary, we can assume that $\Delta^{\ell+1} f(a) < \delta$ for every $a \in D$ (keeping in mind that removing a finite set does not affect the density $d_W(D)$). Define $B := \Delta^\ell g(D)$; from Lemma 3.5, it follows that B has positive upper density.

Next let $s \in B$ be arbitrary. Let $a \in D$ be such that $\Delta^\ell g(a) = s$. We claim that conditions (1) and (2) are satisfied. In view of Lemma 3.6 we have that $\Delta^\ell g(m) = s$ for every $m \in [a, a + N]$, establishing condition (2).

Observe that condition (2) also implies that $\Delta^{\ell+1} g(m) = 0$ on $[a, a + N]$. Using Lemma 2.11, we get that

$$g(a + h) = \sum_{i=0}^{h} \binom{h}{i} \Delta^i g(a) = \sum_{i=0}^{\min(h,\ell)} \binom{h}{i} \Delta^i g(a).$$

Therefore, for all $n \in \{0, \ldots, N\}$, we obtain

$$
\begin{aligned}
\max_{x \in F} \|g(a+n)x\|_{\mathbb{T}} \ &\leqslant \ \|g(a+n)\alpha\|_{\mathbb{T}^d} \\
&\leqslant \ \sum_{i=0}^{\ell} \binom{n}{i} \left\| \Delta^i g(a)\alpha \right\|_{\mathbb{T}^d} \\
&< \ \sum_{i=0}^{\ell} N^\ell \left\| \Delta^i g(a)\alpha \right\|_{\mathbb{T}^d}.
\end{aligned}
$$

Finally, using (3.3) and the fact that $a \in D$, we deduce that

$$\sum_{i=0}^{\ell} N^\ell \left\| \Delta^i g(a)\alpha \right\|_{\mathbb{T}^d} \leqslant \varepsilon.$$

This finishes the proof. $\qquad \square$

3.3. Proof of Theorem A

We now combine the main results of the two previous subsections to prove Theorem A.

Lemma 3.8. *Let $f \in \mathcal{F}_{\ell+1}$, let $g(n) = \lfloor f(n) \rfloor$ and $F \subset \mathbb{T}$ be a finite set. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $h_{wm} \in L^2(X, \mathcal{B}, \mu)$ be a weakly mixing function. Then for any $\delta > 0$ and any sufficiently large $N \in \mathbb{N}$ there exists $a \in \mathbb{N}$ such that*
  (1) $\|g(a+n)x\|_{\mathbb{T}} < \delta$ for all $x \in F$ and $n \in \{0, 1, \ldots, N\}$,
  (2) $\frac{1}{N} \sum_{n=1}^{N} \left| \langle T^{g(a+n)} h_{wm}, h_{wm} \rangle \right| \leqslant \delta$.

Proof. First, we apply Theorem 3.3 to find a set $D \subset \mathbb{N}$ with $d_*(D) = 1$ such that for every polynomial $p \in \mathbb{Z}[x]$ of degree ℓ with the constant $\Delta^\ell p$ in D, one has

$$\frac{1}{N} \sum_{n=1}^{N} \left| \langle T^{p(n)} h_{wm}, h_{wm} \rangle \right| \leqslant \delta.$$

Then we apply Theorem 3.7 to find a set $B \subset \mathbb{N}$ with positive upper Banach density such that for any $s \in B$ there exists $a \in \mathbb{N}$ such that for every $n \in \{0, 1, \ldots, N\}$ we have $\sup_{x \in F} \|g(a+n)x\|_{\mathbb{T}} < \delta$ and $\Delta^\ell g(a+n) = s$.

Note that $D \cap B \neq \emptyset$ and let s be in this intersection. Since $\Delta^\ell g(a+n) = s$ for all $n \in [0, N]$, we can use Lemma 2.11 to deduce that for all n in that interval $g(a+n) = p(n)$ for some polynomial p of degree ℓ and with $\Delta^\ell p(n) = s$. Therefore, $\frac{1}{N} \sum_{n=1}^{N} \left| \langle T^{g(a+n)} h_{wm}, h_{wm} \rangle \right| \leqslant \delta$. $\square$

We also need the following well known fact whose proof we include for completeness. Recall that a Følner sequence in $\mathbb{N}$ is a sequence $(F_N)_{N \in \mathbb{N}}$ of finite sets such that $\left| (F_N + x) \cap F_N \right| / |F_N| \to 1$ as $N \to \infty$. Given a Følner sequence $(F_N)_{N \in \mathbb{N}}$, the upper density of a set $E \subset \mathbb{N}$ relative to $(F_N)_{N \in \mathbb{N}}$ is

$$\bar{d}_{(F_N)}(E) := \limsup_{N \to \infty} \frac{|E \cap F_N|}{F_N}.$$

Lemma 3.9. *Let x_n be a sequence of non-negative real numbers and let $L > 0$. If there exists a Følner sequence $(F_N)_{N \in \mathbb{N}}$ such that*

$$\lim_{N \to \infty} \frac{1}{|F_N|} \sum_{n \in F_N} |x_n - L| = 0,$$

then for every $\varepsilon > 0$ the set $E := \{n : x_n > L - \varepsilon\}$ is thick.

Proof. Denote by $D := \mathbb{N} \backslash E$

$$\sum_{n \in F_N} |x_n - L| = \sum_{n \in F_N \cap E} |x_n - L| + \sum_{n \in F_N \cap D} |x_n - L| \geqslant \varepsilon |F_N \cap D|$$

Dividing by $|F_N|$ and letting $N \to \infty$ we deduce that $\bar{d}_{(F_N)}(D) = 0$ and hence $d_{(F_N)}(E) = 1$, which implies that E is thick. $\square$

The following theorem is used in the proof of Theorem A to show that the set $R_{\varepsilon, A}$ defined in (1.1) is W-syndetic. It can be viewed as an analogue of the classical von Neumann ergodic theorem.

Theorem 3.10. *Let $U \colon \mathscr{H} \to \mathscr{H}$ be a unitary operator on a Hilbert space $\mathscr{H}$ and let $P \colon \mathscr{H} \to \mathscr{H}$ denote the orthogonal projection onto the subspace of U-invariant elements in $\mathscr{H}$. Let $\ell \in \mathbb{N} \cup \{0\}$, $f \in \mathcal{F}_{\ell+1}$ and define $W := \Delta^\ell f$. Then for any $h \in \mathscr{H}$ we have*

$$UW\text{-}\lim_{n \to \infty} U^{\lfloor f(n) \rfloor} h = Ph \qquad \text{in norm.}$$

Proof. Using the spectral theorem for unitary operators, we can assume without loss of generality that $\mathscr{H} = L^2(\mathbb{T}, \nu)$ for some Borel probability measure ν on $\mathbb{T}$, and $Uh = e(x)h(x)$ for all $h \in L^2(\mathbb{T}, \nu)$, where $e(x) = e^{2\pi i x}$. Also, note that the orthogonal projection onto the subspace of U-invariant elements in $L^2(\mathbb{T}, \nu)$ is given by

$$Ph(x) = \begin{cases} h(0), & \text{if } x = 0; \\ 0, & \text{if } x \in \mathbb{T} \backslash \{0\}. \end{cases}$$

Fix $h \in L^2(\mathbb{T}, \nu)$. Then, in light of Corollary 2.17, we have for any $x \in \mathbb{T} \backslash \{0\}$,

$$UW\text{-}\lim_{n \to \infty} U^{\lfloor f(n) \rfloor} h(x) = h(x) \left(UW\text{-}\lim_{n \to \infty} e(x \lfloor f(n) \rfloor) \right) = 0.$$

On the other hand, $UW\text{-}\lim_{n \to \infty} U^{\lfloor f(n) \rfloor} h(0) = h(0)$. It now follows from the dominated convergence theorem that $UW\text{-}\lim_{n \to \infty} U^{\lfloor f(n) \rfloor} h = Ph$ in L^2-norm and the proof is completed. $\square$

We are finally in position to prove Theorem A.

Proof of Theorem A. Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system, let $A \in \mathcal{B}$ and let $\varepsilon > 0$. Also, let $\ell \in \mathbb{N} \cup \{0\}$, let $f \in \mathcal{F}_{\ell+1}$ and let $W = \Delta^\ell f$. Using

Theorem 3.10 and the ergodic decomposition we conclude that

$$(3.4) \qquad U\!\underset{n\to\infty}{W\text{-}\lim}\ \mu(A\cap T^{-\lfloor f(n)\rfloor}A)\geqslant \mu^2(A).^{11}$$

This proves that the set $R_{\varepsilon,A}$ defined by (1.1) is W-syndetic.

We now move to show that $R_{\varepsilon,A}$ is also thick. In view of the Jacobs-de Leeuw-Glicksberg Decomposition, Theorem 3.1, we can write $1_A = h_c + h_{wm} + h_\varepsilon$, where h_c, h_{wm} and h_ε are pairwise orthogonal, h_{wm} is a weakly mixing function, h_ε satisfies $\|h_\varepsilon\|_{L^2} \leqslant \varepsilon/2$ and

$$h_c = \sum_{j=1}^{d} c_j f_j$$

for some $d \in \mathbb{N}$, $c_j \in \mathbb{C}$ and eigenfunctions $f_j \in L^2(X,\mathcal{B},\mu)$ satisfying $Tf_j = e(\theta_j)f_j$ for some $\theta_j \in \mathbb{T}$. Moreover, $h_c \geqslant 0$ and $\int h_c\, \mathsf{d}\mu = \mu(A)$. In view of Cauchy-Schwarz inequality we have that $\|h_c\|_{L^2}^2 \geqslant \mu(A)^2$.

Observe that $R_{\varepsilon,A}$ contains the set

$$J := \left\{ n \in \mathbb{N} : \langle T^{g(n)}h_c, h_c \rangle + \langle T^{g(n)}h_{wm}, h_{wm} \rangle > \|h_c\|_{L^2}^2 - \tfrac{\varepsilon}{2} \right\},$$

so it suffices to show that J is thick. In view of Lemma 3.9 it suffices to find a Følner sequence $(F_N)_{N\in\mathbb{N}}$ of $\mathbb{N}$ for which

$$(3.5) \qquad \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{n\in F_N} \left| \langle T^{g(n)}h_c, h_c \rangle + \langle T^{g(n)}h_{wm}, h_{wm} \rangle - \|h_c\|_{L^2}^2 \right| = 0.$$

As a matter of fact, we will show that there exists a Følner sequence $(F_N)_{N\in\mathbb{N}}$ such that simultaneously

$$(3.6) \qquad \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{n\in F_N} \left| \langle T^{g(n)}h_c, h_c \rangle - \|h_c\|_{L^2}^2 \right| = 0,$$

$$(3.7) \qquad \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{n\in F_N} \left| \langle T^{g(n)}h_{wm}, h_{wm} \rangle \right| = 0.$$

It is obvious that (3.6) and (3.7) together imply (3.5). The existence of such a Følner sequence is equivalent to the statement that, for every $N_0 \in \mathbb{N}$ and every $\delta > 0$ there exist $a, N \in \mathbb{N}$ with $N \geqslant N_0$ such that

$$(3.8) \qquad \frac{1}{N} \sum_{n=0}^{N} \left| \langle T^{g(a+n)}h_c, h_c \rangle - \|h_c\|_{L^2}^2 \right| \leqslant \delta,$$

$$(3.9) \qquad \frac{1}{N} \sum_{n=0}^{N} \left| \langle T^{g(a+n)}h_{wm}, h_{wm} \rangle \right| \leqslant \delta.$$

Let $N_0 \in \mathbb{N}$ and $\delta > 0$ be arbitrary. Recall that $h_c = \sum_{j=1}^{d} c_j f_j$ where $f_j \in L^2(X,\mathcal{B},\mu)$ satisfy $Tf_j = e(\theta_j)f_j$. Define $F := \{\theta_1,\ldots,\theta_d\} \subset \mathbb{T}$. Applying Lemma 3.8 we find $a, N \in \mathbb{N}$ such that (3.9) is satisfied and $\|g(a+n)\theta_j\|_{\mathbb{T}} < \delta^2/(2\pi\|h_c\|_{L^2}^4)$ for all $\theta_j \in F$ and every $n \in \{0,1,\ldots,N\}$.

---

[11]The inequality (3.4) can be viewed as an analogue of the classical result due to Khintchine [32] which states that for every measure preserving system $(X,\mathcal{B},\mu,T)$ and every $A \in \mathcal{B}$, $\displaystyle\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(A\cap T^{-n}A) \geqslant \mu^2(A)$.

Since $T^{g(a+n)} f_j = e(\theta_j g(a+n)) f_j$ and since $|e(x) - 1| < 2\pi \|x\|_{\mathbb{T}}$ for every $x \in \mathbb{T}$, we deduce that $\|h_c - T^{g(a+n)} h_c\|_{L^2} \leqslant \delta / \|h_c\|_{L^2}$. It follows that

$$(3.10) \qquad \left| \langle T^{g(a+n)} h_c, h_c \rangle - \|h_c\|_{L^2}^2 \right| \leqslant \delta, \qquad \forall n \in \{1, \ldots, N\}.$$

From (3.10) it follows that (3.8) holds and this finishes the proof. $\qquad \square$

4. Multiple recurrence along sequences in $\mathcal{F}$

In this section we provide a proof of Theorem D, modulo a result concerning equidistribution on nilmanifolds (see Theorem 1.13) to be proved in Section 5. Actually, we prove a slightly stronger result (see Theorem 4.2 below), that deals with the class of sequences $\mathcal{S}(f)$ which we will presently introduce.

Definition 4.1. Let $\ell \in \mathbb{N}$ and $f \in \mathcal{F}_\ell$. Denote by $\mathcal{S}(f)$ the collection

$$\mathcal{S}(f) := \left\{ \sum_{i=0}^{k} c_i \Delta^i f : k \geqslant 0, c_0, \ldots, c_k \in \mathbb{Q} \right\}.$$

Note that $\mathcal{S}(f)$ is closed under pointwise addition and under (discrete) differentiation Δ. In view of Lemma 2.11, the set $\mathcal{S}(f)$ is also closed under the shift $n \mapsto n + 1$. Observe that every $g \in \mathcal{S}(f)$ tends either to ∞ or to 0 as $n \to \infty$. Therefore if $g, \tilde{g} \in \mathcal{S}(f)$ are such that $\lim_{n \to \infty} |g(n) - \tilde{g}(n)| = c < \infty$, then $c = 0$.

If $g \in \mathcal{S}(f)$ is such that $g(n) \to 0$ as $n \to \infty$ we say that the *degree* of g is 0. For every other $g \in \mathcal{S}(f)$ there exists a unique $d \in \mathbb{N}$ and $\tilde{g} \in \mathcal{F}_d$ such that $|g(n) - \tilde{g}(n)| \to 0$ as $n \to \infty$. We call d the *degree* of g.

Theorem 4.2. *Let $k \in \mathbb{N}$, $\ell \in \mathbb{N}$, $f \in \mathcal{F}_\ell$, $W := \Delta^{\ell-1} f$ and $f_1, \ldots, f_k \in \mathcal{S}(f)$. Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system.*

(1) For any $h_1, \ldots, h_k \in L^\infty(X, \mathcal{B}, \mu)$ the limit

$$UW\text{-}\lim_{n \to \infty} T^{\lfloor f_1(n) \rfloor} h_1 \cdot T^{\lfloor f_2(n) \rfloor} h_2 \cdots T^{\lfloor f_k(n) \rfloor} h_k$$

exists in $L^2(X, \mathcal{B}, \mu)$.

(2) For any $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$(4.1) \qquad UW\text{-}\lim_{n \to \infty} \mu\big(A \cap T^{-\lfloor f_1(n) \rfloor} A \cap T^{-\lfloor f_2(n) \rfloor} A \cap \cdots \cap T^{-\lfloor f_k(n) \rfloor} A\big) > 0.$$

The structure of Section 4 is as follows. In Subsection 4.1 we provide a summary of relevant background material. In Subsection 4.2 we prove Theorem 4.5 which essentially shows that nilfactors are characteristic for the averages appearing in Theorem 4.2. In Subsection 4.3 we use Theorem 4.5 to reduce Theorem 4.2 to nilsystems; this is the content of Theorem 4.11. We then show how Theorem 4.11 follows from an equidistribution result, Theorem 4.12. Finally, in Subsection 4.4, we further reduce Theorem 4.12 to a more streamlined statement concerning equidistribution on nilmanifolds, Theorem 1.13, which is proved in Section 5.

4.1. Background

In this subsection we collect some facts which will be used throughout Section 4.

Nilsystems: Let G be an s-step nilpotent Lie group and let $\Gamma \subset G$ be a discrete closed subgroup such that $X := G/\Gamma$ is compact. The space X is called a *nilmanifold*. A closed subset $Y \subset X$ is a *sub-nilmanifold* if there exists a closed subgroup $H \subset G$ such that $Y = \pi(H)$, where $\pi \colon G \to X$ is the natural projection.

For each $b \in G$ we define a translation on X (which we also denote by b) as follows: given $g\Gamma \in X$ we let $b \cdot g\Gamma = (bg)\Gamma$. Let us denote by μ_X the normalized Haar measure on X, i.e., the unique Borel probability measure invariant under the natural action of G on X (cf. [42]). Given

$b \in G$, the triple (X, μ_X, b) is called an *s-step nilsystem*. A measure preserving system is called an *s-step pro-nilsystem* if it is (isomorphic in the category of measure preserving systems to) an inverse limit of s-step nilsystems. In other words, $(X, \mathcal{B}, \mu, T)$ is an s-step pro-nilsystem if there exist σ-algebras $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{B}$ such that $\mathcal{B} = \bigcup \mathcal{B}_n$ and for all n, the system $(X, \mathcal{B}_n, \mu, T)$ is isomorphic to an s-step nilsystem.

Uniformity seminorms: We will make use of the uniformity seminorms introduced in [30] for ergodic systems. As was observed in [12], the ergodicity of the system is not necessary.

Given a bounded sequence $a \colon \mathbb{N} \to \mathbb{C}$ we define the *Cesàro mean* of a as

$$C\text{-}\lim_{n \to \infty} a(n) := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a(n)$$

whenever this limit exists. Note that $C\text{-}\lim$ coincides with $W\text{-}\lim$ for $W(n) = n$.

Definition 4.3. Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system. We define the *uniformity seminorms* on $L^\infty(X)$ recursively as follows.

$$\|h\|_0 = \int_X h \, \mathsf{d}\mu \qquad \text{and} \qquad \|h\|_s^{2^s} = C\text{-}\lim_{n \to \infty} \|\bar{h} \cdot T^n h\|_{s-1}^{2^{s-1}} \text{ for every } s \in \mathbb{N}.$$

The existence of the limits in this definition was established in [30] for ergodic systems and in [12, Section 2.2] in general.

The following result from [30] gives the relation between the uniformity seminorms and nil-systems.

Theorem 4.4. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system. For each $s \in \mathbb{N}$ there exists a factor $\mathcal{Z}_s \subset \mathcal{B}$ with the following properties:
  (1) The measure preserving system $(X, \mathcal{Z}_s, \mu, T)$ is an s-step pro-nilsystem;
  (2) $L^\infty(\mathcal{Z}_s) = \left\{ h \in L^\infty(\mathcal{B}) : \int_X h \cdot h' \, \mathsf{d}\mu = 0 \quad \forall h' \in L^\infty(X, \mathcal{B}, \mu) \text{ with } \|h'\|_{s+1} = 0 \right\}.$

4.2. Characteristic Factors

Here is the main theorem of this subsection.

