POLYNOMIAL CONFIGURATIONS IN THE PRIMES

THÁI HOÀNG LÊ AND JULIA WOLF

ABSTRACT. The Bergelson-Leibman theorem states that if $P_1, \dots, P_k \in \mathbf{Z}[x]$, then any subset of the integers of positive upper density contains a polynomial configuration $x + P_1(m), \dots, x + P_k(m)$, where $x, m \in \mathbf{Z}$. Various generalizations of this theorem are known. Wooley and Ziegler showed that the variable m can in fact be taken to be a prime minus 1, and Tao and Ziegler showed that the Bergelson-Leibman theorem holds for subsets of the primes of positive relative upper density. Here we prove a hybrid of the latter two results, namely that the step m in the Tao-Ziegler theorem can be restricted to the set of primes minus 1.

1. INTRODUCTION

Roughly twenty years after the ergodic theoretic proof of Szemerédi's theorem on long arithmetic progressions in dense subsets of the integers by Furstenberg [4], Bergelson and Leibman [1] proved the following celebrated polynomial generalization.

Theorem 1 (Bergelson-Leibman)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbf{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the integers of positive relative upper density contains a configuration of the form $a + P_1(d), \dots, a + P_k(d)$, where a, d are integers, $d \neq 0$.*

More recently, Tao and Ziegler [11] proved Theorem 1 for dense subsets of the primes, using the general transference strategy of Green and Tao [5].

Theorem 2 (Tao-Ziegler)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbf{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the primes of positive relative upper density contains a configuration of the form $a + P_1(a), \dots, a + P_k(a)$, where a, d are integers, $d \neq 0$.*

Here for any subset A of the set of primes $\mathcal{P}$, the relative upper density $\overline{d}_{\mathcal{P}}(A)$ of A in $\mathcal{P}$ is defined as

$$\overline{d}_{\mathcal{P}}(A) = \overline{\lim}_{N \to \infty} \frac{|A \cap [N]|}{|\mathcal{P} \cap [N]|}$$

In a recent preprint, Wooley and Ziegler [12] showed that the step d of the polynomial progression in Theorem 1 can be taken to be a shifted prime.

Theorem 3 (Wooley-Ziegler)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbf{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the integers of positive relative upper density contains a configuration of the form $a + P_1(p-1), \dots, a + P_k(p-1)$, where a is an integer and p is prime. The same is true if we replace $p - 1$ with $p + 1$.*

A generalization to polynomials in several variables, with a simpler proof, was subsequently obtained by Frantzikinakis, Host and Kra [3].

Our goal in this paper is to establish the following hybrid of Theorems 2 and 3.

Theorem 4. *Let $P_1, \dots, P_k$ be polynomials in $\mathbf{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the primes of positive relative upper density contains a configuration of the form $a + P_1(p-1), \dots, a + P_k(p-1)$, where a is an integer and p is prime. The same is true if we replace $p-1$ with $p+1$.*

In other words, we claim that the step d in Theorem 2 can be restricted to be of the form $p-1$ (or $p+1$). In fact, our proof shows that there are infinitely many such configurations, and we are able to give a lower bound on their number which is of the order of magnitude predicted by the Bateman-Horn conjecture. Previously, the question about the existence of such configurations has also been posed as Conjecture 1.2 in [9].

Our method is very similar to that employed in [3], in the sense that we compare an average over the integers to an average along the shifted primes using multiple applications of van der Corput's lemma and a PET induction scheme. However, we proceed quantitatively in the spirit of [11], and rely on a refined analysis of the correlation properties of the pseudorandom measure from that paper.

The rest of the article is structured as follows. In Section 2 we set up our notation, and in Section 3 we reduce Theorem 4 to the more technical Propositions 1 and 2. We study a simple example of Proposition 1 in Section 4, and follow it up in Section 5 with a discussion of a modified polynomial forms condition that arises from the example, together with an outline of the proof of Proposition 2. The technical details of the proof of Proposition 2 can be found in an appendix. Finally, the general case of Proposition 1 is proved in Section 6 using the now standard PET induction scheme.

Acknowledgements. Work on this project began during the first author's visit to École polytechnique, and he would like to thank the Centre de Mathématiques Laurent Schwartz for its hospitality. The authors would also like to thank Terence Tao for helpful discussions.

2. Preliminaries

We assume some familiarity with the work of Green and Tao [5] and Tao and Ziegler [11], as well as with the definition and basic properties of the Gowers uniformity norms. Here we only briefly remind the reader of the most important definitions, lemmas and parameter settings from those papers. The experienced reader is encouraged to skip this section and consult it later as the need arises.

Landau's O, o and Vinogradov's $\ll, \gg$ notation are given their usual asymptotic meaning. That is, for two quantities X, Y, we write $X \ll Y, Y \gg X$, or $X = O(Y)$ if we have a bound $|X| \leq CY$ for some constant C. If C depends on other parameters such as k, then this dependence is indicated as $X \ll_k Y, Y \gg_k X$, or $X = O_k(Y)$. By $o(1)$ we denote a quantity

that goes to 0 as $N \to \infty$. If this quantity depends on other parameters such as k, then this dependence is sometimes indicated as $o_k(1)$.

Throughout the paper, we fix a system of polynomials $P_1, \dots, P_k$ with integer coefficients, each vanishing at 0. We can certainly assume that these polynomials are distinct. We also fix a subset $A \subseteq \mathcal{P}$ satisfying $\overline{d}_{\mathcal{P}}(A) = \delta_0 > 0$. All implicit constants are allowed to depend on $\delta_0, P_1, \dots, P_k$.

To get around the fact that the primes are not equidistributed with respect to small moduli, we let $w \ll \log\log\log N$ be any sufficiently slowly growing function in N, and let $W = \prod_{p<w} p$ be the product of the primes less than w, so that $W \ll \log\log N$. Eventually, just as in [5] and [11], we will be able to take w be a sufficiently large constant, see the discussion in Section 7.

It follows from the assumption on the density of A that there is an infinite sequence of integers N' going to infinity such that

$$|A \cap [N']| > \frac{1}{2}\delta_0 \frac{N'}{\log N'}.$$

We set $N = \lfloor N'/2W \rfloor$, and observe that the asymptotic limit as $N \to \infty$ is equivalent to the asymptotic limit as $N' \to \infty$. By the pigeonhole principle, we can find $b = b(N) \in [W]$ coprime to W such that

$$|\{x \in [N/2] : Wx + b \in A\}| \gg \frac{W}{\phi(W)} \frac{N}{\log N},$$

where ϕ is Euler's totient function.

The expression $\mathbf{E}_{y \in Y} f(y)$ denotes the average of a function f over a finite set Y. Borrowing notation from ergodic theory, we also write $\int_X$ for $\mathbf{E}_{x \in X} f(x)$ and $Tf(x) = f(x-1)$. For convenience we set X equal to the cyclic group $\mathbf{Z}_N$. The fact that the elements x we consider are restricted to lie in the interval $[N/2]$ ensures that there is no problem with wrap-around in X.

For a modulus W and a residue $1 \leq b \leq W$ coprime to W, let us define

$$\Lambda_{W,b;N}(n) = \begin{cases} \frac{\phi(W)}{W} \log(Wn+b), & \text{if } Wn+b \text{ is prime and } 1 \leq n \leq N, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For the purposes of this paper, a *measure* is a non-negative function $\nu : X \to [0, \infty)$ satisfying $\int_X \nu = 1 + o(1)$ and the pointwise bound $\nu = O_\epsilon(N^\epsilon)$ for any $\epsilon > 0$. The measures we will be working with are of the form

$$\nu_{W,b}(n) = \frac{\phi(W)}{W} \log R \left(\sum_{m | Wn+b} \mu(m) \chi\left(\frac{\log m}{\log R}\right) \right)^2, \tag{2}$$

where μ is the Möbius function and χ is an even smooth function supported on $[-1,1]$ satisfying

$$\int_0^1 |\chi'(t)|^2 dt = 1.$$

In [11], Tao and Ziegler defined a *pseudorandom measure* to be a measure satisfying two technical conditions known as the *polynomial forms* condition and the *polynomial correlation* condition, and they showed that $\nu_{W,b}$ as defined in (2) satisfies both of these. We refer the reader to the precise definitions of the polynomial forms and the polynomial correlation condition in [11, Definitions 3.6 and 3.9]. In this paper, we will need a variant of the polynomial forms condition for pairs of pseudorandom measures, which we call the *extra condition*. It will be given in Section 5, where we also verify that this extra condition is satisfied by a pair ν_{W,b_1}, ν_{W,b_2} for potentially distinct b_1, b_2.

Let us list the important remaining parameters.

- Let $d_0 = \max_{1\leq i\leq k} \deg P_i$ denote the maximal degree of the polynomials.
- Let $M = N^{\eta_0}$ be the "coarse scale", which serves as a bound for the step of the polynomial progression. We can take η_0 to be any positive number less than $1/2d_0$.
- Let $0 < \eta_1 \ll \eta_0$ be a tiny parameter, depending on $P_1, \dots, P_k$, which controls the degree of pseudorandomness of a measure ν.
- Let $0 < \eta_2 \ll \eta_1/d_0$, and $R = N^{\eta_2}$ be the sieve level which is used in the construction of ν.

