THE POLYNOMIAL MULTIDIMENSIONAL SZEMERÉDI THEOREM ALONG SHIFTED PRIMES

NIKOS FRANTZIKINAKIS, BERNARD HOST, AND BRYNA KRA

Abstract. If $\vec{q}_1,\dots,\vec{q}_m\colon \mathbb{Z}\to\mathbb{Z}^\ell$ are polynomials with zero constant terms and $E\subset\mathbb{Z}^\ell$ has positive upper Banach density, then we show that the set $E\cap(E-\vec{q}_1(p-1))\cap\ldots\cap(E-\vec{q}_m(p-1))$ is nonempty for some prime p. We also prove mean convergence for the associated averages along the prime numbers, conditional to analogous convergence results along the full integers. This generalizes earlier results of the authors, of Wooley and Ziegler, and of Bergelson, Leibman and Ziegler.

1. Introduction

1.1. **Background and new results.** Recent advances in ergodic theory and number theory have lead to numerous results on patterns in subsets of the integers with positive upper density, with descriptions of possible restrictions on differences between successive terms. In this vein, we show that the parameters in the polynomial multidimensional Szemerédi Theorem of Bergelson and Leibman [4] can be restricted to the shifted primes. Let $\mathbb{P}$ denote the set of prime numbers.

Theorem 1.1. *Let $\ell,m\in\mathbb{N}$, $\vec{q}_1,\dots,\vec{q}_m\colon\mathbb{Z}\to\mathbb{Z}^\ell$ be polynomials with $\vec{q}_i(0)=\vec{0}$ for $i=1,\dots,m$, and let $E\subset\mathbb{Z}^\ell$ with upper Banach density[1] $d^*(E)>0$. Then the set of integers n such that*

$$d^*\big(E\cap(E-\vec{q}_1(n))\cap\ldots\cap(E-\vec{q}_m(n))\big)>0$$

has nonempty intersection with $\mathbb{P}-1$ and $\mathbb{P}+1$.

In fact, our argument shows this intersection has positive relative density in the shifted primes.

The first result in this direction was due to Sárközy [17], who used analytic number theory to show that the difference set $E-E$ for a set E of positive upper Banach density contains a shifted prime $p-1$ for some $p\in\mathbb{P}$ (and similarly, as for all the results stated here, a shifted prime of the form $p+1$). In [7], relying on strong uniformity results of [11] related to the primes combined with a bit of ergodic theory, we took a first step towards a multiple version, showing that such E contains an arithmetic progression

2000 *Mathematics Subject Classification.* Primary: 11B30; Secondary: 37A45, 28D05, 05D10.

Key words and phrases. Arithmetic progressions, higher degree uniformity, multiple recurrence.

The first author was partially supported by Marie Curie IRG 248008, the second author by the Institut Universitaire de France, and the third by NSF grant 0900873.

[1]The *upper Banach density* $d^*(E)$ of a set $E\subset\mathbb{Z}^\ell$ is defined by $d^*(E)=\limsup_{|I|\to\infty}\frac{|E\cap I|}{|I|}$, where the lim sup is taken over all parallelepipeds $I\subset\mathbb{Z}^\ell$ whose side lengths tend to infinity.

of length 3 whose common difference is a shifted prime. This was generalized in two ways. First, Wooley and Ziegler [20] proved Theorem 1.1 for $\ell = 1$, relying on a deep ergodic structure theorem and milder number theoretic input than used in [7]. More recently, Bergelson, Leibman, and Ziegler [5], proved Theorem 1.1 for linear polynomials $\vec{q}_1, \dots, \vec{q}_m$, by combining the ergodic results on IP-recurrence of [9] and the uniformity results related to the primes of [11], [12] and [13] (their proof also gives the partition version of our main result in full generality). Theorem 1.1 generalizes the results of [20] and [5], and is in the spirit of [7], with the main ingredient being the number theoretic uniformity results of [11], [12] and [13].

By the Furstenberg Correspondence Principle (see Section 2.1 below), Theorem 1.1 is equivalent to an ergodic version and this is the version that we prove:

Theorem 1.2. *Let $\ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and let $T_1, \dots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations. Let $m \in \mathbb{N}$, $q_{i,j} \colon \mathbb{Z} \to \mathbb{Z}$ be polynomials with $q_{i,j}(0) = 0$ for $i = 1, \dots, \ell$ and $j = 1, \dots, m$. Then for any $A \in \mathcal{X}$ with $\mu(A) > 0$, the set of integers n such that*

$$\mu\big(A \cap (\prod_{i=1}^{\ell} T_i^{q_{i,1}(n)} A) \cap \ldots \cap (\prod_{i=1}^{\ell} T_i^{q_{i,m}(n)} A)\big) > 0$$

has nonempty intersection with $\mathbb{P} - 1$ and $\mathbb{P} + 1$.

We also prove mean convergence results for the corresponding multiple ergodic averages over the primes, conditional on the convergence of the corresponding averages over the full set of natural numbers (in some cases these results are not known). To keep notation to a minimum, we only state the result for polynomials taking values in $\mathbb{Z}$, but the analogous statement in $\mathbb{Z}^\ell$ can be proved in a similar way.

