OPTIMAL LOWER BOUNDS FOR MULTIPLE RECURRENCE

SEBASTIÁN DONOSO, ANH NGOC LE, JOEL MOREIRA, AND WENBO SUN

Abstract. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system, $A \in \mathcal{B}$ and $\epsilon > 0$. We study the largeness of sets of the form

$$S = \{n \in \mathbb{N} \colon \mu(A \cap T^{-f_1(n)}A \cap T^{-f_2(n)}A \cap \ldots \cap T^{-f_k(n)}A) > \mu(A)^{k+1} - \epsilon\}$$

for various families $(f_1, \ldots, f_k)$ of functions $f_i : \mathbb{N} \to \mathbb{Z}$.

For $k \leq 3$ and $f_i(n) = if(n)$, we show that S has positive density if $f(n) = q(p_n)$ where $q \in \mathbb{Z}[x]$ satisfies $q(1) = 0$ and (p_n) is the sequence of primes; or when f is a Hardy field sequence. If T^q is ergodic for some $q \in \mathbb{N}$, then for all $r \in \mathbb{Z}$, S is syndetic if $f(n) = qn + r$.

For $f_i(n) = a_i n$, where a_i are distinct integers, we show that S can be empty for $k \geq 4$, and for $k = 3$ we found an interesting relation between the largeness of S and the existence (and abundance) of solutions to certain linear equations in sparse sets of integers. We also provide several partial results when the f_i are distinct polynomials.

1. Introduction

1.1. Historical background. The classical Poincaré recurrence theorem states that for every measure preserving system $(X, \mathcal{B}, \mu, T)$ and every set $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists some $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$. This result was improved by Khintchine in [17], who showed that under the same conditions, for every $\epsilon > 0$, the set

$$S := \{n : \mu(A \cap T^{-n}A) > \mu(A)^2 - \epsilon\}$$

is *syndetic*, meaning that it has bounded gaps. Taking a mixing system, one sees that the bound $\mu(A)^2$ is optimal.

In [14], Furstenberg established a multiple recurrence theorem, showing that for every measure preserving system $(X, \mathcal{B}, \mu, T)$, every $k \in \mathbb{N}$ and set $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists a syndetic set $S \subset \mathbb{N}$ such that for all $n \in S$, we have

$$\mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{1}$$

One could hope to improve Furstenberg's multiple recurrence theorem in the same way that Khintchine's theorem strengthens Poincaré's. Since for a system mixing of all orders, the left hand side of (1) approaches $\mu(A)^{k+1}$ as $n \to \infty$, one could hope that under the same conditions as Furstenberg's multiple recurrence theorem, for every $\epsilon > 0$, the set

$$\{n : \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > \mu(A)^{k+1} - \epsilon\} \tag{2}$$

is syndetic. This was showed to be true by Furstenberg when the system is weakly mixing, and the general case was finally settled by Bergelson Host and Kra in [3], who showed that if the system $(X, \mathcal{B}, \mu, T)$ is ergodic, then the set in (2) is syndetic when $k = 1, 2, 3$ (with the case $k = 1$ following from Khintchine's theorem). However, the set in (2) may be empty if the system is not ergodic or if $k \geq 4$:

Theorem 1.1 (See [3, Theorems 2.1 and 1.3])**.** *There exist a (non ergodic) measure preserving system $(X,\mathcal{B},\mu,T)$ and for each $\ell\in\mathbb{N}$ a set $A\in\mathcal{B}$ with $\mu(A)>0$ such that for every $n\in\mathbb{N}\setminus\{0\}$*

$$\mu(A\cap T^{-n}A\cap T^{-2n}A)\leq\frac{1}{2}\mu(A)^{\ell}$$

There exist a totally ergodic measure preserving system $(X,\mathcal{B},\mu,T)$ and for each $\ell\in\mathbb{N}$ a set $A\in\mathcal{B}$ with $\mu(A)>0$ such that for every $n\in\mathbb{N}\setminus\{0\}$

$$\mu(A\cap T^{-n}A\cap\cdots\cap T^{-4n}A)\leq\frac{1}{2}\mu(A)^{\ell}$$

The first part of this theorem explains why one needs to focus on ergodic systems when studying optimal recurrence.

Furstenberg's multiple recurrence theorem has been extended in several different directions, each leading to the question of whether (or under which conditions) can optimal recurrence be achieved. In this paper, we are mostly concerned with expressions of the form

$$\mu(A\cap T^{f_1(n)}A\cap T^{f_2(n)}A\cap\cdots\cap T^{f_k(n)}A)$$

for various families $(f_1,\ldots,f_k)$ of functions $f_i:\mathbb{N}\to\mathbb{Z}$. In most cases where recurrence has been established, optimal recurrence can be obtained for weakly mixing systems (cf. [1] when the f_i are polynomials and [2, 8, 9] for more general f_i), or when the functions are "independent" (see [12, 13] for the case of linearly independent polynomials and [9] for more general f_i with different growth). In the general case, besides the aforementioned paper [3], the main progress was obtained by Frantzikinakis in [6], where the case when $k\leq 3$ and the f_i are polynomials is studied in detail.

1.2. Optimal recurrence along $(T^{f(n)},T^{2f(n)},\ldots,T^{kf(n)})$.

Our first result concerns the sequence $(p_n)_{n\in\mathbb{N}}$ of primes and answers a question of Kra. Multiple recurrence along polynomials evaluated at primes was established by Frantzikinakis, Host and Kra in [10, 11]. Our result states that one can also obtain optimal recurrence in this setting.

Theorem 1.2. *Let $(p_n)_{n\in\mathbb{N}}$ be the (increasing) enumeration of the primes, let $(X,\mathcal{B},\mu,T)$ be an invertible ergodic measure preserving system and let $f\in\mathbb{Z}[x]$ be such that $f(1)=0$. Then for every $A\in\mathcal{B}$, $\epsilon>0$ and $k\in\{1,2,3\}$, the set*

$$\big\{n\in\mathbb{N}\colon\mu(A\cap T^{f(p_n)}A\cap T^{2f(p_n)}A\cap\cdots\cap T^{kf(p_n)}A)>\mu(A)^{k+1}-\epsilon\big\}\tag{3}$$

has positive lower density[1].

Theorem 1.2 follows from the stronger Theorem 3.2 below. We remark that the set in (3) is not syndetic in general. In fact, it follows from [22] that when $f(x)=x-1$, for every non-trivial finite system, there exists $A\in\mathcal{B}$ such that the set in (3) has unbounded gaps.

A similar result can be obtained if the sequence $f(p_n)$ is replaced with the sequence $\lfloor f(n)\rfloor$, where $\lfloor x\rfloor$ is the largest integer not greater than x, and f is a function belonging to a Hardy field with polynomial growth and sufficiently far away from $\mathbb{Q}[x]$. More precisely, denote by $\mathcal{G}$ the set of all equivalence classes of smooth functions $\mathbb{R}\to\mathbb{R}$, where $f\sim g$ if there exists a constant $c>0$ such that $f(x)=g(x)$ for all $x>c$. A *Hardy field* is a subfield of the ring $(\mathcal{G},+,\times)$ which is closed under differentiation. Let $\mathcal{H}$ be the union of all Hardy fields. We say that a function $a(x)$ has polynomial growth if there exists $d\in\mathbb{N}$ such that $a(x)/x^d\to 0$.

[1]The lower density $\underline{d}(E)$ of a set $E\subset\mathbb{N}$ is the number $\underline{d}(E)=\liminf\limits_{N\to\infty}\frac{|E\cap\{1,\ldots,N\}|}{N}$.

Theorem 1.3. *Let $a \in \mathcal{H}$ have polynomial growth and satisfy $\left|a(x) - cp(x)\right| / \log(x) \to \infty$ for every $c \in \mathbb{R}$, $p \in \mathbb{Z}[x]$. Then for every invertible ergodic measure preserving system $(X, \mathcal{B}, \mu, T)$, every $A \in \mathcal{B}$, every $\epsilon > 0$ and every $k \in \{1, 2, 3\}$, the set*

$$\left\{n \in \mathbb{N} \colon \mu(A \cap T^{\lfloor a(n) \rfloor} A \cap T^{2\lfloor a(n) \rfloor} A \cap \cdots \cap T^{k\lfloor a(n) \rfloor} A) > \mu(A)^{k+1} - \epsilon\right\} \tag{4}$$

has positive lower density.

Theorem 1.3 is proved in Section 3.2. Examples of functions that satisfy the conditions in the previous theorem are $a(x) = x^c$ where $c > 0$, $c \notin \mathbb{Z}$, $a(x) = x \log x$, $a(x) = x^2\sqrt{2} + x\sqrt{3}$, and $a(x) = x^3 + (\log x)^3$. We point out that in Theorem 1.3 we cannot replace "has positive density" by "is syndetic". This is easy to see for certain functions $a(x)$ growing slowly (for instance $a(x) = x^c$ when $c < 1$). For such functions, $\lfloor a(n) \rfloor$ is constant in arbitrarily long intervals and takes every value which is large enough. Therefore there are gaps of the set (4) which are arbitrarily long. On the other hand, we expect the set in (4) to be *thick*, i.e. contain arbitrarily long intervals, whenever $a \in \mathcal{H}$ has polynomial growth. Some evidence in this direction is given in [4], where the set $\left\{n \in \mathbb{N} \colon \mu(A \cap T^{\lfloor a(n) \rfloor} A \cap \cdots \cap T^{k\lfloor a(n) \rfloor} A) > 0\right\}$ is shown to be thick, as well as the set in (4) when $k = 1$.

Our third result concerns sequences of the form $f(n) = qn + r$ for fixed $q, r \in \mathbb{Z}$ and was suggested by Kra.

Theorem 1.4. *Let $q, r \in \mathbb{Z}$, with $q > 0$, and $(X, \mathcal{B}, \mu, T)$ be a measure preserving system with T^q ergodic. Let $A \in \mathcal{B}$, $\epsilon > 0$ and $k \in \{1, 2, 3\}$. Then the set*

$$\left\{n \in \mathbb{N} \colon \mu(A \cap T^{-(qn+r)} A \cap T^{-2(qn+r)} A \cap \cdots \cap T^{-k(qn+r)} A) > \mu(A)^{k+1} - \epsilon\right\} \tag{5}$$

is syndetic.

Theorem 1.4 follows from Theorem 3.5 below, which deals with a more general situation involving Beatty sequences. Observe that the conclusion of Theorem 1.4 is equivalent to the statement that the intersection

$$\left\{n \in \mathbb{N} \colon \mu(A \cap T^{-n} A \cap T^{-2n} A \cap \cdots \cap T^{-kn} A) > \mu(A)^{k+1} - \epsilon\right\} \cap \left(q\mathbb{Z} + r\right)$$

is syndetic.

If in (5) one replaces the optimal lower bound $\mu(A)^{k+1} - \epsilon$ with 0, then the set is syndetic for any $k \in \mathbb{N}$. This was proved in [15] for $k = 2$ and $k = 3$, and for larger k this is essentially the content of [5, Corollary 6.5]; see also [21].

1.3. **Optimal recurrence along $(T^{-a_1 n}, T^{-a_2 n}, \ldots, T^{-a_d n})$.** Next we study obtaining optimal recurrence for the expression

$$\mu(T^{-a_1 n} A \cap T^{-a_2 n} A \cap \cdots \cap T^{-a_d n} A)$$

where $a_1, \ldots, a_d$ are distinct integer numbers. In particular, if $a_i = i$, then the results of Bergelson, Host and Kra tell us that we have optimal recurrence if and only if $d \leq 4$. More generally, in [6] it is proved that if $d \leq 3$, or $d = 4$ and $a_2 + a_3 = a_1 + a_4$, then optimal recurrence holds, but any other case is not known.

Expanding an argument of Ruzsa, presented in the appendix of [3], we prove that for $d \geq 5$, one does not have optimal recurrence.

Theorem 1.5. *Let $a_1 < \ldots < a_5$ be pairwise distinct integers. There exists an ergodic system $(X, \mathcal{B}, \mu, T)$ such that for every $\ell > 0$, there exists a set $A \in \mathcal{B}$ with $\mu(A) > 0$ such that*

$$\mu(T^{-a_1 n}A \cap T^{-a_2 n}A \cap \ldots \cap T^{-a_5 n}A) \leq \frac{1}{2}\mu(A)^\ell$$

for every non-zero integer n.

Theorem 1.5 is proved in Section 4.1. The cases not covered by the above results seem difficult to address. For instance, it is not known whether for every ergodic measure preserving system, every set A and every $\epsilon > 0$ there exists (a syndetic set of) n for which

$$\mu(A \cap T^{-2n}A \cap T^{-3n}A \cap T^{-4n}A) > \mu(A)^4 - \epsilon.$$

In [6], Frantzikinakis showed that a positive answer to this question would imply the existence of solutions to a certain linear equation in sparse sets. We obtain a converse result, showing a tight connection between optimal lower bounds for multiple recurrence and solutions to linear equations in sparse sets. In order to formulate our result, we need to introduce some notation.

Definition 1.6. Let $m, d, N \in \mathbb{N}$. Denote $[N] := \{0, 1, \ldots, N-1\}$. Given a set $E \subseteq [N]^m$ and a subspace $V \subseteq \mathbb{Q}^{d\times m}$, denote

$$D_{m,N}(V, E) = \frac{|V \cap E^d|}{|V \cap [N]^{d\times m}|} \qquad d_{m,N}(E) = \frac{|E|}{N^m}.$$

Observe that a point $(x_1, \ldots, x_{dm}) \in [N]^{d\times m}$ belongs to V if and only if the coordinates $x_1, \ldots, x_{dm}$ satisfy some system of linear equations. The reader should think of $D_{m,N}(V, E)$ as the proportion of solutions to that system of equations with all variables in E.