Theorem 4.5. Let $\ell, k \in \mathbb{N}$, $f \in \mathcal{F}_\ell$ and $W = \Delta^{\ell-1} f$. For every $f_1, \ldots, f_k \in \mathcal{S}(f)$ with $\lim |f_i(n) - f_j(n)| = \lim |f_i(n)| = \infty$ whenever $i \neq j$, there exists $s \in \mathbb{N}$ such that for any measure preserving system (X, μ, T) and any functions $h_1, \ldots, h_k \in L^\infty(X)$ with the property $\|h_i\|_s = 0$ for some i, one has

$$UW\text{-}\lim_{n \to \infty} \prod_{i=1}^{k} T^{\lfloor f_i(n) \rfloor} h_i = 0,$$

in $L^2(X, \mathcal{B}, \mu)$.

The following lemma represents a useful intermediate step in obtaining Theorem 4.5.

Proposition 4.6. Let $\ell, k \in \mathbb{N}$, let $f \in \mathcal{F}_\ell$ and let $W = \Delta^{\ell-1} f$. Suppose $f_0, f_1, \ldots, f_k \in \mathcal{S}(f)$ satisfy $\lim |f_i(n) - f_j(n)| = \lim |f_i(n)| = \infty$ whenever $i \neq j$ and that the degree of f_k is greater or equal than the degree of any other f_i. Then there exists $s \in \mathbb{N}$ such that for any measure preserving system (X, μ, T), any $h_k \in L^\infty(X)$ with $\|h_k\|_s = 0$ and any bounded sequence $(a_n)_{n \in \mathbb{N}}$ in $\mathbb{C}$ we have

$$\lim_{W(N) - W(M) \to \infty} \sup_{\|h_0\|_\infty, \ldots, \|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}_{n \in [M,N]}^{W} a_n \prod_{i=0}^{k} T^{\lfloor f_i(n) \rfloor} h_i \right\|_{L^2} = 0.$$

Proof that Proposition 4.6 implies Theorem 4.5. After permuting the f_i's and h_i's we assume that $\|h_k\|_s = 0$. Also, after multiplying by a constant we will assume that $\|h_i\|_\infty \leqslant 1$ for all i.

The desired conclusion can be rephrased as follows: let $(M_t)_{t\in\mathbb{N}}$, $(N_t)_{t\in\mathbb{N}}$ be sequences in $\mathbb{N}$ satisfying $W(N_t) - W(M_t) \to \infty$. Then

$$(4.2) \qquad \lim_{t\to\infty} \left\| \mathbb{E}^W_{n\in[M_t,N_t]} \prod_{i=1}^{k} T^{\lfloor f_i(n)\rfloor} h_i \right\|_{L^2} = 0.$$

Letting $h_0^{(t)} := \overline{\mathbb{E}^W_{n\in[M_t,N_t]} \prod_{i=1}^{k} T^{\lfloor f_i(n)\rfloor} h_i}$, we can rewrite (4.2) as

$$(4.3) \qquad \lim_{t\to\infty} \int_X \mathbb{E}^W_{n\in[M_t,N_t]} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)\rfloor} h_i \, \mathrm{d}\mu = 0.$$

Let $\tilde{f} \in \mathcal{F}_{\ell+1}$ be such that $\Delta\tilde{f} = f$. Observe that $\mathcal{S}(f) \subset \mathcal{S}(\tilde{f})$. Since T preserves the measure, we have

$$\int_X \mathbb{E}^W_{n\in[M_t,N_t]} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)\rfloor} h_i \, \mathrm{d}\mu = \int_X \mathbb{E}^W_{n\in[M_t,N_t]} T^{\lfloor\tilde{f}(n)\rfloor} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)\rfloor + \lfloor\tilde{f}(n)\rfloor} h_i \, \mathrm{d}\mu.$$

Observe that $\lfloor x \rfloor + \lfloor y \rfloor = \lfloor x + y \rfloor + e$ for some $e = e(x,y) \in \{0,1\}$. Therefore we can write $\lfloor f_i(n)\rfloor + \lfloor\tilde{f}(n)\rfloor = \lfloor f_i(n) + \tilde{f}(n)\rfloor + e_{i,n}$ where $e_{i,n} \in \{0,1\}$. Let $v_n := (e_{1,n}, \ldots, e_{k,n}) \in \{0,1\}^k$ and, for each $v = (v_1, \ldots, v_k) \in \{0,1\}^k$, let $A_v = \{n \in \mathbb{N} : v_n = v\}$ be the set of n's for which $(e_{1,n}, \ldots, e_{k,n}) = v$. Since $1 = \sum_{v\in\{0,1\}^k} 1_{A_v}$, (4.3) becomes

$$\sum_{v\in\{0,1\}^k} \lim_{t\to\infty} \mathbb{E}^W_{n\in[M_t,N_t]} 1_{A_v}(n) \int_X T^{\lfloor\tilde{f}(n)\rfloor} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)+\tilde{f}(n)\rfloor} \left(T^{v_i} h_i\right) \, \mathrm{d}\mu = 0.$$

After replacing h_i with $T^{v_i} h_i$ (and using the fact that $\|h_i\|_s = \|T^{v_i} h_i\|_s$ for every $s \in \mathbb{N}$) it suffices to show that, for any fixed bounded sequence (a_n),

$$(4.4) \qquad \lim_{t\to\infty} \mathbb{E}^W_{n\in[M_t,N_t]} a_n \int_X T^{\lfloor\tilde{f}(n)\rfloor} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)+\tilde{f}(n)\rfloor} h_i \, \mathrm{d}\mu = 0.$$

Finally, observe that now all the functions $f_i(n) + \tilde{f}(n)$ have the same degree $\ell+1$, so that we satisfy all the conditions of Proposition 4.6. Using the Cauchy-Schwarz inequality we conclude that indeed

$$\left| \mathbb{E}^W_{n\in[M_t,N_t]} a_n \int_X T^{\lfloor\tilde{f}(n)\rfloor} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)+\tilde{f}(n)\rfloor} h_i \, \mathrm{d}\mu \right|$$

$$\leqslant \left\| \mathbb{E}^W_{n\in[M_t,N_t]} a_n T^{\lfloor\tilde{f}(n)\rfloor} h_0^{(t)} \prod_{i=1}^{k} T^{\lfloor f_i(n)+\tilde{f}(n)\rfloor} h_i \right\|_{L^2}$$

$$\leqslant \sup_{\|\tilde{h}_0\|_\infty, \ldots, \|\tilde{h}_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n\in[M_t,N_t]} a_n \prod_{i=0}^{k} T^{\lfloor f_i(n)+\tilde{f}(n)\rfloor} \tilde{h}_i \right\|_{L^2}$$

$$\to \ 0 \qquad \text{as } t \to \infty,$$

where in the last step we invoked Proposition 4.6 (and we set $f_0 \equiv 0$ and $\tilde{h}_k := h_k$). $\qquad\square$

Let us now turn to the proof of Proposition 4.6. We start with the following lemma whose simple proof is omitted.

Lemma 4.7. *Let $W \in \mathcal{F}_1$ and let $(N_t)_{t\in\mathbb{N}}$ and $(M_t)_{t\in\mathbb{N}}$ be sequences of positive integers with $W(N_t) - W(M_t) \to \infty$ as $t \to \infty$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ and $t_0 > 1$ such that for all $t \geqslant t_0$ and $a\colon \mathbb{N}^2 \to \mathbb{C}$ with $\sup_{(n,m)\in\mathbb{N}^2} |a(n,m)| \leqslant 1$, if*

$$\mathbb{E}^W_{n,m\in[M_t,N_t]} |a_t(n,m)|^2 \leqslant \delta$$

then

$$\mathbb{E}^W_{n,m\in[M_t,N_t]} |a_t(n,m)| \leqslant \varepsilon.$$

The following lemma is based on [5, Lemma 3.2]; see also [18].

Lemma 4.8. *Let $\{f_0, f_1, \ldots, f_k\} \subset \mathcal{F}$, let $W \in \mathcal{F}_1$ and let $s \in \mathbb{N}$. Suppose for any measure preserving system (X, μ, T) and any $h_k \in L^\infty(X)$ with $\|h_k\|_{s-1} = 0$ we have*

$$\lim_{W(N)-W(M)\to\infty} \sup_{\|h_0\|_\infty,\ldots,\|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n\in[M,N]} \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \right\|_{L^2} = 0.$$

Then for every bounded sequence $a\colon \mathbb{N} \to \mathbb{C}$, any measure preserving system (X, μ, T) and any $h_k \in L^\infty(X)$ with $\|h_k\|_s = 0$ we also have

$$\lim_{W(N)-W(M)\to\infty} \sup_{\|h_0\|_\infty,\ldots,\|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n\in[M,N]} a_n \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \right\|_{L^2} = 0.$$

Proof. Let c be an upper bound on $|a(n)|$. We have

$$\left\| \mathbb{E}^W_{n\in[M,N]} a(n) \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \right\|_{L^2}^2$$

$$= \int_X \mathbb{E}^W_{n,m\in[M,N]} a(n)\overline{a(m)} \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \cdot T^{\lfloor f_i(m)\rfloor} \bar{h}_i \, \mathsf{d}\mu$$

$$\leqslant c^2 \mathbb{E}^W_{n,m\in[M,N]} \left| \int_X \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \cdot T^{\lfloor f_i(m)\rfloor} \bar{h}_i \, \mathsf{d}\mu \right|.$$

In view of Lemma 4.7, in order to show that the limit as $W(N) - W(M) \to \infty$ of the expression in the previous line is 0 uniformly over all $h_0, \ldots, h_{k-1} \in L^\infty$ with $\|h_i\|_\infty \leqslant 1$, it suffices to show that the limit as $W(N) - W(M) \to \infty$ of $\Psi(M, N)$ is 0 uniformly over $h_0, \ldots, h_{k-1}$, where

$$\Psi(M,N) := \mathbb{E}^W_{n,m\in[M,N]} \left| \int_X \prod_{i=0}^{k} T^{\lfloor f_i(n)\rfloor} h_i \cdot T^{\lfloor f_i(m)\rfloor} \bar{h}_i \, \mathsf{d}\mu \right|^2.$$

We have

$$\Psi(M,N) = \mathbb{E}^W_{n,m\in[M,N]} \int_{X\times X} \prod_{i=0}^{k} (T\times T)^{\lfloor f_i(n)\rfloor} h_i \otimes \bar{h}_i \cdot (T\times T)^{\lfloor f_i(m)\rfloor} \bar{h}_i \otimes h_i \, \mathsf{d}(\mu\otimes\mu)$$

$$= \int_{X\times X} \left| \mathbb{E}^W_{n\in[M,N]} \prod_{i=0}^{k} (T\times T)^{\lfloor f_i(n)\rfloor} (h_i \otimes \bar{h}_i) \right|^2 \, \mathsf{d}(\mu\otimes\mu)$$

$$= \left\| \mathbb{E}^W_{n\in[M,N]} \prod_{i=0}^{k} (T\times T)^{\lfloor f_i(n)\rfloor} (h_i \otimes \bar{h}_i) \right\|_{L^2(\mu\otimes\mu)}^2.$$

A quick induction argument gives the inequality $\|h \otimes \bar{h}\|_{s-1} \leqslant \|h\|_s^2$. In particular, if $\|h_k\|_s = 0$ then $\|h_k \otimes \overline{h_k}\|_{s-1} = 0$, and so the assumptions of the lemma imply that the limit as $W(N)-$

$W(M) \to \infty$ of this last expression is 0 uniformly over $h_0 \otimes \bar{h}_0, \ldots, h_{k-1} \otimes \bar{h}_{k-1} \in L^\infty(X \times X)$ with $\|h_i \otimes \bar{h}_i\| \leqslant 1$. $\qquad\square$

We also need the following result.

Lemma 4.9 ([31, Proposition 21.7, page 327])**.** *Let* (X, μ, T) *be a measure preserving system, and* $h_k \in L^\infty(X)$ *with* $\|h_k\|_k = 0$. *Then for all* $c_0, c_1, \ldots, c_k \in \mathbb{Z}$ *with* $c_k \neq 0$ *we have*

$$\lim_{N-M\to\infty} \sup_{\|h_0\|_\infty, \ldots, \|h_{k-1}\|_\infty \leqslant 1} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} T^{c_0 n} h_0 \, T^{c_1 n} h_1 \cdot \ldots \cdot T^{c_k n} h_k \right\|_{L^2} = 0.$$

We shall prove Proposition 4.6 by an induction scheme described below. The base case of this induction is covered by the following lemma.

Lemma 4.10 (Base case)**.** *Let* $W \in \mathcal{F}_1$, *let* $f_0, f_1, \ldots, f_k \colon \mathbb{N} \to \mathbb{R}$ *be functions such that each* $f_i(n) \to 0$ *as* $n \to \infty$, *and let* $c_0, c_1, \ldots, c_k \in \mathbb{Z} \backslash \{0\}$ *be pairwise distinct. Then for every measure preserving system* (X, μ, T), *every* $h_k \in L^\infty(X)$ *with* $\|h_k\|_{k+1} = 0$, *and every bounded sequence* $(a_n)_{n \in \mathbb{N}}$ *in* $\mathbb{C}$ *we have*

$$\lim_{W(N)-W(M)\to\infty} \sup_{\|h_0\|_\infty, \ldots, \|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n \in [M,N]} a_n \prod_{i=0}^{k} T^{\lfloor c_i W(n) + f_i(n) \rfloor} h_i \right\|_{L^2} = 0.$$

Proof. In view of Corollary 2.9, the set $\{n \in \mathbb{N} : \lfloor W(n) + f_i(n) \rfloor \neq \lfloor W(n) \rfloor\}$ has zero Banach W-density, and hence

$$UW\text{-}\lim_{n\to\infty} \left\| \prod_{i=0}^{k} T^{\lfloor c_i W(n) + f_i(n) \rfloor} h_i - \prod_{i=0}^{k} T^{\lfloor c_i W(n) \rfloor} h_i \right\|_{L^2} = 0.$$

Since each c_i is an integer, iterating the fact that $\lfloor x + y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor + e$ for some $e = e(x, y) \in \{-1, 0\}$, we can write $\lfloor c_i W(n) \rfloor = c_i \lfloor W(n) \rfloor + e_{i,n}$ where $e_{i,n}$ takes only finitely many values. Thus the vectors $(e_{1,n}, \ldots, e_{k,n})$ take only finitely many values as n goes through $\mathbb{N}$. For each of those finitely many vectors $v = (v_1, \ldots, v_k)$, let A_v be the set of n's for which $(e_{1,n}, \ldots, e_{k,n}) = v$. Since $1 = \sum_v 1_{A_v}$, after replacing h_i with $T^{v_i} h_i$ (and using the fact that $\|h_k\|_s = \|T^{v_k} h_k\|_s$ for every $s \in \mathbb{N}$) it suffices to show that

$$\lim_{W(N)-W(M)\to\infty} \sup_{\|h_0\|_\infty, \ldots, \|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n \in [M,N]} 1_{A_v}(n) a_n \prod_{i=0}^{k} T^{c_i \lfloor W(n) \rfloor} h_i \right\|_{L^2} = 0.$$

for every v. In view of Lemma 4.8 it thus suffices to show that

$$\lim_{W(N)-W(M)\to\infty} \sup_{\|h_0\|_\infty, \ldots, \|h_{k-1}\|_\infty \leqslant 1} \left\| \mathbb{E}^W_{n \in [M,N]} \prod_{i=0}^{k} T^{c_i \lfloor W(n) \rfloor} h_i \right\|_{L^2} = 0.$$

whenever $\|h_k\|_k = 0$.

Let $N_m = \max\{n \in \mathbb{N} : \lfloor W(n) \rfloor = m\}$. Observe that $\sum_{n=N_{m-1}+1}^{N_m} \Delta W(n) = 1 + o_{m\to\infty}(1)$. If $N_{p-1} < M \leqslant N_p$ and $N_{q-1} < N \leqslant N_q$ then

$$\sum_{n=M}^{N} \Delta W(n) \prod_{i=0}^{k} T^{c_i \lfloor W(n) \rfloor} h_i =$$

$$= \sum_{j=p}^{q-1} \left(1 + o_{W(N)-W(M)\to\infty}(1)\right) \prod_{i=0}^{k} T^{c_i j} h_i + O_{W(N)-W(M)\to\infty}(1) \prod_{i=0}^{k} T^{c_i m} h_i$$

and so

$$\lim_{W(N)-W(M)\to\infty}\sup_{\|h_0\|_\infty,\dots,\|h_{k-1}\|_\infty\leqslant 1}\left\|\frac{1}{W(N)-W(M)}\sum_{n=M}^N \Delta W(n)\prod_{i=0}^k T^{c_i\lfloor W(n)\rfloor}h_i\right\|_{L^2}$$

$$=\lim_{p-q\to\infty}\sup_{\|h_0\|_\infty,\dots,\|h_{k-1}\|_\infty\leqslant 1}\left\|\frac{1}{p-q}\sum_{j=p}^{q-1}\prod_{i=0}^k T^{c_ij}h_i\right\|_{L^2}.$$

In view of Lemma 4.9, this limit is 0. $\qquad\square$

The induction that we will use to prove Proposition 4.6 is similar to the PET-induction scheme which was utilized in [2].

Let $f\in\mathcal{F}_\ell$. Given $\phi(n)=\beta(n)+c_1\Delta^{\ell-1}f+\cdots+c_{\ell-1}\Delta f+c_\ell f\in\mathcal{S}(F)$ (cf. Definition 4.1), the highest i for which $c_i\neq 0$ coincides with the degree of ϕ. We call c_i the *leading coefficient of* ϕ. We say that $\phi,\tilde\phi\in\mathcal{S}(f)$ are *equivalent* if they have the same degree and leading coefficient. Now fix a finite set $P\subset\mathcal{S}(f)$ and, for each $j=1,\dots,\ell$, let m_j denote the number of equivalence classes in P of degree j. The vector $(m_1,\dots,m_\ell)$ is called the *characteristic vector* of P. We order characteristic vectors by letting $(m_1,\dots,m_\ell)<(\tilde m_1,\dots,\tilde m_\ell)$ if the maximum j for which $m_j\neq\tilde m_j$ satisfies $m_j<\tilde m_j$.

Proof of Proposition 4.6. We prove the theorem by induction on the characteristic vector of $P=\{f_0,f_1,\dots,f_k\}$. The case when the characteristic vector is of the form $(m,0,\dots,0)$ (i.e., all functions have degree 1) follows from Lemma 4.10. Assume now that some function in P has degree at least 2 and that the theorem has been proved for all families whose characteristic vector is strictly smaller than that of P.

If not all functions in P have the same degree then pick $i_0\in\{0,1,\dots,k-1\}$ such that the degree of f_{i_0} is minimal among all degrees present in P. If all functions in P have the same degree and there is more than one equivalence class in P then pick $i_0\in\{0,1,\dots,k-1\}$ such that f_{i_0} is not equivalent to f_k. Finally, if there is only one equivalency class in P then we can pick $i_0\in\{0,1,\dots,k-1\}$ arbitrarily.