Theorem 1.3. *Let $\ell \in \mathbb{N}$, $(X, \mathcal{B}, \mu)$ be a probability space, $T_1, \dots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations, and let $q_1, \dots, q_\ell \colon \mathbb{Z} \to \mathbb{Z}$ be polynomials such that the averages*

$$\frac{1}{N} \sum_{n=1}^{N} f_1(T_1^{q_1(an+b)} x) \cdot \ldots \cdot f_\ell(T_\ell^{q_\ell(an+b)} x) \tag{1}$$

converge in $L^2(\mu)$ as $N \to \infty$ for all integers $a, b \geq 1$. Then the averages

$$\frac{1}{\pi(N)} \sum_{p \in \mathbb{P} \cap [1,N]} f_1(T_1^{q_1(p)} x) \cdot \ldots \cdot f_\ell(T_\ell^{q_\ell(p)} x), \tag{2}$$

where $\pi(N)$ denotes the number of primes up to N, also converge in $L^2(\mu)$ as $N \to \infty$.

Convergence of (2) for a single linear polynomial was proved by Wierdl [19] (more generally he showed pointwise convergence, an issue that we do not address here). When all the transformations are equal and one restricts to linear polynomials, we proved convergence of (2) in [7], but for $\ell \geq 3$ this was conditional upon the results of [12] and [13] that were subsequently proven. In the case where all the transformations are equal, convergence of (2) was proved by Wooley and Ziegler in [20]. Combined with the convergence results of [14] and [16], Theorem 1.3 recovers the convergence results of [20]. Using the convergence results of [18], we obtain the new result of mean convergence for the linear averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]} f_1(T_1^p x)\cdot\ldots\cdot f_\ell(T_\ell^p x),$$

and combined with the results of [6], we have mean convergence for other new cases, for example the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]} f_1(T_1^p x)\cdot f_2(T_2^{p^2} x)\cdot\ldots\cdot f_\ell(T_\ell^{p^\ell} x).$$

Combining the $\mathbb{Z}^\ell$ version of Theorem 1.3 with the convergence results of [1] and [2], we have mean convergence of the averages

$$\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]} f_1(T_1^{p^2} x)\cdot f_2(T_1^{p^2} T_2^p x).$$

1.2. **Strategy of the proof.** We prove Theorems 1.2 and 1.3 by reducing the problem to a deep result on the uniformity of the modified von Mangoldt function (Theorem 2.2 below). The main idea is to compare the multiple ergodic averages along the primes with the corresponding ones along the natural numbers, and show that the difference between the two converges to zero in mean. Some variation of this idea holds and is given in Proposition 3.6. The proof of this follows by successive applications of the van der Corput lemma and a straightforward PET induction argument, reducing the problem to the aforementioned uniformity result. Given the comparison result of Proposition 3.6, the proof of Theorem 1.3 follows in a straightforward manner, and the proof of Theorem 1.2 follows similarly, with the additional input of a uniform version of the polynomial Szemerédi theorem.

1.3. **Further directions.** Combining the method of this paper with the multiple recurrence result and methods of [15], one can show that Theorem 1.2 holds under the relaxed assumption that the transformations $T_1,\ldots,T_\ell$ generate a nilpotent group (and thus obtain further combinatorial implications, as in [15]). Likewise the obvious extension of Theorem 1.3 to the nilpotent case holds. In both cases, the missing ingredient is an extension of the uniformity estimate of Lemma 3.5 to the case that the transformations $T_1,\ldots,T_\ell$ generate a nilpotent group, which can be proved using the PET induction scheme in [15]. We do not carry this out here.

A more challenging problem is the extension of Theorems 1.2 and 1.3 to sequences involving fractional powers. For example, one could hope to show that for any positive real numbers a and b, any $E\subset\mathbb{Z}$ with $d^*(E)>0$ contains patterns of the form $m, m+[p^a], m+2[p^a]$, or patterns of the form $m, m+[p^a], m+[p^b]$ for some $m\in\mathbb{N}$ and $p\in\mathbb{P}$. If one is to use the methods of this paper, the missing ingredient is the appropriate variant of Lemma 3.5, a seemingly nontrivial result.

Lastly, let us mention that for two or more transformations, even the simplest pointwise variants of the mean convergence results we have established remain open. For example, it is not known whether for every probability space $(X, \mathcal{X}, \mu)$, measure preserving transformation $T\colon X \to X$, and functions $f_1, f_2 \in L^\infty(\mu)$, the averages $\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\cap[1,N]} f_1(T^p x)\cdot f_2(T^{2p}x)$, or the averages $\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\cap[1,N]} f_1(T^p x)\cdot f_2(T^{p^2}x)$, converge pointwise as $N \to \infty$. For such problems, the methods of the current paper do not seem to be applicable.

1.4. **General conventions and notation.** We denote the positive integers by $\mathbb{N} = \{1, 2, \ldots\}$ and write $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. If f is a measurable function on a measure space X with transformation $T\colon X \to X$, we write $Tf = f \circ T$. If S is a finite set and $a\colon S \to \mathbb{C}$, then we write $\mathbb{E}_{n\in S} a(n) = \frac{1}{|S|}\sum_{n\in S} a(n)$. We use the symbol $\ll$ when some expression is majorized by a constant multiple of some other expression. If this constant depends on variables $k_1, \ldots, k_\ell$, we write $\ll_{k_1,\ldots,k_\ell}$. We use $o_N(1)$ to denote a quantity that converges to zero when $N \to \infty$ and all other parameters are fixed.