We do not explicitly specify the parameters η_1, η_2, but insist that they depend only on the system $P_1, \dots, P_k$ and are chosen sufficiently small to accommodate all our estimates (notably those arising from the PET induction). Note that Tao and Ziegler also needed the "fine scale" $H = N^{\eta_7}$, but we shall not need it here. In this sense our work is much simpler than [11].

For completeness, we state two basic and well-known inequalities we shall use repeatedly.

Lemma 1 (Cauchy-Schwarz)**.** *Let A, B be sets, let f, F be functions on A and let g be a function on $A \times B$. If $|f| \leq F$ pointwise, then*

$$|\mathbf{E}_{a\in A, b\in B} f(a)g(a,b)|^2 \leq \mathbf{E}_{a\in A} F(a) \mathbf{E}_{a\in A} F(a) \, |\mathbf{E}_{b\in B} g(a,b)|^2 .$$

Lemma 2 (van der Corput)**.** *Let $(x_m)_{m\in\mathbf{Z}}$ be a real-valued sequence satisfying $x_m = 0$ outside the interval $[M]$. Then*

$$\left|\mathbf{E}_{m\in[M]} x_m\right|^2 \ll \mathbf{E}_{|h|<M} \mathbf{E}_{m\in[M]} x_m x_{m+h}.$$

This lemma follows by simply expanding out the square, and is reminiscent of [11, Lemma A.1]. Note that, in contrast with [11, Lemma A.1] where m and h are on different scales, in our situation h and m are on the same scale. This fact is important for us since it will make the Gowers norms appear.

3. Overview of the proof

The main result of Tao and Ziegler is the following [11, Theorem 2.3]. We shall use it as a black box in the sequel, although we will need to delve into the details of the proof in a later part of the argument.

Theorem 5 (Uniform polynomial Szemerédi theorem in the primes)**.** *Let ν be a pseudorandom measure on X. If f is a function on X such that $0 \leq f \leq \nu$, $\int_X f \geq \delta$, then*

$$\mathbf{E}_{m\in[M]} \int_X T^{P_1(Wm)/W} f \dots T^{P_k(Wm)/W} f \geq c(\delta) - o(1)$$

for some constant $c(\delta) > 0$ depending on δ.

Just as in [3], we will also need the following deep result from Green and Tao's programme of counting linear patterns in primes (see [6, 7, 8], but also [3, Theorem 2.2]).

Theorem 6 (Green-Tao, Green-Tao-Ziegler)**.** *For every $d \in \mathbf{Z}^+$, we have*

$$\lim_{N\to\infty} \max_{\substack{1\leq b<W,\\(b,W)=1}} \|\Lambda_{W,b;N} - 1_{[N]}\|_{U^d(\mathbf{Z}_{(2d+1)N})} = 0.$$

Our main result will be deduced from two statements, the first of which is analogous to [3, Lemma 3.5]. The "extra condition" mentioned in the hypotheses of Proposition 1 below is quite technical, and will be defined in Section 5 (Definition 2).

Proposition 1. *Let ν_1, ν_2 be a pair of pseudorandom measures on X satisfying the extra condition. Suppose that $f_1, \dots, f_k$ are functions on X with $|f_i| \leq \nu_1$ for $i = 1, \dots, k$, and that a is a weight on X with support in $[M]$ such that $|a| \leq 1 + \nu_2$. Then*

$$\mathbf{E}_{m\in[M]} \int_X a(m) T^{P_1(Wm)/W} f_1 \cdots T^{P_k(Wm)/W} f_k = O(\|a\|_{U^d(\mathbf{Z}_{(2d+1)M})}) + o(1),$$

where d is an integer depending only on the system of polynomials.

We shall also show that a pair of measures satisfying the hypotheses of Proposition 1 actually exists.

Proposition 2. *For any $b \neq 0$ coprime to W, the pair $\nu_1 = \nu_{W,b}, \nu_2 = \nu_{W,1}$ of pseudorandom measures satisfies the extra condition.*

Remark 3.1. If we were only interested in configurations inside the full set of primes (rather than subsets of positive relative density), this proposition and the needed extra condition would be slightly simpler to state, and to prove. However, we need to be able to take potentially distinct distinct residue classes for the pseudorandom measures governing a and the f_is since the residue class b mod W on which the set A is dense was chosen by the pigeonhole principle.

To conclude this section, we show how Theorem 4 follows from Propositions 1 and 2.

Proof of Theorem 4 assuming Propositions 1 and 2: Suppose that we are given a subset $A \subseteq \mathcal{P}$ of relative upper density δ_0. We shall let $f_1 = \cdots = f_k = f$, where

$$f(x) = \begin{cases} \frac{\phi(W)}{W} \log R & \text{if } R \leq x \leq N/2 \text{ and } Wx + b \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

As remarked in Section 2, by the pigeonhole principle we can choose b such that $\int_X f \gg \delta_0$ provided that N is sufficiently large. Set $\nu_1 = \nu_{W,b}$ so that $0 \leq f \leq \nu_1$. Let

$$g(x) = \begin{cases} \frac{\phi(W)}{W} \log(Wx + 1) & \text{if } R \leq x \leq M \text{ and } Wx + 1 \text{ is prime,} \\ 0 & \text{otherwise.} \end{cases}$$

(in other words, g is the same function as $\Lambda_{W,1;M}$ except on $[R]$), then there is a constant α such that $0 \leq \alpha g \leq \nu_2 = \nu_{W,1}$.

Set $a = \alpha(g - 1_{[M]})$, so that $|a| \leq 1 + \nu_2$. Proposition 2 states that the pair ν_1, ν_2 satisfies the extra condition.

Applying Proposition 1 with these choices yields

$$\mathbf{E}_{m \in [M]} \int_X a(m) T^{P_1(Wm)/W} f \cdots T^{P_k(Wm)/W} f = O(\|g - 1_{[M]}\|_{U^d(\mathbf{Z}_{(2d+1)M})}) + o(1).$$

Since g and $\Lambda_{W,1;M}$ differ on a negligible subset of $\mathbf{Z}_{(2d+1)M}$, Theorem 6 tells us that the right-hand side can be made arbitrarily small if N is sufficiently large. By Theorem 5 we also have

$$\mathbf{E}_{m \in [M]} \int_X T^{P_1(Wm)/W} f \cdots T^{P_k(Wm)/W} f \geq c(\delta_0) - o(1).$$

Since $P_i(Wm)$ is much less than $N/2$ for $m \in [M]$ and the progressions therefore cannot wrap around the group $\mathbf{Z}_N$, we can replace the average over X with the average over [N] and find that

$$\mathbf{E}_{m \in [M]} \mathbf{E}_{x \in [N]} g(m) f(x + P_1(Wm)/W) \cdots f(x + P_k(Wm)/W) \geq c(\delta_0) - o(1). \tag{4}$$

It remains to replace g by a suitable indicator function for the primes congruent to 1 mod W. Since we are only looking for a lower bound, this is straightforward to accomplish. Indeed, we see that the left-hand side of (4) is bounded above by

$$\frac{1}{MN} \frac{\phi(W)}{W} \log(WM + 1) \cdot \left(\frac{\phi(W)}{W} \right)^k (\log R)^k$$

times the number of pairs $(m, x) \in [M] \times [N]$ such that $Wm + 1 \in \mathcal{P}$, $x + P_i(Wm)/W \in [N/2]$ and $Wx + b + P_i(Wm) \in A$ for all $i = 1, \ldots, k$. This is equivalent to saying that for sufficiently large N, the number of pairs $(p, x) \in [WM + 1] \times [N]$ satisfying $p \in \mathcal{P}, p \equiv 1(W)$ such that $x + P_i(p - 1)/W \in [N/2]$ and $Wx + b + P_i(p - 1) \in A$ for all $i = 1, \ldots, k$ is at least

$$(c(\delta_0) - o(1)) \frac{MN}{(\log M)(\log R)^k} \left(\frac{W}{\phi(W)} \right)^{k+1}.$$

But the right-hand side tends to infinity with N, concluding the proof of Theorem 4. □

It thus suffices to prove Propositions 1 and 2.

4. A toy example

In this section we will study the toy example of the configuration $x, x+(p-1)^2$, and use it to motivate the definition of the extra condition in the subsequent section. (Note, however, that the existence of this particular configuration in the primes already follows from the work of Li and Pan [9]. A more general result was recently proved by Rice [10].) For simplicity we assume here that $W = 1$.