We now use the van der Corput lemma. Let $u_n=a_n\prod_{i=0}^k T^{\lfloor f_i(n)\rfloor}h_i$. In view of Proposition 2.5, instead of

$$\lim_{W(N)-W(M)\to\infty}\sup_{\|h_0\|_\infty,\dots,\|h_{k-1}\|_\infty\leqslant 1}\left\|\mathbb{E}_{n\in[M,N]}^W u_n\right\|_{L^2}=0,$$

it suffices to show that

$$\lim_{W(N)-W(M)\to\infty}\sup_{\|h_0\|_\infty,\dots,\|h_{k-1}\|_\infty\leqslant 1}\mathbb{E}_{n\in[M,N]}^W\langle u_{n+m},u_n\rangle=0.$$

We have

$$\langle u_{n+m},u_n\rangle = a_{n+m}\overline{a_n}\int_X\prod_{i=0}^k T^{\lfloor f_i(n+m)\rfloor}h_i\cdot T^{\lfloor f_i(n)\rfloor}\overline{h_i}\ \mathsf{d}\mu$$

$$= a_{n+m}\overline{a_n}\int_X\prod_{0\leqslant i\leqslant k} T^{\lfloor f_i(n+m)\rfloor-\lfloor f_{i_0}(n)\rfloor}h_i\cdot T^{\lfloor f_i(n)\rfloor-\lfloor f_{i_0}(n)\rfloor}\overline{h_i}\ \mathsf{d}\mu.$$

We now repeat an argument that was already used in the proof of Lemma 4.10. Since $\lfloor a\rfloor-\lfloor b\rfloor=\lfloor a-b\rfloor+e$ for some $e\in\{0,1\}$, we can write $\lfloor f_i(n+m)\rfloor-\lfloor f_{i_0}(n)\rfloor=\lfloor f_i(n+m)-f_{i_0}(n)\rfloor+e_{i,n}$ where $e_{i,n}\in\{0,1\}$ for every $i=1,\dots,k$ and $\lfloor f_i(n)\rfloor-\lfloor f_{i_0}(n)\rfloor=\lfloor f_i(n)-f_{i_0}(n)\rfloor+\tilde e_{i,n}$ where $\tilde e_{i,n}\in\{0,1\}$ for every $i=2,\dots,k$. For each vector $v=(v_1,\dots,v_k,\tilde v_2,\dots\tilde v_k)\in\{0,1\}^{2k-1}$, let A_v be the set of n's for which $(e_{1,n},\dots,e_{k,n},\tilde e_{2,n},\dots\tilde e_{k,n})=v$. Since $1=\sum_v 1_{A_v}$, we have that

$\langle u_{n+m}, u_n \rangle$ can be written as

$$\sum_v 1_{A_v}(n) a_{n+m} \overline{a_n} \int_X \prod_{0 \leqslant i \leqslant k} T^{\lfloor f_i(n+m) \rfloor - \lfloor f_{i_0}(n) \rfloor} h_i \cdot T^{\lfloor f_i(n) \rfloor - \lfloor f_{i_0}(n) \rfloor} \overline{h_i} \, \mathsf{d}\mu.$$

We will show that each summand goes to 0 as $W(N) - W(M) \to \infty$ uniformly over h_0. Fix $v \in \{0,1\}^{2k-1}$ and let

$$\tilde{f}_i(n) := \begin{cases} f_i(n) - f_{i_0}(n), & \text{if } i < i_0 \\ f_{i-1}(n) - f_{i_0}(n), & \text{if } i > i_0 \end{cases} \quad \text{and} \quad \tilde{h}_i := \begin{cases} T^{v_i} h_i, & \text{if } i < i_0 \\ T^{v_{i-1}} h_{i-1}, & \text{if } i > i_0 \end{cases},$$

$$\tilde{f}_{k+i} := f_i(n+m) - f_{i_0}(n) \quad \text{and} \quad \tilde{h}_{k+i} := T^{\tilde{v}_i} \overline{h_i} \qquad \text{for each } i \in \{0, 1, \dots, k\}.$$

Observe that $\|\tilde{h}_{2k}\|_s = \|h_k\|_s = 0$. Since f_k has the highest degree within P and due to our choice of i_0, it follows that $\tilde{f}_{2k}$ has the highest degree within $\tilde{P} := \{\tilde{f}_1, \dots, \tilde{f}_{2k}\}$. Indeed $\tilde{f}_{2k}$ either has the same degree as f_k (in the case f_{i_0} and f_k are not equivalent) or its degree is that of f_k minus 1 (in the case that there is only one equivalence class in P, which implies that every function in $\tilde{P}$ has the same degree)[12]. Moreover, since the degree of f_k is at least 2, the degree of $\tilde{f}_{2k}$ is at least 1.

It remains to show that

$$\lim_{W(N)-W(M) \to \infty} \sup_{\|h_0\|_\infty, \dots, \|h_{k-1}\|_\infty \leqslant 1} \mathbb{E}^W_{n \in [M,N]} 1_{A_v}(n) a_{n+m} \overline{a_n} \int_X \prod_{i=0}^{2k} T^{\lfloor \tilde{f}_i(n) \rfloor} \tilde{h}_i \, \mathsf{d}\mu = 0$$

for all v and m. Replacing $\sup_{\|h_0\|_\infty, \dots, \|h_{k-1}\|_\infty \leqslant 1}$ with $\sup_{\|h_0\|_\infty, \dots, \|h_{2k-1}\|_\infty \leqslant 1}$, replacing the convergence in measure with norm convergence, and using Lemma 4.8, it thus suffices to prove that

$$(4.5) \qquad \lim_{W(N)-W(M) \to \infty} \sup_{\|h_0\|_\infty, \dots, \|h_{2k-1}\|_\infty \leqslant 1} \mathbb{E}^W_{n \in [M,N]} \prod_{i=0}^{2k} T^{\lfloor \tilde{f}_i(n) \rfloor} \tilde{h}_i = 0$$

whenever $\tilde{h}_{2k} \in L^\infty(X)$ with $\|\tilde{h}_{2k}\|_{s-1} = 0$. If some function $\tilde{f}_i$ in $\tilde{P}$ has degree 0, then $\lim_{n \to \infty} \tilde{f}_i(n) = 0$ and so we can remove it from $\tilde{P}$ and ignore the corresponding term $T^{\lfloor \tilde{f}_i(n) \rfloor} \tilde{h}_i$ in (4.5). We can therefore assume that $\tilde{P} \subset \mathcal{S}(f)$. Thus (4.5) will follow by induction after we show that the characteristic vector $(\tilde{m}_1, \dots, \tilde{m}_\ell)$ of $\tilde{P}$ is strictly smaller than the characteristic vector $(m_1, \dots, m_\ell)$ of P.

Indeed, for each f_i which is not equivalent to f_{i_0}, the functions $f_i(n) - f_{i_0}(n)$ and $f_i(n+m) - f_{i_0}(n)$ are equivalent to each other, and have the same degree as f_i. Moreover, if f_i and f_j are not equivalent to f_{i_0}, then $f_i - f_{i_0}$ is equivalent to $f_j - f_{i_0}$ if and only if f_i is equivalent to f_j. Letting d be the degree of f_{i_0}, this shows that $\tilde{m}_j = m_j$ for all $j > d$. Finally, if f_i is equivalent to f_{i_0}, then both $f_i(n) - f_{i_0}$ and $f_i(n+m) - f_{i_0}$ (when $i \neq i_0$) have degree smaller than that of f_i. This shows that $\tilde{m}_d < m_d$.

Therefore $(\tilde{m}_1, \dots, \tilde{m}_\ell)$ is strictly smaller than $(m_1, \dots, m_\ell)$ and by induction (4.5) holds. $\qquad \square$

4.3. Reduction to an equidistribution result

In this subsection we prove Theorem 4.2 modulo a result about equidistribution of certain sequences in nilmanifolds. Let $k, \ell \in \mathbb{N}$, let $f \in \mathcal{F}_\ell$, let $W = \Delta^{\ell-1} f$, let $f_1, \dots, f_k \in \mathcal{S}(f)$, let (X, μ, T) be an invertible probability measure preserving system and let $A \subset X$ have positive

---

[12]For some values of m the degree of $\tilde{f}_k$ may be smaller than the degree of f_k minus one; for instance if $f_{i_0}(n) = f_k(n+1)$ and $m = 1$. Nevertheless, for all but finitely many values of m the degree of $\tilde{f}_k$ is precisely one less than the degree of f_k.

measure. We need to show that

$$(4.6) \qquad UW\text{-}\lim_{n\to\infty} \mu\big(A \cap T^{\lfloor f_1(n)\rfloor}A \cap T^{\lfloor f_2(n)\rfloor}A \cap \cdots \cap T^{\lfloor f_k(n)\rfloor}A\big) > 0.$$

and that for any $h_1, \ldots, h_k \in L^\infty(X, \mathcal{B}, \mu)$ the limit

$$(4.7) \qquad UW\text{-}\lim_{n\to\infty} T^{\lfloor f_1(n)\rfloor}h_1 \cdot T^{\lfloor f_2(n)\rfloor}h_2 \cdots T^{\lfloor f_k(n)\rfloor}h_k$$

exists in $L^2(X, \mathcal{B}, \mu)$.

We start by making several standard reductions whose details are largely omitted (cf. [8, 18] where similar reductions were performed). First, employing the ergodic decomposition (see, for instance, [15, Theorem 4.8]) we may assume that (X, μ, T) is ergodic. Indeed, assuming that Theorem 4.2 holds for each ergodic component of μ it follows that (4.6) holds (cf. [15, Section 7.2.3]), and that (4.7) holds (cf. the last paragraph of [31, Section 21.2.3]). Next, let $s \in \mathbb{N}$ be given by Theorem 4.5 and let h be the projection of 1_A onto the $(s-1)$-step nilfactor $\mathcal{Z}_{s-1}$. In view of Theorem 4.5 and Theorem 4.4, the left hand side of (4.6) is the same as

$$(4.8) \qquad UW\text{-}\lim_{n\to\infty} \int_X h \cdot T^{\lfloor f_1(n)\rfloor}h \cdot T^{\lfloor f_2(n)\rfloor}h \cdots T^{\lfloor f_k(n)\rfloor}h \; \mathsf{d}\mu.$$

Similarly, letting $\tilde{h}_i$ be the projection of h_i onto the $(s-1)$-step nilfactor, the limit in (4.7) becomes

$$(4.9) \qquad UW\text{-}\lim_{n\to\infty} T^{\lfloor f_1(n)\rfloor}\tilde{h}_1 \cdot T^{\lfloor f_2(n)\rfloor}\tilde{h}_2 \cdots T^{\lfloor f_k(n)\rfloor}\tilde{h}_k.$$

The s-step nilfactor is an inverse limit of s-step nilsystems. A standard approximation argument (see [24, Lemma 3.2]) shows that it suffices to establish positivity of (4.8) and show the limit in (4.9) exists under the additional assumptions that the system is a nilsystem and that $h, \tilde{h}_1, \ldots, \tilde{h}_k$ are continuous functions with $h \geqslant 0$ and $\int_X h \; \mathsf{d}\mu = \mu(A) > 0$.

We have now reduced Theorem 4.2 to the following statement.

Theorem 4.11. *Let G be a nilpotent Lie group, let $\Gamma \subset G$ be a co-compact discrete subgroup, let $X = G/\Gamma$ and let $b \in G$. Let $\ell, k \in \mathbb{N}$, let $f \in \mathcal{F}_{\ell+1}$, let $W := \Delta^\ell f$ and let $f_1, \ldots, f_k \in \mathcal{S}(f)$.*

(1) *For any $h_1, \ldots, h_k \in C(X)$ and $x \in X$ the limit*

$$UW\text{-}\lim_{n\to\infty} h_1\big(b^{\lfloor f_1(n)\rfloor}x\big) \cdot h_2\big(b^{\lfloor f_2(n)\rfloor}x\big) \cdots h_k\big(b^{\lfloor f_k(n)\rfloor}x\big)$$

   exists.

(2) *For every non-negative function $h \in C(X)$ with $\int_X h \; \mathsf{d}\mu_X > 0$*

$$UW\text{-}\lim_{n\to\infty} \int_X h(x) \cdot h\big(b^{\lfloor f_1(n)\rfloor}x\big) \cdot h\big(b^{\lfloor f_2(n)\rfloor}x\big) \cdots h\big(b^{\lfloor f_k(n)\rfloor}x\big) \; \mathsf{d}\mu(x) > 0.$$

In order to prove Theorem 4.11 we will need the following result, which is proved in the next subsection.

Theorem 4.12. *Let G be a nilpotent Lie group, let $\Gamma \subset G$ be a co-compact discrete subgroup and let $X = G/\Gamma$. Let $k, \ell \in \mathbb{N}$, let $f \in \mathcal{F}_{\ell+1}$, let $W = \Delta^\ell f$ and let $f_1, \ldots, f_k \in \mathcal{S}(f)$. Then there exists a constant $C > 0$ and for every $b \in G$ there exists a measure ν on X^k such that the sequence*

$$(4.10) \qquad n \mapsto \Big(b^{\lfloor f_1(n)\rfloor}, b^{\lfloor f_2(n)\rfloor}, \ldots, b^{\lfloor f_k(n)\rfloor}\Big)\Gamma^k$$

is ν-w.d. with respect to W-lim. Moreover, there are linear maps $\phi_1, \ldots, \phi_k \colon \mathbb{Z}^{\ell+1} \to \mathbb{Z}$ such that the Haar measure μ_Y of the subnilmanifold[13]

$$Y := \overline{\left\{ \left(b^{\phi_1(\mathbf{m})}\Gamma, \ldots, b^{\phi_k(\mathbf{m})}\Gamma \right) : \mathbf{m} \in \mathbb{Z}^{\ell+1} \right\}}$$

satisfies $\nu \geqslant C\mu_Y$.

Proof of Theorem 4.11 conditionally on Theorem 4.12. For each $x \in X$, let $g_x \in G$ be such that $x = g_x\Gamma$. We have

$$\prod_{i=1}^{k} h_i\big(b^{\lfloor f_i(n)\rfloor}x\big) = \prod_{i=1}^{k} h_i\Big(g_x\big(g_x^{-1}bg_x\big)^{\lfloor f_i(n)\rfloor}\Gamma \Big).$$

Now let C be the constant given by Theorem 4.12 and let ν_x be the measure on X^k given by the same theorem with $g_x^{-1}bg_x$ in place of b, noting crucially that C does not depend on x. Finally, let $H_x(x_1, x_2, \ldots, x_k) = h_1(g_x x_1)h_2(g_x x_2)\cdots h_k(g_x x_k)$ We then have

$$(4.11) \qquad W\text{-}\lim_{n\to\infty} \prod_{i=1}^{k} h_i\big(b^{\lfloor f_i(n)\rfloor}x\big) = \int_{X^k} H_x \, \mathsf{d}\nu_x$$

and in particular the limit exists, proving part (1).

Let now $h_1 = \cdots = h_k = h$ be a non-negative function with $\int_X h \, \mathsf{d}\mu_X > 0$. Let

$$Y_x = \overline{\left\{ \left((g_x^{-1}bg_x)^{\phi_1(\mathbf{m})}\Gamma, \ldots, (g_x^{-1}bg_x)^{\phi_k(\mathbf{m})}\Gamma \right) : \mathbf{m} \in \mathbb{Z}^{\ell+1} \right\}}$$

and let μ_{Y_x} be the Haar measure on Y_x. Thus Theorem 4.12 and (4.11) imply that

$$(4.12) \qquad W\text{-}\lim_{n\to\infty} \prod_{i=1}^{k} h\big(b^{\lfloor f_i(n)\rfloor}x\big) \geqslant C \int_{Y_x} H_x \, \mathsf{d}\mu_{Y_x}.$$

Observe that if $h(x) > 0$, then $H_x(1_{G^k}\Gamma^k) = h(x)^k > 0$. Since H_x is a continuous function and $1_{G^k}\Gamma^k \in Y_x$ for every $x \in X$, it follows that whenever $h(x) > 0$, also $\int_{Y_x} H_x \, \mathsf{d}\mu_{Y_x} > 0$. Now let $\tilde{X} \subset X$ be the set of points x for which $h(x) > 0$. Since $\int_X h \, \mathsf{d}\mu_X > 0$ we have that $\mu_X(\tilde{X}) > 0$. Let $I \colon \tilde{X} \to \mathbb{R}$ be the function defined by

$$I(x) := h(x) \int_{Y_x} H_x \, \mathsf{d}\mu_{Y_x}.$$

Since $I(x) > 0$ for every $x \in \tilde{X}$ we have $\int_{\tilde{X}} I(x) \, \mathsf{d}\mu_X(x) > 0$. Combining (4.12) with the dominated convergence theorem we conclude that

$$UW\text{-}\lim_{n\to\infty} \int_X h(x) \cdot \prod_{i=1}^{k} h\big(b^{\lfloor f_i(n)\rfloor}x\big) \, \mathsf{d}\mu(x) \geqslant \int_X I(x) \, \mathsf{d}\mu_X(x) > \int_{\tilde{X}} I(x) \, \mathsf{d}\mu_X(x) > 0.$$

$\square$

4.4. An additional reduction

Given a connected and simply connected nilpotent Lie group G, for each $b \in G$ and $t \in \mathbb{R}$, the element b^t is well defined in G. In particular, for every discrete and co-compact subgroup $\Gamma \subset G$, every nilrotation on G/Γ is embedable in a flow. Under these assumptions we are able to derive results about sequences of the form $b^{\lfloor f(n)\rfloor}$ from results about (simpler) sequences $b^{f(n)}$.

In this subsection we derive Theorem 4.12 from Theorem 1.13, which is proved in Section 5 and whose statement we now recall.

---

[13]The fact that Y is a subnilmanifold of X follows from [36] or [46].

Theorem 1.13. *Let G be a connected and simply connected nilpotent Lie group, let $\Gamma \subset G$ be a co-compact discrete subgroup and let $X = G/\Gamma$. Let $k, \ell \in \mathbb{N}$ with $k \leqslant \ell$, let $f \in \mathcal{F}_\ell$, let $W = \Delta^{\ell-1} f$, let $b_0, \ldots, b_{k-1} \in G$ be commuting elements, let*

$$Y = \overline{\left\{ b_0^{t_0} \cdots b_{k-1}^{t_{k-1}} \Gamma : t_0, \ldots, t_{k-1} \in \mathbb{R} \right\}} \subset X,$$

and let μ_Y be the normalized Haar measure on Y. Then for every continuous function $H \in C(X)$ we have

$$UW\text{-}\lim_{n \to \infty} H\left(b_0^{f(n)} b_1^{\Delta f(n)} \cdots b_{k-1}^{\Delta^{k-1} f(n)} \Gamma \right) = \int_Y H(y) \, \mathsf{d}\mu_Y(y).$$

To achieve our goal we actually need the following corollary of Theorem 1.13.

Corollary 4.13. *Let $k, \ell \in \mathbb{N}$, let $f \in \mathcal{F}_{\ell+1}$, let $W = \Delta^\ell f$ and let $f_1, \ldots, f_k \in \mathcal{S}(f)$. Then there are linear maps $\phi_1, \ldots, \phi_k \colon \mathbb{R}^{\ell+1} \to \mathbb{R}$ with integer coefficients which satisfy the following property. Let G be a connected and simply connected nilpotent Lie group, let $\Gamma \subset G$ be a co-compact discrete subgroup, let $b \in G$, $X = G/\Gamma$ and let $H \in C(X^k)$.*
   Then we have

$$(4.13) \qquad UW\text{-}\lim_{n \to \infty} H\left(\left(b^{f_1(n)}, b^{f_2(n)}, \ldots, b^{f_k(n)} \right) \Gamma^k \right) = \int_Y H(y) \, \mathsf{d}\mu_Y,$$

where μ_Y is the normalized Haar measure on the subnilmanifold $Y \subset X^k$ defined as

$$Y := \overline{\left\{ \left(b^{\phi_1(\mathbf{t})} \Gamma, \ldots, b^{\phi_k(\mathbf{t})} \Gamma \right) : \mathbf{t} \in \mathbb{R}^{\ell+1} \right\}}.$$

Proof. For each $j = 1, \ldots, k$, write

$$f_j(n) = c_{j,0} f(n) + c_{j,1} \Delta f(n) + \cdots + c_{j,\ell} \Delta^\ell f(n) + \beta_j(n),$$

where $\beta_j(n) \to 0$ as $n \to \infty$ and $(c_{j,0}, \ldots, c_{j,\ell}) \in \mathbb{Q}^{\ell+1} \backslash \{\mathbf{0}\}$. Multiplying f by a common multiple of the denominators of all the $c_{j,i}$ if necessary we can and will assume that all the $c_{j,i}$ are integers.