2. Background

2.1. **Furstenberg correspondence principle.** We state a modification of the correspondence principle of Furstenberg (the formulation given is similar to the one in [4]):

Furstenberg Correspondence Principle ([8]). *Let $\ell \in \mathbb{N}$ and $\mathbb{E} \subset \mathbb{Z}^\ell$. There exists a probability space $(X, \mathcal{X}, \mu)$, commuting invertible measure preserving transformations $T_1, \ldots, T_\ell\colon X \to X$, and set $A \in \mathcal{X}$ with $\mu(A) = d^*(E)$, such that*

$$d^*(E \cap (E - \vec{n}_1) \cap \ldots \cap (E - \vec{n}_\ell)) \geq \mu\big(A \cap (\prod_{i=1}^{\ell} T_i^{n_{i,1}} A) \cap \ldots \cap (\prod_{i=1}^{\ell} T_i^{n_{i,m}} A)\big)$$

for all $m \in \mathbb{N}$ and $\vec{n}_j = (n_{1,j}, \ldots, n_{\ell,j}) \in \mathbb{Z}^\ell$ for $j = 1, \ldots, m$.

In particular, this correspondence shows that Theorem 1.1 follows from Theorem 1.2.

2.2. **Averages along the primes and weighted averages.** Let $\Lambda\colon \mathbb{N} \to \mathbb{R}$ denote the von Mangoldt function, taking the value $\log p$ on a prime p and its powers and 0 elsewhere. Also let

$$\Lambda'(n) = \mathbf{1}_{\mathbb{P}}(n) \cdot \Lambda(n)$$

for $n \in \mathbb{N}$. Throughout, the roles of Λ and Λ' are interchangeable, and all the results can be proven for either function (as the contribution from prime powers greater than 1 is negligible in our averages).

The following lemma is classical (for a proof, see for example [7]) and allows us to relate averages over the primes with weighted averages over the integers:

Lemma 2.1. *If $a\colon \mathbb{N} \to \mathbb{C}$ is bounded, then*

$$\Big|\frac{1}{\pi(N)} \sum_{p\in\mathbb{P}, p\leq N} a(p) - \frac{1}{N}\sum_{n=1}^{N} \Lambda'(n) \cdot a(n)\Big| = o_N(1).$$

In particular, the average in (2) is asymptotically equal to the weighted average over the natural numbers:

$$\frac{1}{N}\sum_{n=1}^{N}\Lambda'(n)\cdot f_1(T_1^{q_1(n)}x)\cdot\ldots\cdot f_\ell(T_\ell^{q_\ell(n)}x).$$

2.3. **Gowers norms.** If $a\colon \mathbb{Z}_N\to\mathbb{C}$, we inductively define:

$$\|a\|_{U_1(\mathbb{Z}_N)}=\left|\mathbb{E}_{n\in\mathbb{Z}_N}a(n)\right|$$

and

$$\|a\|_{U_{d+1}(\mathbb{Z}_N)}=\left(\mathbb{E}_{h\in\mathbb{Z}_N}\|a_h\cdot\bar{a}\|_{U_d(\mathbb{Z}_N)}^{2^d}\right)^{1/2^{d+1}},$$

where $a_h(n)=a(n+h)$. Gowers [10] showed that for $d\geq 2$ this defines a norm on $\mathbb{Z}_N$.

2.4. **Uniformity of the modified von Mangoldt function.** For $w>2$ let

$$W=\prod_{p\in\mathbb{P},p<w}p$$

denote the product of the primes bounded by w. For $r\in\mathbb{N}$ let

$$\Lambda'_{w,r}(n)=\frac{\phi(W)}{W}\cdot\Lambda'(Wn+r),$$

where ϕ denotes the Euler function.

The next result is key for our study. It was obtained in [11] (Theorem 7.2), conditional upon results on the Möbius function later obtained in [12] (Theorem 1.1) and the inverse conjecture for the Gowers norms (recently proved in [13]):

Theorem 2.2 (Green and Tao ([11], [12]), Green, Tao, and Ziegler [13])**.** *With the previous notation, for every $d\in\mathbb{N}$, the maximum, taken over those r between 1 and W satisfying $(r,W)=1$, of*

$$\|(\Lambda'_{w,r}-1)\cdot\mathbf{1}_{[1,N]}\|_{U_d(\mathbb{Z}_{dN})}$$

converges to 0 as $N\to\infty$ and then $w\to\infty$.

Note that in [11] (Theorem 7.2), the result is stated with w being a specific slowly growing function of N, but the authors also note any sufficiently slowly growing function of N works too, and this implies our version. Furthermore, in [11] the theorems are stated without the indicator function $\mathbf{1}_{[1,N]}$, but the results of [11], [12] and [13], also imply this version.

3. Comparing averages

3.1. **PET (polynomial exhaustion technique) induction.** We describe the inductive scheme from [4] and follow the notation and implementation used in [6]. Let $\ell,m\in\mathbb{N}$. Given ℓ ordered families of polynomials

$$\mathcal{Q}_1=(q_{1,1},\ldots,q_{1,m}),\ldots,\mathcal{Q}_\ell=(q_{\ell,1},\ldots,q_{\ell,m}),$$

we define an *ordered family $(\mathcal{Q}_1,\ldots,\mathcal{Q}_\ell)$ of m polynomial ℓ-tuples* by

$$(\mathcal{Q}_1,\ldots,\mathcal{Q}_\ell)=\big((q_{1,1},\ldots,q_{\ell,1}),\ldots,(q_{1,m},\ldots,q_{\ell,m})\big).$$

This gives a concise way of recording the polynomial iterates that appear in the average of

$$f_1(T_1^{q_{1,1}(n)} \cdots T_\ell^{q_{\ell,1}(n)} x) \cdot \ldots \cdot f_m(T_1^{q_{1,m}(n)} \cdots T_\ell^{q_{\ell,m}(n)} x).$$

The maximum of the degrees of the polynomials in the families $\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell$ is called *the degree of the family* $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$.