Let ν_1, ν_2 be a pair of pseudorandom measures. Suppose that f_0, f_1 are positive functions satisfying $f_0, f_1 \leq \nu_1$ and suppose further that the weight a, which is supported on $[M]$, satisfies $|a| \leq \nu_2$. We shall show that, under an additional assumption on ν_1, ν_2, we can prove the estimate

$$E = \int_X \mathbf{E}_{m\in[M]} a(m) f_0(x) T^{m^2} f_1(x) = O(\|a\|_{U^3(\mathbf{Z}_{7M})}) + o(1). \tag{5}$$

Let us first eliminate f_0 from the average E. By Cauchy-Schwarz, we have

$$E^2 \leq \int_X \nu_1(x) \int_X \nu_1(x) \left| \mathbf{E}_{m\in[M]} a(m) T^{m^2} f_1(x) \right|^2.$$

Recalling that $\int_X \nu_1 = 1 + o(1)$ and using van der Corput, we have

$$\begin{aligned} E^2 &\ll (1+o(1)) \int_X \mathbf{E}_{\substack{m\in[M],\\ |h|<M}} a(m)a(m+h)\nu_1 T^{m^2} f_1 T^{(m+h)^2} f_1 + o(1) \\ &= (1+o(1)) \int_X \mathbf{E}_{\substack{m\in[M],\\ |h|<M}} a(m)a(m+h) T^{-m^2} \nu_1 f_1 T^{2mh+h^2} f_1 + o(1), \end{aligned}$$

where in the second line we shift the variable x by $R_1(m) = -m^2$, thus making the term f_1 appear, rather than a shift of f_1. Let

$$E_1 = \int_X \mathbf{E}_{\substack{m\in[M],\\ |h|<M}} a(m)a(m+h) T^{R_1(m)} \nu_1 f_1 T^{2mh+h^2} f_1.$$

Note that the system of shifts of f_1 appearing in E_1, namely $0, 2mh+h^2$, is "simpler" than the system in E in the sense that the polynomials are now linear in m. Next, we want to eliminate the new shift of f_1 from E_1. Again, by Cauchy-Schwarz, we have

$$E_1^2 \leq \int_X \nu_1(x) \int_X \nu_1(x) \left| \mathbf{E}_{\substack{m\in[M],\\ |h|<M}} a(m)a(m+h) T^{R_1(m)} \nu_1 T^{2mh+h^2} f_1 \right|^2,$$

and by van der Corput this is

$$\begin{aligned} &\ll (1+o(1)) \int_X \nu_1 \mathbf{E}_{\substack{m\in[M],\\ |h|,|k|<M}} a(m)a(m+k)a(m+h)a(m+h+k) \\ &\quad \times T^{R_1(m)} \nu_1 T^{R_1(m+k)} \nu_1 T^{2mh+h^2} f_1 T^{2(m+k)h+h^2} f_1 + o(1). \end{aligned}$$

Let the last integral be E_2. Shifting x by $R_2(m,h) = -(2mh+h^2)$ to make f_1 appear, we obtain

$$\begin{aligned} E_2 &= \int_X \mathbf{E}_{\substack{m\in[M],\\ |h|,|k|<M}} a(m)a(m+k)a(m+h)a(m+h+k) \\ &\qquad \times T^{R_2(m,h)}\nu_1 T^{R_1(m)+R_2(m,h)}\nu_1 T^{R_1(m+k)+R_2(m,h)}\nu_1 f_1 T^{2kh} f_1. \end{aligned}$$

Again, the system of shifts of f_1 is now simpler than the previous one, in that it does not depend on m at all. (Tao and Ziegler deduced from this step their generalized von Neumann inequality, which bounds E in terms of an averaged local Gowers norm of f_1.) We repeat the same process one more time to eliminate f_1 completely from the average. Indeed, by Cauchy-Schwarz, E_2^2 is less than or equal to

$$\begin{aligned} &\Big(\mathbf{E}_{|h|,|k|<M}\int_X \nu_1 T^{2kh}\nu_1\Big)\Big(\mathbf{E}_{|h|,|k|<M}\int_X \nu_1 T^{2kh}\nu_1 \Big|\mathbf{E}_{m\in[M]} a(m)a(m+k)a(m+h)a(m+h+k) \\ &\quad \times T^{R_2(m,h)}\nu_1 T^{R_1(m)+R_2(m,h)}\nu_1 T^{R_1(m+k)+R_2(m,h)}\nu_1\Big|^2\Big). \end{aligned}$$

Since ν_1 satisfies the polynomial forms condition [11, Definition 3.6], the first factor is $1+o(1)$. By van der Corput, the second factor is at most

$$\begin{aligned} &\mathbf{E}_{\substack{m\in[M],\\ |h|,|k|,|l|<M}} \prod_{\omega\in\{0,1\}^3} a(m+\omega\cdot(l,k,h)) \int_X T^{R_2(m,h)}\nu_1 T^{R_1(m)+R_2(m,h)}\nu_1 T^{R_1(m+k)+R_2(m,h)}\nu_1 \\ &\quad \times T^{R_2(m+l,h)}\nu_1 T^{R_1(m+l)+R_2(m+l,h)}\nu_1 T^{R_1(m+k+l)+R_2(m+l,h)}\nu_1. \end{aligned}$$

If it were not for the presence of the integral, then this would be equal to

$$\mathbf{E}_{\substack{m\in[M],\\ |h|,|k|,|l|<M}} \prod_{\omega\in\{0,1\}^3} a(m+\omega\cdot(l,k,h)),$$

which would give us the desired estimate, since the latter quantity is bounded above by a constant times $\|a\|_{U^3(\mathbf{Z}_{7M})}^8$. Indeed, we trivially have

$$\|a\|_{U^3(\mathbf{Z}_{7M})}^8 \gg \frac{1}{M^4}\sum_{m,h,k,l\in\mathbf{Z}_{7M}} \prod_{\omega\in\{0,1\}^3} a(m+\omega\cdot(l,k,h)).$$

Identifying $\mathbf{Z}_{7M}$ with the integers in $(-3M,4M]$, we see that since a is supported on $[M]$, the term $\prod_{\omega\in\{0,1\}^3} a(m+\omega\cdot(l,k,h))$ is non-zero only if $m\in[M]$ and $|h|,|k|,|l|<M$. For these m,h,k,l, the representative of $m+\omega\cdot(l,k,h)$ in $(-3M,4M]$ is $m+\omega\cdot(l,k,h)$ itself, for any $\omega\in\{0,1\}^3$. Thus

$$\|a\|_{U^3(\mathbf{Z}_{7M})}^8 \gg \frac{1}{M^4}\sum_{\substack{m\in[M],\\ |h|,|k|,|l|<M}} \prod_{\omega\in\{0,1\}^3} a(m+\omega\cdot(l,k,h)),$$

where m,h,k,l are now elements of $\mathbf{Z}$, and the claimed bound follows after renormalization.

To continue, let us write $\vec{m} = (m, h, k, l) \in \mathbf{Z}^4$, the integral as $\int_X \prod_{i=1}^6 T^{Q_i(\vec{m})}\nu_1$, where $Q_i \in \mathbf{Z}[\vec{m}]$ for $i = 1, \ldots, 6$, and the weight in front of the integral as $\prod_{j=1}^8 a(L_j(\vec{m}))$, where for $j = 1, \ldots, 8$ the L_j are linear forms defining the eight vertices of the parallelepiped in the U^3 norm. Also, let $\Omega_M = \{(m, h, k, l) \in \mathbf{Z}^4 : m \in [M], |h|, |k|, |l| < M\}$. We want to show that

$$F = \mathbf{E}_{\vec{m}\in\Omega_M} \prod_{j=1}^8 a(L_j(\vec{m})) \times \left(\int_X \prod_{i=1}^6 T^{Q_i(\vec{m})}\nu_1(x) - 1 \right) = o(1). \tag{6}$$

Recalling that $|a| \leq \nu_2$, by Cauchy-Schwarz we have that

$$|F|^2 \ll \|\nu_2\|^8_{U^3(\mathbf{Z}_{7M})} \mathbf{E}_{\vec{m}\in\Omega_M} \prod_{j=1}^8 \nu_2(L_j(\vec{m})) \times \left(\int_X \prod_{i=1}^6 T^{Q_i(\vec{m})}\nu_1(x) - 1 \right)^2.$$

Since ν_2 is pseudorandom, we have $\|\nu_2\|_{U^3(\mathbf{Z}_{7M})} = 1 + o(1)$. (Note that a priori ν_2 is a pseudorandom measure with respect to N, but by choosing R, the sieve level in the definition of ν_2, sufficiently small, we can ensure that ν_2 is also pseudorandom with respect to M.)

By squaring out $\left(\int_X \prod_{i=1}^6 T^{Q_i(\vec{m})}\nu_1(x) - 1 \right)^2$, we see that it suffices to show that

$$\mathbf{E}_{\vec{m}\in\Omega_M} \prod_{j=1}^8 \nu_2(L_j(\vec{m})) \times \left(\int_X \prod_{i=1}^6 T^{Q_i(\vec{m})}\nu_1(x) \right)^k = 1 + o(1) \tag{7}$$

for $k = 0, 1, 2$. But it is precisely expressions of this type that will be governed by our new "extra condition", which we shall formally introduce in the next section.

Remark 4.1. It is well known that if ν_2 is a pseudorandom measure, then so is $(\nu_2 + 1)/2$. It will be easy to see that if the pair ν_1, ν_2 satisfies the extra condition, then so does the pair $\nu_1, (\nu_2 + 1)/2$. The above result therefore also applies to the case where $|a| \leq \nu_2 + 1$, which is what we need in the proof of Theorem 4.

5. A discussion of the polynomial forms condition

Let us recall Tao and Ziegler's definition of the polynomial forms condition [11, Definition 3.6], of which the extra condition will be a variant.