For each $i = 0, \ldots, \ell$ let $b_i \in G^k$ be the point $b_i = \left(b^{c_{1,i}}, b^{c_{2,i}}, \ldots, b^{c_{k,i}} \right)$. Also let $g(n) = (b^{\beta_1(n)}, \ldots, b^{\beta_k(n)})$. We have

$$(4.14) \qquad \left(b^{f_1(n)}, b^{f_2(n)}, \ldots, b^{f_k(n)} \right) = b_0^{f(n)} b_1^{\Delta f(n)} \cdots b_\ell^{\Delta^\ell f(n)} g(n).$$

Since each $\beta_j(n) \to 0$ as $n \to \infty$ it follows that $g(n) \to 1_G$ as $n \to \infty$. Since X is compact, H is uniformly continuous and hence

$$\lim_{n \to \infty} \left| H\left(\left(b^{f_1(n)}, b^{f_2(n)}, \ldots, b^{f_k(n)} \right) \Gamma^k \right) - H\left(b_0^{f(n)} b_1^{\Delta f(n)} \cdots b_\ell^{\Delta^\ell f(n)} \Gamma^k \right) \right| = 0.$$

For each $j = 1, \ldots, k$ and $\mathbf{t} = (t_0, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}$, define $\phi_j(\mathbf{t}) = c_{j,0} t_0 + \cdots + c_{j,\ell} t_\ell$. Notice that $b_0^{t_0} \cdots b_\ell^{t_\ell} = \left(b^{\phi_1(\mathbf{t})} \Gamma, \ldots, b^{\phi_k(\mathbf{t})} \right)$. Appealing to Theorem 1.13 it follows that (4.13) holds with

$$Y = \overline{\left\{ b_0^{t_0} \cdots b_\ell^{t_\ell} \Gamma^k : \mathbf{t} \in \mathbb{R}^{\ell+1} \right\}} = \overline{\left\{ \left(b^{\phi_1(\mathbf{t})} \Gamma, \ldots, b^{\phi_k(\mathbf{t})} \Gamma \right) : \mathbf{t} \in \mathbb{R}^{\ell+1} \right\}} \subset X^k.$$

$\square$

We are now ready to prove Theorem 4.12.

Proof of Theorem 4.12. Every nilpotent Lie group G is the quotient of a simply connected nilpotent Lie group $\tilde{G}$ by a discrete subgroup. In turn every simply connected nilpotent Lie group is (isomorphic as a Lie group to) a closed subgroup of a connected and simply connected nilpotent Lie group $\hat{G}$ (cf. [36, Subsection 1.11]). Therefore it suffices to prove the theorem assuming G is connected and simply connected.

Let $\tilde{G} = G \times \mathbb{R}$, let $\tilde{\Gamma} = \Gamma \times \mathbb{Z}$, let $\tilde{X} = \tilde{G}/\tilde{\Gamma}$ and let $a = (b, 1) \in \tilde{G}$. Recall the notation $\{t\} = t - \lfloor t \rfloor$ and let $\pi \colon \tilde{X}^k \to X^k$ be the map

$$\pi\big((g_1, t_1, \ldots, g_k, t_k)\tilde{\Gamma}^k\big) = (b^{-\{t_1\}}g_1, \ldots, b^{-\{t_k\}}g_k)\Gamma^K.$$

Observe that π is well defined (i.e., the choice of the co-set representative does not matter in the definition of π) and that $\pi(a^{t_1}, \ldots, a^{t_k}) = (b^{\lfloor t_1 \rfloor}, \ldots, b^{\lfloor t_k \rfloor})$ for every $t_1, \ldots, t_k \in \mathbb{R}$. We warn the reader that π is not a continuous map; indeed π is discontinuous at the points $(g_1, t_1, \ldots, g_k, t_k)\tilde{\Gamma}^k$ where at least one $t_i \in \mathbb{Z}$, but continuous elsewhere. In particular π is continuous almost everywhere (with respect to the Haar measure on $\tilde{X}^k$).

Let $\phi_1, \ldots, \phi_k$ be given by Corollary 4.13, let

$$\tilde{Y} := \overline{\left\{ \big(a^{\phi_1(\mathbf{t})}, \ldots, a^{\phi_k(\mathbf{t})}\big)\tilde{\Gamma}^k : \mathbf{t} \in \mathbb{R}^{\ell+1} \right\}},$$

let $\mu_{\tilde{Y}}$ be the Haar measure on $\tilde{Y}$ and let $\nu := \pi_* \mu_{\tilde{Y}}$ be the pushforward measure. Notice that for any $H \in C(X^k)$,

$$H\left(\big(b^{\lfloor f_1(n) \rfloor}, b^{\lfloor f_2(n) \rfloor}, \ldots, b^{\lfloor f_k(n) \rfloor}\big)\Gamma^k \right) = H \circ \pi\left(\big(a^{f_1(n)}, a^{f_2(n)}, \ldots, a^{f_k(n)}\big)\tilde{\Gamma}^k \right)$$

so in view of Corollary 4.13, and using the fact that π is continuous $\mu_{\tilde{Y}}$-a.e., it follows that the sequence defined in (4.10) is indeed ν-w.d. with respect to $W\text{-}\lim$.

Finally we show that $\nu \geqslant C\mu_Y$ for some constant C. Let $H \in C(X^k)$, assume $H(x) \geqslant 0$ for every $x \in X^k$. In view of [35, Theorem B]

$$\int_{X^k} H(x) \, \mathsf{d}\mu_Y(x) \;=\; \lim_{N \to \infty} \frac{1}{N^{\ell+1}} \sum_{\mathbf{n} \in [0,N]^{\ell+1}} H\left(\big(b^{\phi_1(\mathbf{n})}, \ldots, b^{\phi_k(\mathbf{n})}\big)\Gamma^k \right)$$

and similarly

$$
\begin{aligned}
\int_{X^k} H(x) \, \mathsf{d}\nu(x) \;&=\; \int_{\tilde{Y}} H \circ \pi(\tilde{y}) \, \mathsf{d}\mu_{\tilde{Y}}(\tilde{y}) \\
&=\; \lim_{T \to \infty} \frac{1}{T^{\ell+1}} \int_{[0,T]^{\ell+1}} H \circ \pi\left(\big(a^{\phi_1(\mathbf{t})}, \ldots, a^{\phi_k(\mathbf{t})}\big)\tilde{\Gamma}^k \right) \mathsf{d}\mathbf{t} \\
&=\; \lim_{T \to \infty} \frac{1}{T^{\ell+1}} \int_{[0,T]^{\ell+1}} H\left(\big(b^{\lfloor \phi_1(\mathbf{t}) \rfloor}, \ldots, b^{\lfloor \phi_k(\mathbf{t}) \rfloor}\big)\Gamma^k \right) \mathsf{d}\mathbf{t} \\
&=\; \lim_{N \to \infty} \frac{1}{N^{\ell+1}} \sum_{\mathbf{n} \in [0,N)^{\ell+1}} \int_{[0,1]^{\ell+1}} H\left(\big(b^{\phi_1(\mathbf{n}) + \lfloor \phi_1(\mathbf{t}) \rfloor}, \ldots, \phi^{\phi_k(\mathbf{n}) + \lfloor \phi_k(\mathbf{t}) \rfloor}\big)\Gamma^k \right) \mathsf{d}\mathbf{t} \\
&\geqslant\; C \lim_{N \to \infty} \frac{1}{N^{\ell+1}} \sum_{\mathbf{n} \in [0,N)^{\ell+1}} H\left(\big(b^{\phi_1(\mathbf{n})}, \ldots, b^{\phi_k(\mathbf{n})}\big)\Gamma^k \right) \\
&=\; C \int_{X^k} H(x) \, \mathsf{d}\mu_Y(x),
\end{aligned}
$$

where $C > 0$ is the Lebesgue measure of the set

$$\left\{ \mathbf{t} \in [0,1]^{\ell+1} : \lfloor \phi_1(\mathbf{t}) \rfloor = \lfloor \phi_2(\mathbf{t}) \rfloor = \cdots = \lfloor \phi_k(\mathbf{t}) \rfloor = 0 \right\}.$$

$\square$

5. Well-distribution with respect to Riesz means on nilmanifolds

In this section we prove Theorem 1.13. Results of similar nature were obtained by Frantzik-inakis in [17] (see Remark 1.8 in the Introduction).

Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and $X := G/\Gamma$. We use $\pi \colon G \to X$ to denote the natural projection of G onto X.

Definition 5.1. Let $m \in \mathbb{N}$. A map $g \colon \mathbb{Z}^m \to G$ is called a *polynomial sequence* if there exist $u \in \mathbb{N}$, $a_1, \ldots, a_u \in G$ and $p_1, \ldots, p_u \in \mathbb{R}[x_1, \ldots, x_m]$ such that

$$g(h) = a_1^{p_1(h)} a_2^{p_2(h)} \cdot \ldots \cdot a_u^{p_u(h)}, \qquad \forall h \in \mathbb{Z}^m.$$

It will also be convenient to introduce the following notation: Given a bounded complex-valued sequence $a \colon \mathbb{N} \to \mathbb{C}$ we define

$$C\text{-}\overline{\lim_{n \to \infty}}\, a(n) := \limsup_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} a(n) \right|$$

and

$$UW\text{-}\overline{\lim_{n \to \infty}}\, a(n) := \sup_{(M_i, N_i)_{i \in \mathbb{N}}} \lim_{i \to \infty} \left| \frac{1}{W(N_i) - W(M_i)} \sum_{n=M_i}^{N_i} \Delta W(n) a(n) \right|,$$

where the supremum is taken of all sequences $(M_i, N_i)_{i \in \mathbb{N}} \subset \mathbb{N}^2$ satsifying $M_i \leqslant N_i$ and $\lim_{i \to \infty} W(N_i) - W(M_i) = \infty$.

We can now state the main theorem of this section. Notice that Theorem 1.13 corresponds to the special case $m = 0$.

Theorem 5.2. *Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and define $X := G/\Gamma$. Let $\ell, u \in \mathbb{N}$, let $m \in \mathbb{Z}$ with $m \geqslant 0$, let $f \in \mathcal{F}_\ell$ and let $g_0, g_1, \ldots, g_{\ell-1} \colon \mathbb{Z}^m \to G$ be polynomial sequences defined by*

$$g_i(h) := a_{i,1}^{p_1(h)} a_{i,2}^{p_2(h)} \cdot \ldots \cdot a_{i,u}^{p_{i,u}(h)}, \qquad \forall h \in \mathbb{Z}^m,$$

where $a_{i,j}$ with $(i, j) \in \{0, 1, \ldots, \ell-1\} \times \{1, \ldots, u\}$ is a collection of commuting elements of G and $p_{i,j} \in \mathbb{R}[x_1, \ldots x_m]$. If $m = 0$ then simply let g_i be commuting elements of G for $i = 0, \ldots, \ell-1$. Set

$$g(n, h) := g_0(h)^{f(n)} g_1(h)^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta^{\ell-1} f(n)}, \qquad \forall n \in \mathbb{N}, \forall h \in \mathbb{Z}^m.$$

Then for all $F \in C(X)$ we have

$$C\text{-}\overline{\lim_{h_m \to \infty}} \cdots C\text{-}\overline{\lim_{h_1 \to \infty}}\, UW\text{-}\overline{\lim_{n \to \infty}} \left(F\big(g(n, h)\Gamma\big) - \int_Y F \, \mathsf{d}\mu_Y \right) = 0,$$

where $h = (h_1, \ldots, h_m)$, $W := \Delta^{\ell-1} f$, and μ_Y is the Haar measure of the connected subnilmanifold Y[14] defined by

$$Y := \overline{\{ g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-1}(h)^{t_{\ell-1}} \Gamma : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell \}}.$$

5.1. The abelian case of Theorem 5.2

Before we proceed to give a proof of Theorem 5.2 in its full generality, let us establish the following special case.

Lemma 5.3 (Theorem 5.2 for the special case $G = \mathbb{R}^d$). *Let $d, \ell \in \mathbb{N}$, let $m \in \mathbb{Z}$ with $m \geqslant 0$, let $f \in \mathcal{F}_\ell$, let $p_0, \ldots, p_{\ell-1} \colon \mathbb{Z}^m \to \mathbb{R}^d$ be polynomials (or just constants if $m = 0$) and set*

$$p(n, h) := p_0(h) f(n) + p_1(h) \Delta f(n) + \ldots + p_{\ell-1}(h) \Delta^{\ell-1} f(n), \qquad \forall n \in \mathbb{N}, \forall h \in \mathbb{Z}^m.$$

---

[14]It is shown in Lemma A.8 that Y is indeed a connected subnilmanifold.

Let $\pi_1 \colon \mathbb{R}^d \to \mathbb{T}^d$ denote the natural projection from $\mathbb{R}^d$ onto $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ and define $W :=
\Delta^{\ell-1}f$. Let Y denote the connected subtorus of $\mathbb{T}^d$ given by

$$Y := \overline{\left\{ \pi_1\big(p_0(h)t_0 + p_1(h)t_1 + \ldots + p_{\ell-1}(h)t_{\ell-1}\big) : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell \right\}}.$$

Then for all $F \in C(\mathbb{T}^d)$ we have

$$\underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} F\big(\pi_1\big(p(n,h)\big)\big) = \int_Y F \, \mathsf{d}\mu_Y,$$

where $h = (h_1, \ldots, h_m)$.

Proof. It suffices to show that for all continuous group characters $\chi \colon \mathbb{T}^d \to \{w \in \mathbb{C} : |w| = 1\}$
with the property that χ restricted to Y is non-trivial, one has

$$(5.1) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} \chi\big(\pi_1\big(p(n,h)\big)\big) = 0,$$

because for any such character we have $\int_Y \chi \, \mathsf{d}\mu_Y = 0$ and linear combinations of continuous
group characters of this kind together with constants are dense in $C(Y)$. The character χ is
naturally associated with an element $\tau \in \mathbb{Z}^d$ such that $\chi\big(\pi(t)\big) = e(\langle t, \tau \rangle)$ for every $t \in \mathbb{R}^d$,
where $e(x) := e^{2\pi i x}$. The condition that χ restricted to Y is non-trivial implies that the map

$$(h, t_0, \ldots, t_{\ell-1}) \mapsto \big\langle p_0(h)t_0 + p_1(h)t_1 + \ldots + p_{\ell-1}(h)t_{\ell-1}, \, \tau \big\rangle \bmod 1$$

is not constant. Therefore, for some $j \in \{0, \ldots, \ell-1\}$, the map $h \mapsto \langle p_j(h), \tau \rangle$ is not identically
0. Choose the smallest such j. Since p_j is a polynomial, $\langle p_j(h), \tau \rangle \neq 0$ for most $h \in \mathbb{Z}^m$, in the
sense that the set of zeros $R := \{h \in \mathbb{Z}^m : \langle p_j(h), \tau \rangle = 0\}$ satisfies

$$\underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} 1_R(h_1, \ldots, h_m) = 0.$$

On the other hand, for each $h \in \mathbb{Z}^m \backslash R$, the function $n \mapsto \big\langle p(n,h), \, \tau \big\rangle$ belongs to $\mathcal{F}$ and hence,
using Theorem 2.12, we get

$$\underset{n \to \infty}{UW\text{-}\lim} e\big(\langle p(n,h), \, \tau \rangle\big) = 0.$$

Therefore

$$
\begin{aligned}
\left| \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} \chi\big(\pi_1\big(p(n,h)\big)\big) \right| &= \left| \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} e\big(\langle p(n,h), \tau \rangle\big) \right| \\
&\leqslant \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \left| \underset{n \to \infty}{UW\text{-}\lim} e\big(\langle p(n,h), \tau \rangle\big) \right| \\
&\leqslant \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} 1_R(h_1, \ldots, h_m) \\
&= 0.
\end{aligned}
$$

$\square$

5.2. Lifting equidistribution from the horizontal torus to the nilmanifold

Definition 5.4. Let G be a connected and simply connected nilpotent Lie group and let $\Gamma \subset G$
be a uniform and discrete subgroup. The *horizontal torus* of the nilmanifold $X := G/\Gamma$ is the
compact abelian group $G/(\Gamma[G,G])$, where $[G,G]$ denotes the commutator subgroup of G (i.e.
the subgroup of G generated by all elements $[g,h] = g^{-1}h^{-1}gh$).

A *horizontal character* is a map of the form $\eta \circ \pi$, where $\pi \colon X \to G/(\Gamma[G,G])$ is the projection
onto the horizontal torus and η is a character of $G/(\Gamma[G,G])$.

The following theorem is the main technical result of this paper. We will show that it implies
Theorem 5.2.

Theorem 5.5. *Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and define $X := G/\Gamma$. Let $\ell, u \in \mathbb{N}$, let $m \in \mathbb{Z}$ with $m \geqslant 0$, let $f \in \mathcal{F}_\ell$, let $a_{i,j}$ with $(i,j) \in \{0, 1, \ldots, \ell-1\} \times \{1, \ldots, u\}$ be a collection of commuting elements of G and, for $(i,j) \in \{0, 1, \ldots, \ell-1\} \times \{1, \ldots, u\}$, let $p_{i,j} \in \mathbb{R}[x_1, \ldots x_m]$. Let $g_0, g_1, \ldots, g_{\ell-1} : \mathbb{Z}^m \to G$ denote the polynomial sequences defined by*

$$g_i(h) := a_{i,1}^{p_1(h)} a_{i,2}^{p_2(h)} \cdot \ldots \cdot a_{i,u}^{p_{i,u}(h)}, \qquad \forall h \in \mathbb{Z}^m,$$

(or let $g_i \in G$ be commuting constants if $m = 0$) and set

$$g(n, h) := g_0(h)^{f(n)} g_1(h)^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta^{\ell-1} f(n)}, \qquad \forall n \in \mathbb{N}, \ \forall h \in \mathbb{Z}^m.$$

Define $W := \Delta^{\ell-1} f$. If for all non-trivial horizontal characters η one has

$$(5.2) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \ \underset{n \to \infty}{UW\text{-}\lim} \ \eta\big(g(n, h)\Gamma\big) = 0,$$

then for all $F \in C(X)$ with $\int_X F \, \mathsf{d}\mu_X = 0$, we have

$$(5.3) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \ \underset{n \to \infty}{UW\text{-}\lim} \ F\big(g(n, h)\Gamma\big) = 0,$$

where $h = (h_1, \ldots, h_m)$.

Outline of the proof of Theorem 5.5: For the proof of Theorem 5.5 we have to distinguish between the cases $\ell = 1$ and $\ell \geqslant 2$. The case $\ell = 1$ is easy and taken care of in Section 5.3. For the case $\ell \geqslant 2$, which is proved in Section 5.4, we use induction on the degree of $g(n, h)$ (see Definition 5.9). The base case of this induction follows immediately from Lemma 5.3. For the proof of the inductive step we use a variant of the classical van der Corput Lemma, which says that instead of (5.3) it suffices to show

$$(5.4) \qquad \underset{h_{m+1} \to \infty}{C\text{-}\lim} \ \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \ \underset{n \to \infty}{UW\text{-}\lim} \ F\big(g(n + h_{m+1}, h)\Gamma\big) \overline{F}\big(g(n, h)\Gamma\big) = 0.$$

We can also assume without loss of generality that F satisfies

$$(5.5) \qquad F(tx) = \chi(t) F(x), \qquad \forall t \in Z(G), \ \forall x \in X,$$

where $Z(G)$ is the center of G and $\chi : Z(G) \to \{w \in \mathbb{C} : |w| = 1\}$ is a continuous group character of $Z(G)$. Indeed, the algebra generated by functions satisfying (5.5) is uniformly dense in $C(X)$. Using ideas form [26], one can show that the projection of the sequence $(g(n + h_{m+1}, h), g(n, h))$ onto the quotient group $(G \times G)/Z(G)^\triangle$, where $Z(G)^\triangle = \{(g, g) : g \in Z(G)\}$, has degree strictly smaller than of $g(n, h)$. Moreover, due to (5.5), the function $F \otimes \overline{F}$ is invariant under $Z(G)^\triangle$. It thus follows from the induction hypothesis that

$$\underset{h_{m+1} \to \infty}{C\text{-}\lim} \ \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \ \underset{n \to \infty}{UW\text{-}\lim} \ F\big(g(n + h_{m+1}, h)\Gamma\big) \overline{F}\big(g(n, h)\Gamma\big) = \int_Y F \otimes \overline{F} \, \mathsf{d}\mu_Y,$$

where Y is a certain sub-nilmanifold of $X \times X$ which we describe explicitly in our proof. By judiciously using (5.5) one can then show that $\int_Y F \otimes \overline{F} \, \mathsf{d}\mu_Y = 0$, which completes the proof.