Fix an integer $s \geq 1$; we restrict ourselves to families of degree $\leq s$. For $i = 1, \ldots, \ell$, define $\mathcal{Q}'_i$ to be the (possibly empty) set given by:

$$\mathcal{Q}'_i = \{\text{nonconstant } q_{i,j} \in \mathcal{Q}_i \colon q_{i',j} \text{ is constant for } i' < i\}.$$

Two polynomials are said to be *equivalent* if they have the same degree and the same leading coefficient. For $i = 1, \ldots, \ell$ and $j = 1, \ldots, s$, we let $w_{i,j}$ denote the number of distinct non-equivalent classes of polynomials of degree j in the family $\mathcal{Q}'_i$.

Define the *(matrix) type* of the family $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$ to be the matrix

$$\begin{pmatrix} w_{1,s} & \ldots & w_{1,1} \\ w_{2,s} & \ldots & w_{2,1} \\ \vdots & \ldots & \vdots \\ w_{\ell,s} & \ldots & w_{\ell,1} \end{pmatrix}.$$

A matrix is said to be of *matrix type zero* if all the $w_{i,j}$ are zero, and this happens exactly when all the polynomials are constant.

We order the types lexicographically: given two $\ell \times s$ matrices $W = (w_{i,j})$ and $W' = (w'_{i,j})$, we say that W is bigger than W', and write $W > W'$, if $w_{1,d} > w'_{1,d}$, or $w_{1,d} = w'_{1,d}$ and $w_{1,d-1} > w'_{1,d-1}$, …, or $w_{1,i} = w'_{1,i}$ for $i = 1, \ldots, d$ and $w_{2,d} > w'_{2,d}$, and so on. We have:

Lemma 3.1. *Every decreasing sequence of types of families of ℓ-tuples of polynomials is stationary.*

Thus applying some operation that reduces the type, after finitely many repetitions, the procedure terminates. Such an operation is described in the next subsection.

3.2. **The van der Corput operation.** Given a family $\mathcal{Q} = \big(q_1, \ldots, q_m\big)$, $q \in \mathbb{Z}[t]$, and $h \in \mathbb{N}$, we define the families $S_h\mathcal{Q}$ and $\mathcal{Q} - q$ as follows:

$$S_h\mathcal{Q} = (S_h q_1, \ldots, S_h q_m) \text{ and } \mathcal{Q} - q = \big(q_1 - q, \ldots, q_m - q\big),$$

where $(S_h q)(n) = q(n+h)$.

Given a family of ℓ-tuples of polynomials $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$, an ℓ-tuple $(q_1, \ldots, q_\ell) \in (\mathcal{Q}_1, \ldots \mathcal{Q}_\ell)$, and $h \in \mathbb{N}$, define the operation

$$(q_1, \ldots, q_\ell, h)\text{-vdC}(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell) = (\tilde{Q}_{1,h}, \ldots \tilde{Q}_{\ell,h}),$$

where

$$\tilde{Q}_{i,h} = (S_h\mathcal{Q}_i - q_i, \mathcal{Q}_i - q_i)$$

for $i = 1, \ldots, \ell$.

Starting with a family $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$, we successively apply appropriate van der Corput operations to arrive at constant families of ℓ-tuples of polynomials. This is achieved using:

Lemma 3.2 (Bergelson and Leibman [4]). *Let $(\mathcal{Q}_1,\dots,\mathcal{Q}_\ell)$ be a family of ℓ-tuples of polynomials with nonzero matrix type. Then there exists $(q_1,\dots,q_\ell)\in(\mathcal{Q}_1,\dots,\mathcal{Q}_\ell)$ such that for every $h\in\mathbb{N}$, the family $(q_1,\dots,q_\ell,h)$-vdC$(\mathcal{Q}_1,\dots,\mathcal{Q}_\ell)$ has strictly smaller type than $(\mathcal{Q}_1,\dots,\mathcal{Q}_\ell)$.*

While this lemma is usually stated to hold for sufficiently large h, this is only in order to maintain extra properties of the polynomial family (such as being essentially distinct), and we do not need these properties here. Thus we are able to phrase this in the slightly stronger, and easier to use for our purposes, setting of all $h\in\mathbb{N}$.

Assuming Lemma 3.2, the proof of the next result is standard:

Lemma 3.3. *Let $(\mathcal{Q}_1,\dots,\mathcal{Q}_\ell)$ be a family of m polynomial ℓ-tuples and nonzero matrix type. Suppose that we successively apply the $(q_1,\dots,q_\ell,h)$-vdC operation for appropriate choices of $q_1,\dots,q_\ell\in\mathbb{Z}[t]$ and $h\in\mathbb{N}$, as described in the previous lemma, each time obtaining a family of ℓ-tuples of polynomials with strictly smaller type. Then after a finite number of operations, depending only on ℓ, m, and the maximum degree of the polynomials (but not on the successive choices of h), we obtain families of ℓ-tuples of polynomials of degree 0.*

3.3. **Controlling averages.** We use the following version of the van der Corput Lemma:

