Proceedings of the Edinburgh Mathematical Society (2019) **62**, 443–457
doi:10.1017/S0013091518000561

SZEMERÉDI'S THEOREM IN THE PRIMES

LUKA RIMANIĆ AND JULIA WOLF

School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK
(luka.rimanic@cantab.net; julia.wolf@cantab.net)

(Received 10 October 2017; first published online 19 November 2018)

In memory of Kevin Henriot

Abstract Green and Tao famously proved in 2005 that any subset of the primes of fixed positive density contains arbitrarily long arithmetic progressions. Green had previously shown that, in fact, any subset of the primes of relative density tending to zero sufficiently slowly contains a three-term progression. This was followed by work of Helfgott and de Roton, and Naslund, who improved the bounds on the relative density in the case of three-term progressions. The aim of this note is to present an analogous result for longer progressions by combining a quantified version of the relative Szemerédi theorem given by Conlon, Fox and Zhao with Henriot's estimates of the enveloping sieve weights.

Keywords: Green–Tao theorem; Szemerédi's theorem; pseudorandomness; arithmetic combinatorics

2010 *Mathematics subject classification:* Primary 11B30

1. Introduction

Let $r_k(N)$ denote the maximal size of a subset of $[N] := \{1, 2, \dots, N\}$ not containing any non-trivial k-term arithmetic progressions, and let $r_k(\mathcal{P}_N)$ denote the maximal size of a subset of the set $\mathcal{P}_N := \mathcal{P} \cap [N]$ of primes less than N not containing any non-trivial k-term arithmetic progressions. Define the corresponding critical densities by

$$\alpha_k(N) := \frac{r_k(N)}{N} \quad \text{and} \quad \alpha_k(\mathcal{P}_N) := \frac{r_k(\mathcal{P}_N)}{|\mathcal{P}_N|},$$

respectively. The current state of the art in the integers is

$$\alpha_3(N) \ll (\log N)^{-1+o(1)} \ [1],$$
$$\alpha_4(N) \ll (\log N)^{-c} \quad \text{for some } c > 0 \ [9],$$
$$\alpha_k(N) \ll (\log\log N)^{-2^{-2^{k+9}}} \quad \text{for all } k \geq 5 \ [5].$$

Regarding relative density in the primes, the current record for progressions of length 3 is

$$\alpha_3(\mathcal{P}_N) \ll (\log\log N)^{-1+o(1)},$$

© 2018 The Edinburgh Mathematical Society

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

arrived at through a series of articles by Green [7], Helfgott and de Roton [10], and finally Naslund [12]. Henriot [11] extended Naslund's result to all linear systems of complexity one, and our result relies crucially on the optimized estimates of the enveloping sieve weights he gave in that paper. The aim of this note is to extend these results to longer arithmetic progressions as follows.

Theorem 1.1. *For every $k \geq 4$ there exists $c_k > 0$ such that the following holds. Let N_0 and b_0 be positive reals such that there exists a surjective, strictly decreasing function $\beta\colon [N_0, \infty) \to (0, b_0]$ with*

$$\alpha_k(N) \leq \beta(N) \quad \text{for all integers } N \geq N_0.$$

Then, for all sufficiently large $N \geq N_0$, we have

$$\alpha_k(N) \ll_k \beta((\log\log N)^{c_k}). \tag{1.1}$$

Taking β to be the best asymptotic upper bound for Szemerédi's theorem in the integers known to date, Theorem 1.1 immediately yields

$$\alpha_4(\mathcal{P}_N) \ll (\log\log\log N)^{-c} \quad \text{for } c > 0 \text{ determined by [9]},$$

$$\alpha_k(\mathcal{P}_N) \ll (\log\log\log\log N)^{-2^{-2^{k+9}}} \quad \text{for all } k \geq 5.$$

For comparison, in their celebrated work on long arithmetic progressions in the primes, Green and Tao [8] obtained bounds on $\alpha_k(\mathcal{P}_N)$ of the form $(\log_{(7)} N)^{-c_k}$ whenever $k \geq 4$, where $\log_{(s)} N$ denotes the s-fold iterated logarithm of N.

The proof of Theorem 1.1 naturally splits into two parts, each of which contributes one logarithm to the bound in (1.1). The first part consists of a relative Szemerédi theorem as proved by Conlon *et al.* [2], which we quantify for our purposes in § 2 (some details have been relegated to the Appendices). In the second part we make use of Henriot's [11] optimized estimates of the usual sieve weights associated with the primes.

2. A Quantitative relative Szemerédi theorem

At the heart of the aforementioned relative Szemerédi theorem lies the idea, already present in [8] and elaborated on in [6, 14, 15] and finally [3], that one can deal with an unbounded function in a pseudorandom setting by approximating it with a bounded function while preserving its density and the count of arithmetic progressions.

Let us first recall Szemerédi's theorem in the dense setting, which can, using Varnavides' averaging trick [16], be rephrased in the following weighted form.[*]

Theorem 2.1 (Szemerédi's theorem, dense setting). *Let $k \geq 4$ be an integer. Suppose that β is defined on $[N_0, \infty)$ as in Theorem 1.1. Let $N > N_0$ be a sufficiently*

[*] When f is the indicator function of a set, Theorem 2.1 follows from a naive application of the idea in [16]. In the general case of a 1-bounded function we use the known lower bound on $\alpha_k(N)$ [13] to absorb the additional loss in α.

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

large integer, and let $\alpha > 0$. *Then every* $f\colon \mathbb{Z}_N \to [0,1]$ *with* $\mathbb{E}f \geq \alpha$ *satisfies*

$$\mathbb{E}_{x,d\in\mathbb{Z}_N}[f(x)f(x+d)\dots f(x+(k-1)\,d)] \gg_k (\beta^{-1}(\alpha/8))^{-4}, \tag{2.1}$$

where β^{-1} *denotes the inverse of* β.