We end this subsection by showing that Theorem 5.5 implies Theorem 5.2.

Proof that Theorem 5.5 implies Theorem 5.2. We assume that $m > 0$, the case $m = 0$ can be proved in the same way. We proceed by induction of the dimension d of G. If $d = 1$ then $G = \mathbb{R}$, in which case Theorem 5.2 follows from Lemma 5.3. Assume therefore that $d > 1$ and that Theorem 5.2 has already been established for all connected simply connected nilpotent Lie groups G' with dimension d' smaller than d.

Recall that Y is defined as

$$Y := \overline{\{g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-1}(h)^{t_{\ell-1}} \Gamma : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell\}}.$$

Let $\pi\colon G \to X$ denote the natural projection from G onto X and, using Lemma A.8, choose a closed, connected and *rational*[15] subgroup H of G such that $Y = \pi(H)$. Let $\vartheta\colon G \to G/[G,G]$ denote the natural projection of G onto $G/[G,G]$ and define $\tilde{G} := \vartheta(G) = G/[G,G]$. If $\vartheta(H)$ is a proper subgroup of $\tilde{G}$, then $G' := \vartheta^{-1}(\vartheta(H)) = H \cdot [G,G]$ is a closed, normal connected and rational subgroup of G with $\dim(G') < \dim(G)$ and $g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-1}(h)^{t_{\ell-1}} \in G'$ for all $h \in \mathbb{Z}^m$ and $(t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell$. In this case we can replace G with G' and Theorem 5.2 follows from the induction hypothesis.

Let us therefore assume that $\vartheta(H) = \tilde{G}$. Define $\tilde{\Gamma} := \vartheta(\Gamma)$, $T_1 := \tilde{G}/\tilde{\Gamma}$ and, for $i = 0, 1, \ldots, \ell - 1$, let $p_i(h) := \vartheta(g_i(h))$. Set

$$p(n, h) := p_0(h)f(n) + p_1(h)\Delta f(n) + \ldots + p_{\ell-1}(h)\Delta^{\ell-1}f(n), \qquad \forall n \in \mathbb{N}, \ \forall h \in \mathbb{Z}^m.$$

Note that $\tilde{G} \cong \mathbb{R}^{d'}$ and $T_1 \cong \mathbb{T}^{d'}$ for some $d' \in \mathbb{N}$. Let $\pi_1\colon \tilde{G} \to T_1$ denote the natural factor map from $\tilde{G}$ onto T_1. Fix a non-trivial horizontal character $\eta\colon X \to \{w \in \mathbb{C} : |w| = 1\}$. Since $[G,G]\cdot\Gamma$ belongs to the kernel of η, there exists a non-trivial group character $\chi\colon T_1 \to \{w \in \mathbb{C} : |w| = 1\}$ such that

$$\chi(\pi_1(\vartheta(a))) = \eta(a\Gamma), \qquad \forall a \in G.$$

Since $\vartheta(H) = \tilde{G}$, the set

$$\left\{\pi_1\big(p_0(h)t_0 + p_1(h)t_1 + \ldots + p_{\ell-1}(h)t_{\ell-1}\big) : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell\right\}$$

is dense in T_1. It thus follows from Lemma 5.3 that

$$C\text{-}\lim_{h_m \to \infty} \cdots C\text{-}\lim_{h_1 \to \infty} UW\text{-}\lim_{n \to \infty} \chi\big(\pi_1(p(n, h))\big) = \int_{T_1} \chi \, \mathsf{d}\mu_{T_1} = 0.$$

This implies that

$$C\text{-}\lim_{h_m \to \infty} \cdots C\text{-}\lim_{h_1 \to \infty} UW\text{-}\lim_{n \to \infty} \eta\big(g(n, h)\Gamma\big) = 0$$

and therefore, using Theorem 5.5, we obtain

$$C\text{-}\overline{\lim}_{h_m \to \infty} \cdots C\text{-}\overline{\lim}_{h_1 \to \infty} UW\text{-}\overline{\lim}_{n \to \infty} F\big(g(n, h)\Gamma\big) = 0$$

for all $F \in C(X)$ with $\int_X F \, \mathsf{d}\mu_X = 0$. It must therefore be the case that $Y = X$, which implies

$$C\text{-}\overline{\lim}_{h_m \to \infty} \cdots C\text{-}\overline{\lim}_{h_1 \to \infty} UW\text{-}\overline{\lim}_{n \to \infty} F\big(g(n, h)\Gamma\big) = 0$$

for all $F \in C(X)$ with $\int_Y F \, \mathsf{d}\mu_Y = 0$ and the proof is completed. $\qquad\square$

5.3. Proof of Theorem 5.5 for $\ell = 1$

In this section we prove the case $\ell = 1$ of Theorem 5.5.

Proposition 5.6. *Let G be a connected and simply connected nilpotent Lie group, let $\Gamma \subset G$ be a uniform discrete subgroup and let $X := G/\Gamma$. Let $b \in G$ and let μ be the Haar measure on the subnilmanifold $\overline{\{b^t : t \in \mathbb{R}\}}$. Then for every $W \in \mathcal{F}_1$,*

$$UW\text{-}\lim_{n \to \infty} \delta_{b^{W(n)}\Gamma} = \mu \qquad \text{in the weak* topology.}$$

Proof. We need to show that for every $F \in C(X)$

$$(5.6) \quad \lim_{W(N)-W(M)\to\infty} \left| \frac{1}{W(N) - W(M)} \sum_{n=M}^{N} \Delta W(n)F(b^{W(n)}\Gamma) - \frac{1}{N - M} \int_M^N F(b^t\Gamma) \, \mathsf{d}t \right| = 0.$$

---

[15] A closed subgroup H of G is called *rational* if the set $H\Gamma$ is a closed subset of G (cf. [37]).

Fix $\varepsilon > 0$ and let $\eta = \eta(\varepsilon) > 0$ be such that whenever $t_1, t_2 \in \mathbb{R}$ with $|t_1 - t_2| < \eta$ then $|F(b^{t_1}\Gamma) - F(b^{t_2}\Gamma)| < \varepsilon$. We can thus replace $F(b^{W(n)}\Gamma)$ with $F(b^{\lfloor W(n)/\eta \rfloor \eta}\Gamma)$ at a cost of ε. Note that

$$\frac{1}{W(N) - W(M)} \sum_{n=M}^{N} \Delta W(n) F(b^{\lfloor W(n)/\eta \rfloor \eta}\Gamma)$$

$$= \frac{1}{W(N) - W(M)} \sum_{k=\lfloor W(M)/\eta \rfloor}^{\lfloor W(N)/\eta \rfloor} \sum_{n:W(n)\in\left[k\eta,(k+1)\eta\right)} \Delta W(n) F(b^{k\eta}\Gamma).$$

Also, since $W \in \mathcal{F}_1$, we have

$$\lim_{k\to\infty} \sum_{n:W(n)\in\left[k\eta,(k+1)\eta\right)} \Delta W(n) = \eta.$$

Therefore

$$\limsup_{W(N)-W(M)\to\infty} \left| \frac{1}{W(N)-W(M)} \sum_{n=M}^{N} \Delta W(n) F(b^{W(n)}\Gamma) - \frac{\eta}{W(N)-W(M)} \sum_{k=\lfloor W(M)/\eta \rfloor}^{\lfloor W(N)/\eta \rfloor} F(b^{k\eta}\Gamma) \right| < \varepsilon.$$

Since ε is arbitrary, (5.6) now follows from Lemma A.7. $\qquad\square$

Lemma 5.7 (Theorem 5.5 for the special case $\ell = 1$)**.** *Let G be a connected simply connected nilpotent Lie group, $\Gamma \subset G$ a uniform and discrete subgroup and define $X := G/\Gamma$. Let $u, m \in \mathbb{N}$, let $W \in \mathcal{F}_1$, let $a_1, \ldots, a_u \in G$ be commuting and let $p_1, \ldots, p_u \in \mathbb{R}[x_1, \ldots x_m]$. Let*

$$g(h) = a_1^{p_1(h)} \cdot \ldots \cdot a_u^{p_u(h)} \qquad and \qquad g(n, h) := g(h)^{W(n)}, \qquad \forall n \in \mathbb{N}, \ \forall h \in \mathbb{Z}^m.$$

If for all non-trivial horizontal characters η one has

$$C\text{-}\lim_{h_m\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} UW\text{-}\lim_{n\to\infty} \eta\big(g(n, h)\Gamma\big) = 0,$$

then the points $g(n, h)\Gamma$, where $h = (h_1, \ldots, h_m)$, equidistribute as follows:

$$C\text{-}\lim_{h_m\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} UW\text{-}\lim_{n\to\infty} \delta_{g(n,h)\Gamma} = \mu_X.$$

Proof. For each $h \in \mathbb{N}^m$, let $Y_h \subset X$ be the subnilmanifold $Y_h := \overline{\{g(h)^t\Gamma : t \in \mathbb{R}\}} \subset X$ and let μ_h be the Haar measure on Y_h. In view of Proposition 5.6,

$$(5.7) \qquad\qquad UW\text{-}\lim_{n\to\infty} \delta_{g(n,h)\Gamma} = \mu_h.$$

On the other hand, Lemma A.7 implies that there exists a time $t \in \mathbb{R}$ such that $C\text{-}\lim_{n\to\infty} \delta_{g(h)^{tn}\Gamma} = \mu_h$, for every $h \in \mathbb{N}^m$, so we need to show that

$$C\text{-}\lim_{h_m\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} C\text{-}\lim_{n\to\infty} \delta_{g(h)^{tn}\Gamma} = \mu_X.$$

Since the map $\tilde{g}: (h_1, \ldots, h_m, n) \mapsto g(h)^{tn}$ is a polynomial sequence, the result now follows directly from Theorem A.5.

$\qquad\square$

5.4. Proof of Theorem 5.5 for general ℓ

Definition 5.8. Let $s \in \mathbb{N}$ and let $G_1, \ldots, G_s, G_{s+1}$ be subgroups of a nilpotent Lie group G. We call $G_\bullet := \{G_1, \ldots, G_s, G_{s+1}\}$ an *s-step pre-filtration* of G if

$$[G_i, G_j] \subset G_{i+j}, \qquad \forall j, i = \{1, \ldots, s\} \text{ with } i + j \leqslant s + 1,$$

and $G_{s+1} = \{1_G\}$. A pre-filtration $G_\bullet$ is called a *filtration* if $G_1 = G$. A *normal pre-filtration* is a pre-filtration consisting only of normal subgroups.

Note that for any connected simply connected nilpotent Lie group G there exists a right-invariant metric $d_G \colon G \times G \to [0, \infty)$ on G. For any uniform and discrete subgroup Γ the metric d_G descends to a metric d_X on $X := G/\Gamma$ via

$$(5.8) \qquad d_X(x\Gamma, y\Gamma) := \inf\{d_G(x\gamma, y\gamma') : \gamma, \gamma' \in \Gamma\}.$$

Given a subset $S \subset G$ and a point $g \in G$ we denote $d(g, S) = \inf_{s \in S} d_G(g, s)$.

Definition 5.9. Let $\ell \in \mathbb{N}$, let $f \in \mathcal{F}_\ell$, let $g_0, g_1, \ldots, g_{\ell-1} \colon \mathbb{Z}^m \to G$ be polynomial sequences and set

$$g(n, h) := g_0(h)^{f(n)} g_1(h)^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta^{\ell-1} f(n)}, \qquad \forall n \in \mathbb{N}, \ \forall h \in \mathbb{Z}^m.$$

For every $n_1 \in \mathbb{N}$ let $\Delta_{n_1} g(n, h)$ denote the *discrete derivative of $g(n, h)$ in direction n_1*, that is, $\Delta_{n_1} g(n, h) := g(n + n_1, h) g(n, h)^{-1}$. We define the degree of $g(\cdot, h)$ to be the smallest number $s \in \mathbb{N}$ such that there exists a s-step normal pre-filtration $G_\bullet = \{G_1, G_2, \ldots, G_s, G_{s+1} = \{1_G\}\}$ with the property that for every fixed $h \in \mathbb{N}^m$ we have

$$\lim_{n \to \infty} d_G\Big(g(n, h), G_1\Big) = 0$$

and, for all $j \in \{1, \ldots, s\}$ and all $n_1, \ldots n_j \in \mathbb{N}$,

$$\lim_{n \to \infty} d_G\Big(\Delta_{n_j} \cdots \Delta_{n_1} g(n, h), G_{j+1}\Big) = 0.$$

In this case we say $G_\bullet$ is a normal pre-filtration of G that *realizes* the step of $g(n, h)$. If there exists no such filtration, then we say that $g \colon \mathbb{N} \times \mathbb{Z}^m \to G$ has infinite degree.

Lemma 5.10. *Let $\ell \in \mathbb{N}$, let $f \in \mathcal{F}_\ell$ and $g_0, g_1, \ldots, g_{\ell-1} \colon \mathbb{Z}^m \to G$ be as in Theorem 5.5 and define*

$$g(n, h) := g_0(h)^{f(n)} g_1(h)^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta^{\ell-1} f(n)}, \qquad \forall n \in \mathbb{N}, \ \forall h \in \mathbb{Z}^m.$$

Let s denote the nilpotency step of G. Then $g \colon \mathbb{N} \times \mathbb{Z}^m \to G$ has finite degree (in fact, the degree is smaller or equal to $\ell(s+1)$).

Proof. Let $C_\bullet := \{G = C_1, C_2, \ldots, C_s, C_{s+1} = \{1_G\}\}$ denote the lower central series of G. Define a new filtration $G_\bullet = \{G_1, G_2, \ldots, G_{\ell(s+1)}, G_{\ell(s+1)+1} = \{1_G\}\}$ by setting $G_{(j-1)\ell+i} := C_j$ for all $j \in \{1, \ldots, s+1\}$ and $i \in \{1, \ldots, \ell\}$.

Using the fact that the polynomial sequences $g_0, \ldots, g_{\ell-1}$ commute we see that for all $n_1, \ldots n_j \in \mathbb{N}$ we have

$$\Delta_{n_j} \cdots \Delta_{n_1} g(n, h) = g_0(h)^{\Delta_{n_j} \cdots \Delta_{n_1} f(n)} g_1(h)^{\Delta_{n_j} \cdots \Delta_{n_1} \Delta f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta_{n_j} \cdots \Delta_{n_1} \Delta^{\ell-1} f(n)}.$$

Since $f \in \mathcal{F}_\ell$ it follows that for all $j \geqslant \ell$ and all $i \geqslant 0$ the expression $\Delta_{n_j} \cdots \Delta_{n_1} \Delta^i f(n)$ converges to 0 as $n \to \infty$. This implies that as long as $j \geqslant \ell$ then for all choices of $n_1, \ldots, n_j$ the sequence $\Delta_{n_j} \cdots \Delta_{n_1} g(n, h)$ converges to the identity 1_G of G. Using this observation it is now straightforward to check that $G_\bullet$ is a normal filtration with the property that for all $h \in \mathbb{N}^m$, all $j \in \{1, \ldots, \ell(s+1)\}$ and all $n_1, \ldots, n_j \in \mathbb{N}$,

$$(5.9) \qquad \lim_{n \to \infty} d_G\Big(\Delta_{n_j} \cdots \Delta_{n_1} g(n, h), \ G_{j+1}\Big) = 0.$$

Indeed, if $j < \ell$ then G_{j+1} equals the entire group G and so (5.9) holds trivially, whereas if $j \geqslant \ell$ then (5.9) holds because $\Delta_{n_j} \cdots \Delta_{n_1} g(n, h)$ converges to 1_G. □

Proof of Theorem 5.5. We proceed by induction on the degree of $g(n, h)$, which is finite according to Lemma 5.10. If the degree of $g(n, h)$ equals 1 then G must be abelian, in which case the claim follows from Lemma 5.3. Let us therefore assume that the degree of $g(n, h)$ is equal to some $s \geqslant 2$ and that Theorem 5.5 has already been proven for all systems $(\tilde{G}, \tilde{\Gamma}, f, \tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_\ell)$ where $\tilde{g}(n, h') = \tilde{g}_0(h')^{f(n)} \tilde{g}_1(h')^{\Delta f(n)} \cdot \ldots \cdot \tilde{g}_{\ell-1}(h')^{\Delta^{\ell-1} f(n)}$ has degree smaller than s.

Define

$$(5.10) \qquad Z := \overline{\{g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-2}(h)^{t_{\ell-2}} \Gamma : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-2}) \in \mathbb{R}^{\ell-1}\}}.$$

If $Z = \{1_X\}$ then $g_0(h) = \ldots = g_{\ell-2}(h) = 1_G$ for all $h \in \mathbb{Z}^m$, in which case Theorem 5.5 follows from Lemma 5.7. Thus we can assume that $Z \neq \{1_X\}$.

Invoking Lemma A.8 we can find a closed rational and connected subgroup H of G such that $\pi(H) = Z$. Let L denote the normal closure of H in G, i.e., the smallest normal subgroup of G containing H. One can show that L is also connected, simply connected, rational and closed (see [38, p. 844]). Also, we remark that L is unique in the sense that if H' is another closed rational and connected subgroup of G with $\pi(H') = Z$ then the normal closure of H' coincides with L. Let $L_\bullet := \{L = L_1, L_2, \ldots, L_r, L_{r+1} = \{1_G\}\}$ denote the lower central series of L and note that all elements in $L_\bullet$ are themselves connected, simply connected, rational, closed and normal subgroups of G (cf. [42]).

Let $C_\bullet := \{G = C_1, C_2, \ldots, C_r, C_{r+1} = \{1_G\}\}$ denote the lower central series of G and let j_0 denote the biggest number in $\{1, \ldots, r\}$ such that $L \cap G_{j_0} \neq \{1_G\}$. Note that j_0 exists, because $L \neq \{1_G\}$. Define $V := L \cap G_{j_0}$.

For any $a \in G$ and any $t \in V$ the element $[a, t] = a^{-1} t^{-1} a t$ belongs to C_{j_0+1}. Since L is normal and $t \in L$, we also have $[a, t] \in L$. Therefore $[a, t]$ belongs to $L \cap C_{j_0+1}$, which implies that $[a, t] = 1_G$. This proves that V is a subgroup of the center $Z(G)$ of G.

Define $T := Z(G)/(Z(G) \cap \Gamma)$. Note that T is a torus (i.e., isomorphic to $\mathbb{T}^d$, where $d := \dim(Z(G))$). Also, since $Z(G) \cap \Gamma$ acts trivially on X, the action of $Z(G)$ on X naturally descents to an action of T on X. Recall that our goal is to show

$$(5.11) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} \left(F\big(g(n, h)\Gamma\big) - \int_X F \, \mathsf{d}\mu_X \right) = 0$$

for all all $F \in C(X)$, where $h = (h_1, \ldots, h_m)$. Note that any continuous group character χ of T lifts to a continuous group homomorphism from $Z(G)$ to $\{z \in \mathbb{C} : |z| = 1\}$ whose kernel contains $Z(G) \cap \Gamma$. By abuse of language we will use χ to denote both those maps.