Lemma 3.4. *Let $N\in\mathbb{N}$ and $v(1),\dots,v(N)$ be elements of a Hilbert space $\mathcal{H}$, with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. Then*

$$\left\|\frac{1}{N}\sum_{n=1}^{N}v(n)\right\|^2\ll\frac{1}{N^2}\sum_{n=1}^{N}\|v(n)\|^2+\frac{1}{N}\sum_{h=1}^{N}\left\|\frac{1}{N}\sum_{n=1}^{N-h}\langle v(n+h),v(n)\rangle\right\|.$$

Before stating the main lemma used to control averages, we give a simple case that illustrates the technique:

Example. Let $a\colon\mathbb{N}\to\mathbb{C}$ be a sequence that satisfies $a(n)/n^{1/4}\to 0$. Let $(X,\mathcal{X},\mu)$ be a probability space, $T\colon X\to X$ be a measure preserving transformation, and $f\in L^\infty(\mu)$ be a function bounded by 1. Then we have that

$$\left\|\frac{1}{N}\sum_{n=1}^{N}a(n)\cdot T^{n^2}f\right\|_{L^2(\mu)}\ll\|a\cdot\mathbf{1}_{[1,N]}\|_{U_3(\mathbb{Z}_{3N})}+o_N(1). \tag{3}$$

To prove this, we apply van der Corput (Lemma 3.4 for $v(n)=a(n)\cdot T^{n^2}f$) and the Cauchy-Schwarz Inequality and we have

$$\left\|\frac{1}{N}\sum_{n=1}^{N}a(n)\cdot T^{n^2}f\right\|^2_{L^2(\mu)}\ll$$
$$\frac{1}{N}\sum_{h_1=1}^{N}\left\|\frac{1}{N}\sum_{n=1}^{N-h_1}\bar{a}(n+h_1)\cdot a(n)\cdot T^{2nh_1+h_1^2}f\right\|_{L^2(\mu)}+\frac{1}{N^2}\sum_{n=1}^{N}|a(n)|^2.$$

By assumption, the second term is $o_N(1)$ and we are left with estimating the first term. For $h_1 = 1, \ldots, N$, rewriting the interior sum as

$$\frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{[1,N]}(n+h_1)\cdot \bar{a}(n+h_1)\cdot a(n)\cdot T^{2nh_1+h_1^2}f,$$

and applying van der Corput and Cauchy-Schwarz once more, we have that

$$\left\|\frac{1}{N}\sum_{n=1}^{N-h_1} \bar{a}(n+h_1)\cdot a(n)\cdot T^{2nh_1+h_1^2}f\right\|_{L^2(\mu)}^2 \ll$$

$$\frac{1}{N}\sum_{h_2=1}^{N}\Big|\frac{1}{N}\sum_{n=1}^{N-h_1-h_2} a(n)\cdot\bar{a}(n+h_1)\cdot\bar{a}(n+h_2)\cdot a(n+h_1+h_2)\Big| + \frac{1}{N^2}\sum_{n=1}^{N}|\bar{a}(n+h_1)\cdot a(n)|^2.$$

Again, by assumption, the average over $h_1 \in \{1, \ldots, N\}$ of the second term is $o_N(1)$. By further applications of Cauchy-Schwarz, we have that the eighth power of the $L^2(\mu)$-norm of the original average is bounded by a constant multiple of

$$\frac{1}{N^2}\sum_{1\le h_1,h_2\le N}\Big|\frac{1}{N}\sum_{n=1}^{N-h_1-h_2} a(n)\cdot\bar{a}(n+h_1)\cdot\bar{a}(n+h_2)\cdot a(n+h_1+h_2)\Big|^2 + o_N(1). \tag{4}$$

On the other hand, letting $a_N(n) = a(n)\cdot\mathbf{1}_{[1,N]}(n)$, for $n = 1, \ldots, 3N$, and thinking of a_N as a function $\mathbb{Z}_{3N} \to \mathbb{C}$, we have that

$$|\!|\!|a_N|\!|\!|_{U_3(\mathbb{Z}_{3N})}^8 = \mathbb{E}_{h_1,h_2\mathbb{Z}_{3N}}|\mathbb{E}_{n\in\mathbb{Z}_{3N}} a_N(n)\cdot\bar{a}_N(n+h_1)\cdot\bar{a}_N(n+h_2)\cdot a_N(n+h_1+h_2)|^2.$$

(The sums $n+h_1$, $n+h_2$, and $n+h_1+h_2$ are taken modulo $3N$, and we make the somewhat less conventional identification of $\mathbb{Z}_{3N}$ with $[1, \ldots, 3N]$.) This is greater than or equal to (eliminating values with $N < h_1, h_2 \le 3N$)

$$\frac{1}{9N^2}\sum_{1\le h_1,h_2\le N}|\mathbb{E}_{n\in\mathbb{Z}_{3N}} a_N(n)\cdot\bar{a}_N(n+h_1)\cdot\bar{a}_N(n+h_2)\cdot a_N(n+h_1+h_2)|^2,$$

where we maintain the same convention on sums. Since in this expression we have $1 \le h_1, h_2 \le N$ and $a_N(n)$ is zero for $n \in \{N+1, \ldots, 3N\}$, we have that all h_1, h_2, n that make a nonzero contribution to this last average satisfy $1 \le n+h_1+h_2 \le 3N$. In particular, there are no circular effects and the last expression is equal to