In order to extend this result to a sparse setting, we shall need to impose a pseudo-randomness condition on the function in question. We shall use the following arithmetic version of a definition given by Conlon *et al.* [**3**, Definition 2.2], which has the additional feature of measuring the speed of convergence.

Definition 2.2 ((k,δ)-linear forms condition, arithmetic setting). Let $k \geq 2$ and $\delta > 0$. We say that a function $\nu\colon \mathbb{Z}_N \to [0,\infty)$ satisfies the (k,δ)-*linear forms condition* (or (k,δ)-*LFC* for short) if

$$\left|\mathbb{E}_{x_1^{(0)},x_1^{(1)},\dots,x_k^{(0)},x_k^{(1)}\in\mathbb{Z}_N}\left[\prod_{j=1}^{k}\prod_{\omega\in\{0,1\}^{[k]\setminus\{j\}}}\nu\Big(\sum_{i=1}^{k}(j-i)x_i^{(\omega_i)}\Big)^{n_{j,\omega}}\right]-1\right|\leq\delta,$$

for any choice of exponents $n_{j,\omega}\in\{0,1\}$.

We are now ready to state the main result of this section.

Theorem 2.3 (quantitative relative Szemerédi theorem). *For every integer* $k \geq 3$ *there exists* $c_k' > 0$ *such that the following holds.*

Suppose that β *is defined on* $[N_0,\infty)$ *as in Theorem 1.1. Let* $N > N_0$ *be a sufficiently large integer and let* $\delta > 0$. *Suppose that* $\nu\colon \mathbb{Z}_N \to [0,\infty)$ *satisfies the* (k,δ)-*linear forms condition. Then, for all* $\alpha > 0$ *and every function* $f\colon \mathbb{Z}_N \to [0,\infty)$ *such that* $0 \leq f \leq \nu$ *and* $\mathbb{E}f \geq \alpha$, *one has*

$$\mathbb{E}_{x,d\in\mathbb{Z}_N}[f(x)f(x+d)\dots f(x+(k-1)d)] \gg_k \frac{1}{(\beta^{-1}(\alpha/8))^4}-\frac{1}{\log^{c_k'}(1/\delta)}. \tag{2.2}$$

A qualitative version of Theorem 2.3 was given by Conlon *et al.* [**3**, Theorem 2.4], so the only novelty here is the explicit error term with respect to the speed of convergence of the linear forms condition.

In the remainder of this section we summarize the main steps in the proof of Theorem 2.3. To begin with we need one further definition.

Definition 2.4. For any positive integer r and any function $f\colon \mathbb{Z}_N \to \mathbb{R}$, we define the *cut norm* of f by

$$\|f\|_{\square,r} := \sup\left|\mathbb{E}_{x_1,\dots,x_r\in\mathbb{Z}_N} f(x_1+\dots+x_r)\prod_{j\in[r]}1_{A_j}(x_{-j})\right|, \tag{2.3}$$

where the supremum is taken over all $A_1,\dots,A_r \subseteq \mathbb{Z}_N^{r-1}$, and x_{-j} stands for the $(r-1)$-tuple $(x_1,\dots,x_{j-1},x_{j+1},\dots,x_r)\in\mathbb{Z}_N^{r-1}$.

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

In Appendix B we show that the (k,δ)-LFC implies control in the cut norm, a qualitative version of which is implicit in [2, 3].

Corollary 2.5. *Suppose that ν satisfies the (k,δ)-linear forms condition. Then*

$$\|\nu - 1\|_{\square,k-1} \leq 2\delta^{1/2^{k-1}}.$$

This allows us to use the following *dense model theorem* taken almost verbatim from [2, Theorem 5.1], which states that an unbounded function with a pseudorandom majorant can be well approximated by a 1-bounded function of the same density, such that their difference behaves well with respect to inner products with certain test functions.

Theorem 2.6 (dense model theorem). *There exists an absolute constant $C > 0$ such that the following holds.*

Let $\varepsilon > 0$, and suppose that $\nu\colon \mathbb{Z}_N \to [0,\infty)$ satisfies $\|\nu - 1\|_{\square,k-1} \leq \varepsilon$. Then, for every $f\colon \mathbb{Z}_N \to [0,\infty)$ such that $0 \leq f \leq \nu$, there exists a function $\widetilde{f}\colon \mathbb{Z}_N \to [0,1]$ such that

$$\|f - \widetilde{f}\|_{\square,k-1} \leq \log^{-1/C}(1/\varepsilon).$$

The second ingredient in the proof of Theorem 2.3 is the so-called *counting lemma*, which states that the count of arithmetic progressions is preserved under the conditions implied by the dense model theorem.

Proposition 2.7 ((k,δ)-relative counting lemma, arithmetic setting). *Suppose that ν satisfies the (k,δ)-linear forms condition and let $f, \widetilde{f}$ be functions on $\mathbb{Z}_N$ such that $0 \leq f \leq \nu$ and $0 \leq \widetilde{f} \leq 1$. If $\|f - \widetilde{f}\|_{\square,k-1} \leq \varepsilon$, then*

$$\begin{aligned}&|\mathbb{E}_{x,d} f(x)f(x+d)\dots f(x+(k-1)d) - \mathbb{E}_{x,d}\widetilde{f}(x)\widetilde{f}(x+d)\dots\widetilde{f}(x+(k-1)d)|\\ &\quad\ll_k \delta^{1/2^{2^k+k-2}} + \varepsilon^{1/2^{2^k-1}}.\end{aligned}$$

A qualitative version of Proposition 2.7 was given in [3, Theorem 2.17] (see also [17, Lemma 4.1]), so again the novelty here is the explicit error term with respect to the speed of convergence of the (k,δ)-LFC and the quality of the approximation of f by $\widetilde{f}$. We provide proofs of Corollary 2.5 and Proposition 2.7 in Appendices B and D, respectively, as they are relatively minor modifications of the proofs in [2, 3].