Definition 1.7. A subset $S \subseteq \mathbb{N}$ is a *Bohr*$_0$ if there exist $d \in \mathbb{N}$, $\rho > 0$ and $\alpha \in \mathbb{T}^d$ such that $S = \{n \in \mathbb{N} : \|n\alpha\|_{\mathbb{T}^d} < \rho\}$, where $\|\cdot\|_{\mathbb{T}^d}$ denotes the distance to the identity in $\mathbb{T}^d$.

Theorem 1.8. *Let $\ell > 4$ and $a_1, \ldots, a_4 \in \mathbb{Z}$ be distinct. Let V be the subspace of $\mathbb{Q}^4$ spanned by $(a_1^i, \ldots, a_4^i)$ for $0 \leq i \leq 2$.*[2] *Let $C > 0$ and suppose that for every $m \in \mathbb{N}$, every sufficiently large N and subset $E \subseteq [N]^m$, we have $D_{m,N}(V^m, E) \geq C d_{m,N}(E)^\ell$. Then for every invertible ergodic system $(X, \mathcal{B}, \mu, T)$ and every $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists a Bohr$_0$ set $S \subseteq \mathbb{N}$ such that*

$$\limsup_{N-M\to\infty} \frac{1}{|S \cap [M, N)|} \sum_{n\in S\cap[M,N)} \mu(T^{a_1 n}A \cap \cdots \cap T^{a_4 n}A) \geq C\left(1 - \frac{4}{\ell}\right)^\ell \mu(A)^\ell.$$

Theorem 1.8 is proved in Section 4.3.

Remark 1.9. It is easy to see that the conclusion of Theorem 1.8 implies that the set

$$\left\{n \in \mathbb{N}\colon \mu(T^{a_1 n}A \cap \cdots \cap T^{a_4 n}A) \geq C\left(1 - \frac{4}{\ell}\right)^\ell \mu(A)^\ell - \epsilon\right\} \tag{6}$$

is syndetic for all $\epsilon > 0$.

Unfortunately, the condition $D_{m,N}(V^m, E) \geq C d_{m,N}(E)^\ell$ seems difficult to verify in concrete instances, even for $m = 1$. We obtain a partial converse to Theorem 1.8 which shows that it is essentially as difficult as establishing optimal lower bounds for the corresponding multiple recurrence problem.

[2] We adopt the convention that if some a_i equals to 0, then $a_i^0 = 0^0 = 1$.

Theorem 1.10. *Let $\ell \geq 4$ and $a_1, \ldots, a_4 \in \mathbb{Z}$ be distinct. Let V be the subspace of $\mathbb{Q}^4$ spanned by $(a_1^i, \ldots, a_4^i)$ for $0 \leq i \leq 2$. Let $C > 0$ and suppose that for every ergodic system $(X, \mathcal{B}, \mu, T)$ and every $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists a Bohr$_0$ set $S \subseteq \mathbb{N}$ such that*

$$\limsup_{N-M\to\infty} \frac{1}{|S \cap [M,N)|} \sum_{n \in S \cap [M,N)} \mu(T^{a_1 n} A \cap \cdots \cap T^{a_4 n} A) \geq C\mu(A)^\ell.$$

Then for every $m \in \mathbb{N}$, every sufficiently large N and every $E \subseteq [N]^m$,

$$D_{m,N}(V^m, E) \geq C\beta^m d_{m,N}(E)^\ell,$$

where $\beta > 0$ is an explicit constant depending only on $a_1, \ldots, a_4$ and ℓ.

Theorem 1.10 is proved in Section 4.2.

We provide some examples to illustrate Theorems 1.8 and 1.10.

Example 1.11. $(a_1, a_2, a_3, a_4) = (0, 2, 3, 4)$. In this case V is the $\mathbb{Q}$-span of $(1, 1, 1, 1)$, $(0, 2, 3, 4)$ and $(0, 4, 9, 16)$. It is not hard to show that

$$V = \big\{(x, y, z, w) \in \mathbb{Q}^4 \colon x - 6y + 8z - 3w = 0\big\}.$$

For convenience set $m = 1$. Then $D_{1,N}(V, E)$ is essentially the density of solutions of the equation $x - 6y + 8z - 3w = 0$ in E, i.e. the proportion of tuples $(x, y, z, w) \in [N]^4$ satisfying $x - 6y + 8z - 3w = 0$ that belong to E^4. The condition in Theorem 1.8 can be rephrased informally as saying that this density can be bounded from below by the ℓ-th power of the density $d_{1,N}(E)$ of the set E.

Example 1.12. Suppose that $a_1 + a_2 = a_3 + a_4$. In this case, an elementary computation shows that

$$V = \big\{(x, y, z, w) \in \mathbb{Q}^4 \colon s(x - y) + t(z - w) = 0\big\},$$

where $s = a_3 - a_4$ and $t = a_2 - a_1$. We can assume, without loss of generality, that both s and t are positive.

Given $N, m \in \mathbb{N}$ and a set $E \subset [N]^m$, denote by $P(n)$ the number of pairs $(x, z) \in E^2$ satisfying $sx + tz = n$ for each $n \in \mathbb{N}^m$. Observe that if $(x, y, z, w) \in E^4 \cap V^m$ then $sx + tz = sy + tw \in \big[(s+t)N\big]^m$. We have

$$\sum_{n \in \left[(s+t)N\right]^m} P(n) = |E|^2 \qquad \text{and} \qquad \sum_{n \in \left[(s+t)N\right]^m} P(n)^2 = |E^4 \cap V^m|.$$

It follows from the Cauchy-Schwarz inequality that $\frac{|E|^4}{N^m (s+t)^m} \leq |E^4 \cap V^m|$, which in turn implies that

$$\beta^m d_{m,N}(E)^4 \leq D_{m,N}(V, E) \tag{7}$$

where $0 < \beta < \lim\limits_{N\to\infty} \frac{N^3}{|V \cap [N]^4|(s+t)}$ (it is easy to see that the limit exists and is positive).

The conclusion (7) also follows from combining Theorem 1.10 with [6, Theorem C].

1.4. **Optimal recurrence along polynomials.** In [6], Frantzikinakis studied in detail the optimal recurrence for polynomial sequences with $k \leq 3$ and dealt with most cases in that regime. However, some stubborn questions remain unanswered. For instance, it is not known if there exists $\ell > 0$ such that for every ergodic system $(X, \mathcal{B}, \mu, T)$, every $A \in \mathcal{B}$ and every $\epsilon > 0$, the set

$$\big\{n \in \mathbb{N} \colon \mu(A \cap T^{-n} A \cap T^{-2n} A \cap T^{-n^2} A) > \mu(A)^\ell - \epsilon\big\} \tag{8}$$

is non-empty (let alone syndetic). Intriguingly enough, if we replace n^2 by n^3 (or by n^d for any $d > 2$) and set $\ell = 4$ in (8), then by Theorem B and Section 4.2 in [6] we deduce that the set obtained is syndetic. We give a partial positive result for a situation where (8) is syndetic.

Proposition 1.13. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system and let $\mathcal{Z}_3$ be the 3-step nilfactor of X (see Section 2 for the definition). Assume that $\mathcal{Z}_3$ is an inverse limit of nilsystems that can be represented as G/Γ, with G is a connected. Then for every $A \in \mathcal{B}$ and $\epsilon > 0$, the set*

$$\left\{n \in \mathbb{N} \colon \mu(A \cap T^{-n}A \cap T^{-2n}A \cap T^{-n^2}A) > \mu(A)^4 - \epsilon\right\}$$

is syndetic.

Remark 1.14. In particular, the hypothesis of Proposition 1.13 is satisfied if $(X = G/\Gamma, \mathcal{B}, \mu, T)$ is an ergodic nilsystem with G being a connected Lie group.

We are unable to remove the connectedness assumption. Hence the general question regarding optimal recurrence for $(0, n, 2n, n^2)$ remains open. However in next result, we provide an example of lack of optimal recurrence for this family in the case of two commuting transformations.

Proposition 1.15. *There exists a system $(X, \mathcal{B}, \mu, T_1, T_2)$, with T_1 ergodic, $T_1T_2 = T_2T_1$ such that for every integer $\ell > 0$, there exists $A \in \mathcal{B}$ with $\mu(A) > 0$ such that*

$$\mu(A \cap T_1^{-n}A \cap T_1^{-2n}A \cap T_2^{-n^2}A) \leq \frac{1}{2}\mu(A)^\ell$$

for every positive integer n.

Proposition 1.13 and Proposition 1.15 are proved in Section 5.

Acknowledgements. We thank Bryna Kra for proposing to us the questions studied in this paper and for many helpful discussions. We thank the Mathematics Research Communities program of the AMS for giving us the optimal environment to start this project. The first author is grateful for the support of Fondecyt Iniciación Grant 11160061, and the third author is supported by the NSF via grant DMS-1700147.

2. Background

2.1. **Nilmanifolds, nilsystems and nilsequences.** Given a group G, we denote its lower central series by $G = G_1 \triangleright G_2 \triangleright \cdots$, where each term is defined by $G_{i+1} = [G_i, G]$, i.e., G_{i+1} is the subgroup of G generated by all the commutators $[a, b] := aba^{-1}b^{-1}$ with $a \in G_i$ and $b \in G$. The group G is a *k-step nilpotent group* if G_{k+1} is the trivial group.

Let G be a k-step nilpotent Lie group and let Γ be a *uniform* (i.e closed and cocompact) subgroup of G. The compact homogeneous space $X := G/\Gamma$ is called a *k-step nilmanifold*. Let $\pi \colon G \to X$ be the standard quotient map. We write $1_X = \pi(1_G)$ where 1_G is the identity element of G. Denote by G^0 the connected component of G containing the identity 1_G. If X is connected, then $X = \pi(G^0)$.

The space X is endowed with a unique probability measure that is invariant under translations by G. This measure is called the *Haar measure* for X, and denoted by μ_X. For every $\tau \in G$, the measure preserving system $(X, \mathcal{B}, \mu_X, T)$ given by $Tx = \tau \cdot x, x \in X$ is called a *k-step nilsystem*, where $\mathcal{B}$ is the Borel σ-algebra of X.

Let $C(X)$ denote the set of continuous functions on X. For $f \in C(X)$ and $x \in X$, the sequence $\psi(n) := f(T^n x)$ is called a *basic k-step nilsequence*. A *k-step nilsequence* is a uniform limit of basic k-step nilsequences.

We say that a sequence $(x_n)_{n\in\mathbb{N}}$ is *equidistributed* on a nilmanifold X if for every $F \in C(X)$, we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} F(x_n) = \int_X F\, d\mu_X.$$

Similarly, we say that $(x_n)_{n\in\mathbb{N}}$ is *well distributed* on X if

$$\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1} F(x_n) = \int_X F\, d\mu_X.$$

for all $F \in C(X)$.

2.2. **Nilfactors.** Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Suppose $(s_j(n))_{n\in\mathbb{N}}$ is an integer valued sequence for $1 \le j \le k$. A factor $(Y, \mathcal{D}, \nu, S)$ of X is said to be *characteristic for* $(s_1(n), \ldots, s_k(n))$ if for any bounded functions $f_1, ..., f_k$ on X, we have

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k} T^{s_j(n)} f_j - \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k} T^{s_j(n)}\mathbb{E}\left(f_j|Y\right)\right) = 0,$$

where $\mathbb{E}\left(f|Y\right)$ denotes the conditional expectation of f onto the factor Y and the limit is taken in $L^2(X,\mu)$. Host and Kra [16] showed that there exists a characteristic factor for $(n, 2n, \ldots, kn)$ which is an inverse limit of $(k-1)$-step nilsystems. We call this factor the $(k-1)$*-step nilfactor of* X and denote it by $\mathcal{Z}_{k-1}(X)$ (or $\mathcal{Z}_{k-1}$ when there is no confusion).

2.3. **Limit formula for multiple averages on nilsystems.** The following description of the limiting distribution of multiple ergodic averages in nilsystems is essentially due to Ziegler [23].

Theorem 2.1. *Let $a_1, \ldots, a_d \in \mathbb{Z}$ be distinct, let $(X = G/\Gamma, \mathcal{B}, \mu, T)$ be a k-step ergodic nilsystem and let $f_1, f_2, \ldots, f_d \in L^\infty(\mu)$. For each $i = 1, \ldots, k$, let $\Gamma_i = \Gamma \cap G_i$ and let μ_i be the Haar measure of G_i/Γ_i. Then for μ-a.e. $x = g\Gamma \in X$, we have*

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1} f_1(T^{a_1 n}x)\ldots f_d(T^{a_d n}x) = \int_{G_1/\Gamma_1}\cdots\int_{G_k/\Gamma_k}\prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}}\ldots g_k^{\binom{a_i}{k}}\Gamma\right) d\mu_k(g_k\Gamma_k)\ldots d\mu_1(g_1\Gamma_1). \tag{9}$$

Remark 2.2. Theorem 2.1 in particular asserts that the right hand side of (9) does not depend on the choice of representative g_i for the co-set $g_i\Gamma_i$.

Remark 2.3. The statement in [23, Theorem 1.2] requires G to be connected and simply connected. These restrictions were removed in [3, Theorem 5.4], although in that paper the limit is described in a different (but equivalent) form; see also [20, Theorem 6.3].