Consider the class of functions $\phi \in C(X)$ with the property that

$$(5.12) \qquad \phi(tx) = \chi(t)\phi(x), \qquad \forall t \in Z(G), \ \forall x \in X,$$

for some continuous group character χ of T. This class of functions separates points in X.[16] Thus, by the Stone-Weierstrass theorem, the span of such functions is dense in $C(X)$. Therefore, to prove (5.11), it suffices to show

$$(5.13) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \underset{n \to \infty}{UW\text{-}\lim} \left(\phi\big(g(n, h)\Gamma\big) - \int_X \phi \, \mathsf{d}\mu_X \right) = 0$$

---

[16]Let x and y be two distinct points in X. To show that functions satisfying (5.12) separate these two points, we distinguish the cases $y \notin Tx$ and $y \in Tx$. If $y \notin Tx$ then $Tx \neq Ty$ and so there exists a continuous function $\phi' \in C(T \backslash X)$ with $\phi'(Tx) \neq \phi'(Ty)$. Then ϕ' lifts to a continuous and T-invariant function $\phi \in C(X)$ with $\phi(x) \neq \phi(y)$. On the other hand, if $y \in Tx$ then there is $t_0 \in T$, which is not the identity, such that $y = t_0 x$. Let χ_0 be any group character of T with the property that $\chi_0(t_0) \neq 1$. Let U be a small neighborhood of x and let ρ be a continuous function on X with the property that $\rho(x) = 1$ and $\rho(z) = 0$ for all $z \notin U$. Now define $\phi(z) := \int_T \rho(sz)\chi_0(s) \, \mathsf{d}\mu_T(s)$, where μ_T is the normalized Haar measure on T. It is straightforward to check that ϕ is a non-zero and continuous function on X satisfying $\phi(tz) = \chi_0(t)\phi(z)$ for all $t \in T$ and $z \in X$. In particular $\phi(y) = \phi(t_0 x) = \chi_0(t_0)\phi(x)$ is not equal to $\phi(x)$.

for all functions ϕ of this kind.

Fix $\phi \in C(X)$ and a continuous group character χ of T such that (5.12) is satisfied. If χ is trivial when restricted to V, then, after we mod out G by V, we have reduced the question to a nilpotent Lie group of smaller dimension. Hence, by induction on the dimension of G, we can assume without loss of generality that χ is non-trivial when restricted to V. This also implies that $\int_X \phi \, \mathsf{d}\mu_X = 0$ and hence (5.13) becomes

$$(5.14) \qquad \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \; \underset{n \to \infty}{UW\text{-}\lim} \; \phi\big(g(n,h)\Gamma\big) = 0.$$

Using a version of van der Corput's lemma proved in the appendix (see Lemma A.9) we see that instead of (5.14) it suffices to show

$$(5.15) \qquad \underset{h_{m+1} \to \infty}{C\text{-}\lim} \; \underset{h_m \to \infty}{C\text{-}\lim} \cdots \underset{h_1 \to \infty}{C\text{-}\lim} \; \underset{n \to \infty}{UW\text{-}\lim} \; \phi\big(g(n+h_{m+1},h)\Gamma\big)\overline{\phi}\big(g(n,h)\Gamma\big) = 0.$$

Using Lemma 2.11 we can rewrite $\Delta^j f(n + h_{m+1})$ as

$$\begin{aligned}
\Delta^j f(n + h_{m+1}) &= \sum_{i=0}^{h_{m+1}} \binom{h_{m+1}}{i} \Delta^{i+j} f(n) \\
&= \sum_{i=j}^{\ell-1} \binom{h_{m+1}}{i-j} \Delta^i f(n) + \varepsilon_j(n),
\end{aligned}$$

where $\varepsilon_j(n) := \sum_{i=\ell}^{h_{m+1}} \binom{h_{m+1}}{i} \Delta^{i+j} f(n)$. Define $\varepsilon(n) := \prod_{j=0}^{\ell} g_j(h_1, \ldots, h_m)^{\varepsilon_j(n)}$. Using the fact that the polynomial sequences $g_0, \ldots, g_{\ell-1}$ commute, we obtain

(5.16)

$$\begin{aligned}
g(n + h_{m+1}, h) &= \prod_{j=0}^{\ell-1} g_j(h)^{\Delta^j f(n+h_{m+1})} \\
&= \prod_{j=0}^{\ell-1} \prod_{i=j}^{\ell-1} g_j(h)^{\binom{h_{m+1}}{i-j}\Delta^i f(n) + \varepsilon_j(n)} \\
&= \varepsilon(n) \prod_{i=0}^{\ell-1} \left(\prod_{j=0}^{i} g_j(h)^{\binom{h_{m+1}}{i-j}} \right)^{\Delta^i f(n)} \\
&= \varepsilon(n)\Big(g_1'(h')^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}'(h')^{\Delta^{\ell-1}f(n)}\Big)\Big(g_0(h)^{f(n)} \cdot \ldots \cdot g_{\ell-1}(h)^{\Delta^{\ell-1}f(n)}\Big),
\end{aligned}$$

where $h = (h_1, \ldots, h_m)$, $h' = (h_1, \ldots, h_m, h_{m+1})$ and

$$(5.17) \qquad g_i'(h') := \prod_{j=0}^{i-1} g_j(h)^{\binom{h_{m+1}}{i-j}} \qquad \text{for all } i = 0, \ldots, \ell-1,$$

where the empty product defining $g_0'(h')$ is to be understood as 1_G. Define

$$g'(n, h') := g_1'(h')^{\Delta f(n)} \cdot \ldots \cdot g_{\ell-1}'(h')^{\Delta^{\ell-1}f(n)} \qquad \forall n \in \mathbb{N}, \; \forall h' \in \mathbb{Z}^{m+1}.$$

Thus from (5.16) we get

$$(5.18) \qquad g(n + h_{m+1}, h) = \varepsilon(n)g'(n, h')g(n, h).$$

Using the definition of the metric d_X given in (5.8) and the right-invariance of d_G we see that

$$d_X\Big(\varepsilon(n)g'(n,h')g(n,h)\Gamma,\ g'(n,h')g(n,h)\Gamma\Big)$$

$$= \inf_{\gamma,\gamma'\in\Gamma} d_G\Big(\varepsilon(n)g'(n,h')g(n,h)\gamma,\ g'(n,h')g(n,h)\gamma'\Big)$$

$$\leqslant d_G\Big(\varepsilon(n)g'(n,h')g(n,h),\ g'(n,h')g(n,h)\Big)$$

$$= d_G\left(\varepsilon(n),\ 1_G\right).$$

Since for fixed $h_1,\ldots,h_{m+1}\in\mathbb{N}$ the error terms $\varepsilon_i(n)$ converge to 0 as $n\to\infty$ for all $i\in\{0,1,\ldots,\ell-1\}$, it follows that

$$\lim_{n\to\infty} d_G\left(\varepsilon(n),\ 1_G\right)=0$$

and hence

(5.19) $$\lim_{n\to\infty} d_X\Big(\varepsilon(n)g'(n,h')g(n,h)\Gamma,\ g'(n,h')g(n,h)\Gamma\Big)=0.$$

Combining (5.18) and (5.19) and using the fact that ϕ is continuous, we obtain

$$\lim_{n\to\infty}\big|\phi\big(g(n+h_{m+1},h)\Gamma\big)-\phi\big(g'(n,h')g(n,h)\Gamma\big)\big|=0.$$

Hence (5.15) is equivalent to

(5.20) $$C\text{-}\varlimsup_{h_{m+1}\to\infty} C\text{-}\varlimsup_{h_m\to\infty}\cdots C\text{-}\varlimsup_{h_1\to\infty} UW\text{-}\varlimsup_{n\to\infty}\phi\big(g'(n,h')g(n,h)\Gamma\big)\overline{\phi}\big(g(n,h)\Gamma\big)=0,$$

where $h=(h_1,\ldots,h_m)$ and $h'=(h_1,\ldots,h_m,h_{m+1})$.

Let $Z(G)^\triangle := \{(a,a):a\in Z(G)\}$ and denote by $\sigma\colon G\times G\to(G\times G)/Z(G)^\triangle$ the natural projection of $G\times G$ onto $(G\times G)/Z(G)^\triangle$. Define

$$G^\triangle \ :=\ \{(a,a):a\in G\};$$
$$K \ :=\ G^\triangle\cdot(L\times L);$$
$$\Gamma_K \ :=\ K\cap(\Gamma\times\Gamma)$$
$$\tilde{G} \ :=\ \sigma(K);$$
$$\tilde{\Gamma} \ :=\ \sigma\big(\Gamma_K\big);$$
$$\tilde{X} \ :=\ \tilde{G}/\tilde{\Gamma};$$
$$(\text{for }0\leqslant i\leqslant\ell-1)\qquad \tilde{g}_i(h') \ :=\ \sigma\big(g_i'(h')g_i(h),g_i(h)\big);$$
$$\tilde{g}(n,h') \ :=\ \tilde{g}_0(h')^{f(n)}\tilde{g}_1(h')^{\Delta f(n)}\cdot\ldots\cdot\tilde{g}_{\ell-1}(h')^{\Delta^{\ell-1}f(n)}.$$

Also, it follows from (5.12) that $\phi\otimes\overline{\phi}\colon X\times X\to\mathbb{C}$ is invariant under the action of $Z(G)^\triangle$. Therefore there exists a continuous function $\tilde{\phi}\colon\tilde{X}\to\mathbb{C}$ with the property that

$$(\phi\otimes\overline{\phi})(a\Gamma,b\Gamma)=\tilde{\phi}(\sigma(a,b)\tilde{\Gamma}),\qquad\forall(a,b)\in K.$$

One can check that (5.20) is equivalent to

(5.21) $$C\text{-}\varlimsup_{h_{m+1}\to\infty} C\text{-}\varlimsup_{h_m\to\infty}\cdots C\text{-}\varlimsup_{h_1\to\infty} UW\text{-}\varlimsup_{n\to\infty}\tilde{\phi}\big(\tilde{g}(n,h')\tilde{\Gamma}\big)=0.$$

We now make three claims:
CLAIM 1: We have $\int_{\tilde{X}}\tilde{\phi}\,\mathsf{d}\mu_{\tilde{X}}=0$.
CLAIM 2: For all non-trivial horizontal characters $\tilde{\eta}$ of $\tilde{G}/\tilde{\Gamma}$ we have

(5.22) $$C\text{-}\varlimsup_{h_{m+1}\to\infty} C\text{-}\varlimsup_{h_m\to\infty}\cdots C\text{-}\varlimsup_{h_1\to\infty} UW\text{-}\varlimsup_{n\to\infty}\tilde{\eta}\big(\tilde{g}(n,h')\tilde{\Gamma}\big)=0.$$

CLAIM 3: The sequence $\tilde{g}(n, h')$ has degree smaller than the sequence $g(n, h)$.

Note that once Claims 1, 2 and 3 have been proven, the proof of Theorem 5.5 is completed. Indeed, Claims 2 and 3 allow us to invoke the induction hypothesis and apply Theorem 5.5 to the system $(\tilde{G}, \tilde{\Gamma}, f, \tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_\ell)$ in order to obtain the identity

$$(5.23) \qquad \underset{h_{m+1}\to\infty}{C\text{-}\lim} \; \underset{h_m\to\infty}{C\text{-}\lim} \; \cdots \underset{h_1\to\infty}{C\text{-}\lim} \; \underset{n\to\infty}{W\text{-}\lim} \; \left(\tilde{\phi}\big(\tilde{g}(n, h')\tilde{\Gamma}\big) - \int_{\tilde{X}} \tilde{\phi} \, \mathsf{d}\mu_{\tilde{X}} \right) = 0.$$

However, due to Claim 1 we have that (5.23) implies (5.21), which finishes the proof.

Proof of Claim 1: Define $\tilde{V} := \sigma(V \times \{1_G\})$ and note that $\tilde{V}$ and V are isomorphic. Also, from (5.12) we can derive that for all $\tilde{t} \in \tilde{V}$ and $\tilde{x} \in \tilde{X}$ one has

$$\tilde{\phi}(\tilde{t}\tilde{x}) = \tilde{\chi}(\tilde{t})\tilde{\phi}(\tilde{x}),$$

where $\tilde{\chi}$ is defined via $\tilde{\chi}(\sigma(t, 1_G)) := \chi(t)$ for all $t \in V$. Since χ is non trivial when restricted to V, and there is an isomorphism from V to $\tilde{V}$ taking χ to $\tilde{\chi}$, we conclude that $\tilde{\chi}$ is also non-trivial. Let $\tilde{t} \in \tilde{V}$ be such that $\tilde{\chi}(\tilde{t}) \neq 1$. Since $\tilde{V}$ is a subgroup of $\tilde{G}$, we have that $\mu_{\tilde{X}}$ is invariant under $\tilde{t}$, which implies that

$$\int_{\tilde{X}} \tilde{\phi}(\tilde{x}) \, \mathsf{d}\mu_{\tilde{X}}(\tilde{x}) = \int_{\tilde{X}} \tilde{\phi}(\tilde{t}\tilde{x}) \, \mathsf{d}\mu_{\tilde{X}}(\tilde{x}) = \tilde{\chi}(\tilde{t}) \int_{\tilde{X}} \tilde{\phi}(\tilde{x}) \, \mathsf{d}\mu_{\tilde{X}}(\tilde{x}).$$

and hence $\int_{\tilde{X}} \tilde{\phi} \, \mathsf{d}\mu_{\tilde{X}} = 0$ as claimed.

Proof of Claim 2: Since $\tilde{g}(n, h') = \sigma\big(g'(n, h')g(n, h), g(n, h)\big)$ and for any horizontal character $\tilde{\eta}$ of $\tilde{G}/\tilde{\Gamma}$ there exists a horizontal character η'' of K/Γ_K such that $\tilde{\eta} \circ \sigma = \eta''$, it suffices to show that for all non-trivial horizontal characters η'' of K/Γ_K we have

$$(5.24) \qquad \underset{h_{m+1}\to\infty}{C\text{-}\lim} \; \underset{h_m\to\infty}{C\text{-}\lim} \; \cdots \underset{h_1\to\infty}{C\text{-}\lim} \; \underset{n\to\infty}{UW\text{-}\lim} \; \eta''\Big(\big(g'(n, h')g(n, h), g(n, h)\big)\Gamma_K\Big) = 0.$$

Note that $[L, G]$ is a normal subgroup of L. Let $\omega \colon G/[G, G] \times L/[L, G] \to K/[K, K]$ denote the map $\omega(a[G, G], b[L, G]) = (ab, a)[K, K]$. The first step in proving (5.24) is to show that ω is a well defined continuous and surjective homomorphism from $G/[G, G] \times L/[L, G]$ onto $K/[K, K]$. It is straightforward to check that the map $(a, b) \mapsto (ab, a)[K, K]$ is a continuous and surjective group homomorphism. So it only remains to show that $[G, G] \times [L, G]$ belongs to the kernel of the homomorphism $(a, b) \mapsto (ab, a)[K, K]$. This, however, follows immediately from Lemma A.3.

Let $\vartheta_1 \colon G \to G/[G, G]$ denote the natural factor map from G onto $G/[G, G]$. Since $G/[G, G]$ is a connected, simply connected and abelian Lie group and $\vartheta_1(\Gamma)$ a uniform and discrete subgroup of $G/[G, G]$, there exists $d_1 \in \mathbb{N}$ such that $G/[G, G]$ is isomorphic to $\mathbb{R}^{d_1}$ and $\vartheta_1(\Gamma)$ is isomorphic to $\mathbb{Z}^{d_1}$. Similarly, if $\vartheta_2 \colon L \to L/[L, G]$ denotes the natural factor map from L onto $L/[L, G]$ and $\vartheta_3 \colon K \to K/[K, K]$ the natural factor map from K onto $K/[K, K]$, then we can identify $L/[L, G]$ with $\mathbb{R}^{d_2}$ and $\vartheta_2(L \cap \Gamma)$ with $\mathbb{Z}^{d_2}$ as well as $K/[K, K]$ with $\mathbb{R}^{d_3}$ and $\vartheta_3(\Gamma_K)$ with $\mathbb{Z}^{d_3}$, for some numbers $d_2, d_3 \in \mathbb{N}$.

Observe that $L \cdot \Gamma$ is a subgroup of G and that L is the connected component of the identity of $L \cdot \Gamma$. Also, since $\pi(H) = Z$ and $H \subset L$, we have that $\pi^{-1}(Z) \subset L \cdot \Gamma$ and hence, using the definition of Z (see. (5.10)), we conclude that the set $\{g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-2}(h)^{t_{\ell-2}} : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-2}) \in \mathbb{R}^{\ell-1}\}$ is a subset of $L \cdot \Gamma$. However, the set $\{g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_{\ell-2}(h)^{t_{\ell-2}} : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-2}) \in \mathbb{R}^{\ell-1}\}$ is connected and contains the identity. It follows that

$$g_i(h)^t \in L, \qquad \forall h \in \mathbb{Z}^m, \; \forall t \in \mathbb{R}, \; \forall i \in \{0, 1, \ldots, \ell-1\}$$

and therefore, using (5.17),

$$g_i'(h') \in L, \qquad \forall h' \in \mathbb{Z}^m, \; \forall i \in \{0, 1, \ldots, \ell-1\}.$$

For $i = 0, 1, \ldots, \ell-1$, define $p_i(h) := \vartheta_1(g_i(h))$, for $i = 0, 1, \ldots, \ell-2$, define $q_i(h) := \vartheta_2(g_i(h))$ and, for $i = 1, \ldots, \ell-1$, define $q_i'(h') := \vartheta_2(g_i'(h'))$. It follows from (5.17) (also cf. (A.3)) that

$$q_i'(h_1, \ldots, h_{m+1}) := \sum_{j=1}^{i} \binom{h_{m+1}}{j} q_{i-j}(h_1, \ldots, h_m).$$

Set

$$p(n, h) := p_0(h)f(n) + p_1(h)\Delta f(n) + \ldots + p_{\ell-1}(h)\Delta^{\ell-1}f(n), \qquad \forall n \in \mathbb{N},\ \forall h \in \mathbb{Z}^m.$$

and

$$q'(n, h') := q_1'(h')\Delta f(n) + \ldots + q_{\ell-1}'(h')\Delta^{\ell-1}f(n), \qquad \forall n \in \mathbb{N},\ \forall h' \in \mathbb{Z}^{m+1}.$$

Denote by T_1 the torus $(G/[G,G])/\vartheta_1(\Gamma)$ and let $\pi_1\colon G/[G,G] \to (G/[G,G])/\vartheta_1(\Gamma)$ denote the corresponding factor map. Similarly, let T_2 and T_3 be the tori $T_2 := (L/[L,G])/\vartheta_2(\Gamma_L)$ and $T_3 := (K/[K,K])/\vartheta_3(\Gamma_K)$ and let π_2 and π_3 be the corresponding factor maps. Using Lemma A.2, we can identify T_2 with $(H \cap [L,G])\backslash Z$. In particular, we have that

$$\pi_2\big(\{q_0(h)t_0 + q_1(h)t_1 + \ldots + q_{\ell-2}(h)t_{\ell-2} : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-2}) \in \mathbb{R}^{\ell-1}\}\big) \text{ is dense in } T_2.$$

Moreover, the hypothesis (5.2) implies that

$$\pi_1\big(\{p_0(h)t_0 + p_1(h)t_1 + \ldots + p_{\ell-1}(h)t_{\ell-1} : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^{\ell}\}\big) \text{ is dense in } T_1.$$

It is therefore guaranteed by Lemma A.11 that

$$(\pi_1 \otimes \pi_2)\big(\{\tilde{p}_0(h')t_0 + \ldots + \tilde{p}_{\ell-1}(h')t_{\ell-1} : h' \in \mathbb{Z}^{m+1}, (t_0, \ldots, t_{\ell-1}) \in \mathbb{R}^{\ell}\}\big) \text{ is dense in } T_1 \times T_2,$$

where $\tilde{p}_0(h_1, \ldots, h_{m+1}) := (p_0(h_1, \ldots, h_m), 0)$ and

$$\tilde{p}_i(h_1, \ldots, h_{m+1}) := (p_i(h_1, \ldots, h_m), q_i'(h_1, \ldots, h_{m+1}))$$

for $i = 1, \ldots, \ell-1$. Using Lemma 5.3 we deduce that
(5.25)
$$C\text{-}\lim_{h_{m+1}\to\infty} C\text{-}\lim_{h_m\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} W\text{-}\lim_{n\to\infty} F_1\big(\pi_1\big(p(n,h)\big)\big) F_2\big(\pi_2\big(q'(n,h')\big)\big) = \int_{T_1} F_1\, \mathsf{d}\mu_{T_1} \int_{T_2} F_2\, \mathsf{d}\mu_{T_2}$$

for all $F_1 \in C(T_1)$ and $F_2 \in C(T_2)$.