Definition 1 (Polynomial forms condition)**.** A measure $\nu : X \to \mathbf{R}^+$ is said to satisfy the *polynomial forms condition* if for any family of polynomials $Q_j \in \mathbf{Z}[m_1, \ldots, m_D]$, $j \in J$, satisfying

- the difference $Q_i - Q_j$ is not constant for $i \neq j$;
- the number of polynomials $|J|$ and the number of variables l are bounded by $1/\eta_1$;
- the total degree of each Q_j for $j \in J$ is at most d_0, and all coefficients are at most CW^{d_0}, where C is a constant (depending on $P_1, \ldots, P_k$);

we have

$$\mathbf{E}_{\vec{h}\in\Omega\cap\mathbf{Z}^D}\int_X\prod_{j\in J}T^{Q_j(\vec{h})}\nu(x)=1+o_\epsilon(1) \tag{8}$$

for any convex body $\Omega\subset\mathbf{R}^D$ of inradius at least N^ϵ and contained in the ball $B(0,M^2)$.

Inequality (7) does not exactly follow from Definition 1, but we still can deduce it using Tao and Ziegler's machinery. We make the following general definition.

Definition 2 (Extra condition)**.** A pair of measures $\nu_1,\nu_2:X\to\mathbf{R}^+$ is said to satisfy the *extra condition* if for any family of polynomials $Q_j\in\mathbf{Z}[m_1,\dots,m_D]$, $j\in J_1$, and any family of linear forms $L_j:\mathbf{Z}^D\to\mathbf{Z}$, $j\in J_2$, satisfying

- the difference of polynomials Q_i-Q_j is not constant for $i\neq j$;
- the number of polynomials $|J_1|$ and the number of variables D are bounded by $1/\eta_1$;
- the total degree of each Q_j for $j\in J_1$ is at most d_0, and all coefficients are at most CW^{d_0}, where C is a constant (depending on $P_1,\dots,P_k$);
- the linear forms L_j, $j\in J_2$ are pairwise linearly independent;
- the number of linear forms $|J_2|$ is bounded by $1/\eta_1$;
- the coefficients of each L_j are 0 or 1;

we have

$$\mathbf{E}_{\vec{m}\in\Omega_{M,D}}\prod_{j\in J_2}\nu_2(L_j(\vec{m}))\left(\int_X\prod_{j\in J_1}\nu_1(x+Q_j(\vec{m}))\right)^k=1+o(1) \tag{9}$$

for $k=0,1,2$, where $\Omega_{M,D}=\{(m,h_1,\dots,h_{D-1})\in\mathbf{Z}^D : m\in[M],|h_i|<M \text{ for any } i=1,\dots,D-1)\}$.

Remark 5.1. The extra condition is tailor-made to suit our needs. One could merge it with the polynomial forms condition to obtain a more general statement, but this appears unnecessary.

Let us now show that for any $b\neq 0$ coprime to W, the pair $\nu_1=\nu_{W,b}$, $\nu_2=\nu_{W,1}$ given by (2) satisfies the extra condition. That is, we shall turn to proving Proposition 2. To begin with, let us recall some more definitions from [11].

Definition 3 (Good, bad and terrible primes)**.** Let $P_j\in\mathbf{Z}[x_1,\dots,x_D]$, $j\in J$, be a family of polynomials. We say a prime p is *good* with respect to the family $P_j, j\in J$, if

- the polynomials $P_j \pmod p$, $j\in J$, (considered as elements of $\mathbf{F}_p[x_1,\dots,x_D]$) are pairwise coprime;
- for each $j\in J$, there is a variable x_i such that P_j can be expressed as $P_j=P_{j,1}x_i+P_{j,0}$ where $P_{j,1},P_{j,0}\in\mathbf{F}_p[x_1,\dots,x_{i-1},x_{i+1},\dots,x_D]$ are such that $P_{j,1}$ is non-zero and coprime to $P_{j,0}$.

We say p is *bad* if it is not good. We say p is *terrible* if at least one of the P_j vanishes identically mod p.

We shall need the following slight variant of the basic correlation estimate [11, Proposition 10.1], in which we now have a pair of pseudorandom measures.

Proposition 3 (Correlation estimate)**.** *Write $\nu_1 = \nu_{W,b_1}$ and $\nu_2 = \nu_{W,b_2}$ with $b_1, b_2 \neq 0$ and coprime to W. Let $J_1, J_2 \subset \mathbf{N}$ be two disjoint indexing sets, and let $J = J_1 \cup J_2$. For $j \in J$, let $P_j \in \mathbf{Z}[x_1, \ldots, x_D]$ have degree at most d. Let Ω be a convex body in $\mathbf{R}^D$ of inradius at least $R^{4|J|+1}$. Let $\mathcal{P}_b$ be the set of primes $w \leq p \leq R^{\log R}$ which are bad with respect to $((WP_j + b_1)_{j\in J_1}, (WP_j + b_2)_{j\in J_2})$, and suppose that there are no terrible primes in the same range. Then*

(10)

$$\mathbf{E}_{x\in\Omega\cap\mathbf{Z}^D} \prod_{j\in J_1} \nu_1(P_j(x)) \prod_{j\in J_2} \nu_2(P_j(x)) = 1 + o_{D,J,d}(1) + O_{D,J,d}\left(\mathrm{Exp}\left(O_{D,J,d}\left(\sum_{p\in\mathcal{P}_b} \frac{1}{p}\right)\right)\right).$$

Here we have written $\mathrm{Exp}(x) = \max(e^x - 1, 0)$, so that $\mathrm{Exp}(x) \ll x$ when $0 \leq x \ll 1$.

The proof of [11, Proposition 10.1] generalizes readily to yield Proposition 3. However, since it is relatively complex and buried in various appendices of a long paper, we give the details for the convenience of the reader in the appendix to this paper.

To conclude, let us deduce the extra condition from the above correlation estimate.

Proof of Proposition 2 assuming Proposition 3: Recall that we want to show that

$$\mathbf{E}_{\vec{m}\in\Omega_{M,D}} \prod_{j\in J_2} \nu_2(L_j(\vec{m})) \left(\int_X \prod_{j\in J_1} \nu_1(x + Q_j(\vec{m}))\right)^k = 1 + o(1) \tag{11}$$

for $k = 0, 1, 2$, where ν_1, ν_2 are defined as in Proposition 3. If $k = 0$, then the integral disappears, and we are left to show that

$$\mathbf{E}_{\vec{m}\in\Omega_{M,D}} \prod_{j\in J_2} \nu_2(L_2(\vec{m})) = 1 + o(1).$$

If we choose R sufficiently small in terms of M, then ν_2 is pseudorandom with respect to M, and (11) simply follows from Green and Tao's linear forms condition in [5, Definition 3.1] in this case.

Let us now discuss the case $k = 2$ (the case $k = 1$ is even simpler). First we make a reduction, replacing the integral on X with the average $\mathbf{E}_{x\in[N]}$, thus regarding ν_1 as a function on $\mathbf{Z}$ rather than X. The values of $\nu_1(x + Q_j(\vec{m}))$ may be different when ν_1 is regarded as a function on $\mathbf{Z}$ because of the wrap-around effect, but they must agree whenever $1 \leq x \leq N - O((WM)^{d_0})$. Recall that we also have the bound $\nu_1 \ll_\epsilon N^\epsilon$ for any $\epsilon > 0$. Thus

$$\int_X \prod_{j\in J_1} \nu_1(x + Q_j(\vec{m})) - \mathbf{E}_{x\in[N]} \prod_{j\in J_1} \nu_1(x + Q_j(\vec{m})) \ll_\epsilon (WM)^{d_0} N^{\epsilon-1}, \tag{12}$$

for any $\epsilon > 0$, and (11) follows if we can show that

$$\mathbf{E}_{\vec{m}\in\Omega_{M,D}} \prod_{j\in J_2} \nu_2(L_j(\vec{m})) \left(\mathbf{E}_{x\in[N]} \prod_{j\in J_1} \nu_1(x + Q_j(\vec{m})) \right)^2 = 1 + o(1). \tag{13}$$

Expanding out (13), we see that it is equivalent to

$$\mathbf{E}_{\vec{m}\in\Omega_{M,D},x,x'\in[N]} \prod_{j\in J_2} \nu_2(L_j(\vec{m})) \prod_{j\in J_1} \nu_1(x + Q_j(\vec{m}))\nu_1(x' + Q_j(\vec{m})) = 1 + o(1). \tag{14}$$

Now this expression falls within the scope of Proposition 3: the polynomials in question are $L_i(\vec{m}), x + Q_j(\vec{m}), x' + Q_j(\vec{m})$, in variables $\vec{m}, x, x'$. For this system, there is no terrible prime greater than w, and the only bad primes greater than w are those dividing $Q_j - Q_{j'}$ for some $j \neq j'$. Therefore, the left hand side of (14) equals $1 + o(1) + O\left(\mathrm{Exp}\left(O\left(\sum_{p\in\mathcal{P}_b} p^{-1}\right)\right)\right)$, where $\mathcal{P}_b$ denotes the set of primes dividing $Q_j - Q_{j'}$ for some $j \neq j'$.