Let $(X = G/\Gamma, \mathcal{B}, \mu, T)$ be an ergodic nilsystem. Then its Kronecker factor $\mathcal{Z}_1$ is $(G/(G_2\Gamma), T)$. Let $\pi\colon X \to \mathcal{Z}_1$ be the natural projection. Suppose that $Tx = \tau\cdot x$ for all $x \in X$ and some $\tau \in G$. Let α be the projection of τ on $\mathcal{Z}_1$. Define

$$S_\delta := \{n \in \mathbb{N}\colon \alpha^n \in B(\delta)\}, \tag{10}$$

where $B(\delta)$ is the ball in $\mathcal{Z}_1$ centered at 0 with radius δ. Observe that S_δ is a Bohr$_0$ set and by ergodicity, the *uniform density* $d(S_\delta)$ of S_δ is

$$d(S_\delta) := \lim_{N-M\to\infty} \frac{|S_\delta \cap [M,N)|}{N-M} = \mu_{\mathcal{Z}_1}(B(\delta)),$$

where $\mu_{\mathcal{Z}_1}$ is the Haar measure on $\mathcal{Z}_1$.

We need the following proposition, whose proof for case $d = 3$ is sketched in [6, Page 35]. The proof for general d is similar and included here for completeness.

Proposition 2.4. *Let $a_1, \dots, a_d \in \mathbb{Z}$ be distinct, let $(X = G/\Gamma, \mathcal{B}, \mu, T)$ be a k-step ergodic nilsystem and let $f_1, f_2, \dots, f_d \in L^\infty(\mu)$. For each $i = 1, \dots, k$, let $\Gamma_i = \Gamma \cap G_i$ and let μ_i be the Haar measure of G_i/Γ_i. Also, for each $\delta > 0$ let S_δ be defined by (10). Then for μ almost every $x = g\Gamma \in X$, we have:*

$$\lim_{\delta\to 0} \lim_{N-M\to\infty} \frac{1}{|S_\delta \cap [M,N)|} \sum_{n\in S_\delta\cap[M,N)} \prod_{i=1}^{d} f_i(T^{a_i n}x) =$$
$$\int_{G_2/\Gamma_2} \int_{G_2/\Gamma_2} \int_{G_3/\Gamma_3} \cdots \int_{G_k/\Gamma_k} \prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}} \cdots g_k^{\binom{a_i}{k}} \Gamma\right) d\mu_k(g_k\Gamma_k) \dots d\mu_2(g_2\Gamma_1) d\mu_2(g_1\Gamma_1).$$

Proof. Let $\pi : X \to \mathcal{Z}_1$ be the natural projection. For any character χ of the compact abelian group $\mathcal{Z}_1 = X/G_2$, the composition $\chi \circ \pi$ is in $L^\infty(\mu)$, and $\chi \circ \pi(T^n x) = \chi\big(n\alpha + \pi(x)\big)$ for all $n \in \mathbb{N}$ and $x \in X$. On the other hand, $\chi \circ \pi(gh\Gamma) = \chi \circ \pi(g\Gamma)$ whenever $h \in G_2$. By Theorem 2.1, for μ-almost every $x = g\Gamma \in X$, we have

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi\big(n\alpha + \pi(x)\big) \prod_{i=1}^{d} f_i(T^{a_i n}x) =$$
$$\int_{G_1/\Gamma_1} \cdots \int_{G_k/\Gamma_k} \chi\big(\pi(gg_1\Gamma)\big) \prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}} \cdots g_k^{\binom{a_i}{k}} \Gamma\right) d\mu_k(g_k\Gamma_k) \dots d\mu_1(g_1\Gamma_1). \tag{11}$$

As χ is a character of Z, we have $\chi\big(n\alpha+\pi(x)\big) = \chi(n\alpha)\chi_i\big(\pi(x)\big)$, and $\chi\big(\pi(gg_1\Gamma)\big) = \chi\big(\pi(g\Gamma)\big)\chi\big(\pi(g_1\Gamma)\big)$. Note that $x = g\Gamma$. After canceling $\chi\big(\pi(x)\big)$ from both sides of (11), we get:

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi(n\alpha) \prod_{i=1}^{d} f_i(T^{a_i n}x) =$$
$$\int_{G_1/\Gamma_1} \cdots \int_{G_k/\Gamma_k} \chi\big(\pi(g_1\Gamma)\big) \prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}} \cdots g_k^{\binom{a_i}{k}} \Gamma\right) d\mu_k(g_k\Gamma_k) \dots d\mu_1(g_1\Gamma_1). \tag{12}$$

We can approximate the Riemann integrable function $\mathbb{1}_{B(\delta)}$ by finite linear combinations of characters, and so we can replace χ in (12) with $\mathbb{1}_{B(\delta)}$ to get:

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mathbb{1}_{B(\delta)}(n\alpha) \prod_{i=1}^{d} f_i(T^{a_i n}x) =$$
$$\int_{G_1/\Gamma_1} \cdots \int_{G_k/\Gamma_k} \mathbb{1}_{B(\delta)}\big(\pi(g_1\Gamma)\big) \prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}} \cdots g_k^{\binom{a_i}{k}} \Gamma\right) d\mu_k(g_k\Gamma_k) \dots d\mu_1(g_1\Gamma_1). \tag{13}$$

The left hand side of (13) is equal to:

$$m_Z\big(B(\delta)\big) \lim_{N-M\to\infty} \frac{1}{|S_\delta \cap [M,N)|} \sum_{n\in S_\delta\cap[M,N)} \prod_{i=1}^{d} f_i(T^{a_i n}x).$$

On the other hand, the right hand side of (13) is equal to:

$$\int_{\pi^{-1}(B(\delta))}\int_{G_2/\Gamma_2}\cdots\int_{G_k/\Gamma_k}\prod_{i=1}^{d} f_i\left(g g_1^{\binom{a_i}{1}}\cdots g_k^{\binom{a_i}{k}}\Gamma\right) d\mu_k(g_k\Gamma_k)\dots d\mu_1(g_1\Gamma_1).$$

Let μ_δ be the probability measure on X defined by

$$\int_X f\,d\mu_\delta = \frac{1}{\mu_{\mathcal{Z}_1}(B(\delta))}\int_{\pi^{-1}(B(\delta))} f\,d\mu_X \qquad \forall f\in C(X).$$

Since μ_X is invariant under the action of G (and hence of G_2) and the set $\pi^{-1}(B(\delta))$ is invariant under G_2, we have that μ_δ is invariant under the action of G_2. Moreover, any limit point of $\{\mu_\delta : \delta > 0\}$ is supported on G_2/Γ_2. This shows that $\mu_\delta \to \mu_{G_2/\Gamma_2}$ as $\delta\to 0$, where μ_{G_2/Γ_2} is the Haar measure on G_2/Γ_2.

Therefore, dividing both sides of (13) by $\mu_{\mathcal{Z}_1}(B(\delta))$ and taking the limit as $\delta \to 0$, we obtain the desired conclusion. □

We also need the following proposition whose proof is sketched in [6, Page 34].

Proposition 2.5. *Let $(X,\mathcal{B},\mu,T)$ be an ergodic system and define S_δ as in (10). Let $a_1,a_2,a_3\in\mathbb{Z}$ be distinct and $f_1,f_2,f_3\in L^\infty(\mu)$. Assume that $\mathbb{E}(f_i|\mathcal{Z}_2)=0$ for some $1\le i\le 3$. Then*

$$\lim_{N-M\to\infty}\frac{1}{|S_\delta\cap[M,N)|}\sum_{n\in S_\delta\cap[M,N)} f_1(T^{a_1n}x)f_2(T^{a_2n}x)f_3(T^{a_3n}x)=0, \tag{14}$$

where the limit is taken in $L^2(\mu)$.

Proof. Without loss of generality, we assume $\mathbb{E}(f_1|\mathcal{Z}_2)=0$. Let L be the limit on the left hand side of (14) and $d(S_\delta)$ be the Banach density of S_δ. Then

$$d(S_\delta)L = \lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}\mathbb{1}_{S_\delta}(n)f_1(T^{a_1n}x)f_2(T^{a_2n}x)f_3(T^{a_3n}x)= \tag{15}$$
$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}\mathbb{1}_{B(\delta)}(n\alpha)f_1(T^{a_1n}x)f_2(T^{a_2n}x)f_3(T^{a_3n}x).$$

Approximating the Riemann integrable function $\mathbb{1}_{B(\delta)}$ by linear combinations of characters, it suffices to show

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}\chi(n\alpha)f_1(T^{a_1n}x)f_2(T^{a_2n}x)f_3(T^{a_3n}x)=0 \tag{16}$$

for all character χ of $\mathcal{Z}_1$. Note that the limit in the left hand side of (16) is equal to

$$\bar{\chi}(x)\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}\chi(n\alpha+x)f_1(T^{a_1n}x)f_2(T^{a_2n}x)f_3(T^{a_3n}x). \tag{17}$$

By [16, Theorem 1.1 and 12.1], the above limit exists in $L^2(\mu)$ and does not change if we replace f_i by $\mathbb{E}(f_i|\mathcal{Z}_3)$. Therefore, by approximation, we can assume that $(X,\mathcal{B},\mu,T)$ is a 3-step nilsystem.

First suppose that $(X,\mathcal{B},\mu,T)$ is totally ergodic. Then we can assume that its Kronecker factor $\mathcal{Z}_1$ has the form $(G,\mathcal{G},m,\alpha)$, where G is a connected compact abelian group, $\mathcal{G}$ is the Borel σ-algebra, m is the Haar measure and α is the rotation defined above. Since G is connected, there exists $g\in G$ such that $a_2g=\alpha$. Let α/a_2 denote that element. Consider the system $Y=(X\times G,\mathcal{B}\times\mathcal{G},\mu\times m,T\times\alpha/a_2)$. Since $\mathbb{E}(f_1|\mathcal{Z}_2(X))=0$, for almost every ergodic component Y_t of Y we have $\mathbb{E}(f_1\otimes 1|\mathcal{Z}_2(Y_t))=0$ (one way to verify is to show $\|f_1 \otimes 1\|_3 = 0$ where $\|\cdot\|_k$ is Host-Kra's seminorm defined in [16]). Hence by [16, Theorem 12.1],

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} (T \times \alpha/a_2)^{a_1 n} f_1 \otimes 1 \cdot (T \times \alpha/a_2)^{a_2 n} f_2 \otimes \chi \cdot (T \times \alpha/a_2)^{a_3 n} f_3 \otimes 1 = 0, \tag{18}$$

where the limit is taken in $L^2(\mu \times m)$. Rewriting the left hand side of (18), we get

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi(n\alpha + y) f_1(T^{a_1 n}x) f_2(T^{a_2 n}x) f_3(T^{a_3 n}x) = \tag{19}$$

$$\chi(y) \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi(n\alpha) f_1(T^{a_1 n}x) f_2(T^{a_2 n}x) f_3(T^{a_3 n}x) = 0$$

for all $y \in G$. Since $\chi(y) \neq 0$ for all $y \in G$, (19) implies (16).

We now return to general situation without the total ergodicity assumption. Let k be the number of connected components of X. Since $(X, \mathcal{B}, \mu, T)$ is ergodic, $(X, \mathcal{B}, \mu, T^k)$ is totally ergodic. For all $0 \leq i \leq k-1$, applying the above argument with $T^k, T^{a_1 i} f_1, T^{a_2 i} f_2, T^{a_3 i} f_3$ replacing T, f_1, f_2, f_3, respectively, we get

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi((kn+i)\alpha) f_1(T^{a_1(kn+i)}x) f_2(T^{a_2(kn+i)}x) f_3(T^{a_3(kn+i)}x) = 0$$

for all character χ of $\mathcal{Z}_1$. Taking the average over all $0 \leq i \leq k-1$, we derive (16). This finishes the proof. $\square$

3. Optimal recurrence along $(T^{-f(n)}, T^{-2f(n)}, \ldots, T^{-kf(n)})$

3.1. **Optimal recurrence for the sequence of shifted primes.** We begin this section by recalling the following classification of certain tuples $(Q_1(n), Q_2(n), Q_3(n))$ of polynomials, introduced in [6].

Definition 3.1. A family of polynomials $Q_1(n), Q_2(n), Q_3(n) \in \mathbb{Z}[n]$ is said to be of *type* (e_1), (e_2) *or* (e_3) if some permutation of them has the form

- (e_1) $\{lq, mq, rq\}$ with $0 \leq l < m < r$ and $l + m \neq r$.
- (e_2) $\{lq, mq, kq^2 + rq\}$
- (e_3) $\{kq^2 + lq, kq^2 + mq, kq^2 + rq\}$

for some $q \in \mathbb{Q}[n]$ and constants $k, l, m, r \in \mathbb{Z}$ with $k \neq 0$.

We prove a stronger version of Theorem 1.2.

Theorem 3.2. *Let $(p_n)_{n\in\mathbb{N}}$ be the increasing enumeration of the primes. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system, $A \in \mathcal{B}$ and $\epsilon > 0$. Suppose Q_1, Q_2, Q_3 are integer polynomials with $Q_i(0) = 0$ for $i = 1, 2, 3$. Then the sets*

$$\{n \in \mathbb{N} : \mu(A \cap T^{-Q_1(p_n-1)}A) > \mu(A)^2 - \epsilon\}$$

and

$$\{n \in \mathbb{N} : \mu(A \cap T^{-Q_1(p_n-1)}A \cap T^{-Q_2(p_n-1)}A) > \mu(A)^3 - \epsilon\}$$

have positive lower density. Moreover, the set

$$\{n \in \mathbb{N} : \mu(A \cap T^{-Q_1(p_n-1)}A \cap T^{-Q_2(p_n-1)}A \cap T^{-Q_3(p_n-1)}A) > \mu(A)^4 - \epsilon\} \tag{20}$$

also has positive lower density unless the polynomials are pairwise distinct and of type $(e_1), (e_2)$ or (e_3).