Fix now a non-trivial horizontal characters η'' of K/Γ_K. Since $[K,K] \cdot \Gamma_K$ belongs to the kernel of η'', there exists a character $\chi''\colon T_3 \to \{w \in \mathbb{C} : |w| = 1\}$ such that $\eta'' = \chi'' \circ \pi_3 \circ \vartheta_3$. Next, using the fact that $\omega\colon G/[G,G] \times L/[L,G] \to K/[K,K]$ is a continuous and surjective homomorphism satisfying $\omega(\vartheta_1(\Gamma) \times \vartheta_2(\Gamma \cap L)) \subset \vartheta_3(\Gamma_K)$, we can find two other characters $\chi\colon T_1 \to \{w \in \mathbb{C} : |w| = 1\}$ and $\chi'\colon T_2 \to \{w \in \mathbb{C} : |w| = 1\}$, such that

$$(\chi'' \circ \pi_3) \circ \omega = (\chi \circ \pi_1) \otimes (\chi' \circ \pi_2).$$

Also, we have

$$\vartheta_3\big(g'(n,h')g(n,h), g(n,h)\big) = \omega(p(n,h), q'(n,h'));$$

Hence (5.24) follows from

(5.26) $$C\text{-}\lim_{h_{m+1}\to\infty} C\text{-}\lim_{h_m\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} UW\text{-}\lim_{n\to\infty} \chi\big(\pi_1(p(n,h))\big)\chi'\big(\pi_2(q'(n,h'))\big) = 0.$$

But (5.26) follows immediately from (5.25), which finishes the proof of Claim 2.
Proof of Claim 3: Let $G_\bullet := \{G_1, G_2, \ldots, G_s, G_{s+1} = \{1_G\}\}$ be a normal pre-filtration of G that realizes the step of $g(n,h)$. Define a new normal pre-filtration $K_\bullet = \{K_1, K_2, \ldots, K_s, K_{s+1} =$

$\{1_G\}\}$ via

$$K_i := K \cap \left(G_i^\triangle \cdot (G_{i+1} \times G_{i+1}) \right) = G_i^\triangle \cdot \left((L \cap G_{i+1}) \times (L \cap G_{i+1}) \right).$$

It is shown in [26, Proposition 7.2] that $[K_i, K_j] \subset K_{i+j}$; in particular, K_i is normal in K and hence $K_\bullet$ is a normal pre-filtration.

From (5.18) we deduce that

$$\varepsilon(n) g'(n, h') = \Delta_{h_{m+1}} g(n, h)$$

and hence

$$\lim_{n \to \infty} \inf_{a \in G_2} d_G(g'(n, h'), a) = 0.$$

Using the metric $d_K\big((a, b), (c, d)\big) := d_G(a, c) + d_G(b, d)$, it follows that

$$\lim_{n \to \infty} \inf_{(a,b) \in K_1} d_K\left(\big(g'(n, h') g(n, h), g(n, h) \big), (a, b) \right)$$

$$\leqslant \quad \lim_{n \to \infty} \inf_{a \in G_2} d_K\left(\big(g'(n, h') g(n, h), g(n, h) \big), \big(a g(n, h), g(n, h) \big) \right)$$

$$= \quad \lim_{n \to \infty} \inf_{a \in G_2} d_G\big(g'(n, h'), a \big)$$

$$= \quad 0.$$

A similar argument shows that for all $j \in \{1, \ldots, s\}$ and all $n_1, \ldots n_j \in \mathbb{N}$,

$$\lim_{n \to \infty} \inf_{(a,b) \in K_{j+1}} d_K\big(\Delta_{n_j} \cdots \Delta_{n_1} (g'(n, h') g(n, h), g(n, h)), (a, b) \big) = 0.$$

To finish the proof of Claim 3 define $\tilde{G}_i := \sigma(K_i)$. Since $K_s = G_s^\triangle$ and G_s is a subset of $Z(G)$, we have $K_s \subset Z(G)^\triangle$. In particular, $\tilde{G}_s = \sigma(K_s)$ is trivial. Therefore $\tilde{G}_\bullet = \{\tilde{G}_1, \tilde{G}_2, \ldots, \tilde{G}_{s-1}, \tilde{G}_s = \{1_{\tilde{G}}\}\}$ is a $(s-1)$-step normal pre-filtration of $\tilde{G}$ with the property that for every fixed $h' \in \mathbb{N}^{m+1}$ we have

$$\lim_{n \to \infty} d_{\tilde{G}}\left(\tilde{g}(n, h'), \ \tilde{G}_1 \right) = 0,$$

where $d_{\tilde{G}}(x, y) := \inf\{d_K(a, b) : a, b \in K, \sigma(a) = x, \sigma(b) = y\}$ and, for all $j \in \{1, \ldots, s-1\}$ and all $n_1, \ldots n_j \in \mathbb{N}$,

$$\lim_{n \to \infty} d_G\left(\Delta_{n_j} \cdots \Delta_{n_1} \tilde{g}(n, h'), \ \tilde{G}_{j+1} \right) = 0.$$

This shows that the degree of $\tilde{g}(n, h')$ is smaller than or equal to $s - 1$.   □

APPENDIX A. SOME AUXILIARY LEMMAS

Lemma A.1. *If $f : [1, \infty) \to \mathbb{R}$ is either a tempered function or a function from a Hardy field satisfying $\lim_{x \to \infty} f^{(\ell)}(x) = \pm\infty$ and $\lim_{x \to \infty} f^{(\ell+1)}(x) = 0$ for some $\ell \in \mathbb{N} \cup \{0\}$, then f restricted to $\mathbb{N}$ belongs to $\mathcal{F}$.*

Proof. If we are in the case when f is tempered, then by definition there exists $\ell \in \mathbb{N} \cup \{0\}$ such that f is $(\ell + 1)$-times continuously differentiable, $f^{(\ell+1)}(x)$ tends monotonically to 0, and $\lim_{x \to \infty} x f^{(\ell+1)}(x) = \pm\infty$. We can rewrite the last condition as $\lim_{x \to \infty} |f^{(\ell+1)}(x)| \succ 1/x$, which implies $\lim_{x \to \infty} f^{(\ell)}(x) = \pm\infty$. On the other hand, since $f^{(\ell+1)}(x)$ tends to 0, $f^{(\ell)}(N) \sim \sum_{n=1}^N f^{(\ell+1)}(n)$, and hence we can conclude that $f^{(\ell+1)}$ restricted to $\mathbb{N}$ belongs to $\mathcal{F}_0$. Since $\big(\Delta(f^{(\ell)})\big)(n) = f^{(\ell+1)}(n)\big(1 + \mathrm{o}_{n \to \infty}(1)\big)$, we conclude that $f^{(\ell)}$ restricted to $\mathbb{N}$ belongs to $\mathcal{F}_1$. Repeating this arguments shows that $f^{(i)}$ restricted to $\mathbb{N}$ belongs to $\mathcal{F}_{\ell+1-i}$; in particular f restricted to $\mathbb{N}$ belongs to $\mathcal{F}_{\ell+1} \subset \mathcal{F}$.

If we are in the case where f belongs to a Hardy field, then f is also $(\ell+1)$-times continuously differentiable, $f^{(\ell+1)}(x)$ tends monotonically to 0, and $\lim_{x\to\infty} f^{(\ell)}(x) = \pm\infty$. Hence, similar arguments can be used to show that f restricted to $\mathbb{N}$ belongs to $\mathcal{F}_{\ell+1}$. $\square$

Lemma A.2. *Let G be a group, let H be a subgroup of G and let L denote the normal closure of H in G. Define $N := [L, G]$. Then $L = H \cdot N$.*

Proof. Since L is a normal subgroup of G, the group $N = [L, G]$ is a subgroup of L and therefore $H \cdot N \subset L$. To show that $H \cdot N \supset L$ it suffices to show that $H \cdot N$ is a normal subgroup of G, because any normal subgroup of G containing H must also contain L, because L is the normal closure of H.

First, let us show that $H \cdot [L, G]$ is a group. Since L is a normal subgroup of G, the commutator group $N = [L, G]$ is also a normal subgroup of G. It is then easy to check that the product $H \cdot N$ of a group H and a normal subgroup N is itself a group.

Finally, we show that $H \cdot N$ is normal. Let $g \in G$ and $h \in H$. Then

$$ghNg^{-1} \;=\; hgg^{-1}h^{-1}ghNg^{-1} \;=\; hg[g, h]Ng^{-1}.$$

Note that $[g, h] \in N$ and that $gNg^{-1} = N$. Therefore

$$hg[g, h]Ng^{-1} \;=\; hN.$$

It follows that for all $g \in G$ one has

$$g(H \cdot N)g^{-1} = g\left(\bigcup_{h \in H} hN\right)g^{-1} = \bigcup_{h \in H} ghNg^{-1} = \bigcup_{h \in H} hN = H \cdot N.$$

This proves that $H \cdot N$ is normal. $\square$

Lemma A.3. *Let G be a group and denote by $G^\triangle$ the diagonal $\{(g, g) : g \in G\} \subset G^2$. Let L be a normal subgroup of G and define $K := G^\triangle \cdot (L \times L) \subset G \times G$. Then $[K, K] \supset [L, G] \times [L, G]$.*

Proof. We will show that $[L, G] \times \{1_K\} \subset [K, K]$. From this it will follow by symmetry that $\{1_K\} \times [L, G] \subset [K, K]$, which implies that $[L, G] \times [L, G] \subset [K, K]$.

$[L, G]$ is the group generated by the set $\{[l, g] : l \in L, g \in G\}$. It thus suffices to show that $([l, g], 1) \in [K, K]$ for all $l \in L$ and $g \in G$. But observe that

$$([l, g], 1) = \big[(l, 1), (g, g)\big].$$

Since $(g, g) \in G^\triangle \subset K$ and $(l, 1) \in L \times L \subset K$, we conclude that $\big[(l, 1), (g, g)\big] \in [K, K]$ and hence $([l, g], 1) \in [K, K]$. $\square$

The following lemma is similar in spirit to [7, Lemma 1.1].

Lemma A.4. *Let $m \in \mathbb{N}$, let K be a compact subset of a Banach space and let $a \colon \mathbb{N}^m \to K$ be a function. Assume that for every $h_1, \ldots, h_m \in \mathbb{N}$, for every $i = 1, \ldots, m$, and every Følner sequence (F_N) on $\mathbb{N}^i$, the limit*

$$\lim_{N\to\infty} \frac{1}{|F_N|} \sum_{h' \in F_N} a(h', h_{i+1}, \ldots, h_m)$$

exists. Then for every Følner sequence $(F_N)_{N \in \mathbb{N}}$ on $\mathbb{N}^m$,

$$C\text{-}\lim_{h_m\to\infty} C\text{-}\lim_{h_{m-1}\to\infty} \cdots C\text{-}\lim_{h_1\to\infty} a(h_1, \ldots, h_m) = \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{h \in F_N} a(h).$$

Proof. First observe that since the limit $\lim_{N\to\infty} \frac{1}{|F_N|} \sum_{h \in F_N} a(h)$ exists for every Følner sequence, its value does not depend on the choice of Følner sequence. Indeed, if two Følner sequences witnessed different limits, then one could form a third Følner sequence by alternating

the sets in each of the given sequences, thus forming a Følner sequence for which the limit would not exist.

We now prove the lemma by induction on m. For $m = 1$ the statement becomes trivial. Next assume that $m > 1$ and that the conclusion has been established for $m - 1$. Fix a Følner sequence $(F_N)_{N \in \mathbb{N}}$ on $\mathbb{N}^{m-1}$. By induction hypothesis, for every $h_m \in \mathbb{N}$,

$$\underset{h_{m-1}\to\infty}{C\text{-}\lim} \cdots \underset{h_1\to\infty}{C\text{-}\lim} a(h_1,\ldots,h_m) = \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{h'\in F_N} a(h',h_m).$$

For each $H \in \mathbb{N}$, let $N(H) \in \mathbb{N}$ be such that for every $h_m \in \{1,\ldots,H\}$,

$$\left\| \lim_{N\to\infty} \frac{1}{|F_N|} \sum_{h'\in F_N} a(h',h_m) \ - \ \frac{1}{|F_{N(H)}|} \sum_{h'\in F_{N(H)}} a(h',h_m) \right\| < \frac{1}{H},$$

where $\|.\|$ is the norm on the Banach space. Letting $\tilde{F}_H = F_{N(H)} \times \{1,\ldots,H\} \subset \mathbb{N}^m$ we have that $(\tilde{F}_H)_{H\in\mathbb{N}}$ form a Følner sequence. Averaging the previous inequality over all $h_m \in \{1,\ldots,H\}$ and taking $H \to \infty$ we conclude that

$$\underset{h_m\to\infty}{C\text{-}\lim} \ \underset{h_{m-1}\to\infty}{C\text{-}\lim} \cdots \underset{h_1\to\infty}{C\text{-}\lim} a(h_1,\ldots,h_m) = \lim_{N\to\infty} \frac{1}{|\tilde{F}_N|} \sum_{h\in\tilde{F}_N} a(h).$$

$\square$

Combining the previous lemma with Theorems B and B* from [35] we obtain the following.

Theorem A.5. *Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and define $X := G/\Gamma$. Let $g \colon \mathbb{Z}^m \to G$ be a polynomial sequence. The following are equivalent:*
- *$g(\mathbb{Z}^m)\Gamma$ is dense in X;*
- *$\underset{h_m\to\infty}{C\text{-}\lim} \cdots \underset{h_1\to\infty}{C\text{-}\lim} F(g(h_1,\ldots,h_\ell)\Gamma) = \int_X F \, \mathsf{d}\mu_X$ for all $F \in C(X)$;*
- *$\underset{h_m\to\infty}{C\text{-}\lim} \cdots \underset{h_1\to\infty}{C\text{-}\lim} \eta(g(h_1,\ldots,h_\ell)\Gamma) = 0$ for all non-trivial horizontal characters η of G/Γ.*

Remark A.6. Lemma A.4 and Theorem A.5 have more general versions involving uniform Cesàro averages, which coincide with the UW-lim when $W(n) = n$. More precisely, every C-lim appearing in either statement could be replaced with UC-lim, which is defined as

$$\underset{n\to\infty}{UC\text{-}\lim} a(n) := \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} a(n).$$

Lemma A.7 (cf. [17, Lemmas 4.3 and 5.2]). *Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and define $X := G/\Gamma$. For every $g_0,\ldots,g_\ell \in G$ there exists a set $R \subset \mathbb{R}^{\ell+1}$ of full measure such that for every $(\xi_0,\ldots,\xi_\ell) \in R$,*

$$\overline{\{g_0^{\xi_0 n_0} \cdots g_\ell^{\xi_\ell n_\ell}\Gamma : n_0,\ldots,n_\ell \in \mathbb{Z}\}} = \overline{\{g_0^{t_0} \cdots g_\ell^{t_\ell}\Gamma : t_0,\ldots,t_\ell \in \mathbb{R}\}}.$$

Proof. Ratner's theorem ([43, Theorem A]) implies that $\overline{\{g_0^{t_0} \cdots g_\ell^{t_\ell}\Gamma : t_0,\ldots,t_\ell \in \mathbb{R}\}} = \tilde{G}/\tilde{\Gamma}$, where $\tilde{G}$ is a connected and simply connected closed subgroup of G containing $g_0,\ldots,g_\ell$ and $\tilde{\Gamma} = \Gamma \cap \tilde{G}$. Let $T = \tilde{G}/(\tilde{\Gamma}[\tilde{G},\tilde{G}])$ and let $\vartheta \colon \tilde{G}/\tilde{\Gamma} \to T$ be the canonical projection map. In view of [35, Theorem C] we have $\overline{\{g_0^{\xi_0 n_0} \cdots g_\ell^{\xi_\ell n_\ell}\Gamma : n_0,\ldots,n_\ell \in \mathbb{Z}\}} = \tilde{G}/\tilde{\Gamma}$ if and only if

$$(A.1) \qquad \overline{\{\vartheta(g_0^{\xi_0 n_0} \cdots g_\ell^{\xi_\ell n_\ell}\Gamma) : n_0,\ldots,n_\ell \in \mathbb{Z}\}} = T.$$

However T is a torus and ϑ is a homomorphism, so a routine argument shows that the set of $\xi = (\xi_0,\ldots,\xi_\ell) \in \mathbb{R}^{\ell+1}$ for which (A.1) holds has full measure. $\square$

Lemma A.8. *Let G be a connected simply connected nilpotent Lie group, Γ a uniform and discrete subgroup of G and define $X := G/\Gamma$. Let $\ell, m \in \mathbb{N}$, let $g_0, g_1, \ldots, g_\ell \colon \mathbb{Z}^m \to G$ be polynomial sequences and define*

$$Y := \overline{\{g_0(h)^{t_0} g_1(h)^{t_1} \cdot \ldots \cdot g_\ell(h)^{t_\ell} \Gamma : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}\}}.$$

Then Y is a connected subnilmanifold of X.

Proof. For each $h \in \mathbb{Z}^m$, let

$$Y_h := \overline{\{g_0(h)^{t_0} \cdots g_\ell(h)^{t_\ell} \Gamma : (t_0, t_1, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}\}}.$$

Using Lemma A.7 we can find $\xi_0, \ldots, \xi_\ell \in \mathbb{R}$ such that for every $h \in \mathbb{Z}^m$,

$$Y_h = \overline{\{g_0(h)^{\xi_0 n_0} \cdots g_\ell(h)^{\xi_\ell n_\ell} \Gamma : n_0, \ldots, n_\ell \in \mathbb{Z}\}}.$$

Thus

$$Y = \overline{\bigcup_{h \in \mathbb{Z}^m} Y_h} = \overline{\{g_0(h)^{\xi_0 n_0} \cdots g_\ell(h)^{\xi_\ell n_\ell} \Gamma : h \in \mathbb{Z}^m, (n_0, \ldots, n_\ell) \in \mathbb{Z}^{\ell+1}\}}.$$

In particular, Y is the closure of the set $\tilde{g}(\mathbb{Z}^m \times \mathbb{Z}^{\ell+1})$ for some polynomial map $\tilde{g} \colon \mathbb{Z}^m \times \mathbb{Z}^{\ell+1} \to X$. Therefore Leibman's theorem [35] implies that Y is a finite union of subnilmanifolds. However, from the definition, it is clear that Y is connected and therefore it must be a connected subnilmanifold as claimed. $\square$

Next, we derive some useful Lemmas needed for the proof of Theorem 5.5.