Proof of Theorem 2.3 (assuming Corollary 2.5 and Proposition 2.7). Using Corollary 2.5 in conjunction with Theorem 2.6 we find $\widetilde{f}\colon \mathbb{Z}_N \to [0,1]$ such that

$$\|f - \widetilde{f}\|_{\square,k-1} \ll_k \frac{1}{\log^{1/C}(1/\delta)},$$

where C is the constant in Theorem 2.6. Now Proposition 2.7 yields

$$\begin{aligned}&|\mathbb{E}_{x,d} f(x)f(x+d)\dots f(x+(k-1)d) - \mathbb{E}_{x,d}\widetilde{f}(x)\widetilde{f}(x+d)\dots\widetilde{f}(x+(k-1)d)|\\ &\quad\ll_k \delta^{1/2^{2^k+k-2}} + \log^{-1/2^{2^k-1}C}(1/\delta) \ll_k \log^{-c'_k}(1/\delta),\end{aligned}$$

and the claim made in Theorem 2.3 now follows from Theorem 2.1. □

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

3. Proof of Theorem 1.1

With Theorem 2.3 in hand, the remainder of the proof of Theorem 1.1 is fairly standard. It is well known that the primes show random-like behaviour once one removes their bias with respect to residue classes modulo small primes. This is achieved using the so-called W-trick, which reduces Theorem 1.1 to the following.

Theorem 3.1 (main theorem in W-tricked primes). *There exists $c_k > 0$ such that the following holds.*

Suppose that β is defined on $[N_0, \infty)$ as in Theorem 1.1. Let $N' > N_0$ be a sufficiently large integer, and let $\omega := c_0 \log N'$ for some $c_0 \in [1/4, 1/2]$. Let $W := \prod_{p \leq \omega} p$, and let b be a positive integer coprime to W. Suppose that $B \subset [N']$ is a set free of k-term arithmetic progressions such that $b + W \cdot B \subset \mathcal{P}$ and

$$|B| \geq \alpha \frac{W}{\phi(W)} \frac{N'}{\log N'}.$$

Then

$$\alpha \ll_k \beta((\log\log N')^{c_k}).$$

Proof of Theorem 1.1 (assuming Theorem 3.1). Let N be a sufficiently large integer, and let $A \subset \mathcal{P} \cap [N]$ be a subset of relative density α free of k-term progressions. We may assume that $\alpha \gg N^{-1/4}$, as otherwise the claim is trivial. Let $\omega := \frac{1}{4} \log N$. It is not difficult to see that for such a choice of ω one has $W = N^{1/4+o(1)}$. Define $N' := \lfloor N/W \rfloor = N^{3/4+o(1)}$. By an averaging argument (see, for example, Lemma 2.1 in [10]) there exists a positive integer b coprime to W such that the set $B := \{n \in [N']\colon b + Wn \in A\}$ satisfies

$$|B| \gg \alpha \frac{W}{\phi(W)} \frac{N'}{\log N'}.$$

Since $\omega \sim \frac{1}{3} \log N'$ as $N \to \infty$ and B inherits the property of not containing any non-trivial k-term arithmetic progressions (by translation-invariance and homogeneity), we can use Theorem 3.1 to conclude that

$$\alpha \ll_k \beta((\log\log N')^{c_k}) \ll_k \beta((\log\log N)^{c_k}),$$

which completes the proof of Theorem 1.1. □

It thus suffices to prove the main theorem in the W-tricked primes. In order to use Theorem 2.3 one needs to construct a majorant for the primes that satisfies the (k, δ)-LFC. The existence of such a pseudorandom majorant was already established in [8] using ingredients from [4]. It is here that we make crucial use of the simplified and optimized estimates from [11], allowing us to choose w as large as $c_0 \log N'$.

Proof of Theorem 3.1. By Bertrand's postulate one can find a prime $M \in [2N', 4N']$. This choice of M helps avoid possible wrap-around issues with the function

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

$\lambda_{b,W} : \mathbb{Z}_M \to \mathbb{R}^+$, defined by

$$\lambda_{b,W}(n) := \frac{\phi(W)}{W}(\log N')1_{[N']}(n)1_{\mathcal{P}}(b + Wn). \tag{3.1}$$

Bringing the indicator function of $B \subseteq [N']$ in line with these weights, we define $f_B : \mathbb{Z}_M \to \mathbb{R}^+$ by

$$f_B(n) := \frac{M}{N'}\frac{\phi(W)}{W}(\log N')1_B(n).$$

Note that $\mathbb{E}_{[M]}f_B \geq \alpha$, and $0 \leq f_B \ll \lambda_{b,W}$ since $b + W \cdot B \subset \mathcal{P}$.

By Proposition 6.2 in [11], since M is large enough and equal to N' up to a constant factor, there exists a pseudorandom majorant $\nu\colon \mathbb{Z}_M \to \mathbb{R}^+$ satisfying the (k, δ)-LFC with $\delta \ll_k (\log N')^{-1+o(1)}$, as well as the inequality

$$0 \leq f_B \ll \lambda_{b,W} \ll_k \nu.$$

Since B does not contain any non-trivial k-term arithmetic progressions, the left-hand side of (2.2) with $N = M$ and $f = f_B$ is $O(1/N')$, which is easily seen to be negligible. It therefore follows from Theorem 2.3 that

$$\frac{1}{\beta^{-1}(\alpha/8)} \ll_k \frac{1}{\log^{c'_k/4}(1/\delta)} \ll_k \frac{1}{(\log\log N')^{c_k}}$$

for some constant c_k. This implies Theorem 3.1 as β is a strictly decreasing function. □

4. Remarks

We do not see any fundamental obstruction to extending the result in this paper to more general systems of translation-invariant linear equations for which a result of Varnadives strength is known, but shall not attempt to do so here. It would arguably be of greater interest to find a more direct approach for longer progressions along the lines of [10] and [12], where the bounded function $\widetilde{f}$ is constructed explicitly.