Proof. We only prove that the set in (20) has positive density under the given hypothesis, as the proofs for the other two sets are similar. Fix $\epsilon > 0$ and assume that the family Q_1, Q_2, Q_3 is not of type $(e_1), (e_2)$ nor (e_3). Denote

$$c(n) = \mu(A \cap T^{-Q_1(n)}A \cap T^{-Q_2(n)}A \cap T^{-Q_3(n)}A)$$

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

By [19, Theorem 4.1], the sequence $c(n)$ can be decomposed as $c(n) = \psi(n) + \delta(n)$, where $\psi(n)$ is a nilsequence and

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} |\delta(n)| = 0. \tag{21}$$

By [18, Theorem 1.1], we also have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\delta(p_n - 1)| = 0. \tag{22}$$

Since a nilsequence is a uniform limit of basic nilsequences, there exists a basic nilsequence $F(b^n 1_Y)$ such that $|\psi(n) - F(b^n 1_Y)| < \epsilon/4$ for all $n \in \mathbb{N}$. Here F is a continuous function on a nilmanifold $Y = G/\Gamma$, $b \in G$ acts ergodically on Y and $1_Y = \pi(1_G) \in Y$. Assume that Y has d connected components and Y_0 is the component containing 1_Y. Observe that $b^{dn}1_Y \in Y_0$ for all $n \in \mathbb{N}$. Since the polynomial family $Q_1(n), Q_2(n), Q_3(n)$ is not of the types $(e_1), (e_2)$ nor (e_3), the polynomial family $P_1(n) = Q_1(dn), P_2(n) = Q_2(dn), P_3(n) = Q_3(dn)$ is also not of these types. Hence by [6, Theorem C], the set $S = \{n \in \mathbb{N} : c(dn) > \mu(A)^4 - \epsilon/4\}$ is syndetic. Together with (21), we get

$$\lim_{N\to\infty} \frac{1}{|[N] \cap S|} \sum_{n\in[N]\cap S} |\delta(dn)| = 0,$$

which implies

$$\limsup_{N\to\infty} \frac{1}{|[N] \cap S|} \sum_{n\in[N]\cap S} |c(dn) - F(b^{dn}1_Y)| < \epsilon/4.$$

We deduce that there exists an n such that $F(b^{dn}1_Y) > \mu(A)^4 - \epsilon/2$.

Since $b^{dn}1_Y \in Y_0$ and F is continuous, there is an open subset U of Y_0 such that $F > \mu(A)^4 - 3\epsilon/4$ on U. By [18, Corollary 1.4], the sequence $b^{p_n-1}1_Y$ is equidistributed on Y_0 when restricted to $p_n \equiv 1 \mod d$. Hence the set $R := \{n \in \mathbb{N} : b^{p_n-1}1_Y \in U\}$ has positive density, and for every $n \in R$ we have $F(b^{p_n-1}1_Y) > \mu(A)^4 - 3\epsilon/4$. On the other hand, from (22) it follows that the set $R' := \{n \in R : c(p_n - 1) < \mu(A)^4 - \epsilon\}$ has 0 density. Therefore the set $R \setminus R'$ has positive density and is contained in the set (20). This finishes the proof. □

3.2. **Optimal recurrence for Hardy sequences.** We prove a slight generalization of Theorem 1.3.

Theorem 3.3. *Let $a \in \mathcal{H}$ have polynomial growth and satisfy $|a(x) - cp(x)|/\log(x) \to \infty$ for every $c \in \mathbb{R}$, $p \in \mathbb{Z}[x]$. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system, $A \in \mathcal{B}$ and $\epsilon > 0$. Let $0 \le l \le m \le r \in \mathbb{Z}$. Then the sets*

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor a(n)\rfloor}A) > \mu(A)^2 - \epsilon\}$$

and

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor a(n)\rfloor}A \cap T^{-m\lfloor a(n)\rfloor}A) > \mu(A)^3 - \epsilon\}$$

have positive lower density. If $r = l + m$ then the set

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor a(n)\rfloor}A \cap T^{-m\lfloor a(n)\rfloor}A \cap T^{-r\lfloor a(n)\rfloor}A) > \mu(A)^4 - \epsilon\} \tag{23}$$

also has positive lower density.

Proof. The proof is similar to that of Theorem 3.2. Let

$$c(n) = \mu(A \cap T^{-ln}A \cap T^{-mn}A \cap T^{-rn}A)$$

and write $c(n) = \psi(n) + \delta(n)$, where $\psi(n)$ is a nilsequence and $\delta(n)$ satisfies

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} |\delta(n)| = 0.$$

Let $F(b^n 1_Y)$ be an approximation of $\psi(n)$ as in proof of Theorem 3.2. By [18, Theorem 1.1],

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\delta(\lfloor a(n)\rfloor)| = 0.$$

This implies

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |c(\lfloor a(n)\rfloor) - F(b^{\lfloor a(n)\rfloor} 1_Y)| < \epsilon/4. \tag{24}$$

As in the proof of Theorem 3.2, there is an open subset U of Y such that $F > \mu(A)^4 - 3\epsilon/4$ on U. By [7, Theorem 1.2], the sequence $(b^{\lfloor a(n)\rfloor} 1_X)$ is equidistributed on Y. Hence the set $\{n \in \mathbb{N} : F(b^{\lfloor a(n)\rfloor} 1_Y) > \mu(A)^4 - 3\epsilon/4\}$ has positive lower density. This fact combined with (24) implies the set $\{n \in \mathbb{N} : c(\lfloor a(n)\rfloor) > \mu(A)^4 - \epsilon\}$ has positive lower density. □

Remark 3.4. In above proof, we do not utilize the fact that the open set U is inside the identity component Y_0. This is because the orbit along $\lfloor a(n)\rfloor$ is equidistributed on the entire Y. On the other hand, the orbit along primes minus 1 is only equidistributed on some connected components of Y (Y_0 is one of them).

3.3. **Optimal recurrence for Beatty sequences.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The *discrete spectrum* $\sigma(T)$ is the set of eigenvalues $\theta \in \mathbb{T} := \mathbb{R}/\mathbb{Z}$ for which there exists a non-zero eigenfunction $f \in L^2(\mu)$ satisfying $f(Tx) = e^{2\pi i\theta} f(x)$ for μ-almost every $x \in X$.

Given a measure preserving system $(X, \mathcal{B}, \mu, T)$, the transformation T^q is ergodic if and only if $\sigma(T) \cap \langle 1/q\rangle = \{0\}$, where $\langle a\rangle$ denotes the abelian group generated by a, as we view $1/q$ as an element of the group $\mathbb{T}$. Theorem 1.4 follows from the next result.

Theorem 3.5. *Let $\theta, \gamma \in \mathbb{R}$ with $\theta > 0$ and $(X, \mathcal{B}, \mu, T)$ be an ergodic system whose discrete spectrum $\sigma(T)$ satisfies $\sigma(T) \cap \langle\theta^{-1}\rangle = \{0\}$. Let $0 \le l \le m \le r \in \mathbb{Z}$. Then for any $A \in \mathcal{B}$ and $\epsilon > 0$, the sets*

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor\theta n+\gamma\rfloor}A) > \mu(A)^2 - \epsilon\}$$

and

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor\theta n+\gamma\rfloor}A \cap T^{-m\lfloor\theta n+\gamma\rfloor}A) > \mu(A)^3 - \epsilon\}$$

are syndetic. If $r = l + m$ then the set

$$\{n \in \mathbb{N} : \mu(A \cap T^{-l\lfloor\theta n+\gamma\rfloor}A \cap T^{-m\lfloor\theta n+\gamma\rfloor}A \cap T^{-r\lfloor\theta n+\gamma\rfloor}A) > \mu(A)^4 - \epsilon\}$$

is also syndetic.

Proof. If $0 < \theta \leq 1$, then the set $S = \{\lfloor \theta n + \gamma \rfloor : n \in \mathbb{N}\}$ is co-finite in $\mathbb{N}$ and the conclusion follows from [3, Theorem 1.2].

Assume $\theta > 1$. Define $c(n), \psi(n), \delta(n), Y = G/\Gamma, F(b^n 1_Y)$ as in the proof of Theorem 3.3. Then by [21, Theorem 2.1], the discrete spectrum of (Y, μ_Y, b) is contained in the discrete spectrum of (X, μ, T). Hence $\sigma(b) \cap \langle \theta^{-1} \rangle = \{0\}$.

Claim 3.6. *The sequence $(b^{\lfloor \theta n + \gamma \rfloor} 1_Y)_{n \in \mathbb{N}}$ is well distributed on Y.*

Proof. Let $F \in C(Y)$. It suffices to show

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} F(b^{\lfloor \theta n + \gamma \rfloor} 1_Y) = \int_Y F \, d\mu_Y. \tag{25}$$

Let $S = \{\lfloor \theta n + \gamma \rfloor : n \in \mathbb{N}\}$. Then an integer m belongs to S if $m = \lfloor \theta n + \gamma \rfloor$ for some $n \in \mathbb{N}$. This is equivalent to

$$\theta n + \gamma - 1 < m \leq \theta n + \gamma$$

or

$$n - \frac{1-\gamma}{\theta} < m\theta^{-1} \leq n + \frac{\gamma}{\theta}$$

for some $n \in \mathbb{N}$. This is equivalent to $\{m\theta^{-1}\} \in J$, where $J = [0, \gamma/\theta] \cup ((1-\gamma)/\theta, 1)$.

Let $W = \overline{\{n\theta^{-1} \mod 1 : n \in \mathbb{N}\}}$. Then W is a closed subgroup of $\mathbb{T}$. Since $\sigma(b) \cap \langle \theta^{-1} \rangle = \{0\}$, for any $F \in C(Y)$ and $G \in C(W)$, we have that

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{m=M}^{N-1} F(b^m 1_Y) G(m\theta^{-1}) = \int_Y F \, d\mu_Y \int_W G \, d\mu_W.$$

Approximating the Riemann integrable function $\mathbb{1}_{J\cap W}$ by continuous functions, we then get

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{m=M}^{N-1} F(b^m 1_Y) \mathbb{1}_{J\cap W}(m\theta^{-1}) = \mu_W(J \cap W) \int_Y F \, d\mu_Y,$$

or equivalently

$$\frac{1}{\mu_W(J\cap W)} \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{m=M}^{N-1} F(b^m 1_Y) \mathbb{1}_{J\cap W}(m\theta^{-1}) = \int_Y F \, d\mu_Y. \tag{26}$$

Note that $\{m\theta^{-1}\} \in J \cap W$ if and only if $m \in S$, and the uniform density of S is exactly $\mu_W(J \cap W)$. Therefore the left hand side of (26) is the same as the left hand side of (25) This proves our claim. $\square$

As pointed out in the proof of Theorem 3.2, there is an open set U of Y such that $F > \mu(A)^4 - 3\epsilon/4$ on U. Since $(b^{\lfloor \theta n + \gamma \rfloor} 1_X)_{n \in \mathbb{N}}$ is well distributed on Y, the set $S = \{n \in \mathbb{N} : F(b^{\lfloor \theta n + \gamma \rfloor} 1_X) > \mu(A)^4 - 3\epsilon/4\}$ is syndetic. Since the sequence $(\delta(n))_{n \in \mathbb{N}}$ tends to zero in the uniform density, and the set $\{\lfloor \theta n + \gamma \rfloor : n \in S\}$ has positive uniform density, we have

$$\lim_{N-M\to\infty} \frac{1}{|S \cap [M,N)|} \sum_{n \in S \cap [M,N)} |\delta(\lfloor \theta n + \gamma \rfloor)| = 0.$$

Hence

$$\limsup_{N-M\to\infty} \frac{1}{|S \cap [M,N)|} \sum_{n \in S \cap [M,N)} |c(\lfloor \theta n + \gamma \rfloor) - F(b^{\lfloor \theta n + \gamma \rfloor} 1_Y)| < \epsilon/4.$$

Since $F(b^{\lfloor \theta n + \gamma \rfloor} 1_Y) > \mu(A)^4 - 3\epsilon/4$ when $n \in S$, we get that the set of $n \in S$ such that $c(\lfloor \theta n + \gamma \rfloor) > \mu(A)^4 - \epsilon$ is syndetic. This finishes the proof. $\square$

4. Optimal recurrence along $(T^{-a_1 n}, T^{-a_2 n}, \ldots, T^{-a_k n})$

4.1. **Poor lower bound for $k = 5$.** In this subsection, we prove Theorem 1.5. We adapt the proof of Theorem 1.3 in [3].

We use the measure preserving system $(X, \mathcal{B}, \mu, T)$ where $X = \mathbb{T}^2$ is the 2-dimensional torus, μ is the Haar measure, and $T(x, y) = (x + \alpha, y + 2x + \alpha)$ for some irrational $\alpha \in \mathbb{R}$. It is well known that this system is totally ergodic. For every $n \in \mathbb{Z}$ and every point $(x, y) \in \mathbb{T}^2$ a quick computation shows that $T^n(x, y) = (x + n\alpha, y + 2nx + n^2\alpha)$.

Let $\ell > 1$. We take a suitably large $L, C \in \mathbb{N}$ and a set $\Lambda \subset \{0, \ldots, L-1\}$ to be chosen later. Let

$$B := \bigcup_{b \in \Lambda} I_b \quad \text{where } I_b := \left[\frac{b}{CL}, \frac{b}{CL} + \frac{1}{C^2 L}\right)$$

and let $A = \mathbb{T} \times B$. For each $n \in \mathbb{Z}$, in order for a point (x, y) to belong to $A \cap T^{-a_1 n} A \cap \cdots \cap T^{-a_4 n} A$ we need $y_i := y + 2a_i nx + a_i^2 n^2 \alpha \in B$ for each $i = 1, \ldots, 5$. Let $b_i \in \Lambda$ be such that $y_i \in I_{b_i}$.