But just as in the proof of [11, Corollary 11.2], if a prime p divides $Q_j - Q_{j'}$ for some $j \neq j'$, then p must divide a non-zero difference of the coefficients of the Q_js (recall that $Q_j - Q_{j'}$ are not constant for any $j \neq j$). These coefficients are bounded by $O(W^{d_0})$, so that the total product of such p is at most $O(W^{O(1)})$. As a result, the number of $p \in \mathcal{P}_b$ (which are greater than w) is at most $\log(O(W^{O(1)}))/\log w = o(\log W)$. But then $\sum_{p\in\mathcal{P}_b} p^{-1} < \sum_{p\in\mathcal{P}_b} w^{-1} = o(1)$, since $\log W \ll w$. □

Remark 5.2. The extra condition is in a sense simpler than Tao and Ziegler's full polynomial forms condition, in that all the variables are at scale M, whereas the polynomial forms condition ([11, Theorem 11.1]) makes a statement about convex bodies of inradius as small as N^ϵ. This explains why Tao and Ziegler had to do some extra work to prove the polynomial forms condition from the correlation estimate, while for us it is almost immediate.

6. The general case

As is to be expected, we proceed by PET induction to prove Proposition 1 in the general case. We follow the notation in [3] with the simplification that the dimension is equal to 1.

Given a family of polynomials $\mathcal{Q} = (q_1, \dots, q_k)$ in a variable n (and possibly in other variables), the maximum of the degrees of the q_i (with respect to n) is called the *degree of the family* $\mathcal{Q}$. We work with families of polynomials whose degree is smaller than or equal to a fixed number s.

For $i = 1, \dots, k$, define $\mathcal{Q}'$ to be the possibly empty set

$$\mathcal{Q}' = \{q_i \in \mathcal{Q} : q_i \text{ is constant in } n \}.$$

Two polynomials are said to be *equivalent* if they have the same degree and the same leading coefficient (in n). For $j = 1, \dots, s$, let w_j denote the number of distinct non-equivalent classes of polynomials of degree j in $\mathcal{Q} \setminus \mathcal{Q}'$. Finally, define the *type* of the family $\mathcal{Q}$ to be the vector

$(w_1, \dots, w_s)$. A family is said to be of type zero if all the w_j are zero, in which case all the polynomials are constant (in n). The set of types can be ordered lexicographically, by stipulating that $w = (w_1, \dots, w_s) < w' = (w'_1, \dots, w'_s)$ if there exists d such that $w_d < w'_d$ and $w_j = w'_j$ for all $j > d$.

It follows that any decreasing sequence of types of families of polynomials is eventually stationary, and thus any inductive process that reduces the type must eventually stop.

Given a family $\mathcal{Q} = (q_1, \dots, q_k)$, $q \in \mathbf{Z}[t]$ and $h \in \mathbf{N}$, define following [2] the *van der Corput operation* (q,h)-vdC$(\mathcal{Q})$ by setting

$$(q,h)\text{-vdC}(\mathcal{Q}) = (S_h\mathcal{Q} - q, \mathcal{Q} - q),$$

where $S_h q(n) = q(n+h)$, $S_h\mathcal{Q} = (S_h q_1, \dots, S_h q_k)$ and $\mathcal{Q} - q = (q_1 - q, \dots, q_k - q)$.

The crucial observation is the following [1, 3].

Lemma 3. *Let $\mathcal{Q}$ be a family of polynomials of non-zero type. Then there exists $q \in \mathcal{Q} \setminus \mathcal{Q}'$ such that for all $h \in \mathbf{N}$, the family (q,h)-vdC$(\mathcal{Q} \setminus \mathcal{Q}')$ has strictly smaller type than $\mathcal{Q}$.*

For example, in the toy example in Section 4, we passed from type $(1,0,1)$ to $(1,1,0)$ to $(2,0,0)$.

In the van der Corput operation, we focus on a single variable n. However, we also need to keep in mind that the polynomials in $\mathcal{Q}$ and (q,h)-vdC$(\mathcal{Q})$ are *multivariate*, with a new variable being introduced at each step of the van der Corput operation. This is important when verifying the hypotheses of the (polynomial forms or extra) condition as we apply them to the averages arising throughout.

Proof of Proposition 1: Let us recall that we have functions $f_i, i = 1, \dots, k$ satisfying $|f_i| \leq \nu_1$, and a weight a supported on $[M]$ satisfying $|a| \leq \nu_2$. We start with the average

$$E = \mathbf{E}_{m \in [M]} \int_X a(m) \prod_{q_i \in \mathcal{Q}} T^{q_i} f_i,$$

where the family $\mathcal{Q}$ initially consists of the polynomials $q_i(m) = P_i(Wm)/W$, $j = 1, \dots, k$. Define $\mathcal{Q}' \subseteq \mathcal{Q}$ to be the subset of polynomials which are constant in the variable m. (By hypothesis on the P_i, this actually means that the $q \in \mathcal{Q}'$ are identically 0, but later on, with additional variables, this is not necessarily the case.)

We first apply the Cauchy-Schwarz inequality to obtain

$$|E|^2 \leq \left(\int_X \prod_{q \in \mathcal{Q}'} T^q \nu_1 \right) \left(\int_X \prod_{q \in \mathcal{Q}'} T^q \nu_1 |\mathbf{E}_{m \in [M]} a(m) \prod_{q_i \in \mathcal{Q} \setminus \mathcal{Q}'} T^{q_i} f_i|^2 \right).$$

By the properties of ν_1, the first integral is $1 + o(1)$ (if $\mathcal{Q}' = \emptyset$, we interpret the empty product as equal to 1), and we can bound the second factor by van der Corput by a constant times

$$\int_X \prod_{q \in \mathcal{Q}'} T^q \nu_1 \mathbf{E}_{\substack{m \in [M], \\ |h_1| < M}} a(m) a(m + h_1) \prod_{q_i \in \mathcal{Q} \setminus \mathcal{Q}'} T^{q_i(m)} f_i T^{q_i(m+h_1)} f_i + o(1)$$

Now by Lemma 3, there exists $q_1 \in \mathcal{Q}\backslash\mathcal{Q}'$ such that for all h_1, the family $\mathcal{Q}_1 := (q^{(1)}, h_1)$-vdC$(\mathcal{Q}\backslash\mathcal{Q}')$ has strictly smaller type than $\mathcal{Q}$. We also write $\mathcal{Q}_1^\dagger := \mathcal{Q}' - q^{(1)}$ for the recently deceased nodes. With this notation, shifting by $q^{(1)}$ gives, up to an error term, the expression

$$E_1 = \int_X \mathbf{E}_{\substack{m\in[M],\\ |h_1|<M}} a(m)a(m+h_1) \prod_{q\in\mathcal{Q}_1^\dagger} T^q\nu_1 \prod_{q_i\in\mathcal{Q}_1} T^{q_i} f_{j_i},$$

where the f_{j_i} belong to the set $\{f_1, \dots, f_k\}$. There is no need to keep track of them, so we shall simply write f with no subscript in the sequel.

Write $\mathcal{Q}_1'$ for those polynomials in $\mathcal{Q}_1$ of degree 0 in m. By Cauchy-Schwarz,

$$|E_1|^2 \leq \left(\int_X \prod_{q\in\mathcal{Q}_1'} T^q\nu_1\right)\left(\int_X \prod_{q\in\mathcal{Q}_1'} T^q\nu_1 \left|\mathbf{E}_{m,h_1\in[M]} a(m)a(m+h_1) \prod_{q\in\mathcal{Q}_1^\dagger} T^q\nu_1 \prod_{q\in\mathcal{Q}_1\backslash\mathcal{Q}_1'} T^q f\right|^2\right).$$

By the polynomial forms condition, the first factor is $1+o(1)$, and the second can be bounded above by van der Corput as a constant times

$$\int_X \mathbf{E}_{\substack{m\in[M],\\ |h_1|,|h_2|<M}} a(m)a(m+h_1)a(m+h_2)a(m+h_1+h_2) \prod_{q\in\mathcal{Q}_1^\dagger} T^{q(m)}\nu_1 T^{q(m+h_2)}\nu_1 \prod_{q\in\mathcal{Q}_1\backslash\mathcal{Q}_1'} T^{q(m)} f T^{q(m+h_2)} f.$$

Of course the dependence of the polynomials on h_1 is suppressed here. By Lemma 3, there is $q^{(2)} \in \mathcal{Q}_1 \setminus \mathcal{Q}_1'$ such that the family $\mathcal{Q}_2 := (q^{(2)}, h_2)$-vdC$(\mathcal{Q}_1 \setminus \mathcal{Q}_1')$ has strictly smaller type. Define also $\mathcal{Q}_2^\dagger := (\mathcal{Q}_1' - q^{(2)}) \cup (\mathcal{Q}_1^\dagger - q^{(2)})$, leading after rearranging to

$$E_2 = \int_X \mathbf{E}_{\substack{m\in[M],\\ |h_1|,|h_2|<M}} a(m)a(m+h_1)a(m+h_2)a(m+h_1+h_2) \prod_{q\in\mathcal{Q}_2^\dagger} T^q\nu_1 \prod_{q\in\mathcal{Q}_2} T^q f.$$

Continuing in this vein, we set at step s $\mathcal{Q}_{s+1} := (q^{(s+1)}, h_{s+1})$-vdC$(\mathcal{Q}_s \setminus \mathcal{Q}_s')$ and $\mathcal{Q}_{s+1}^\dagger := (\mathcal{Q}_s' - q^{(s+1)}) \cup (\mathcal{Q}_s^\dagger - q^{(s+1)})$, all the while reducing the type. We also set $\mathcal{Q}_0 = \mathcal{Q}, \mathcal{Q}_0^\dagger = \emptyset$, so that the above recursive definition is valid for all $s \geq 0$. By Lemma 3 we reach a point, at step t say, where $\mathcal{Q}_t = \mathcal{Q}_t'$. In other words, the system is of zero type and all active polynomials are of degree 0 in m.