$$\frac{1}{9N^2}\sum_{1\le h_1,h_2\le N}\Big|\frac{1}{3N}\sum_{n=1}^{3N} a_N(n)\cdot\bar{a}_N(n+h_1)\cdot\bar{a}_N(n+h_2)\cdot a_N(n+h_1+h_2)\Big|^2$$

$$= \frac{1}{27N^2}\sum_{1\le h_1,h_2\le N}\Big|\frac{1}{N}\sum_{n=1}^{N-h_1-h_2} a(n)\cdot\bar{a}(n+h_1)\cdot\bar{a}(n+h_2)\cdot a(n+h_1+h_2)\Big|^2,$$

where the sums $n+h_1$, $n+h_2$, and $n+h_1+h_2$ are taken in $\mathbb{N}$, without reduction modulo $3N$. But this expression is exactly $1/27$ of the average in (4). Combining these estimates, we have that the eighth power of the $L^2(\mu)$-norm of the original averages is bounded by a constant times $\|a_N\|^8_{U_3(\mathbb{Z}_{3N})}$ plus an $o_N(1)$ term. Thus we have estimate (3).

We now turn to the general case:

Lemma 3.5. *Let $\ell, m \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, \ldots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations, $f_1, \ldots, f_m \in L^\infty(\mu)$ be functions bounded by 1, and $q_{i,j} \colon \mathbb{Z} \to \mathbb{Z}$, $i \in \{1, \ldots, \ell\}$, $j \in \{1, \ldots, m\}$, be polynomials. Let $a \colon \mathbb{N} \to \mathbb{C}$ be a sequence of complex numbers satisfying $a(n)/n^c \to 0$ for every $c > 0$. Then there exists $d \in \mathbb{N}$, depending only on the maximum degree of the polynomials $q_{i,j}$ and the integers ℓ and m, such that*

$$\Big\| \frac{1}{N} \sum_{n=1}^{N} a(n) \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{i,1}(n)}\big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{i,m}(n)}\big) f_m \Big\|_{L^2(\mu)} \ll_d \|a \cdot \mathbf{1}_{[1,N]}\|_{U_d(\mathbb{Z}_{dN})} + o_N(1).$$

Furthermore, the implicit constant is independent of the sequence $(a(n))_{n\in\mathbb{N}}$, and the $o_N(1)$ term depends only the integer d and on the sequence $(a(n))_{n\in\mathbb{N}}$.

Proof. For $i = 1, \ldots, \ell$, let $\mathcal{Q}_i = (q_{i,1}, \ldots, q_{i,m})$. If the matrix type of the family $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$ is zero, then all the polynomials are constant, in which case the conclusion holds trivially for $d = 1$. If the matrix type is nonzero, then by Lemma 3.2 there exists $(q_1, \ldots, q_\ell) \in (\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$ such that for $h_1 \in \mathbb{N}$, the family $(q_1, \ldots, q_\ell, h_1)$-vdC$(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$ has type strictly smaller than that of $(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$.

As in the model example, using van der Corput and Cauchy-Schwarz, we have that

$$\Big\| \frac{1}{N} \sum_{n=1}^{N} a(n) \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{i,1}(n)}\big) f_1 \cdot \ldots \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{i,m}(n)}\big) f_m \Big\|_{L^2(\mu)}^{2^{d+1}} \tag{5}$$

is bounded by an $o_N(1)$ term plus a constant multiple of

$$\frac{1}{N} \sum_{h_1=1}^{N} \Big\| \frac{1}{N} \sum_{n=1}^{N-h_1} \bar{a}(n+h_1) \cdot a(n) \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{h_1,i,1}(n)}\big) g_1 \cdot \ldots \cdot \big(\prod_{i=1}^{\ell} T_i^{q_{h_1,i,2m}(n)}\big) g_{2m} \Big\|_{L^2(\mu)}^{2^d},$$

where $(q_{h_1,1,j}, \ldots q_{h_1,\ell,j}) \in (q_1, \ldots, q_\ell, h_1)$-vdC$(\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell)$ for every $h_1 \in \mathbb{N}$ and $j = 1, \ldots, 2m$. If the new family of polynomials has zero matrix type, we stop. If not, as in the model example, we continue to use van der Corput and Cauchy-Schwarz to bound the average over n. By Lemma 3.3, after a finite number of steps, depending only on the maximum degree of the polynomials $q_{i,j}$ and the integers ℓ and m, we have families of polynomials with zero matrix type. Assume that this takes d steps. We deduce that the expression (5) is bounded by a $o_N(1)$ term (using the assumption that $a(n)/n^c \to 0$ for every $c > 0$ to control the lower order terms) plus a constant multiple of

$$\frac{1}{N^d} \sum_{1 \le h_1, \ldots, h_d \le N} \Big| \frac{1}{N} \sum_{n=1}^{N-h_1-\cdots-h_d} a(n) \cdot \bar{a}(n+h_1) \cdot \bar{a}(n+h_2) \cdot \ldots \cdot a(n+h_1+\cdots+h_d) \Big|^2.$$

(Note that the last occurrence of a in this expression may actually be $\bar{a}$, depending on the parity of d.) As in the model example, we see that this last average is bounded by a constant (equal to d^d) times

$$\| a \cdot \mathbf{1}_{[1,N]} \|_{U_{d+1}(\mathbb{Z}_{(d+1)N})}^{2^{d+1}},$$

completing the proof. □

3.4. **Comparing averages.** The key result needed to compare averages over the primes and over the integers is (recall that $W = \prod_{p\in\mathbb{P}, p<w} p$ denotes the product of the primes bounded by w):