We now need the following elementary lemma.

Lemma 4.1. *Let $a_1, \ldots, a_4 \in \mathbb{Z}$ be distinct and let M be the 4×3 matrix whose (i, j) entry is a_i^j for $i = 1, \ldots, 4$ and $j = 0, 1, 2$. For each $i = 1, \ldots, 4$, let v_i be $(-1)^i$ times the determinant of the matrix obtained from M by deleting the ith row. Then for every quadratic polynomial $f \in \mathbb{R}[x]$,*

$$v_1 f(a_1) + v_2 f(a_2) + v_3 f(a_3) + v_4 f(a_4) = 0.$$

Proof. The claim amounts to the statement that the matrix

$$\begin{bmatrix} f(a_1) & 1 & a_1 & a_1^2 \\ f(a_2) & 1 & a_2 & a_2^2 \\ f(a_3) & 1 & a_3 & a_3^2 \\ f(a_4) & 1 & a_4 & a_4^2 \end{bmatrix}$$

has determinant 0. But this follows from the fact that any quadratic polynomial is a linear combination of the polynomials $1, x, x^2$. $\square$

In view of Lemma 4.1, there exist integers $v_1, \ldots, v_4$ and $\tilde{v}_2, \ldots \tilde{v}_5$ (depending only on $a_1, \ldots, a_5$) such that $v_1 y_1 + \cdots + v_4 y_4 = 0$ and $\tilde{v}_2 y_2 + \cdots + \tilde{v}_5 y_5 = 0$. Therefore, if C is large enough, then also $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$, as it will be an integer which can be made smaller than 1 when C is large enough.

Suppose now that Λ does not contain any solution to $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$ except when $b_1 = \cdots = b_5$. Then, if $(x, y) \in A \cap T^{-a_1 n} A \cap \cdots \cap T^{-a_4 n} A$, all the y_i must belong to the same I_b, which implies that $x \in X_n$, where X_n is the set of points $x \in \mathbb{T}$ satisfying $\|2n(a_2 - a_1)x\|_{\mathbb{T}} < 1/C^2 L$. Since $y_1 \in B$, the point y must belong to the set $B - 2a_1 nx - a_1^2 n^2 \alpha$, which being a shift of B has the same measure as B. We conclude that

$$\mu(A \cap T^{-a_1 n} A \cap \cdots \cap T^{-a_4 n} A) \leq \mu_{\mathbb{T}}(X_n) \mu_{\mathbb{T}}(B) = \frac{2}{C^4 L^2} |\Lambda|.$$

Since $\mu(A) = |\Lambda| \frac{1}{C^2 L}$, a quick computation now shows that the proof will be complete once we construct a set $\Lambda \subset \{0, \ldots, L-1\}$ with $|\Lambda| > L^{1-1/\ell}$ and without non-constant solutions to $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$. The existence of such a set Λ is provided by the following lemma.

Lemma 4.2. *Let $a_1, \ldots, a_5 \in \mathbb{Z}$ be pairwise distinct and let v_i and $\tilde{v}_i$ be described in the paragraph after Lemma 4.1. For every $\epsilon > 0$ and every large enough $L \in \mathbb{N}$, there exists a set $\Lambda \subset \{0, \ldots, L-1\}$ with*

$|\Lambda| > L^{1-\epsilon}$ *such that the only* $b_1, \ldots, b_5 \in \Lambda$ *satisfying* $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$ *also satisfy* $b_1 = \cdots = b_5$.

Lemma 4.2 is a generalization of [3, Theorem 2.4], due to I. Ruzsa, corresponding to $a_i = i$. The key to proving Lemma 4.2 is the following intermediate result.

Lemma 4.3. *Let* $a_1 < \cdots < a_5$ *be integers and let* v_i *and* $\tilde{v}_i$ *be as described above. Let* $d \in \mathbb{N}$ *and let* $b_1, \ldots, b_5 \in \mathbb{R}^d$ *all have the same Euclidean norm. If* $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$ *then* $b_1 = \cdots = b_5$.

Unfortunately Lemma 4.3 does not hold for arbitrary v_i and $\tilde{v}_i$, as seen by the example $v_1 = v_3 = \tilde{v}_3 = \tilde{v}_5 = 1$ and $v_2 = v_4 = \tilde{v}_2 = \tilde{v}_4 = -1$ which would provide a counterexample with $d = 1$ and $b_1 = b_2 = b_5 = 1$ and $b_3 = b_4 = -1$. Indeed we will need to use the description of the v_i and $\tilde{v}_i$ given by Lemma 4.1 and this makes the proof somewhat cumbersome.

Proof. The condition $a_1 < \cdots < a_5$ implies that $v_1, v_3 > 0$ and $v_2, v_4 < 0$. Let

$$S := \frac{v_1 b_1 + v_3 b_3}{v_1 + v_3}, \qquad A := b_1 - S, \qquad B = b_2 - S.$$

Applying Lemma 4.1 to a constant polynomial, we get that $v_1 + v_2 + v_3 + v_4 = 0$ and hence, together with $v_1 b_1 + \cdots + v_4 b_4 = 0$, that $S = (v_2 b_2 + v_4 b_4)/(v_2 + v_4)$. Then we have

$$b_1 = S + A, \qquad b_2 = S + B, \qquad b_3 = S - \frac{v_1}{v_3} A, \qquad b_4 = S - \frac{v_2}{v_4} B.$$

Our goal is to show that $b_1 = b_2 = b_3 = b_4 = S$, and so it suffices to show that $A = B = 0$ (the fact that also $b_5 = S$ would then immediately follow from the equation $\tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$). Since the quantity $\|b_i\|^2 - \|S\|^2$ does not depend on i, we find that the following 4 numbers are equal

$$\|A\|^2 + 2\langle S, A\rangle, \qquad \|B\|^2 + 2\langle S, B\rangle, \qquad \frac{v_1^2}{v_3^2}\|A\|^2 - \frac{2v_1}{v_3}\langle S, A\rangle, \qquad \frac{v_2^2}{v_4^2}\|B\|^2 - \frac{2v_2}{v_4}\langle S, B\rangle. \tag{27}$$

Equality between the first and third gives $2\langle S, A\rangle = \|A\|^2 \left(\frac{v_1}{v_3} - 1\right)$; equality between the second and fourth gives $2\langle S, B\rangle = \|B\|^2 \left(\frac{v_2}{v_4} - 1\right)$ and then equality between the first two numbers implies

$$\|A\|^2 v_1 v_4 = \|B\|^2 v_2 v_3. \tag{28}$$

In order to show that $A = B = 0$, we first show that B is a positive scalar multiple of A. Once we do that, we have from (28) that $B = \sqrt{\frac{v_1 v_4}{v_2 v_3}} A$ and hence, equality between the first and last quantities from (27) (together with $2\langle S, A\rangle = \|A\|^2 \left(\frac{v_1}{v_3} - 1\right)$) gives

$$\|A\|^2 \frac{v_1}{v_3} = \frac{v_1 v_2}{v_3 v_4}\|A\|^2 - 2\sqrt{\frac{v_1 v_2}{v_3 v_4}}\langle S, A\rangle \quad \iff \quad \|A\|^2 \sqrt{\frac{v_1}{v_3}} \left(\sqrt{\frac{v_1}{v_3}} - \sqrt{\frac{v_2}{v_4}}\right)\left(1 + \sqrt{\frac{v_1 v_2}{v_3 v_4}}\right) = 0.$$

This implies that $A = 0$ unless $\frac{v_1}{v_3} = \frac{v_2}{v_4}$. Using the description of each v_i from Lemma 4.1 as a Vandermonde determinant, this is equivalent to $(a_1 - a_4)^2 = (a_2 - a_3)^2$. Since we are assuming that $a_1 < a_2 < a_3 < a_4$ this can not happen and hence $A = 0$.

We have reduced the proof to showing that B is a positive scalar multiple of A. It is now the time to use the fact that also $\tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$. From Lemma 4.1, we deduce that $\tilde{v}_5 = v_1$, and so we can write b_5 in terms of S, A, B as

$$b_5 = \frac{1}{v_1}(-\tilde{v}_2 b_2 - \tilde{v}_3 b_3 - \tilde{v}_4 b_4) = S + \frac{\tilde{v}_3}{v_3} A + \left(\frac{v_2 \tilde{v}_4 - \tilde{v}_2 v_4}{v_4 v_1}\right) B = S + \alpha A + \beta B,$$

where $\alpha := \frac{\tilde{v}_3}{v_3}$ and $\beta := \frac{v_2\tilde{v}_4 - \tilde{v}_2 v_4}{v_4 v_1}$. Using the relations established above to write $\|B\|^2$, $\langle S, A\rangle$ and $\langle S, B\rangle$ in terms of $\|A\|^2$ we compute

$$\|b_5\|^2 - \|S\|^2 = 2\alpha\beta\langle A, B\rangle + \|A\|^2\left[\alpha^2 + \alpha\left(\frac{v_1}{v_3} - 1\right) + \left(\beta^2 + \beta\left(\frac{v_2}{v_4} - 1\right)\right)\frac{v_1 v_4}{v_3 v_2}\right].$$

Since $\|b_5\| = \|b_1\|$, we deduce that $\|b_5\|^2 - \|S\|^2 = \|A\|^2\frac{v_1}{v_3}$. After a somewhat tedious computation, we eventually arrive at $\langle A, B\rangle = \|A\|^2\sqrt{\frac{v_1 v_2}{v_3 v_4}} = \|A\| \cdot \|B\|$. But this implies that B must be a positive scalar multiple of A as desired, finishing the proof. ☐

Proof of Lemma 4.2. Let $C = |v_1| + \cdots + |v_4| + |\tilde{v}_2| + \cdots + |\tilde{v}_5|$, let $d > 2/\epsilon$ be a natural number and then let $m \in \mathbb{N}$ be large enough multiple of C depending only on C, d and ϵ (in fact, we need that $m^{d\epsilon - 2} > C^{d-2}d$). Set $L = m^d$. We can express each number in $\{0, \ldots, L-1\}$ using d digits in base m expansion. Let

$$F := \left\{x_0 + x_1 m + \ldots + x_{d-1}m^{d-1} : x_i \in \left[0, \ldots, \frac{m}{C}\right)\right\}.$$

We have $|F| = (m/C)^d$. Let $r : F \to \mathbb{N}$ be the sum of the squares of the digits in base m, in other words, $r(x_0 + x_1 m + \ldots + x_{d-1}m^{d-1}) = x_0^2 + \cdots + x_{d-1}^2$. Then $r(F) \subset [0, dC^2/m^2)$. Therefore there exists $r_0 \in [0, dC^2/m^2)$ such that

$$\Lambda := \{x \in F : r(x) = r_0\}$$

has cardinality $|\Lambda| \geq (m/C)^{d-2}/d$. The choice of parameters above yields $|\Lambda| > L^{1-\epsilon}$.

Finally, suppose that $b_1, \ldots, b_5 \in \Lambda$ satisfy $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$. We identify each b_i with the vector in $\mathbb{R}^d$ obtained from its digits in base m. Then $\|b_1\| = \cdots = \|b_5\|$. Since each digit in b_i is at most m/C, there is no carryover when multiplying by v_i or $\tilde{v}_i$ and thus, the equations $v_1 b_1 + \cdots + v_4 b_4 = \tilde{v}_2 b_2 + \cdots + \tilde{v}_5 b_5 = 0$ apply even when multiplication and addition is being performed in $\mathbb{R}^d$. Applying Lemma 4.3, we conclude that indeed $b_1 = \cdots = b_5$ as desired. ☐

4.2. **Lack of solutions implies poor lower bounds for $k = 4$.** In this subsection we prove Theorem 1.10. We need the following well known equidistribution result whose short proof we include for completeness.

Lemma 4.4. *For every Bohr$_0$ set S, every $\alpha \in \mathbb{R}^m$ whose coordinates are rationally independent, and every cube $I \subseteq \mathbb{T}^m$, we have that*

$$\lim_{N-M\to\infty} \frac{\left|\left\{n \in S \cap [M, N) \colon n^2\alpha \bmod \mathbb{Z}^m \in I\right\}\right|}{|S \cap [M, N)|} = \mu_{\mathbb{T}^m}(I).$$

Proof. By assumption, we can write $S = \{n \in \mathbb{N} \colon nx \in U\}$, where K is a compact abelian group, $U \subseteq K$ is a neighborhood of 1_K such that $\mathbb{1}_U$ is Riemann integrable, and $x \in U$ is a point such that $\overline{\{nx : n \in \mathbb{Z}\}} = K$ and $S = \{n \in \mathbb{N} \colon nx \in U\}$. Then it suffices to show that as $N - M \to \infty$,

$$\frac{|\{n \in S \cap [M, N) \colon n^2\alpha \in I\}|}{N - M} = \frac{|\{n \in [M, N) : (nx, n^2\alpha) \in U \times I\}|}{N - M}$$

converges to $\mu_K(U) \times \mu_{\mathbb{T}^m}(I) = \mu_{K\times\mathbb{T}^m}(U \times I)$, where μ_K, $\mu_{\mathbb{T}^m}$ and $\mu_{K\times\mathbb{T}^m}$ are the Haar measures on K, $\mathbb{T}^m$ and $K \times \mathbb{T}^m$, respectively. This follows once we show that the sequence $(nx, n^2\alpha)_{n\in\mathbb{N}}$ is well distributed on $K \times \mathbb{T}^m$. Since $\mathbb{1}_U$ is Riemann integrable, it suffices to show that for every character $\chi : K \to S^1 \subset \mathbb{C}$ of K and every $\mathbf{b} \in \mathbb{Z}^m$, if either χ is non-trivial or $\mathbf{b} \neq \mathbf{0}$, then

$$\lim_{N-M\to\infty} \frac{1}{N - M}\sum_{n=M}^{N-1} e^{2\pi i\mathbf{b}\cdot\alpha n^2}\chi(x)^n = 0. \tag{29}$$

Let $\theta \in \mathbb{T}$ be such that $\chi(x) = e^{2\pi i\theta}$. Since the group generated by x is dense in K, $\theta \notin \mathbb{Z}$ unless χ is trivial. Then we can write $e^{2\pi i \mathbf{b}\cdot\alpha n^2}\chi(x)^n = e^{2\pi i(n^2\mathbf{b}\cdot\alpha+n\theta)}$. If $\mathbf{b} \neq \mathbf{0}$ then (29) follows from Weyl's equidistribution theorem, and if $\mathbf{b} = \mathbf{0}$ then $\theta \notin \mathbb{Z}$ and (29) follows quickly from evaluating the resulting geometric series. □

Proof of Theorem 1.10. Let $\ell, a_1, \ldots, a_4, C$ and V be as in the statement of the theorem. Since V has a basis of rational vectors, there exists a positive constant ϵ such that any point $\mathbf{x} \in \mathbb{Z}^4$ at a distance (say in the ℓ^∞ norm) less than ϵ from the closure $\overline{V} \subset \mathbb{R}^4$ must in fact belong to V. Fix $m, N_0 \in \mathbb{N}$ and $E \subset [N_0]^m$.