Appendix A. Hypergraph notation

In the subsequent appendices we prove the results claimed in § 2 by quantifying the arguments in [2, 3]. These are given in the language of hypergraphs rather than linear systems of equations, and since the notation in the former setting is more compact and closer to the original source of the arguments we shall use it here. The purpose of the present section is therefore to set up the hypergraph notation used in [2], and to indicate how it relates to that used throughout the main body of the note.

In what follows, unless otherwise stated, we always suppose that k is a fixed positive integer and that the hypergraphs in question are weighted $(k-1)$-uniform k-partite hypergraphs on $X := X_1 \cup X_2 \cup \ldots \cup X_k$. When applied to the arithmetic setting, each

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

set X_i corresponds to a copy of $\mathbb{Z}_N$. To keep the notation as light as possible, write

$$X_{-i} := X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_k$$

and

$$x_{-i} := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \in X_{-i}.$$

Our weighted $(k-1)$-uniform k-partite hypergraph then consists of a k-tuple of functions $g = (g_{-i})_{i=1,\ldots,k}$, with each $g_{-i} : X_{-i} \to \mathbb{R}$. For two such hypergraphs g and ν we say that $g \leq \nu$ whenever $g_{-i}(x_{-i}) \leq \nu_{-i}(x_{-i})$ for all $1 \leq i \leq k$ and all $x_{-i} \in X_{-i}$. We shall often omit the index of the weight when it is clear from the argument it takes. For example, by $\prod_{i=1}^k g(x_{-i})$ we mean $\prod_{i=1}^k g_{-i}(x_{-i})$.

Given a weighted r-uniform hypergraph h on $X_1 \times \cdots \times X_r$, that is, a function $h : X_1 \times \cdots \times X_r \to \mathbb{R}$, we define the *cut norm* of h by

$$\|h\|_{\square,r} := \sup \left| \mathbb{E}_{x_1 \in X_1, \ldots, x_r \in X_r} h(x_1, \ldots, x_r) \prod_{j \in [r]} 1_{A_j}(x_{-j}) \right|, \tag{A.1}$$

where the supremum is taken over all $A_j \subseteq X_{-j}$. Finally, given a weighted $(k-1)$-uniform k-partite hypergraph g on the vertex set X as above, we define the *cut norm* of g by

$$\|g\|_{\square} := \max\{\|g_{-1}\|_{\square,k-1}, \ldots, \|g_{-k}\|_{\square,k-1}\}. \tag{A.2}$$

It is not difficult to see that both $\|.\|_{\square,r}$ and $\|.\|_{\square}$ are indeed norms. They are related to their arithmetic counterparts in § 2 by setting $h(x_1, \ldots, x_r) = f(x_1 + \cdots + x_r)$.

With this notation, we have the following analogue of the (k,δ)-LFC introduced in Definition 2.2.

Definition A.1 ((k,δ)-linear forms condition, hypergraph setting). Let $\delta > 0$. We say that a weighted $(k-1)$-uniform k-partite hypergraph ν satisfies the (k,δ)-*linear forms condition* ((k,δ)-*LFC*) if

$$\left| \mathbb{E}_{x_1^{(0)}, x_1^{(1)} \in X_1, \ldots, x_k^{(0)}, x_k^{(1)} \in X_k} \left[\prod_{j=1}^k \prod_{\omega \in \{0,1\}^{[k] \setminus \{j\}}} \nu((x^{(\omega)})_{-j})^{n_{j,\omega}} \right] - 1 \right| \leq \delta, \tag{A.3}$$

for any choice of exponents $n_{j,\omega} \in \{0,1\}$, with $x^{(\omega)} = (x_1^{(\omega_1)}, \ldots, x_k^{(\omega_k)})$.

In other words, satisfying the (k,δ)-LFC in a hypergraph amounts to containing the expected count of every subgraph of the 2-blow-up of $K_k^{(k-1)}$. Observe also that the (k,δ)-LFC in Definition 2.2 is easily recovered from the above by making the substitution

$$(x^{(\omega)})_{-j} \longmapsto \sum_{i=1}^k (j-i) x_i^{(\omega_i)}. \tag{A.4}$$

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

Appendix B. Controlling the cut norm

We begin by proving the following auxiliary result, which will be useful when attempting to bound the cut norm of $\nu - 1$, as well as when establishing the so-called *strong linear forms condition* in Appendix C.