In order to be able to use the polynomial forms condition at every step, we need to ensure that no two polynomials in $\mathcal{Q}_s'$ differ by constants. We will in fact prove a slightly stronger statement, which we shall need later.

Claim 1. *For any $0 \leq s \leq t$, no two polynomials in $\mathcal{Q}_s \cup \mathcal{Q}_s^\dagger$ differ by constants.*

Proof of Claim 1: It is easy to see that all our polynomials are 0 when all the variables are 0. Therefore, it suffices to show that all the polynomials in $\mathcal{Q}_s \cup \mathcal{Q}_s^\dagger$ are distinct as multivariate polynomials. We prove this by induction on s. When $s = 0$, this follows from our assumption that the polynomials P_i are distinct. Suppose we know already that all the polynomials in

$\mathcal{Q}_s \cup \mathcal{Q}_s^\dagger$ are distinct. By definition, the family $\mathcal{Q}_{s+1} \cup \mathcal{Q}_{s+1}^\dagger$ is obtained by subtracting $q^{(s+1)}$ from all the polynomials in the family $S_{h_{s+1}}(\mathcal{Q}_s \setminus \mathcal{Q}_s') \cup (\mathcal{Q}_s \setminus \mathcal{Q}_s') \cup \mathcal{Q}_s' \cup \mathcal{Q}_s^\dagger$. Thus it suffices to show that all polynomials in the latter family are distinct. Recall that the polynomials in $\mathcal{Q}_s \setminus \mathcal{Q}_s'$ are distinct and non-constant in m. Thus the polynomials in $S_{h_{s+1}}(\mathcal{Q}_s \setminus \mathcal{Q}_s')$ are distinct from each other and from the rest, since they have a new variable, namely h_{s+1}, and are non-constant in this variable. The remaining polynomials from $(\mathcal{Q}_s \setminus \mathcal{Q}_s') \cup \mathcal{Q}_s' \cup \mathcal{Q}_s^\dagger$ are distinct by induction hypothesis since $(\mathcal{Q}_s \setminus \mathcal{Q}_s') \cup \mathcal{Q}_s' = \mathcal{Q}_s$. □

Claim 1 shows that our calculations so far have been valid. Returning to step t, we have

$$E_t = \mathbf{E}_{\substack{m\in[M],\\ |h_1|,\dots,|h_t|<M}} \prod_{\omega\in\{0,1\}^t} a(m+\omega\cdot(h_1,\dots,h_t)) \prod_{q\in\mathcal{Q}_t^\dagger} T^q\nu_1 \prod_{q\in\mathcal{Q}_t'} T^q f,$$

and we apply Cauchy-Schwarz one more time to get

$$|E_t|^2 \le \left(\int_X \mathbf{E}_{|h_1|,\dots,|h_t|<M} \prod_{q\in\mathcal{Q}_t'} T^q\nu_1\right) \times \left(\int_X \mathbf{E}_{|h_1|,\dots,|h_t|<M} \prod_{q\in\mathcal{Q}_t'} T^q\nu_1 \left|\mathbf{E}_{m\in[M]} \prod_{\omega\in\{0,1\}^t} a(m+\omega\cdot(h_1,\dots,h_t)) \prod_{q\in\mathcal{Q}_t^\dagger} T^q\nu_1\right|^2\right).$$

A final van der Corput gives

$$\mathbf{E}_{\substack{m\in[M],\\ |h_1|,\dots,|h_{t+1}|<M}} \prod_{\omega\in\{0,1\}^{t+1}} a(m+\omega\cdot(h_1,\dots,h_{t+1})) \int_X \prod_{q\in\mathcal{Q}_t'} T^{q(m)}\nu_1 \prod_{q\in\mathcal{Q}_t^\dagger} T^{q(m)}\nu_1 T^{q(m+h_{t+1})}\nu_1,$$

and we write F for the difference between this expression and $\mathbf{E}_{\substack{m\in[M],\\ |h_1|,\dots,|h_t|<M}} \prod_{\omega\in\{0,1\}^{t+1}} a(m+\omega\cdot(h_1,\dots,h_{t+1}))$. As in the example in Section 4, by Cauchy-Schwarz we have that F^2 is bounded above by $\|\nu_2\|^{2^{t+1}}_{U^{t+1}(\mathbf{Z}_{(2t+3)M})}$ times the average

$$\mathbf{E}_{m,h_1,\dots,h_{t+1}\in[M]} \prod_{\omega\in\{0,1\}^{t+1}} \nu_2(m+\omega\cdot(h_1,\dots,h_{t+1})) \times \left(\int_X \prod_{q\in\mathcal{Q}_t'} T^{q(m)}\nu_1 \prod_{q\in\mathcal{Q}_t^\dagger} T^{q(m)}\nu_1 T^{q(m+h_{t+1})}\nu_1 - 1\right)^2. \tag{15}$$

In order to use the extra condition on the last expression, we need to verify that no two polynomials in $\mathcal{Q}_t^\dagger \cup S_{h_{t+1}}\mathcal{Q}_t^\dagger \cup \mathcal{Q}_t'$ differ by constants. Again, since all these polynomials are 0 when evaluated at 0, it suffices to show that they are distinct. Before seeing this, let us make some observations.

Claim 2. *For any $0 \le s \le t$, for any polynomials $p \in \mathcal{Q}_s^\dagger$ and $q \in \mathcal{Q}_s \setminus \mathcal{Q}_s'$, $p - q$ is not constant in m.*

Proof of Claim 2: For $s = 0$ there is nothing to prove. Suppose the claim is true for $s \leq t-1$. Let $p \in \mathcal{Q}_{s+1}^{\dagger}$ and $q \in \mathcal{Q}_{s+1} \setminus \mathcal{Q}'_{s+1}$. Write $p = u - q^{(s+1)}$ for $u \in \mathcal{Q}'_s \cup \mathcal{Q}_s^{\dagger}$ and $q = v - q^{(s+1)}$ for $v \in (\mathcal{Q}_s \setminus \mathcal{Q}'_s) \cup S_{h_{s+1}}(\mathcal{Q}_s \setminus \mathcal{Q}'_s)$. It remains to see that $u - v$ is not constant in m. If $u \in \mathcal{Q}'_s$, then u is constant in m, but none of the polynomials in $(\mathcal{Q}_s \setminus \mathcal{Q}'_s) \cup S_{h_{s+1}}(\mathcal{Q}_s \setminus \mathcal{Q}'_s)$ are constant in m. Suppose $u \in \mathcal{Q}_s^{\dagger}$. If $v \in \mathcal{Q}_s \setminus \mathcal{Q}'_s$ then $u - v$ is not constant in m by induction hypothesis. If $v \in S_{h_{s+1}}(\mathcal{Q}_s \setminus \mathcal{Q}'_s)$, we write $v(m) = w(m + h_{s+1})$ for some $w \in \mathcal{Q}_s \setminus \mathcal{Q}'_s$. Then $u(m) - w(m + h_{s+1})$ is not constant in m, since it is already not constant in m upon setting $h_{s+1} = 0$. This proves Claim 2. □

Claim 3. *For any $0 \leq s \leq t$, the polynomials in $\mathcal{Q}_s^{\dagger}$ are not constant in m.*

Proof of Claim 3: Indeed, if $s \geq 1$ then by definition we have

$$\mathcal{Q}_s^{\dagger} = (\mathcal{Q}_{s-1}^{\dagger} - q^{(s)}) \cup (\mathcal{Q}'_{s-1} - q^{(s)})$$

for some $q^{(s)} \in \mathcal{Q}_s \setminus \mathcal{Q}'_s$. Since the polynomials in $\mathcal{Q}'_{s-1}$ are constant in m, the polynomials in $\mathcal{Q}'_{s-1} - q^{(s)}$ are not constant in m. From Claim 2, we know that the polynomials in $\mathcal{Q}_{s-1}^{\dagger} - q^{(s)}$ are not constant in m. □

From Claim 1 we know that the polynomials in $\mathcal{Q}_t^{\dagger} \cup \mathcal{Q}'_t$ are distinct. The polynomials in $S_{h_{t+1}}\mathcal{Q}_t^{\dagger}$ have a new variable, namely h_{t+1}. From Claim 3 we know that they are not constant in h_{s+1}, hence distinct from $\mathcal{Q}_t^{\dagger} \cup \mathcal{Q}'_t$.