Let $X = \mathbb{T}^{2m}$ be the $2m$ dimensional torus endowed with the Lebesgue measure μ. Define the map $T\colon X \to X$ via the formula $T(\mathbf{x}, \mathbf{y}) = (\mathbf{x}+\alpha, \mathbf{y}+2\mathbf{x}+\alpha), \mathbf{x}, \mathbf{y} \in \mathbb{T}^m$ for some $\alpha \in \mathbb{T}^m$ whose coordinates are rationally independent. Then $(X, \mathcal{B}, \mu, T)$ is ergodic, and for each $n \in \mathbb{N}$,

$$T^n(\mathbf{x}, \mathbf{y}) = (\mathbf{x}+n\alpha, \mathbf{y}+2n\mathbf{x}+n^2\alpha).$$

For $\mathbf{i} = (c_1, \ldots, c_m) \in [N_0]^m$, denote

$$B_{\mathbf{i}} = \left[\frac{c_1}{N_0}, \frac{c_1+\epsilon}{N_0}\right) \times \cdots \times \left[\frac{c_m}{N_0}, \frac{c_m+\epsilon}{N_0}\right), \qquad A = \bigcup_{\mathbf{i}\in E} \mathbb{T}^m \times B_{\mathbf{i}}.$$

We can directly compute that $\mu(A) = \frac{\epsilon}{N_0^m}|E| = \epsilon^m d_{m,N_0}(E)$. Let $(\mathbf{x}, \mathbf{y}) \in \mathbb{T}^{2m}$ and $n \in \mathbb{N}$. Note that for all $1 \leq j \leq 4$, $T^{a_j n}(\mathbf{x}, \mathbf{y}) \in A$ if and only if

$$u_j = u_j(\mathbf{x}, \mathbf{y}, n) := \mathbf{y} + 2a_j n\mathbf{x} + a_j^2 n^2\alpha \in B_{\mathbf{i}_j} \tag{30}$$

for some $\mathbf{i}_j \in E$. Fix such a point $(\mathbf{x}, \mathbf{y}, n) \in \mathbb{T}^{2m} \times \mathbb{N}$. Then the vector

$$(u_1, \ldots, u_4) = \mathbf{y}(1, \ldots, 1) + 2n\mathbf{x}(a_1, \ldots, a_4) + n^2\alpha(a_1^2, \ldots, a_4^2) \in B_{\mathbf{i}_1} \times \cdots \times B_{\mathbf{i}_4}$$

belongs to the closure in $\mathbb{R}^{m\times 4}$ of V^m. Since $N_0 u_j$ is at most ϵ away from the integer vector $\mathbf{i}_j$ (in the $\ell^\infty([m])$ distance), from the definition of ϵ we deduce that $(\mathbf{i}_1, \ldots, \mathbf{i}_4)$ belongs to V^m as well. Let W denote the collection of all tuples $(\mathbf{i}_1, \ldots, \mathbf{i}_4) \in V^m$ with $\mathbf{i}_j \in E$. By definition, $D_{m,N_0}(V^m, E) = \frac{|W|}{|V\cap[N_0]^4|^m}$.

By the discussion above,

$$\mu(T^{a_1 n}A \cap \cdots \cap T^{a_4 n}A) = \sum_{(\mathbf{i}_1,\ldots,\mathbf{i}_4)\in W} \int_X \prod_{j=1}^4 \mathbb{1}_{B_{\mathbf{i}_j}}\left(u_j(\mathbf{x}, \mathbf{y}, n)\right) d\mu(\mathbf{x}, \mathbf{y}). \tag{31}$$

Fix $(\mathbf{i}_1, \ldots, \mathbf{i}_4) \in W$. If $n \in \mathbb{N}$ is such that (30) holds for all $1 \leq i \leq 4$, then considering (30) as a linear equation system with $4m$ equations and the coordinates of $\mathbf{y}, n\mathbf{x}, n^2\alpha$ as unknowns (i.e. $3m$ unknowns in total), we deduce that there exists 3 cubes I_1, I_2 and I_3 in $\mathbb{T}^m$ with side length at most $\frac{1}{N_0}$ such that $\mathbf{y} \in I_1, n\mathbf{x} \in I_2, n^2\alpha \in I_3$.

Lemma 4.4 implies that for any Bohr_0 set S,

$$\lim_{N-M\to\infty} \frac{|\{n \in S \cap [M, N)\colon n^2\alpha \in I_3\}|}{|S \cap [M, N)|} = \mu(I_3) \leq \frac{1}{N_0^m}.$$

Since $\mu(I_1) \leq \frac{1}{N_0^m}$ and $\mu(\{\mathbf{x} \in \mathbb{T}^m : n\mathbf{x} \in I_2\}) = \mu(I_2) \leq \frac{1}{N_0^m}$ for any $n \neq 0$, by (31) we conclude that

$$D_{m,N_0}(V^m, E) = \sum_{(\mathbf{i}_1,\dots,\mathbf{i}_4)\in W} \frac{1}{|V \cap [N_0]^4|^m}$$

$$\geq \frac{N_0^{3m}}{|V \cap [N_0]^4|^m} \sum_{(\mathbf{i}_1,\dots,\mathbf{i}_4)\in W} \limsup_{N-M\to\infty} \frac{1}{|S \cap [M,N)|} \sum_{n\in S\cap[M,N)} \int_X \prod_{j=1}^4 \mathbb{1}_{B_{\mathbf{i}_j}}(u_j(\mathbf{x},\mathbf{y},n))\, d\mu(\mathbf{x},\mathbf{y})$$

$$\geq \frac{N_0^{3m}}{|V \cap [N_0]^4|^m} \limsup_{N-M\to\infty} \frac{1}{|S \cap [M,N)|} \sum_{n\in S\cap[M,N)} \mu(T^{a_1 n}A \cap \cdots \cap T^{a_d n}A)$$

$$\geq C \frac{N_0^{3m}}{|V \cap [N_0]^4|^m} \mu(A)^\ell = C \frac{\epsilon^{m\ell} N_0^{3m}}{|V \cap [N_0]^4|^m} d_{m,N_0}(E)^\ell,$$

which finishes the proof by taking $\beta < \lim_{N\to\infty} \frac{\epsilon^\ell N_0^3}{|V \cap [N_0]^4|}$. ☐

4.3. **Solutions to linear equations imply optimal lower bounds for $k = 4$.** In this section we prove Theorem 1.8. We first need to reformulate the assumptions in terms of functions on a torus; this is the content of Lemma 4.6 below. We start with an estimate from harmonic analysis. Let A be a finite set, $f : A \to \mathbb{R}$ a function and $p > 0$. We denote by $\|f\|_{L^p}$ its usual L^p quasinorm when A is endowed with the normalized counting probability measure, i.e.

$$\|f\|_{L^p} := \left(\frac{1}{|A|} \sum_{a\in A} \big|f(a)\big|^p \right)^{1/p}.$$

We will also make use of the weak L^p quasinorm:

$$\|f\|_{L^p_w} := \sup_{s>0}\ s \cdot \left(\frac{|\{a \in A : |f(a)| > s\}|}{|A|} \right)^{1/p}.$$

We remark that when $p < 1$ these quasinorms do not satisfy the triangle inequality. We will only use these quasinorms with $p < 1$ to invoke the following well known interpolation lemma. We include its short proof for completeness.

Lemma 4.5. *Let $0 < p < r < \infty$ and let A be a finite set. For every function $f : A \to \mathbb{R}$ we have*

$$\|f\|_{L^r}^r \leq \frac{r}{r-p} \|f\|_{L^p_w}^p \|f\|_{L^\infty}^{r-p}.$$

Proof. Combining the identity

$$x^r = \int_0^x r s^{r-1}\, ds = r \int_0^\infty s^{r-1} \mathbb{1}_{[0,x]}(s)\, ds = r \int_0^\infty s^{r-1} \mathbb{1}_{\{x>s\}}\, ds$$

with the definition of L^r norm, we deduce the formula

$$\|f\|_{L^r}^r = r \int_0^\infty s^{r-1} \frac{1}{|A|} \sum_{a\in A} \mathbb{1}_{\{|f(a)|>s\}}\, ds = r \int_0^\infty s^{r-1} \frac{\big|\{a \in A : |f(a)| > s\}\big|}{|A|}\, ds.$$

Finally, using the definition of the weak L^p norm we conclude

$$\|f\|_{L^r}^r \leq r \int_0^\infty s^{r-1-p} \|f\|_{L^p_w}^p\, ds = \frac{r}{r-p} \|f\|_{L^p_w}^p \|f\|_{L^\infty}^{r-p}.$$

☐

The following lemma makes use of the quantities $d_{m,N}(E)$ and $D_{m,N}(V,E)$ introduced in Definition 1.6.

Lemma 4.6 (Equivalent inequalities)**.** *Let $m, d, \ell \in \mathbb{N}$ with $\ell > d$, let $C > 0$, $V \subseteq \mathbb{Q}^d$ be a subspace containing the vector $(1,\dots,1)$ and $\overline{V} \in \mathbb{R}^d$ be its closure in $\mathbb{R}^d$. Then (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4):*

(1) *For every large enough N and every subset $E \subseteq [N]^m$, we have $D_{m,N}(V^m, E) \geq C d_{m,N}(E)^\ell$.*
(2) *For every large enough N and every function $c\colon [N]^m \to [0,1]$, we have*

$$\frac{1}{|V \cap [N]^d|^m} \sum_{a_i \in [N]^m, (a_1,\ldots,a_d) \in V^m} c(a_1)c(a_2)\ldots c(a_d) \geq C\|c\|^d_{L^{d/\ell}_w}.$$

(3) *For every large enough N and every function $c\colon [N]^m \to [0,1]$, we have*

$$\frac{1}{|V \cap [N]^d|^m} \sum_{a_i \in [N]^m, (a_1,\ldots,a_d) \in V^m} c(a_1)c(a_2)\ldots c(a_d) \geq C\left(1 - \frac{d}{\ell}\right)^\ell \|c\|^\ell_{L^1}.$$

(4) *Let $Y = \left(\overline{V}/\mathbb{Z}^d\right)^m$ be a subtorus of $\mathbb{T}^{d\times m}$. For every measurable function $f\colon \mathbb{T}^m \to [0,1]$,*

$$\int_Y f(y_1)f(y_2)\ldots f(y_d)\, d\mu_Y(y_1,\ldots,y_d) \geq C\left(1 - \frac{d}{\ell}\right)^\ell \left(\int_{\mathbb{T}^m} f\, d\mu_{\mathbb{T}^m}\right)^\ell.$$

Remark 4.7.

- Whenever we have a point x in $[N]^{dm}$ (or analogously for $\mathbb{Q}^{dm}$, $\mathbb{T}^{dm}$, etc.) we consider $x = (x_{i,j})_{i=1,\ldots,m,\, j=1,\ldots,d}$ with each $x_{i,j} \in [N]$. We then write $x = (x_1,\ldots,x_d)$ where each $x_i \in [N]^m$ is the vector $x_i = (x_{i,j})_{j=1}^m$. Depending on the context, we may also write $x = (x_1,\ldots,x_m)$, where now each $x_j \in [N]^d$ is the vector $x_j = (x_{i,j})_{i=1}^d$ (it should be clear at any point which vectors we are referring to).

  For instance if $v = (v_{i,j})_{i=1,\ldots,m,\, j=1,\ldots,d} \in \mathbb{Q}^{dm}$ then v is in V^m if for every i the vector $(v_{i,j})_{j=1}^d$ of $\mathbb{Q}^d$ belongs to V; and v is in E^d if for every j the vector $(v_{i,j})_{i=1}^m$ is in E. Similarly in (2) and (3), the statement that $(a_1,\ldots,a_d) \in V^m$ should be interpreted by writing each a_j as $(a_{i,j})_{i=1}^m$ and requiring that each vector $(a_{i,j})_{j=1}^d$ is in V.
- It might be true that (3) and (4) are also equivalent to (1) and (2), but we don't have a proof and it is not needed in this paper.

Proof. (2)⇒ (1). Take $c(a) = \mathbb{1}_E(a)$.