Lemma B.1. *Suppose that ν satisfies the (k,δ)-linear forms condition. For each $j \in [k]$ and each $\omega \in \{0,1\}^{[k]\setminus\{j\}}$, let $h_{j,\omega} \in \{1, \nu, \nu-1\}$. Let K denote the number of pairs (j,ω) for which $h_{j,\omega} = \nu - 1$ and suppose that $K \geq 1$. Then*

$$S((h_{j,\omega})_{j,\omega}) := \mathbb{E}_{x_1^{(0)},x_1^{(1)}\in X_1,\dots x_k^{(0)},x_k^{(1)}\in X_k} \prod_{j=1}^{k} \prod_{\omega\in\{0,1\}^{[k]\setminus\{j\}}} h_{j,\omega}((x^{(\omega)})_{-j})$$

satisfies the inequality

$$|S((h_{j,\omega})_{j,\omega})| \leq 2^K \delta. \tag{B.1}$$

Proof. If all $h_{j,\omega}$ are equal to 1 or ν, then, since ν satisfies the (k,δ)-LFC, we have $S((h_{j,\omega})_{j,\omega}) \in [1-\delta, 1+\delta]$. In the general case, when expanding $S((h_{j,\omega})_{j,\omega})$ we get to decide whether to choose ν or -1 for each pair (j,ω) such that $h_{j,\omega} = \nu - 1$. The terms that make a positive contribution are exactly those in which the number of -1s is even; the terms in which the number of -1s is odd make a negative contribution instead. Thus

$$|S((h_{j,\omega})_{j,\omega})| \leq \left[\binom{K}{0} + \binom{K}{2} + \cdots\right](1+\delta) - \left[\binom{K}{1} + \binom{K}{3} + \cdots\right](1-\delta) = 2^K\delta. \quad \square$$

Instead of Corollary 2.5 we shall actually prove the slightly stronger statement that the (k,δ)-LFC yields control in an appropriate Gowers uniformity norm, which dominates the cut norm and is defined as follows. Given an r-uniform hypergraph h on $X_1 \times \cdots \times X_r$, define the *Gowers U^r-norm* by

$$\|h\|_{U^r}^{2^r} := \mathbb{E}_{x_1^{(0)},x_1^{(1)}\in X_1,\dots,x_r^{(0)},x_r^{(1)}\in X_r} \prod_{j=1}^{r} \prod_{\omega\in\{0,1\}^{[r]}} h(x^{(\omega)}).$$

For $k = 3$, the following statement appears in qualitative form as Lemma 6.3 in [**2**].

Corollary B.2 ((k,δ)-LFC implies uniformity). *Suppose that ν satisfies the (k,δ)-linear forms condition. Then*

$$\|\nu - 1\|_\square \leq \|\nu - 1\|_{U^{k-1}} \leq 2\delta^{1/2^{k-1}}.$$

Note that we use Corollary 2.5 in the proof of Theorem 3.1 with N being a prime. In this case the substitution (A.4) facilitates a straightforward transition between the hypergraph setting in Corollary B.2 and the arithmetic setting in Corollary 2.5.

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

Proof. Upon recalling that $\|.\|_{\square}$ is the maximum over all $\|\cdot\|_{\square,k-1}$, the first inequality is a straightforward application of the Gowers–Cauchy–Schwarz inequality [5] formulated for hypergraphs (see, for example, [3, Lemma 6.2]). Note that the U^{k-1}-norm consists of 2^{k-1} factors, each equal to $\nu-1$, so that the result follows from Lemma B.1 with $K=2^{k-1}$. □

Appendix C. Strong linear forms condition

The proof of the counting lemma in [3], which we shall follow in Appendix D, proceeds by induction on the number of majorants that are not identically 1. This requires us to be able to replace ν by 1 under certain assumptions, which we shall be able to do as a result of the following lemma, known as the *strong linear forms condition*.

Lemma C.1 ((k,δ)-strong linear forms). *Let $\delta>0$. Suppose that ν satisfies the (k,δ)-linear forms condition and let $0\le g\le\nu$, $0\le\widetilde{g}\le 1$. Then*

$$\Bigg|\mathbb{E}_{\substack{x_1\in X_1,\dots,x_{k-1}\in X_{k-1}\\ x_k^{(0)},x_k^{(1)}\in X_k}}(\nu(x_{-k})-1)\prod_{j=1}^{k-1}\prod_{\omega\in\{0,1\}}h_{j,\omega}((x^{(\omega)})_{-j})\Bigg| \le 2(1+\delta)^{1-1/2^{k-1}}\delta^{1/2^{k-1}}, \tag{C.1}$$

where $x^{(\omega)}:=(x_1,\dots,x_{k-1},x_k^{(\omega)})$ and each $h_{j,\omega}\in\{g_{-j},\widetilde{g}_{-j}\}$.

It is the explicit form of the error term, given here as a function of the speed of convergence in the (k,δ)-LFC, which distinguishes this statement from [3, Lemma 6.3].

Proof. We may without loss of generality suppose that each $h_{j,\omega}=g_{-j}$, as having $\widetilde{g}$ instead of g is in fact advantageous, allowing us to replace any upper bound of ν by 1. Denote the expectation in (C.1) by $S_k(g)$. Applying the Cauchy–Schwarz inequality for the first time, isolating the x_{k-1} variable, we obtain

$$\begin{aligned}|S_k(g)|^2 \le{}& \Bigg|\mathbb{E}_{\substack{x_1\in X_1,\dots,x_{k-2}\in X_{k-2}\\ x_k^{(0)},x_k^{(1)}\in X_k}}\prod_{\omega\in\{0,1\}}g(x^{(\omega)}_{-(k-1)})\Bigg|\\ &\times\Bigg|\mathbb{E}_{\substack{x_1\in X_1,\dots,x_{k-2}\in X_{k-2}\\ x_k^{(0)},x_k^{(1)}\in X_k}}\prod_{\omega\in\{0,1\}}g(x^{(\omega)}_{-(k-1)})(\mathbb{E}_{x_{k-1}\in X_{k-1}}(\nu(x_{-k})-1)\\ &\times\prod_{j=1}^{k-2}\prod_{\omega\in\{0,1\}}g(x^{(\omega)}_{-j}))^2\Bigg|.\end{aligned}$$