Proposition 3.6. *Let $\ell, m \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, \ldots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations, $f_1, \ldots, f_m \in L^\infty(\mu)$ be functions, and $q_{i,j} \colon \mathbb{Z} \to \mathbb{Z}$, $i \in \{1, \ldots, \ell\}$, $j \in \{1, \ldots, m\}$, be polynomials. Then the maximum, taken over those r between 1 and W satisfying $(r, W) = 1$, of the $L^2(\mu)$-norm of*

$$\frac{1}{N}\sum_{n=1}^{N} (\Lambda'_{w,r}(n) - 1) \cdot (\prod_{i=1}^{\ell} T_i^{q_{i,1}(Wn+r)}) f_1 \cdot \ldots \cdot (\prod_{i=1}^{\ell} T_i^{q_{i,m}(Wn+r)}) f_m$$

converges to 0 as $N \to \infty$ and then $w \to \infty$.

Proof. We can assume that all functions are bounded by 1. We apply Lemma 3.5 for $a_{w,r}(n) = \Lambda'_{w,r}(n) - 1$ for $w, r \in \mathbb{N}$, and the family of polynomials $q_{i,j}(Wn + r)$. Let $\mathbb{Z}_W^* = \{r \in [1, W] \colon (r, W) = 1\}$. We get that there exists $d \in \mathbb{N}$, independent of w and r, such that

$$\max_{r\in\mathbb{Z}_W^*} \left\| \frac{1}{N}\sum_{n=1}^{N} (\Lambda'_{w,r}(n) - 1) \cdot (\prod_{i=1}^{\ell} T_i^{q_{i,1}(Wn+r)}) f_1 \cdot \ldots \cdot (\prod_{i=1}^{\ell} T_i^{q_{i,m}(Wn+r)}) f_m \right\|_{L^2(\mu)} \ll_d$$
$$\max_{r\in\mathbb{Z}_W^*} \| (\Lambda'_{w,r} - 1) \cdot \mathbf{1}_{[1,N]} \|_{U_d(\mathbb{Z}_{dN})} + o_N(1)$$

where the term $o_N(1)$ depends only on the integers d and w. The result now follows from Theorem 2.2. □

4. Proof of the main results

4.1. **Proof of Theorem 1.2.** We use the following uniform multiple recurrence result, proved in the same way as Theorem 3.2 is proved in [3]:

Theorem 4.1. *Let $(X, \mathcal{X}, \mu)$ be a probability space and $T_1, \ldots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations. Let $q_{i,j} \colon \mathbb{Z} \to \mathbb{Z}$ be polynomials with $q_{i,j}(0) = 0$ for $i = 1, \ldots, \ell$ and $j = 1, \ldots, m$. Then for any $A \in \mathcal{X}$ with $\mu(A) > 0$, there exists a positive constant c, depending only on $\mu(A)$ and the polynomials $q_{i,j}$, such that*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\big(A \cap (\prod_{i=1}^{\ell} T_i^{q_{i,1}(n)} A) \cap \ldots \cap (\prod_{i=1}^{\ell} T_i^{q_{i,m}(n)} A)\big) \geq c.$$

It is important to note that the constant c does not depend on the transformations $T_1, \ldots, T_\ell$. This observation enables us to prove a uniform multiple recurrence result more suitable for our purposes (the uniformity in W is crucial):

Corollary 4.2. *Let $(X, \mathcal{X}, \mu)$ be a probability space and $T_1, \ldots, T_\ell \colon X \to X$ be commuting invertible measure preserving transformations. Let $q_{i,j} \colon \mathbb{Z} \to \mathbb{Z}$ be polynomials with $q_{i,j}(0) = 0$ for $i = 1, \ldots, \ell$ and $j = 1, \ldots, m$. Then for any $A \in \mathcal{X}$ with $\mu(A) > 0$, there exists a positive constant c, depending on $\mu(A)$ and the polynomials $q_{i,j}$, such that for every $W \in \mathbb{N}$, we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\big(A \cap (\prod_{i=1}^{\ell} T_i^{q_{i,1}(Wn)} A) \cap \ldots \cap (\prod_{i=1}^{\ell} T_i^{q_{i,m}(Wn)} A)\big) \geq c.$$

Proof. We write the proof for $\ell, m = 1$, as the general case follows in an analogous manner. Let $(X, \mathcal{X}, \mu)$ be a probability space and let $T \colon X \to X$ be an invertible measure preserving transformation. Let $q(n) = c_1 n + \cdots + c_d n^d$, where $c_1, \ldots, c_d \in \mathbb{Z}$ and $d \in \mathbb{N}$. Given $A \in \mathcal{X}$ and $W \in \mathbb{N}$, we have that

$$\mu(A \cap T^{p(Wn)} A) = \mu\big(A \cap (\prod_{i=1}^{d} S_i^{n^i} A)\big)$$

where $S_i = T^{c_i W^i}$ for $i = 1, \ldots, d$. The result now follows from Theorem 4.1. $\square$

Combining Proposition 3.6 and Corollary 4.2, we have for that for sufficiently large $w \in \mathbb{N}$,

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Lambda'_{w,1}(n) \cdot \mu\big(A \cap (\prod_{i=1}^{\ell} T_i^{q_{i,1}(Wn)} A) \cap \ldots \cap (\prod_{i=1}^{\ell} T_i^{q_{i,m}(Wn)} A)\big) > 0.$$

By Lemma 2.1, the conclusion of Theorem 1.2 is satisfied for a set of n with positive relative density in the shifted primes $\mathbb{P} - 1$.