(1)⇒ (2). Let $p := d/\ell$, observe that $\|c\|_{L^p_w} = \frac{1}{N^{m\ell/d}} \sup_{s\geq 0} s|\{a \in [N]^m \colon c(a) > s\}|^{\ell/d}$ and assume that the maximum is obtained at $s = t$. Let $E = \{a \in [N]^m \colon c(a) > t\}$. Since $c \geq t\mathbb{1}_E$, we have

$$\frac{1}{|V \cap [N]^d|^m} \sum_{a_i \in [N]^m, (a_1,\ldots,a_d) \in V} c(a_1)c(a_2)\ldots c(a_d) \geq t^d D_{m,N}(V^m, E).$$

Invoking (1), we get $t^d D_{m,N}(V^m, E) \geq C\frac{t^d|E|^\ell}{N^{\ell m}} = C\|c\|^d_{L^p_w}$.

(2)⇒ (3). We only need to show (3) for $c \neq 0$. By Lemma 4.5,

$$\|c\|^d_{L^p_w} \geq \left(1 - \frac{d}{\ell}\right)^\ell \frac{\|c\|^\ell_{L^1}}{\|c\|^{\ell-d}_{L^\infty}} \geq \left(1 - \frac{d}{\ell}\right)^\ell \|c\|^\ell_{L^1}.$$

(3)⇒(4). Let

$$Y_N = \bigcup_{a \in (V\cap[N]^d)^m} \prod_{i=1}^d \prod_{j=1}^m \left[\frac{a_{i,j}-1}{N}, \frac{a_{i,j}}{N}\right) \subset \mathbb{T}^{dm}$$

and let μ_N be the normalized probability measure supported on Y_N. We claim that $\mu_{N!} \to \mu_Y$ as $N \to \infty$. Indeed, any limit point of the sequence $(\mu_N)_{N\in\mathbb{N}}$ must be supported on Y. Moreover, for any $v \in (V/\mathbb{Z}^d)^m$, if N is larger than the denominators of all coordinates of v, then $\mu_{N!}$ is invariant under v. We conclude that any limit point of the sequence $(\mu_{N!})_{N\in\mathbb{N}}$ is supported on Y and invariant under Y, hence it must be μ_Y.

Now given $f : \mathbb{T}^m \to [0,1]$, let $c : [N]^m \to [0,1]$ be the function

$$c(a) = N^m \int_{[-1/N,0)^m} f(a/N + x)dx.$$

When N is large enough, we have that

$$\begin{aligned}\int_{Y_N} \prod_{i=1}^d f(y_i)\, d\mu_N(y_1,\ldots,y_d) &= \frac{N^{dm}}{|V\cap[N]^d|^m} \sum_{a\in(V\cap[N]^d)^m} \int_{\left[\frac{-1}{N},0\right)^{dm}} \prod_{i=1}^d f\left(\frac{a_i}{N}+x_i\right) d(x_1,\ldots,x_d)\\ &\geq \frac{1}{|V\cap[N]^d|^m} \sum_{a_i\in[N]^m,(a_1,\ldots,a_d)\in V^m} c(a_1)c(a_2)\ldots c(a_d)\\ &\geq C\left(1-\tfrac{d}{\ell}\right)^\ell \|c\|_{L^1}^\ell = C\left(1-\tfrac{d}{\ell}\right)^\ell \left(\int_{\mathbb{T}^m} f\, d\mu_{\mathbb{T}^m}\right)^\ell.\end{aligned}$$

□

We are now ready to prove Theorem 1.8.

Proof of Theorem 1.8. Fix an ergodic system $(X,\mathcal{B},\mu,T)$ and a set $A\in\mathcal{B}$ with $\mu(A)>0$. Let $\mathcal{Z}_2$ be the 2-step nilfactor of X, defined in Section 2.2. Using a standard approximation argument, we can assume that $\mathcal{Z}_2$ is a 2-step nilsystem, so that $\mathcal{Z}_2 = (G/\Gamma, \mu_{\mathcal{Z}_2}, \tau)$, where G is a 2-step nilpotent Lie group, $\Gamma\subset G$ is a uniform subgroup and $\tau\in G$. By a slight abuse of notation, we use $\mathcal{Z}_2$ to denote the measure preserving system, as well as the underlying topological dynamical system and the underlying nilmanifold.

In view of ergodicity, the topological system $\mathcal{Z}_2$ is minimal (see, for instance, [3, Theorem 4.1.1]). We can assume that G is generated by the connected component of the identity and τ. Indeed, the projection of the connected component of G onto $\mathcal{Z}_2 = G/\Gamma$ is an open subset of $\mathcal{Z}_2$ (as its pre-image under the natural map $G\to\mathcal{Z}_2$ is the union of all connected components of G having non-empty intersection with Γ and hence it is open) and by minimality of $\mathcal{Z}_2$ its orbit under τ is all of $\mathcal{Z}_2$. Therefore, if we let $\tilde{G}$ be the subgroup of G generated by the connected component of the identity and τ, it follows that $\mathcal{Z}_2 = \tilde{G}/(\Gamma\cap\tilde{G})$.

Since G is a 2-step nilpotent group, the commutator $G_2=[G,G]$ is inside the center of G, and hence the subgroup $\Gamma_2 = G_2\cap\Gamma$ is normal in G. Therefore $\mathcal{Z}_2 = (G/\Gamma_2)/(\Gamma/\Gamma_2)$ and thus after modding out by Γ_2 we can assume that $G_2\cap\Gamma=\{e\}$, which implies that G_2 is a compact abelian Lie group. From [3, Theorem 4.1.4], it follows that G_2 is connected, and so G_2 must be a finite dimensional torus.

Let K be the quotient $K := \mathcal{Z}_2/G_2 = G/(\Gamma G_2)$ and note that it is also a compact abelian Lie group (but it may be disconnected). Let $\pi\colon G\to K$ be the natural projection, let $a=\pi(\tau)$ and define

$$S_\delta = \{n\in\mathbb{N}\colon a^n\in B(\delta)\},$$

where $B(\delta)$ is the ball in K centered at the identity of K with radius δ. It suffices to show that

$$\lim_{\delta\to 0}\lim_{N-M\to\infty} \frac{1}{|S_\delta\cap[M,N)|}\sum_{n\in S_\delta\cap[M,N)} \int_X \prod_{i=1}^4 f(T^{a_i n}x)\, d\mu(x) \geq C\left(1-\frac{4}{\ell}\right)^\ell \left(\int_X f\, d\mu\right)^\ell \tag{32}$$

for all $0\leq f\leq 1$. By Proposition 2.5, the left hand side of (32) is 0 if we replace at least one of the four f's with $f-\mathbb{E}(f|\mathcal{Z}_2)$. Since $0\leq\mathbb{E}(f|\mathcal{Z}_2)\leq 1$, it suffices to prove (32) under the assumption that $X=\mathcal{Z}_2=G/\Gamma$.

Since G is 2 step nilpotent, by Proposition 2.4, the left hand side of (32) equals to

$$\int_{\mathcal{Z}_2}\int_{G_2}\int_{G_2}\prod_{i=1}^{4} f\left(g g_1^{\binom{a_i}{1}} g_2^{\binom{a_i}{2}}\Gamma\right)\,d\mu_{G_2}(g_2)\,d\mu_{G_2}(g_1)\,d\mu_{\mathcal{Z}_2}(g\Gamma), \tag{33}$$

where μ_X and μ_{G_2} are the Haar measures on X and G_2. Recall that G_2 is a torus, say $G_2 = \mathbb{T}^m$. Consider the subgroup

$$Y := \left\{(y_1,\dots,y_4)\in(\mathbb{T}^m)^4\colon (\exists g_1,g_2\in\mathbb{T}^m)\ y_i = \binom{a_i}{1}g_1 + \binom{a_i}{2}g_2\right\}\subset\mathbb{T}^{4m},$$

where we now changed to the additive notation. Then we may rewrite

$$(33) = \int_{\mathcal{Z}_2}\int_{Y}\prod_{i=1}^{4} f(gy_i\Gamma)\,d\mu_Y(y_1y_2,y_3,y_4)\,d\mu_{\mathcal{Z}_2}(g\Gamma),$$

where μ_Y is the Haar measure on Y. We can also describe Y in terms of V as $Y = \left(\overline{V}/\mathbb{Z}^4\right)^m$, where $\overline{V}$ is the closure V in $\mathbb{R}^4$ (or, equivalently, its $\mathbb{R}$-span).

For each $g\in G$ let $f_g\colon G_2\to\mathbb{R}$ be the function defined by the formula $f_g(g_2\Gamma) = f(gg_2\Gamma)$ for all $g_2\in G_2$. Then by Lemma 4.6, (1) $\Rightarrow$ (4), and then Jensen's inequality, we conclude that

$$\begin{aligned}&\int_{\mathcal{Z}_2}\int_{Y}\prod_{i=1}^{4} f(gy_i\Gamma)\,d\mu_Y(y_1y_2,y_3,y_4)\,d\mu_{\mathcal{Z}_2}(g\Gamma)\geq C\left(1-\frac{4}{\ell}\right)^{\ell}\int_{Z}\left(\int_{G_2} f_g\,d\mu_{G_2}\right)^{\ell}\,d\mu_Z(g\Gamma)\\ &\geq C\left(1-\frac{4}{\ell}\right)^{\ell}\left(\int_{\mathcal{Z}_2}\int_{G_2} f_g\,d\mu_{G_2}\,d\mu_{\mathcal{Z}_2}(g\Gamma)\right)^{\ell} = C\left(1-\frac{4}{\ell}\right)^{\ell}\left(\int_{\mathcal{Z}_2} f\,d\mu_{\mathcal{Z}_2}\right)^{\ell}.\end{aligned}$$

□

5. Optimal recurrence along polynomials

To state our results, we need to introduce a notion defined and studied in detail by Leibman in [20]. The *C-complexity* of a family of integer-valued polynomials $\{p_1,\dots,p_d\}$ is the minimum integer k for which the factor $\mathcal{Z}_k$ is characteristic for this family in every ergodic nilsystem $(G/\Gamma,\mathcal{B},\mu,T)$ with G being connected. Note that the minimum value of k for general ergodic systems is an upper bound of the C-complexity and in some cases it is strictly larger.

Proposition 5.1. *Let $(G/\Gamma,\mathcal{B},\mu,T)$ be an ergodic nilsystem where G is connected. Let $\mathcal{Z}_1$ be its Kronecker factor and let $\alpha\in\mathcal{Z}_1$ be the rotation induced by T. Let $q_1(n)$, $q_2(n)$ be two linearly independent (over $\mathbb{Q}$) integer polynomials with 0 constant term and set $p_1 = aq_1$, $p_2 = bq_2$ and $p_3 = cq_1 + dq_2$, $a,b,c,d\in\mathbb{Z}$. Assume that the C-complexity of the family $\{p_1,p_2,p_3\}$ is equal to one.*

For $\delta>0$, let B_δ be the ball in $\mathcal{Z}_1$ centered at 0 of radius δ and define $S_\delta = \{n\in\mathbb{N} : (q_1(n)\alpha, q_2(n)\alpha)\in B(\delta)\times B(\delta)\}$. Let $f_1,f_2,f_3\in L^\infty(\mu)$ and assume $\mathbb{E}(f_i|\mathcal{Z}_1)=0$ for some $1\leq i\leq 3$. Then

$$\lim_{N-M\to\infty}\frac{1}{|S_\delta\cap[M,N)|}\sum_{n\in S_\delta\cap[M,N)} f_1(T^{p_1(n)}x)f_2(T^{p_2(n)}x)f_3(T^{p_3(n)}x) = 0 \tag{34}$$

where the limit is taken in $L^2(\mu)$.

Proof. The proof is similar to that of Proposition 2.5, subject to some minor changes that we write explicitly. In this proof, all the limits are taken in $L^2(\mu)$. Without loss of generality, we assume $\mathbb{E}(f_1|\mathcal{Z}_1) =$

0. Let L be the limit on the left hand side of (34) and $d(S_\delta)$ be the uniform density of S_δ. Since $\{(q_1(n)\alpha, q_2(n)\alpha)\}$ is well distributed on $\mathcal{Z}_1 \times \mathcal{Z}_1$ (because q_1 and q_2 are independent), we have

$$d(S_\delta)L = \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mathbb{1}_{S_\delta}(n) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x) = \tag{35}$$

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mathbb{1}_{B(\delta)\times B(\delta)}(q_1(n)\alpha, q_2(n)\alpha) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x).$$

Approximating the Riemann integrable function $\mathbb{1}_{B(\delta)\times B(\delta)}$ by finite linear combination of characters, it suffices to show

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi_1(q_1(n)\alpha)\chi_2(q_2(n)\alpha) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x) = 0 \tag{36}$$

for all characters (χ_1, χ_2) of $\mathcal{Z}_1 \times \mathcal{Z}_1$. Note that the limit in the left hand side of (36) is equal to

$$\bar{\chi}_1(y)\bar{\chi}_2(z) \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi_1(q_1(n)\alpha + y)\chi_2(q_2(n)\alpha + z) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x) \tag{37}$$

for every $y, z \in \mathcal{Z}_1$. Since G is connected, there exist $g, h \in G$ such that $ag = \alpha$ and $bh = \alpha$. Let α/a and α/b denote the elements g and h respectively. Consider the system $Y = (X \times \mathcal{Z}_1 \times \mathcal{Z}_1, \mathcal{B} \times \mathcal{G}, \mu \times m \times m, \tilde{T})$, where $\tilde{T} = T \times (\alpha/a) \times (\alpha/b)$. We can write then

$$\begin{aligned} &\bar{\chi}_1(y)\bar{\chi}_2(z) \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \chi_1(q_1(n)\alpha + y)\chi_2(q_2(n)\alpha + z) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x) \\ &= \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \tilde{T}^{p_1(n)} f_1 \otimes 1 \otimes 1 \cdot \tilde{T}^{p_2(n)} f_2 \otimes \chi_1 \otimes 1 \cdot \tilde{T}^{p_3(n)} f_3 \otimes 1 \otimes \chi_2. \end{aligned} \tag{38}$$

Since $\mathbb{E}(f_1|\mathcal{Z}_1(X)) = 0$, for almost every ergodic component Y_t of Y, we have $\mathbb{E}(f_1 \otimes \chi_1 \otimes 1|\mathcal{Z}_1(Y_t)) = 0$ (one way to verify is to show $\|f_1 \otimes \chi_1 \otimes 1\|_2 = 0$ where $\|\cdot\|_k$ is Host-Kra's seminorm defined in [16]).