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

Bounding instances of $g_{-(k-1)}$ above by ν, the above expression is at most

$$\left|\mathbb{E}_{\substack{x_1\in X_1,\dots,x_{k-2}\in X_{k-2}\\ x_k^{(0)},x_k^{(1)}\in X_k}}\prod_{\omega\in\{0,1\}}\nu(x_{-(k-1)}^{(\omega)})\right|$$
$$\times\left|\mathbb{E}_{\substack{x_1\in X_1,\dots,x_{k-2}\in X_{k-2}\\ x_k^{(0)},x_k^{(1)}\in X_k}}\prod_{\omega\in\{0,1\}}\nu(x_{-(k-1)}^{(\omega)})(\mathbb{E}_{x_{k-1}\in X_{k-1}}(\nu(x_{-k})-1)\right.$$
$$\left.\times\prod_{j=1}^{k-2}\prod_{\omega\in\{0,1\}}g(x_{-j}^{(\omega)}))^2\right|.$$

Denoting the second expectation by $S_{k-1}(g)$, the (k,δ)-LFC now implies that

$$|S_k(g)|^2\le(1+\delta)|S_{k-1}(g)|.$$

In order to bound $S_{k-1}(g)$, expand the square and use the Cauchy–Schwarz inequality on x_{k-2}. Continuing inductively we obtain

$$|S_k(g)|^{2^{k-1}}\le(1+\delta)^{2^{k-2}}|S_{k-1}(g)|^{2^{k-2}}\le\ldots\le(1+\delta)^{2^{k-1}-1}|S_1(g)|,$$

where

$$S_1(g):=\mathbb{E}_{x_1^{(0)},x_1^{(1)}\in X_1,\dots,x_k^{(0)},x_k^{(1)}\in X_k}\prod_{\omega\in\{0,1\}^{[k-1]}}(\nu(x_{-k}^{(\omega)})-1)\prod_{j=1}^{k-1}\prod_{\omega\in\{0,1\}^{[k]\setminus\{j\}}}\nu(x_{-j}^{(\omega)})$$

and the exponent of $(1+\delta)$ arises as the sum of 2^{k-j} from $j=2$ to k. There are 2^{k-1} terms of the form $\nu-1$, so by Lemma B.1 we have that

$$|S_k(g)|\le 2^{2^{k-1}}\delta,$$

from which the claim easily follows. □

Appendix D. The counting lemma

In this section we shall prove the relative counting lemma, Proposition 2.7, whose hypergraph version is of the following form.

Proposition D.1 ((k,δ)-relative counting lemma, hypergraph setting). *Let $\varepsilon>0$, and let ν be a weighted hypergraph satisfying the (k,δ)-linear forms condition. Suppose that $0\le g\le\nu$, $0\le\widetilde{g}\le 1$. If $\|g-\widetilde{g}\|_\square\le\varepsilon$, then*

$$\left|\mathbb{E}_{x_1\in X_1,\dots,x_k\in X_k}\left[\prod_{j=1}^{k}g(x_{-j})-\prod_{j=1}^{k}\widetilde{g}(x_{-j})\right]\right|\ll_k\delta^{1/2^{2^k+k-2}}+\varepsilon^{1/2^{2^k-1}}.\tag{D.1}$$

Proposition 2.7 follows from Proposition D.1 by applying the substitution (A.4) to $(x_{-j})_{j=1}^k$, which produces k-term progressions with multiplicity N^{k-2}. The purpose of what follows is to keep track of the dependence on δ and ϵ in the argument as given in [**3**].

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

In fact, it is not difficult to see that this dependence is polynomial, but we shall be a little more precise here (and attempt to make this paper at least somewhat self-contained).

The proof proceeds by induction on the number of $\nu_{-1}, \dots, \nu_{-k}$ that are not identically 1. Denote this number by m, and for every m denote the least upper bound on the left-hand side of (D.1) by $C_m(\varepsilon, \delta)$. The case $m = 0$ is addressed by the following statement.

Proposition D.2 (dense counting lemma, hypergraph setting). *Let $0 \le g, \widetilde{g} \le 1$. If $\|g - \widetilde{g}\|_\square \le \varepsilon$, then*

$$\left| \mathbb{E}_{x_1 \in X_1, \dots, x_k \in X_k} \left[\prod_{j=1}^{k} g(x_{-j}) - \prod_{j=1}^{k} \widetilde{g}(x_{-j}) \right] \right| \le k\varepsilon. \tag{D.2}$$

For the sake of completeness, we include a proof of Proposition D.2 for general k, given in [**2**, Proposition 6.1] for $k = 3$.

Proof. For $i \ne j$, let

$$X_{-i,j} := X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_{j-1} \times X_{j+1} \times \cdots \times X_k,$$

and define $x_{-i,j}$ in a similar vein. Recall that $\|g\|_\square$ is the maximum of $\|g_{-1}\|_{\square,k-1}, \dots, \|g_{-k}\|_{\square,k-1}$, so that $\|g - \widetilde{g}\|_\square \le \varepsilon$ implies that for all functions $a_j \colon X_{-j,k} \to [0,1]$ we have

$$|\mathbb{E}_{x_1 \in X_1, \dots, x_{k-1} \in X_{k-1}} (g(x_{-k}) - \widetilde{g}(x_{-k})) a_1(x_{-1,k}) \dots a_{k-1}(x_{-(k-1),k})| \le \varepsilon.$$

Indeed, it is not difficult to see that this condition is equivalent to that for $\{0,1\}$-valued functions, which are given by the definition of the cut norm. Now, for fixed $x_k \in X_k$, we can set $a_i(x_{-i,k}) := g(x_{-i})$ in the above expectation, and thus

$$\left| \mathbb{E}_{x_1 \in X_1, \dots, x_k \in X_k} (g(x_{-k}) - \widetilde{g}(x_{-k})) \prod_{j=1}^{k-1} g(x_{-j}) \right| \le \varepsilon.$$

Similarly, we get

$$\left| \mathbb{E}_{x_1 \in X_1, \dots, x_k \in X_k} \widetilde{g}(x_{-k}) (g(x_{-(k-1)}) - \widetilde{g}(x_{-(k-1)})) \prod_{j=1}^{k-2} g(x_{-j}) \right| \le \varepsilon$$

and so on. Continuing to telescope in this way, and using the triangle inequality, the claimed statement follows. □

We are now ready to prove the relative counting lemma.