It is also easy to see that the total degrees of the polynomials appearing in this process are not increased, so they are always at most d_0. Also, all of their coefficients can be bounded by a constant C times the maximum of the absolute values of the coefficients of the original polynomials (namely $P_i(Wm)/W$), where C depends only on $P_1, \ldots, P_k$. It follows that all polynomial expressions in (15) satisfy the hypotheses of the extra condition. As in Section 4, expanding out the expression in (15) and using the extra condition to see that it equals $o(1)$ concludes the proof of Proposition 1. □

7. Concluding remarks

Our proof actually gives a lower bound for the number of desired configurations. More precisely, it shows that the number of pairs $(n, p) \in [N] \times [M]$ for which $n + P_1(p-1), \ldots, n + P_k(p-1)$ and p are all prime is at least $cNM/(\log N)^{k+1}$, as long as M grows like a power of N that is at most $N^{1/2d_0}$. Just as in [11], this follows from the proof of Theorem 4 since we can choose w to be arbitrarily slowly growing (see also the more detailed discussion at the start of [5, Section 11]). This is of the correct order of magnitude if one assumes the Bateman-Horn conjecture.

One can also see that Proposition 1 remains true if we apply it to functions f_i satisfying $|f_i| \leq 1$ instead of $|f_i| \leq \nu$. Indeed, under this condition, the proof of Proposition 1 is even more straightforward: at each step, in place of the factor $\left(\int_X \prod_{q \in \mathcal{Q}'_i} T^q \nu_1\right) = 1 + o(1)$, one simply has a constant 1. Thus, taking $f_1 = \ldots = f_k$ to be the characteristic functions of a

set $A \subseteq [N]$ of density δ and using the uniform Bergelson-Leibman theorem [11, Theorem 3.2], one actually obtains a slightly different proof of Theorem 3. It is, of course, in the same spirit as [3, Theorem 1.2], but has the advantage that it gives a lower bound on the number of configurations. Moreover, an inspection of the proof of [11, Theorem 3.2] shows that it is valid as long as M goes to infinity with N. In particular, we do not have to require M to be as big as a small power of N. Of course, the constant $c(\delta)$ now has to depend on the growth of M. More precisely, we have the following result.

Proposition 4. *Let $F(N)$ be any function that goes to infinity with N. Suppose $F(N) < M < N^{1/2d_0}$. Then for any $\delta > 0$, there is a constant $c(F,\delta) > 0$ such that the following holds. Let A be any subset of $[N]$ of density δ. Then A contains at least $c(F,\delta)NM/\log M$ configurations of the form $a + P_1(p-1), \ldots, a + P_k(p-1)$, where $p \leq M$ is a prime.*

This bound does not follow from [3]. On the other hand, the proof of Frantzikinakis, Host and Kra shows that if $A \subset \mathbf{Z}$ and $\overline{d}(A) > 0$, then

$$\overline{d}(A \cap (A - P_1(p-1)) \cdots \cap (A - P_k(p-1))) > 0$$

for p in a set of positive relative density in the primes.

Appendix: The generalized correlation estimate

In this appendix we point out the modifications that need to be made to the proof of [11, Proposition 10.1] to obtain Proposition 3 (correcting some misprints from [11] in the process). Recall that we had $\nu_1 = \nu_{W,b_1}$ and $\nu_2 = \nu_{W,b_2}$ and two disjoint indexing sets $J_1, J_2 \subset \mathbf{N}$, $J = J_1 \cup J_2$. For $j \in J$, we have polynomials $P_j \in \mathbf{Z}[x_1, \ldots, x_D]$ of degree at most d. The convex body $\Omega \subset \mathbf{R}^D$ was assumed to have inradius at least $R^{4|J|+1}$. We denoted by $\mathcal{P}_b$ the set of primes $w \leq p \leq R^{\log R}$ which are bad with respect to $((WP_j + b_1)_{j\in J_1}, (WP_j + b_2)_{j\in J_2})$, and assumed that there are no terrible primes in the same range.

Proof of Proposition 3: We wish to estimate

$$\mathbf{E}_{x \in \Omega \cap \mathbf{Z}^D} \prod_{j \in J_1} \nu_1(P_j(x)) \prod_{j \in J_2} \nu_2(P_j(x)).$$

Expanding this out in terms of the definitions of ν_1, ν_2, we find that

(16)

$$\left(\frac{\phi(W)}{W}\log R\right)^{|J|} \sum_{j\in J} \sum_{m_j, m_j' \geq 1} \left(\prod_{j\in J} \mu(m_j)\mu(m_j')\chi\left(\frac{\log m_j}{\log R}\right)\chi\left(\frac{\log m_j'}{\log R}\right)\right)$$
$$\mathbf{E}_{x\in\Omega\cap\mathbf{Z}^D} \prod_{j\in J_1} 1_{\mathrm{lcm}(m_j,m_j')|WP_j(x)+b_1} \prod_{j\in J_2} 1_{\mathrm{lcm}(m_j,m_j')|WP_j(x)+b_2},$$

where $\mathrm{lcm}(a,b)$ denotes the least common multiple of two integers a and b. Setting $M = \mathrm{lcm}((m_j)_{j\in J},(m'_j)_{j\in J})$, we observe that M can be assumed to be square-free (due to the presence of the Möbius function) and of size at most $R^{2|J|}$ (due to the restrictions on each m_j, m'_j imposed by the cutoff χ). Each function $x \mapsto 1_{\mathrm{lcm}(m_j,m'_j)|WP_j(x)+b_1}$, $x \mapsto 1_{\mathrm{lcm}(m_j,m'_j)|WP_j(x)+b_2}$ is periodic with respect to $M\cdot\mathbf{Z}^D$, and can therefore be defined on $\mathbf{Z}^D_M$. By [11, Corollary C.3], we have

$$\begin{aligned}\mathbf{E}_{x\in\Omega\cap\mathbf{Z}^D}&\prod_{i=1,2}\prod_{j\in J_i}1_{\mathrm{lcm}(m_j,m'_j)|WP_j(x)+b_i}\\ &=\left(1+O\left(R^{-2|J|-1}\right)\right)\mathbf{E}_{y\in\Omega\cap\mathbf{Z}^D_M}\prod_{i=1,2}\prod_{j\in J_i}1_{\mathrm{lcm}(m_j,m'_j)|WP_j(y)+b_i},\end{aligned}$$

where the O error term is easily seen to result in an additive $o(1)$ error, and will therefore be negligible. Setting

$$\alpha_{(a_j)_{j\in J}}=\mathbf{E}_{y\in\mathbf{Z}^D_{\mathrm{lcm}((a_j)_{j\in J})}}\prod_{i=1,2}\prod_{j\in J_i}1_{a_j|WP_j(y)+b_i},$$

it therefore suffices to show that

(17)

$$\begin{aligned}\left(\frac{\phi(W)}{W}\log R\right)^{|J|}&\sum_{j\in J}\sum_{m_j,m'_j\geq 1}\left(\prod_{j\in J}\mu(m_j)\mu(m'_j)\chi\left(\frac{\log m_j}{\log R}\right)\chi\left(\frac{\log m'_j}{\log R}\right)\right)\alpha_{(\mathrm{lcm}(m_j,m'_j))_{j\in J}}\\ &=1+o_{D,J,d}(1)+O_{D,J,d}\left(\mathrm{Exp}\left(O_{D,J,d}\left(\sum_{p\in\mathcal{P}_b}\frac{1}{p}\right)\right)\right).\end{aligned}$$

By the Chinese remainder theorem α is multiplicative in the sense that if $\mathrm{lcm}(m_j,m'_j)=\prod_p p^{r_j(p)}$, then

$$\alpha_{(\mathrm{lcm}(m_j,m'_j))_{j\in J}}=\prod_p\alpha_{(p^{r_j(p)})_{j\in J}},$$

where the latter is a finite product. Since the m_j are assumed to be squarefree, $r_j(p)$ is either 0 or 1 for each j and each p, and we obtain

$$\alpha_{(\mathrm{lcm}(m_j,m'_j))_{j\in J}}=\prod_p c_p((WP_j+b_1)_{j\in J_1,r_j(p)=1},(WP_j+b_2)_{j\in J_2,r_j(p)=1}),$$

where the local factor $c_p(P_1,\dots,P_k)$ is defined by

$$c_p(P_1,\dots,P_k)=\mathbf{E}_{y\in\mathbf{F}_p^D}\prod_{j\in J}1_{P_j(y)\equiv 0(p)}.$$

So the left-hand side of (17) becomes

$$\left(\frac{\phi(W)}{W}\log R\right)^{|J|}\sum_{j\in J}\sum_{m_j,m_j'\geq 1}\left(\prod_{j\in J}\mu(m_j)\mu(m_j')\chi\left(\frac{\log m_j}{\log R}\right)\chi\left(\frac{\log m_j'}{\log R}\right)\right) \prod_{p\leq R^{\log R}} c_p((WP_j+b_1)_{j\in J_1,r_j(p)=1},(WP_j+b_2)_{j\in J_2,r_j(p)=1}), \tag{18}$$

where we were able to restrict the product to primes less than $R^{\log R}$ because each m_j is bounded by R.