Since almost every ergodic component Y_t can be written as G_t/Γ_t with G_t being connected, using the assumption that the C-complexity of the family of polynomials $\{p_1, p_2, p_3\}$ is one, we get

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \tilde{T}^{p_1(n)} f_1 \otimes 1 \otimes 1 \cdot \tilde{T}^{p_2(n)} f_2 \otimes \chi_1 \otimes 1 \cdot \tilde{T}^{p_3(n)} f_3 \otimes 1 \otimes \chi_2 = 0 \tag{39}$$

for almost every t. It follows that (37) equals to 0 in $L^2(\mu \times m \times m)$, which implies that the left hand side of (34) is equal to 0 in $L^2(\mu)$. This finishes the proof. $\square$

Proposition 5.2. *Let $(G/\Gamma, \mu, T)$ be a nilsystem with G being connected. Let p_1, p_2, p_3 be three polynomials as in Proposition 5.1. Then for all $A \in \mathcal{B}$ and every $\epsilon > 0$, the set*

$$\{n \in \mathbb{N} \colon \mu(A \cap T^{-p_1(n)}A \cap T^{-p_2(n)}A \cap T^{-p_3(n)}A) > \mu(A)^4 - \epsilon\}$$

is syndetic.

Proof. Let $\epsilon > 0$ and $A \in \mathcal{B}$ and set $f = \mathbb{E}(\mathbb{1}_A|\mathcal{Z}_1)$. Let $\epsilon > 0$ and let $\delta' > 0$ such that the translation $f_t(\cdot) = f(\cdot + t)$ satisfies that $\|f - f_t\|_{L^1(m)} < \frac{\epsilon}{3}$ if $t \in B(\delta')$. Let $\delta' > \delta > 0$ such that if $q_1(n)\alpha$ and $q_2(n)\alpha$ are in $B(\delta)$ then $p_1(n)\alpha, p_2(n)\alpha$ and $p_3(n)\alpha$ are in $B(\delta')$. Then, for $n \in S_\delta$, we have that $\|f - T^{p_i(n)}f\|_{L^1(m)} < \frac{\epsilon}{3}$ for $i = 1, 2, 3$ and thus

(40)
$$\int f \cdot T^{p_1(n)} f \cdot T^{p_2(n)} f \cdot T^{p_3(n)} f \; dm > \int f^4 dm - 3\frac{\epsilon}{3} \geq \mu(A)^4 - \epsilon.$$

By (40) and Proposition 5.1, we get that

(41)
$$\lim_{N-M\to\infty} \frac{1}{|S_\delta \cap [M,N)|} \sum_{n \in S_\delta \cap [M,N)} \mu(A \cap T^{-p_1(n)}(A) \cap T^{-p_2(n)} A \cap T^{-p_3(n)} A) \geq \mu(A)^4 - \epsilon,$$

which finishes the proof. □

Proposition 5.3. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system and let $\mathcal{Z}_3$ be the 3-step nilfactor of X. Assume that $\mathcal{Z}_3$ is the inverse limit of nilsystems that can be represented as G/Γ, where G is a connected 3-step nilpotent Lie group and Γ is a discrete cocompact subgroup. If the C-complexity of the family of polynomials $\{p_1, p_2, p_3\}$ is equal to one, then the set*

$$\{n \in \mathbb{N} \colon \mu(A \cap T^{-p_1(n)} A \cap T^{-p_2(n)} A \cap T^{-p_3(n)} A) > \mu(A)^4 - \epsilon\}$$

is syndetic.

Proof. For $A \in \mathcal{B}$, let $a(n) = \int \mathbb{1}_A \cdot \mathbb{1}_A \circ T^{p_1(n)} \cdot \mathbb{1}_A \circ T^{p_2(n)} \cdot \mathbb{1}_A \circ T^{p_3(n)} d\mu$ and $\tilde{a}(n) = \int \mathbb{E}(\mathbb{1}_A|\mathcal{Z}_3) \cdot \mathbb{E}(\mathbb{1}_A|\mathcal{Z}_3) \circ T^{p_1(n)} \cdot \mathbb{E}(\mathbb{1}_A|\mathcal{Z}_3) \circ T^{p_2(n)} \cdot \mathbb{E}(\mathbb{1}_A|\mathcal{Z}_3) \circ T^{p_3(n)} d\mu$. We claim that the sequence $a(n) - \tilde{a}(n)$ is uniformly-null, meaning that

$$\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} |a(n) - \tilde{a}(n)|^2 = 0.$$

The proof is essentially given in [3, Corollary 4.5]. Using a telescoping difference between $a(n)$ and $\tilde{a}(n)$, it suffices to show that if some $f_i, 0 \leq i \leq 3$ has 0 conditional with respect to $\mathcal{Z}_3(X)$, then

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \left(\int f_0(x) f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) f_3(T^{p_3(n)}x) d\mu \right)^2 = 0.$$

We assume without loss of generality that $\mathbb{E}(f_0|\mathcal{Z}_3(X)) = 0$. Let $\mu \times \mu = \int_Z d\mu_s dm(s)$ be the ergodic decomposition of $\mu \times \mu$ under $T \times T$. By [3, Proposition 4.3], for almost every s, $\mathbb{E}(f_0 \otimes f_0|\mathcal{Z}_2(X \times X)) = 0$, where $X \times X$ is endowed with the measure μ_s and the transformation $T \times T$. By [6, Theorem B], the 2-step nilfactor is characteristic for the average $\limsup\limits_{N-M\to\infty} \frac{1}{N-M} \sum_{i=M}^{N-1} f_1 \circ T^{p_1(n)} \cdot f_2 \circ T^{p_2(n)} \cdot f_3 \circ T^{p_3(n)}$, for any bounded measurable functions f_1, f_2, f_3 of any measure preserving system. Therefore, the limit as $N - M$ goes to infinity of

(42)
$$\frac{1}{N-M} \sum_{n=M}^{N} \int f_0(x) f_0(x') f_1(T^{p_1(n)}x) f_1(T^{p_1(n)}x') f_2(T^{p_2(n)}x) f_2(T^{p_2(n)}x') f_3(T^{p_3(n)}x) f_2(T^{p_3(n)}x') d\mu_s(x, x')$$

is equal to 0 for almost every s. Integrating (42) with respect to s we deduce the claim.

By the claim, it suffices to prove the result under the assumption that $X = \mathcal{Z}_3$. By an approximation argument we can assume that $(X = G/\Gamma, \mathcal{B}, \mu, T)$ where G is connected. Proposition 5.2 give us the desired conclusion. □

Proof of Proposition 1.13. It follows immediately from Proposition 5.3, since the C-complexity of the family $\{n, 2n, n^2\}$ is equal to one. This is computed for instance in [20, Section 9.8]. □

Proof of Proposition 1.15. Let $X=\mathbb{T}^2$, μ be the Lebesgue measure on $\mathbb{T}^2$, $T_1\colon (x,y)\mapsto(x+\alpha,y+2x+\alpha)$, and $T_2\colon(x,y)\mapsto(x,y-2\alpha)$. Then T_1 and T_2 commute, preserve the measure μ, and moreover, T_1 is ergodic for μ.

We have that $T_1^n(x,y)=(x+n\alpha,y+2nx+n^2\alpha)$, $T_1^{2n}(x,y)=(x+2n\alpha,y+4nx+4n^2\alpha)$ and $T_2^{n^2}(x,y)=(x,y-2n^2\alpha)$. Write $u=y+2nx+n^2\alpha$, $v=y+4nx+4n^2\alpha$ and $w=y+2n^2\alpha$. Then $v-2u+w=0$. Let $\Lambda\subseteq[N]$ be a subset with no arithmetic progression of length 3 and set $A=\mathbb{T}\times\bigcup_{a\in\Lambda}(\frac{a}{N}-\frac{1}{4N},\frac{a}{N}+\frac{1}{4N})$.

If $\mathbb{1}_A(x,y)\mathbb{1}_A(T_1^n(x,y))\mathbb{1}_A(T_1^{2n}(x,y)\mathbb{1}_A(T_2^{n^2}(x,y))>0$, then there exist $a_0,a_1,a_2\in\Lambda$ such that $u\in(a_0-\frac{1}{4n},a_0+\frac{1}{4N})$, $v\in(a_1-\frac{1}{4N},a_1+\frac{1}{4N})$, $w\in(a_2-\frac{1}{4N},a_2+\frac{1}{4N})$. Thus

$$a_1-2a_0+a_2+t=0$$

for some $|t|\le\frac{1}{N}$. So $a_1-2a_0+a_2=0$ and then $a_0=a_1=a_2$. It follows that $nx\in(1/4N,1/4N)$ and

$$\int\mathbb{1}_A(x,y)\mathbb{1}_A(T_1^n(x,y))\mathbb{1}_A(T_1^{2n}(x,y)\mathbb{1}_A(T_2^{n^2}(x,y))d\mu(x,y)\le\frac{1}{N^2}|\Lambda|.$$

A quick computation shows that

$$\frac{1}{N^2}|\Lambda|\le\frac{1}{2}\left(\frac{|\Lambda|}{N}\right)^\ell=\frac{1}{2}\mu(A)^\ell$$

as long as $l\le\frac{\log(|\Lambda)|-2\log(N)+\log(2)}{(\log(|\Lambda|)-\log(N))}$. Taking Λ of cardinality $N^{1-\epsilon}$, the right hand side can be arbitrarily large, finishing the proof.

□

References

[1] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems*, 7(3):337–349, 1987. 2

[2] 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. 2

[3] V. Bergelson, B. Host, and B. Kra. Multiple recurrence and nilsequences. *Invent. Math.*, 160(2):261–303, 2005. With an appendix by I. Ruzsa. 1, 2, 3, 7, 13, 14, 15, 20, 23

[4] V. Bergelson, J. Moreira, and F. K. Richter. Single and multiple recurrence along non-polynomial sequences. arXiv:1711.05729, 2017. 3

[5] N. Frantzikinakis. The structure of strongly stationary systems. *J. Anal. Math.*, 93:359–388, 2004. 3

[6] N. Frantzikinakis. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.*, 360(10):5435–5475, 2008. 2, 3, 4, 5, 6, 8, 9, 10, 11, 23

[7] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Anal. Math.*, 109:353–395, 2009. 12

[8] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. Anal. Math.*, 112:79–135, 2010. 2

[9] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Trans. Amer. Math. Soc.*, 367(8):5653–5692, 2015. 2

[10] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.*, 611:131–144, 2007. 2

[11] N. Frantzikinakis, B. Host, and B. Kra. The polynomial multidimensional Szemerédi theorem along shifted primes. *Israel J. Math.*, 194(1):331–348, 2013. 2

[12] N. Frantzikinakis and B. Kra. Polynomial averages converge to the product of integrals. *Israel J. Math.*, 148:267–276, 2005. Probability in mathematics. 2

[13] N. Frantzikinakis and B. Kra. Ergodic averages for independent polynomials and applications. *J. London Math. Soc. (2)*, 74(1):131–142, 2006. 2

[14] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977. 1

[15] B. Host and B. Kra. An odd Furstenberg-Szemerédi theorem and quasi-affine systems. *J. Anal. Math.*, 86:183–220, 2002. 3

[16] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005. 7, 9, 10, 22

[17] A. Khintchine. Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.*, 109(1):604–615, 1934. 1

[18] A. N. Le. Nilsequences and multiple correlations along subsequences. arXiv:1708.01361v2, 2019. 11, 12

[19] A. Leibman. Multiple polynomial correlation sequences and nilsequences. *Ergodic Theory Dynam. Systems*, 30(3):841–854, 2010. 11

[20] A. Leibman. Orbit of the diagonal in the power of a nilmanifold. *Trans. Amer. Math. Soc.*, 362(3):1619–1658, 2010. 7, 21, 23

[21] J. Moreira and F. Richter. A spectral refinement of the Bergelson-Host-Kra decomposition and new multiple ergodic theorems. *Ergodic Theory Dynam. Systems*, 39(4):1042–1070, 2019. 3, 13

[22] D. K. L. Shiu. Strings of congruent primes. *J. London Math. Soc. (2)*, 61(2):359–373, 2000. 2

[23] T. Ziegler. A non-conventional ergodic theorem for a nilsystem. *Ergodic Theory Dynam. Systems*, 25(4):1357–1370, 2005. 7

Instituto de Ciencias de la Ingeniería, Universidad de O'Higgings, Av. Libertador Bernardo O'Higgins 611, Rancagua, Chile

E-mail address: sebastian.donoso@uoh.cl

Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730, USA

E-mail address: anhle@math.northwestern.edu

Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730, USA

E-mail address: joel.moreira@northwestern.edu

Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus OH, 43210-1174, USA

E-mail address: sun.1991@osu.edu