Proof of Proposition D.1. Proposition D.2 implies that $C_0(\varepsilon, \delta) \le k\varepsilon$. Suppose that we have calculated upper bounds for all $C_m(\varepsilon, \delta)$ for $m = 0, 1, \dots, M$, where m denotes the number of $\nu_{-1}, \dots, \nu_{-k}$ that are not identically 1, as above. Suppose now that $m = M + 1$,

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

and without loss of generality that ν_{-1} is not identically 1. We define auxiliary weighted $(k-1)$-uniform hypergraphs $\nu', g', \widetilde{g}' \colon X_{-1} \to [0,\infty)$ by

$$\begin{aligned}\nu'(x_{-1}) &:= \mathbb{E}_{x_1 \in X_1}[\nu(x_{-2}) \dots \nu(x_{-k})],\\ g'(x_{-1}) &:= \mathbb{E}_{x_1 \in X_1}[g(x_{-2}) \dots g(x_{-k})],\\ \widetilde{g}'(x_{-1}) &:= \mathbb{E}_{x_1 \in X_1}[\widetilde{g}(x_{-2}) \dots \widetilde{g}(x_{-k})].\end{aligned}$$

Unlike $\widetilde{g}'$, the new functions g' and ν' may not be bounded by 1. We therefore define $g'_{\wedge 1} := \max\{g', 1\}$ and $\nu'_{\wedge 1} := \max\{\nu', 1\}$. As noted in [3], the main idea is that g' and ν' behave like dense graphs, so capping them by 1 produces only a small error. In what follows, all expectations will be taken over $X_{-1} = X_2 \times \dots \times X_k$ unless otherwise stated. First, note that the (k,δ)-LFC implies that

$$1 - \delta \le \mathbb{E}[\nu'], \mathbb{E}[\nu'^2] \le 1 + \delta,$$

from which it follows, by Cauchy–Schwarz, that

$$(\mathbb{E}[|\nu' - 1|])^2 \le \mathbb{E}[(\nu' - 1)^2] \le 3\delta. \tag{D.3}$$

Claim. *We have*

$$\|g'_{\wedge 1} - \widetilde{g}'\|_{\square, k-1} \le (3\delta)^{1/2} + C_M(\varepsilon, \delta). \tag{D.4}$$

Proof of claim. Since $0 \le g' \le \nu'$, we have

$$0 \le g' - g'_{\wedge,1} = \max\{g' - 1, 0\} \le \max\{\nu' - 1, 0\} \le |\nu' - 1|. \tag{D.5}$$

Note that for any $A_2 \subseteq X_{-1,2}, \dots, A_k \subseteq X_{-1,k}$ we can write

$$\begin{aligned}&\mathbb{E}[(g'_{\wedge 1} - \widetilde{g}')(x_{-1}) 1_{A_2}(x_3, \dots, x_k) \dots 1_{A_k}(x_2, \dots, x_{k-1})]\\ &\quad = \mathbb{E}[(g'_{\wedge 1} - g')(x_{-1}) 1_{A_2}(x_{-1,2}) \dots 1_{A_k}(x_{-1,k})]\\ &\qquad + \mathbb{E}[(g' - \widetilde{g}')(x_{-1}) 1_{A_2}(x_{-1,2}) \dots 1_{A_k}(x_{-1,k})].\end{aligned}$$

It follows from (D.5) and (D.3) that the first expectation is at most $(3\delta)^{1/2}$ in magnitude. In order to estimate the second term, rewrite it as

$$\mathbb{E}_{x_1,\dots,x_k}\left[1_{A_2,\dots,A_k}(x_{-1}) \prod_{j=2}^{k} g(x_{-j}) - 1_{A_2,\dots,A_k}(x_{-1}) \prod_{j=2}^{k} \widetilde{g}(x_{-j})\right], \tag{D.6}$$

where

$$1_{A_2,\dots,A_k}(x_{-1}) := \prod_{j=2}^{k} 1_{A_j}(x_{-1,j}).$$

In this form, it is easy to see that the number of factors whose majorant is not identically 1 is at most M, so by the inductive hypothesis (D.6) is bounded by $C_M(\varepsilon, \delta)$. This completes the proof of the claim. □

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

Returning to the expression we set out to bound, we see that

$$\mathbb{E}_{x_1,\dots,x_k}\Big[\prod_{j=1}^{k} g(x_{-j}) - \prod_{j=1}^{k} \widetilde{g}(x_{-j})\Big] = \mathbb{E}_{x_2,\dots,x_k}[g(x_{-1})g'(x_{-1}) - \widetilde{g}(x_{-1})\widetilde{g}'(x_{-1})],$$

which in turn we can rewrite as

$$\mathbb{E}[g(g' - \widetilde{g}')] + \mathbb{E}[(g - \widetilde{g})\widetilde{g}'].$$