We can now replace χ by terms which are multiplicative in m_j, m_j', using the Fourier expansion

$$\chi(x)=e^{-x}\int_{-\infty}^{\infty}\phi(\xi)e^{-ix\xi}d\xi$$

for a smooth and rapidly decaying function ϕ. We have

$$\chi\left(\frac{\log m_j}{\log R}\right)=\int_{-\infty}^{\infty}\phi(\xi_j)m_j^{-z_j}d\xi\ ,\quad \text{where}\quad z_j=\frac{1+i\xi_j}{\log R}.$$

Setting up the corresponding notation involving m_j', z_j' and ξ_j', (18) becomes

$$\left(\frac{\phi(W)}{W}\log R\right)^{|J|}\sum_{j\in J}\sum_{m_j,m_j'\geq 1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{j\in J}\mu(m_j)\mu(m_j')m_j^{-z_j}m_j'^{-z_j'}\phi(\xi_j)\phi(\xi_j')d\xi_j d\xi_j'\right) \prod_{p\leq R^{\log R}} c_p((WP_j+b_1)_{j\in J_1,r_j(p)=1},(WP_j+b_2)_{j\in J_2,r_j(p)=1}), \tag{19}$$

which can be rewritten in the form

$$\left(\frac{\phi(W)}{W}\log R\right)^{|J|}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{p\leq R^{\log R}}E_p\left(\prod_{i\in J}\phi(\xi_j)\phi(\xi_j')d\xi_j d\xi_j'\right),$$

where

$$E_p=\sum_{j\in J}\sum_{m_j,m_j'\in\{1,p\}}\left(\prod_{j\in J}\mu(m_j)\mu(m_j')m_j^{-z_j}m_j'^{-z_j'}\right)c_p((WP_j+b_1)_{j\in J_1,r_j(p)=1},(WP_j+b_2)_{j\in J_2,r_j(p)=1}).$$

One now approximates the Euler factor E_p by

$$E_p'=\prod_{j\in J}\frac{(1-p^{-(1+z_j)})(1-p^{-(1+z_j')})}{1-p^{-(1+z_j+z_j')}}$$

using a series of claims for different types of primes p, whose proofs we shall postpone until the end of the section.

Claim 4 (Small primes)**.**

$$\prod_{p<w}\frac{E_p}{E'_p}=\left(\frac{W}{\phi(W)}\right)^{|J|}(1+o(1))$$

Claim 5 (Bad but not terrible primes)**.**

$$\prod_{\substack{w<p\leq R^{\log R}\\ p \text{ bad but not terrible}}}\frac{E_p}{E'_p}=1+O\left(\mathrm{Exp}\left(O\left(\sum_{p\in\mathcal{P}_b}p^{-1}\right)\right)\right)$$

Claim 6 (Good primes)**.**

$$\prod_{\substack{w<p\leq R^{\log R}\\ p \text{ good}}}\frac{E_p}{E'_p}=1+o(1)$$

Together Claims 4, 5 and 6 imply that if there are no terrible primes $>w$, then

$$\prod_{w<p\leq R^{\log R}}\frac{E_p}{E'_p}=\left(\frac{W}{\phi(W)}\right)^{|J|}\left(1+o(1)+O\left(\mathrm{Exp}\left(O\left(\sum_{p\in\mathcal{P}_b}p^{-1}\right)\right)\right)\right).$$

The proof of Proposition 3 is now completed, exactly as in [11, Proposition 10.1], using some elementary theory of the Riemann ζ function, as well as the rapid decay of ϕ. □

Finally, for completeness, we give the proofs of Claims 4, 5 and 6. They rely on the rather elementary [11, Lemma 9.5], which itself is proved with the help of a combinatorial Nullstellensatz [11, Appendix D].

Proof of Claim 4: If $p<w$, then for all $j\in J_i$ we have $WP_j+b_i\equiv b_i\not\equiv 0(p)$, so for such p the local factor $c_p((WP_j+b_1)_{j\in J_1,r_j(p)=1},(WP_j+b_2)_{j\in J_2,r_j(p)=1})$ is equal to 0 unless the set $\{j\in J: r_j(p)=1\}$ is the empty set, in which case it equals 1. The former case happens if and only if all m_j are equal to 1, so that $E_p=1$. A direct computation, using the fact that w goes to infinity much more slowly than R, shows that $E'_p=(1-p^{-1})^{|J|}+o(1)$. The estimate for the product over all primes $p<w$ of E_p/E'_p then follows from the fact that $\prod_{p<w}(1-p^{-1})^{-1}=W/\phi(W)$. □

Proof of Claim 5: If $p>w$ is bad but not terrible, then [11, Lemma 9.5 (b)] implies that the local factor c_p of any non-trivial family is $O(p^{-1})$. Hence the sum defining E_p has a contribution of 1 from $m_j=m'_j=1$ for all $j\in J$, and a contribution of $O(p^{-1})$ from all other terms, so that $E_p=1+O(p^{-1})$. Also, we find by Taylor expanding that

$$E'_p=1-\frac{1}{p}\sum_{j\in[J]}\left(\frac{1}{p^{z_j}}+\frac{1}{p^{z'_j}}\right)+\frac{1}{p}\sum_{j\in[J]}\frac{1}{p^{z_j+z'_j}}+O\left(\frac{1}{p^2}\right)=1+O\left(\frac{1}{p}\right),\tag{20}$$

where the last inequality follows from the fact that z_j, z_j' have real part $1/\log R > 0$. It follows that

$$\prod_{\substack{w<p\leq R^{\log R}\\ p \text{ bad but not terrible}}} \frac{E_p}{E_p'} = \prod_{\substack{w<p\leq R^{\log R}\\ p \text{ bad but not terrible}}} \left(1+O\left(\frac{1}{p}\right)\right) = \prod_{\substack{w<p\leq R^{\log R}\\ p \text{ bad but not terrible}}} \exp\left(O\left(\frac{1}{p}\right)\right),$$

which equals

$$\exp\left(O\left(\sum_{p\in\mathcal{P}_b} p^{-1}\right)\right) = 1+O\left(\mathrm{Exp}\left(O\left(\sum_{p\in\mathcal{P}_b} p^{-1}\right)\right)\right).$$

□

Proof of Claim 6: If $p > w$ is good, then by [11, Lemma 9.5 (c)] the local factor c_p of a non-trivial family consisting of precisely 1 polynomial is $p^{-1} + O(p^{-2})$. Such a non-trivial family arises in the case when $m_j = 1, m_j' = p$ or $m_j = p, m_j' = 1$ or $m_j = p, m_j' = p$ for exactly one $j \in [J]$. In all remaining cases, $c_p = O(p^{-2})$ by [11, Lemma 9.5 (d)]. This implies that

$$E_p = 1 - \left(\frac{1}{p}+O\left(\frac{1}{p^2}\right)\right)\sum_{j\in[J]}\left(\frac{1}{p^{z_j}}+\frac{1}{p^{z_j'}}\right) + \left(\frac{1}{p}+O\left(\frac{1}{p^2}\right)\right)\sum_{j\in[J]}\frac{1}{p^{z_j+z_j'}} + O\left(\frac{1}{p^2}\right),$$

so that by (20) for good primes p,

$$\frac{E_p}{E_p'} = 1+O\left(\frac{1}{p^2}\right).$$

The statement in Claim 6 now follows from the fact that the product $\prod_p(1+O(p^{-2}))$ is convergent and w tends to infinity. □

References

[1] V. Bergelson, A. Leibman. *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, Journal of AMS **9** (1996), no. 3, 725–753.

[2] Q. Chu, N. Frantzikinakis, B. Host. *Ergodic averages of commuting transformations with distinct degree polynomial iterates*, Proc. London Math. Soc. (3), 102, (2011), 801–842.

[3] N. Frantzikinakis, B. Host, B. Kra. *The polynomial multidimensional Szemerédi theorem along shifted primes*, to appear in Israel. J. Math.

[4] H. Furstenberg. *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions.* Journal d'Analyse Mathématique, 31 (1977), 204–256.

[5] B. Green, T. Tao. *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.

[6] B. Green, T. Tao. *Linear equations in primes*, Ann. Math. 2, 171(3), (2010) 1753–1850.

[7] B. Green, T. Tao. *The Möbius function is strongly orthogonal to nilsequences*, Ann. Math. no. 2, 175 (2012), 541–566.

[8] B. Green, T. Tao, T. Ziegler. *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, to appear, Ann. Math. (2010).

[9] H. Li, H. Pan. *Difference sets and Polynomials of prime variables*, Acta Arith, no.1, **138** (2009), 25-52.
[10] A. Rice. *Sárközy's theorem for P-intersective polynomials*, arXiv:1111.6559 (2011).
[11] T. Tao, T. Ziegler. *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), 213–305.
[12] T. Wooley, T. Ziegler. *Multiple recurrence and convergence along the primes*, to appear in American J. of Math.

T. H. Lê, Department of Mathematics, The University of Texas at Austin, 1 University Station, C1200, Austin, TX 78712, USA

E-mail address: leth@math.utexas.edu

J. Wolf, École Polytechnique, Centre de Mathématiques Laurent Schwartz, 91128 Palaiseau, France

E-mail address: julia.wolf@cantab.net