A similar argument holds for the shifted primes $\mathbb{P} + 1$.

4.2. **Proof of Theorem 1.3.** To complete the proof, we follow the method used in [7]. By Lemma 2.1, it suffices to prove convergence in $L^2(\mu)$ for the corresponding weighted averages

$$A(N) = \frac{1}{N} \sum_{n=1}^{N} \Lambda'(n) \cdot T_1^{q_1(n)} f_1 \cdot \ldots \cdot T_\ell^{q_\ell(n)} f_\ell.$$

Equivalently, it suffices to show that the sequence of functions $(A(N))_{N\in\mathbb{N}}$ is Cauchy in $L^2(\mu)$.

Let $\varepsilon > 0$. Fix $w, r \in \mathbb{N}$, and let

$$B_{w,r}(N) = \frac{1}{N} \sum_{n=1}^{N} T_1^{q_1(Wn+r)} f_1 \cdot \ldots \cdot T_\ell^{q_\ell(Wn+r)} f_\ell.$$

(As before, W denotes the product of primes bounded by w.) By Proposition 3.6, we have that for some $w_0 \in \mathbb{N}$ (and corresponding $W_0 \in \mathbb{N}$), if N is large enough, then

$$\left\| A(W_0N) - \frac{1}{\phi(W_0)} \sum_{1\le r\le W_0, (r,W_0)=1} B_{w_0,r}(N) \right\|_{L^2(\mu)} \le \varepsilon/6. \tag{6}$$

By assumption, for $r = 1, \dots, W_0$, the sequence $(B_{w_0,r}(N))_{N\in\mathbb{N}}$ converges in $L^2(\mu)$. Therefore, if M and N are sufficiently large, then for $r = 1, \dots, W_0$ we have

$$\|B_{w_0,r}(N) - B_{w_0,r}(M)\|_{L^2(\mu)} \le \varepsilon/6. \tag{7}$$

Combining (6) and (7) we have that if M and N are sufficiently large, then

$$\|A(W_0N) - A(W_0M)\|_{L^2(\mu)} \le \varepsilon/2. \tag{8}$$

Lastly, for $r = 1, \dots, W_0$, we have

$$\lim_{N\to\infty} \|A(W_0N + r) - A(W_0N)\|_{L^2(\mu)} = 0. \tag{9}$$

Combining (8) and (9), it follows that if M and N are sufficiently large, then

$$\|A(N) - A(M)\|_{L^2(\mu)} \le \varepsilon.$$

Therefore, the sequence $(A(N))_{N\in\mathbb{N}}$ is Cauchy in $L^2(\mu)$, completing the proof of Theorem 1.3.

References

[1] T. Austin. Pleasant extensions retaining algebraic structure, I. Available at arxiv:0905.0518.
[2] T. Austin. Pleasant extensions retaining algebraic structure, II. Available at arxiv:0910.0907.
[3] V. Bergelson, B. Host, R. McCutcheon, F. Parreau. Aspects of uniformity in recurrence. *Colloq. Math.* **84/85** (2000), no. 2, 549–576.
[4] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, **9** (1996), 725–753.
[5] V. Bergelson, A. Leibman, T. Ziegler. The shifted primes and the multidimensional Szemerédi and polynomial van der Waerden Theorems. Preprint 2010. Available at arXiv:1007.1839.
[6] C. Chu, N. Frantzikinakis, B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. To appear, *Proc. London Math. Soc.* Available at arXiv:0912.2641.
[7] N. Frantzikinakis, B. Host, B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.* **611** (2007), 131–144.
[8] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **71** (1977), 204–256.
[9] H. Furstenberg, Y. Katznelson. IP$_r$ sets, Szemerédi's Theorem, and Ramsey Theory. *Bull. Amer. Math. Soc.* **14** (1986), 275–278.
[10] W. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
[11] B. Green, T. Tao. Linear equations in the primes. To appear, *Annals. Math.* Available at arXiv:math/0606088.
[12] B. Green, T. Tao. The Möbius function is strongly orthogonal to nilsequences. To appear, *Annals. Math.* Available at arXiv:0807.1736.
[13] B. Green, T. Tao, T. Ziegler. An inverse theorem for the Gowers U^{s+1}-norm. In preparation. Announcement available at arXiv:1006.0205.
[14] B. Host, B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.

[15] A. Leibman. Multiple recurrence theorem for measure preserving actions of a nilpotent group. *Geom. Funct. Anal.* **8** (1998), 853–931.

[16] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–315.

[17] A. Sárközy. On difference sets of sequences of integers, III. *Acta Math. Acadm. Sci. Hungar.* **31** (1978), 355–386.

[18] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Erg. Th. & Dyn. Sys.* **28** (2008), 657–688.

[19] M. Wierdl. Pointwise ergodic theorem along the prime numbers. *Israel J. Math.* **64** (1988), 315–336.

[20] T. Wooley, T. Ziegler. Multiple recurrence and convergence along the primes. Preprint 2010. Available at arXiv:1001.4081.

University of Crete, Department of mathematics, Knossos Avenue, Heraklion 71409, Greece

E-mail address: frantzikinakis@gmail.com

Laboratoire d'analyse et de mathématiques appliquées, Université de Marne la Vallée & CNRS UMR 8050, 5 Bd. Descartes, Champs sur Marne, 77454 Marne la Vallée Cedex 2, France

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

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

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