Lemma A.9 (Van der Corput lemma)**.** *Let $m \in \mathbb{N} \cup \{0\}$, let $W \in \mathcal{F}_1$ and suppose that $x \colon \mathbb{N}^{m+1} \to \mathbb{C}$ is a bounded complex-valued sequence. If*

$$\underset{h_{m+1} \to \infty}{C\text{-}\overline{\lim}} \, \underset{h_m \to \infty}{C\text{-}\overline{\lim}} \, \cdots \underset{h_1 \to \infty}{C\text{-}\overline{\lim}} \, \underset{n \to \infty}{UW\text{-}\overline{\lim}} \, x(n + h_{m+1}, h_1, \ldots, h_m) \overline{x(n, h_1, \ldots, h_m)} = 0$$

then

$$\underset{h_m \to \infty}{C\text{-}\overline{\lim}} \, \cdots \underset{h_1 \to \infty}{C\text{-}\overline{\lim}} \, \underset{n \to \infty}{UW\text{-}\overline{\lim}} \, x(n, h_1, \ldots, h_m) = 0.$$

Proof. Notice that for any bounded sequence $a \colon \mathbb{N} \to \mathbb{C}$ and $H \in \mathbb{N}$ we have

$$\frac{1}{H^2} \sum_{h_1, h_2 = 1}^{H} a(h_2 - h_1) = \frac{1}{H^2} \sum_{h_1=1}^{H} \sum_{d=1}^{H-h_1} a(d) + a(-d) + O(1/H)$$

$$= \frac{1}{H^2} \sum_{d=1}^{H-1} (H - d)\big(a(d) + a(-d)\big) + O(1/H)$$

where the $O(1/H)$ term accounts for the case $h_1 = h_2$. Hence,

$$\text{(A.2)} \qquad \frac{1}{H^2} \sum_{h_1, h_2 = 1}^{H} a(h_2 - h_1) = \mathbb{E}_{d \in [H]} \left(1 - \frac{d}{H}\right) \big(a(d) + a(-d)\big) + O(1/H).$$

Next, let $(u_n)_{n \in \mathbb{N}}$ be a bounded sequence of complex numbers. Notice that, since ΔW is eventually decreasing, for every $H \leqslant N - M$

$$\mathbb{E}_{n \in [M,N]}^{W} u_n = \mathbb{E}_{n \in [M,N]}^{W} \mathbb{E}_{h \in [H]} u_{n+h} + O\left(\frac{H}{W(N) - W(M)}\right)$$

and so the Cauchy-Schwarz inequality implies that

$$\left|\mathbb{E}^W_{n\in[M,N]}u_n\right|^2 \leqslant \mathbb{E}^W_{n\in[M,N]}\frac{1}{H^2}\sum_{h_1,h_2=1}^{H}u_{n+h_1}\overline{u_{n+h_2}} + \mathrm{O}\left(\frac{H}{W(N)-W(M)}\right)$$

$$= \frac{1}{H^2}\sum_{h_1,h_2=1}^{H}\mathbb{E}^W_{n\in[M,N]}u_n\overline{u_{n+h_2-h_1}} + \mathrm{O}\left(\frac{H}{W(N)-W(M)}\right)$$

where the last equality follows from the general fact that $\mathbb{E}^W_{n\in[M,N]}a(n+h) = \mathbb{E}^W_{n\in[M,N]}a(n) + \mathrm{O}(h/(W(N)-W(M)))$ for any bounded sequence $a:\mathbb{N}\to\mathbb{C}$. Combining the above with (A.2) for $a(d) = \mathbb{E}^W_{n\in[M,N]}u_n\overline{u_{n+d}}$ and observing that $a(-d) = \overline{a(d)} + \mathrm{O}(d/(W(N)-W(M)))$ we get

$$\left|\mathbb{E}^W_{n\in[M,N]}u_n\right|^2 \leqslant \mathbb{E}_{d\in[H]}\left(1-\frac{d}{H}\right)2\mathrm{Re}\mathbb{E}^W_{n\in[M,N]}u_n\overline{u_{n+d}} + \mathrm{O}\left(\frac{H}{W(N)-W(M)}+\frac{1}{H}\right)$$

$$\leqslant 2\mathbb{E}_{d\in[H]}\left|\mathbb{E}^W_{n\in[M,N]}u_n\overline{u_{n+d}}\right| + \mathrm{O}\left(\frac{H}{W(N)-W(M)}+\frac{1}{H}\right).$$

Now, denote by $h=(h_1,\ldots,h_m)$ and let $x(n,h;h_{m+1}) := x(n+h_{m+1},h)\overline{x(n,h)}$. We get

$$U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h) \leqslant \sqrt{2\mathbb{E}_{d\in[H]}U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h;d) + \mathrm{O}\left(\frac{1}{H}\right)}$$

and hence applying again the Cauchy-Schwarz inequality we obtain

$$C\text{-}\overline{\lim_{h_m\to\infty}}\cdots C\text{-}\overline{\lim_{h_1\to\infty}}\,U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h) \leqslant \sqrt{2\mathbb{E}_{d\in[H]}C\text{-}\overline{\lim_{h_m\to\infty}}\cdots C\text{-}\overline{\lim_{h_1\to\infty}}\,U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h;d) + \mathrm{O}\left(\frac{1}{H}\right)}.$$

Since the left hand side does not depend on H, we can let $H\to\infty$ in the right hand side and use the hypothesis to conclude that

$$C\text{-}\overline{\lim_{h_m\to\infty}}\cdots C\text{-}\overline{\lim_{h_1\to\infty}}\,U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h) \leqslant \sqrt{2\lim_{H\to\infty}\mathbb{E}_{d\in[H]}C\text{-}\overline{\lim_{h_m\to\infty}}\cdots C\text{-}\overline{\lim_{h_1\to\infty}}\,U\!W\text{-}\overline{\lim_{n\to\infty}}\, x(n,h;d)} = 0.$$

$\square$

Remark A.10. We observe that the "uniform Cesàro" version of Lemma A.9 (where every instance of C-$\lim$ is replaced with UC-$\lim$) is also true and can be proved in a very similar way.

Lemma A.11. *Suppose $d_1,d_2,\ell,m\in\mathbb{N}$, $T_1 := \mathbb{R}^{d_1}/\mathbb{Z}^{d_1}$ and $T_2 := \mathbb{R}^{d_2}/\mathbb{Z}^{d_2}$. Let $\pi_1:\mathbb{R}^{d_1}\to T_1$ and $\pi_2:\mathbb{R}^{d_2}\to T_2$ denote the corresponding natural factor maps and let $p_0,p_1,\ldots,p_\ell:\mathbb{Z}^m\to \mathbb{R}^{d_1}$ and $q_0,q_1,\ldots,q_{\ell-1}:\mathbb{Z}^m\to\mathbb{R}^{d_2}$ be polynomial sequences such that*

$$\pi_1\big(\{p_0(h)t_0 + p_1(h)t_1 + \ldots + p_\ell(h)t_\ell : h\in\mathbb{Z}^m, (t_0,t_1,\ldots,t_\ell)\in\mathbb{R}^{\ell+1}\}\big) \text{ is dense in } T_1,$$

and

$$\pi_2\big(\{q_0(h)t_0 + q_1(h)t_1 + \ldots + q_{\ell-1}(h)t_{\ell-1} : h\in\mathbb{Z}^m, (t_0,t_1,\ldots,t_{\ell-1})\in\mathbb{R}^{\ell}\}\big) \text{ is dense in } T_2.$$

Define new polynomial sequences $q_1',\ldots,q_\ell':\mathbb{Z}^{m+1}\to\mathbb{R}^{d_2}$ as

$$(A.3)\qquad q_i'(h_1,\ldots,h_{m+1}) := \sum_{j=1}^{i}\binom{h_{m+1}}{j}q_{i-j}(h_1,\ldots,h_m).$$

Set $\tilde{p}_0(h_1, \ldots, h_{m+1}) \coloneqq (p_0(h_1, \ldots, h_m), 0)$ *and, for* $i = 1, \ldots, \ell$,

$$\tilde{p}_i(h_1, \ldots, h_{m+1}) \coloneqq (p_i(h_1, \ldots, h_m), q'_i(h_1, \ldots, h_{m+1})).$$

Then

$$(\pi_1 \otimes \pi_2)\big(\big\{\tilde{p}_0(h')t_0 + \tilde{p}_1(h')t_1 + \ldots + \tilde{p}_\ell(h')t_\ell : h' \in \mathbb{Z}^{m+1}, (t_0, t_1, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}\big\}\big).$$

is dense in $T_1 \times T_2$

Proof. Let

$$T_3 \coloneqq \overline{(\pi_1 \otimes \pi_2)\big(\big\{\tilde{p}_0(h')t_0 + \ldots + \tilde{p}_\ell(h')t_\ell : h' \in \mathbb{Z}^{m+1}, (t_0, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}\big\}\big)}.$$

According to Lemma A.8, T_3 is a subtorus of $T_1 \times T_2$. If T_3 is a proper subtorus of $T_1 \times T_2$ then there exists a non-trivial continuous group character $\chi \colon T_1 \times T_2 \to \{w \in \mathbb{C} : |w| = 1\}$ whose kernel contains T_3. So in order to show that $T_3 = T_1 \times T_2$ it suffices to show that the only continuous group characters χ of $T_1 \times T_2$ whose kernel contains T_3 is the trivial character. Equivalently, it suffices to show that if the function $\psi_\chi \colon \mathbb{Z}^{m+1} \times \mathbb{R}^{\ell+1} \to \mathbb{C}$ given by

$$\psi_\chi(h', t_0, \ldots, t_\ell) = \chi \circ (\pi_1 \otimes \pi_2)\big(\tilde{p}_0(h')t_0 + \tilde{p}_1(h')t_1 + \ldots + \tilde{p}_\ell(h')t_\ell\big)$$

is constant equal to 1, then χ is the trivial character.

Any character χ of $T_1 \times T_2$ can be written as $\chi = \chi_1 \otimes \chi_2$, where χ_1 is a continuous group character of T_1 and χ_2 is a continuous group character of T_2. Hence

$$\begin{aligned} \psi_\chi(h', t_0, \ldots, t_\ell) &= \big((\chi_1 \circ \pi_1) \otimes (\chi_2 \circ \pi_2)\big)\big(\tilde{p}_0(h')t_0 + \tilde{p}_1(h')t_1 + \ldots + \tilde{p}_\ell(h')t_\ell\big) \\ &= (\chi_1 \circ \pi_1)\big(p_0(h)t_0 + \ldots + p_\ell(h)t_\ell\big)(\chi_2 \circ \pi_2)\big(q'_1(h')t_1 + \ldots + q'_\ell(h')t_\ell\big), \end{aligned}$$

where $h' = (h_1, \ldots, h_{m+1})$ and $h = (h_1, \ldots, h_m)$. Pick $\tau^{(1)} \in \mathbb{Z}^{d_1}$ such that $\chi_1 \circ \pi_1(x) = e^{2\pi i \langle x, \tau^{(1)} \rangle}$ for all $x \in \mathbb{R}^{d_1}$ and, similarly, pick $\tau^{(2)} \in \mathbb{Z}^{d_2}$ such that $\chi_2 \circ \pi_2(x) = e^{2\pi i \langle x, \tau^{(2)} \rangle}$ for all $x \in \mathbb{R}^{d_2}$. Since $\psi_\chi(h', t_0, \ldots, t_\ell) = 1$ for all $h' = (h_1, \ldots, h_{m+1}) \in \mathbb{Z}^{m+1}$ and $(t_0, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}$, we have that

$$\langle p_0(h)t_0 + p_1(h)t_1 + \ldots + p_\ell(h)t_\ell, \tau^{(1)} \rangle + \langle q'_1(h')t_1 + \ldots + q'_\ell(h')t_\ell, \tau^{(2)} \rangle \in \mathbb{Z}$$

for all $h' = (h_1, \ldots, h_{m+1}) \in \mathbb{Z}^{m+1}$ and all $(t_0, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}$. It follows that

$$(A.4) \qquad\qquad\qquad \langle p_0(h), \tau^{(1)} \rangle = 0$$

and

$$(A.5) \qquad\qquad \langle p_i(h), \tau^{(1)} \rangle + \langle q'_i(h'), \tau^{(2)} \rangle = 0, \qquad \forall i \in \{1, \ldots, \ell\}.$$

Note that (A.3) and (A.5) together imply

$$(A.6) \qquad \langle p_i(h), \tau^{(1)} \rangle = \sum_{j=1}^{i} \binom{h_{m+1}}{j} \langle -q_{i-j}(h), \tau^{(2)} \rangle, \qquad \forall h \in \mathbb{Z}^m, \ \forall h_{m+1} \in \mathbb{Z}.$$

By comparing coefficients of the variable h_{m+1} we see that (A.6) can only hold if

$$(A.7) \quad \langle p_i(h), \tau^{(1)} \rangle = 0 \quad \text{and} \quad \langle q_{i-j}(h), \tau^{(2)} \rangle = 0, \qquad \forall i \in \{1, \ldots, \ell\}, \forall j \in \{1, \ldots, i\}.$$

It follows that $\langle p_i(h), \tau^{(1)} \rangle = 0$ for all $i = 0, 1, \ldots, \ell$ and $\langle q_i(h), \tau^{(2)} \rangle = 0$ for all $i = 0, 1, \ldots, \ell-1$. Since $\pi_1\big(\big\{p_0(h)t_0 + p_1(h)t_1 + \ldots + p_\ell(h)t_\ell : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_\ell) \in \mathbb{R}^{\ell+1}\big\}\big)$ is dense in T_1, we conclude that $\tau^{(1)} = 0$ and, likewise, since $\pi_2\big(\big\{q_0(h)t_0 + q_1(h)t_1 + \ldots + q_{\ell-1}(h)t_{\ell-1} : h \in \mathbb{Z}^m, (t_0, t_1, \ldots, t_{\ell-1}) \in \mathbb{R}^\ell\big\}\big)$ is dense in T_2, we have that $\tau^{(2)} = 0$. This proves that $\chi = \chi_1 \otimes \chi_2$ is the trivial character. $\qquad\square$

References

[1] V. Bergelson. Ergodic Ramsey theory. In *Logic and combinatorics (Arcata, Calif., 1985)*, volume 65 of *Contemp. Math.*, pages 63–87. Amer. Math. Soc., Providence, RI, 1987.

[2] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems*, 7(3):337–349, 1987.

[3] V. Bergelson. Ergodic Ramsey theory–an update. In *Ergodic theory of $\mathbb{Z}^d$ actions*, volume 228 of *London Math. Soc. Lecture Note Ser.*, pages 1–61. Cambridge Univ. Press, Cambridge, 1996.

[4] V. Bergelson, H. Furstenberg, and R. McCutcheon. IP-sets and polynomial recurrence. *Ergodic Theory Dynam. Systems*, 16(5):963–974, 1996.

[5] V. Bergelson and I. J. Håland Knutson. Weak mixing implies weak mixing of higher orders along tempered functions. *Ergodic Theory Dynam. Systems*, 29(5):1375–1416, 2009.

[6] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[7] V. Bergelson and A. Leibman. Cubic averages and large intersections. In *Recent trends in ergodic theory and dynamical systems*, volume 631 of *Contemp. Math.*, pages 5–19. Amer. Math. Soc., Providence, RI, 2015.

[8] V. Bergelson, A. Leibman, and E. Lesigne. Intersective polynomials and the polynomial Szemerédi theorem. *Adv. Math.*, 219(1):369–388, 2008.

[9] V. Bergelson and R. McCutcheon. Uniformity in the polynomial Szemerédi theorem. In *Ergodic theory of $\mathbb{Z}^d$ actions (Warwick, 1993–1994)*, volume 228 of *London Math. Soc. Lecture Note Ser.*, pages 273–296. Cambridge Univ. Press, Cambridge, 1996.

[10] V. Bergelson and R. McCutcheon. An ergodic IP polynomial Szemerédi theorem. *Mem. Amer. Math. Soc.*, 146(695):viii+106, 2000.

[11] M. Boshernitzan, G. Kolesnik, A. Quas, and M. Wierdl. Ergodic averaging sequences. *J. Anal. Math.*, 95:63–103, 2005.

[12] Q. Chu, N. Frantzikinakis, and B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. Lond. Math. Soc. (3)*, 102(5):801–842, 2011.

[13] J. Cigler. Some remarks on the distribution mod 1 of tempered sequences. *Nieuw Arch. Wisk. (3)*, 16:194–196, 1968.

[14] M. Drmota and R. F. Tichy. A note on Riess means and well distribution. *Math. Nachr.*, 139:267–279, 1988.

[15] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

[16] N. Frantzikinakis. Some open problems on multiple ergodic averages. Preprint, available at https://arxiv.org/abs/1103.3808v1.

[17] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Anal. Math.*, 109:353–395, 2009.

[18] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. Anal. Math.*, 112:79–135, 2010.

[19] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Trans. Amer. Math. Soc.*, 367(8):5653–5692, 2015.

[20] N. Frantzikinakis and M. Wierdl. A Hardy field extension of Szemerédi's theorem. *Adv. Math.*, 222(1):1–43, 2009.

[21] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.

[22] H. Furstenberg. Poincaré recurrence and number theory. *Bull. Amer. Math. Soc. (N.S.)*, 5(3):211–234, 1981.

[23] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J., 1981.

[24] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978.

[25] R. Giuliano Antonini and G. Grekos. Weighted uniform densities. *J. Théor. Nombres Bordeaux*, 19(1):191–204, 2007.

[26] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.

[27] G. H. Hardy. Properties of Logarithmico-Exponential Functions. *Proc. London Math. Soc. (2)*, 10:54–90, 1912.

[28] G. H. Hardy. *Orders of infinity. The Infinitärcalcül of Paul du Bois-Reymond*. Hafner Publishing Co., New York, 1971. Reprint of the 1910 edition, Cambridge Tracts in Mathematics and Mathematical Physics, No. 12.

[29] E. Hlawka. Zur formalen Theorie der Gleichverteilung in kompakten Gruppen. *Rend. Circ. Mat. Palermo (2)*, 4:33–47, 1955.

[30] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[31] B. Host and B. Kra. *Nilpotent structures in ergodic theory*, volume 236 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2018.

[32] A. Khintchine. Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.*, 109(1):604–615, 1934.

[33] Ulrich Krengel. *Ergodic theorems*, volume 6 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.

[34] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. John Wiley & Sons, New York-London-Sydney, 1974.

[35] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):215–225, 2005.

[36] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):201–213, 2005.

[37] A. Leibman. Rational sub-nilmanifolds of a compact nilmanifold. *Ergodic Theory Dynam. Systems*, 26(3):787–798, 2006.

[38] A. Leibman. Multiple polynomial correlation sequences and nilsequences. *Ergodic Theory Dynam. Systems*, 30(3):841–854, 2010.

[39] A. Ostrowski. Über ganzwertige Polynome in algebraischen Zahlkörpern. *J. Reine Angew. Math.*, 149:117–124, 1919.

[40] G. M. Petersen. 'Almost convergence' and uniformly distributed sequences. *Quart. J. Math. Oxford Ser. (2)*, 7:188–191, 1956.

[41] G. Pólya. Über ganzwertige Polynome in algebraischen Zahlkörpern. *J. Reine Angew. Math.*, 149:97–116, 1919.

[42] M. S. Raghunathan. *Discrete subgroups of Lie groups*. Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68.

[43] M. Ratner. Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.*, 63(1):235–280, 1991.

[44] A. Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978.

[45] P. Schatte. On well-distribution with respect to weighted means. *Math. Nachr.*, 138:125–130, 1988.

[46] N. A. Shah. Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements. In *Lie groups and ergodic theory (Mumbai, 1996)*, volume 14 of *Tata Inst. Fund. Res. Stud. Math.*, pages 229–271. Tata Inst. Fund. Res., Bombay, 1998.

[47] E. Szemerédi. On the sets of integers containing no k elements in arithmetic progressions. *Acta Arith.*, 27:299–345, 1975.

Vitaly Bergelson
The Ohio State University
bergelson.1@osu.edu

Joel Moreira
University of Warwick
joel.moreira@warwick.ac.uk

Florian K. Richter
Northwestern University
fkr@northwestern.edu