Recall that $0 \leq \widetilde{g} \leq 1$, so the second term is at most

$$\|g - \widetilde{g}\|_{\square} \leq \varepsilon, \tag{D.7}$$

as in the proof of Proposition D.2. Concerning the first term, the Cauchy–Schwarz inequality yields

$$(\mathbb{E}[g(g' - \widetilde{g}')])^2 \leq \mathbb{E}[g(g' - \widetilde{g}')^2]\mathbb{E}[g] \leq \mathbb{E}[\nu(g' - \widetilde{g}')^2]\mathbb{E}[\nu],$$

which, upon expanding the square, using the (k, δ)-LFC and the strong linear forms condition (Lemma C.1), is bounded above by

$$(1+\delta)(8(1+\delta)^{1-1/2^{k-1}}\delta^{1/2^{k-1}} + \mathbb{E}[(g' - \widetilde{g}')^2]). \tag{D.8}$$

Expanding the final term further as

$$\mathbb{E}[(g' - \widetilde{g}')^2] = \mathbb{E}[(g' - \widetilde{g}')(g' - g'_{\wedge 1})] + \mathbb{E}[(g' - \widetilde{g}')(g'_{\wedge 1} - \widetilde{g}')],$$

we observe that since $0 \leq g' \leq \nu'$ and $0 \leq \widetilde{g}' \leq 1$ the first term is bounded by

$$\mathbb{E}[\nu'|\nu' - 1|] = \mathbb{E}[(\nu' - 1)|\nu' - 1|] + \mathbb{E}[|\nu' - 1|] \leq 3\delta + (3\delta)^{1/2},$$

by (D.3). Finally, we rewrite the second term as

$$\mathbb{E}[(g' - \widetilde{g}')(g'_{\wedge 1} - \widetilde{g}')] = \mathbb{E}[g'g'_{\wedge 1}] - \mathbb{E}[g'\widetilde{g}'] - \mathbb{E}[\widetilde{g}'g'_{\wedge 1}] + \mathbb{E}[(\widetilde{g}')^2]. \tag{D.9}$$

We claim that each of the four summands is close to $\mathbb{E}[(\widetilde{g}')^2]$. Indeed, we can write

$$\mathbb{E}[g'g'_{\wedge 1}] - \mathbb{E}[(\widetilde{g}')^2] = \mathbb{E}_{x_1,\dots,x_k}\Big[g'_{\wedge 1}(x_{-1})\prod_{j=2}^{k} g(x_{-j}) - \widetilde{g}'(x_{-1})\prod_{j=2}^{k} \widetilde{g}(x_{-j})\Big], \tag{D.10}$$

and observe that the tuples $(g'_{\wedge 1}, g_2, \dots, g_k)$ and $(\widetilde{g}', \widetilde{g}_{-2}, \dots, \widetilde{g}_{-k})$ satisfy the box-norm condition in Proposition D.1 with ε replaced by the upper bound in (D.4), while the number of majorants that are not identically 1 is at most M, so that by the inductive

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

hypothesis (D.10) is bounded above by

$$C_M((3\delta)^{1/2} + C_M(\varepsilon, \delta), \delta). \tag{D.11}$$

Bounding the difference between $\mathbb{E}[(\widetilde{g}')^2]$ and each of the remaining summands in (D.9) in a similar fashion, we obtain

$$\mathbb{E}[(g' - \widetilde{g}')^2] \leq 3\delta + (3\delta)^{1/2} + 3C_M((3\delta)^{1/2} + C_M(\varepsilon, \delta), \delta). \tag{D.12}$$

Collecting all the error terms, namely (D.7), (D.8) and (D.12), we see that

$$C_{M+1}(\varepsilon, \delta) \ll_k \varepsilon + (\delta^{1/2^{k-1}} + C_M(3\delta^{1/2} + C_M(\varepsilon, \delta), \delta))^{1/2}. \tag{D.13}$$

For $M \geq 1$ this is satisfied whenever

$$C_{M+1}(\varepsilon, \delta) \ll_k C_M(C_M(\varepsilon, \delta), \delta)^{1/2},$$

which in turn implies the bound

$$C_k(\varepsilon, \delta) \ll_k \delta^{1/2^{2^k+k-2}} + \varepsilon^{1/2^{2^k-1}}.$$

This completes the proof of Proposition D.1. □

Acknowledgements. The authors would like to thank Ben Green, Sean Prendiville, Terence Tao and Aled Walker for useful discussions, and the anonymous referee for a careful reading of the manuscript and several helpful suggestions. The second author is grateful to Régis de la Bretèche for an invitation to Paris in 2016 which prompted this work.

References

1. T. Bloom, A quantitative improvement for Roth's theorem on arithmetic progressions, *J. London Math. Soc. (2)* **93**(3) (2016), 643–663.
2. D. Conlon, J. Fox and Y. Zhao, The Green–Tao theorem: an exposition, *EMS Surv. Math. Sci.* **1**(2) (2014), 249–282.
3. D. Conlon, J. Fox and Y. Zhao, A relative Szemerédi theorem, *Geom. Funct. Anal.* **25**(3) (2015), 733–762.
4. D. A. Goldston, J. Pintz and C. Y. Yildirim, Primes in tuples. I, *Ann. of Math. (2)* **170**(2) (2009), 819–862.
5. W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11**(3) (2001), 465–588.
6. W. T. Gowers, Decompositions, approximate structure, transference, and the Hahn–Banach theorem, *Bull. London Math. Soc.* **42**(4) (2010), 573–606.
7. B. Green, Roth's theorem in the primes, *Ann. of Math. (2)* **161**(3) (2005), 1609–1636.
8. B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Ann. of Math. (2)* **167**(2) (2008), 481–547.
9. B. Green and T. Tao, New bounds for Szemerédi's theorem, III: A polylogarithmic bound for $r_4(n)$, preprint (arXiv:1705.01703, 2017).
10. H. A. Helfgott and A. de Roton, Improving Roth's theorem in the primes, *Int. Math. Res. Not. IMRN* **2011**(4) (2011), 767–783.

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press

11. K. Henriot, On systems of complexity one in the primes, *Proc. Edinb. Math. Soc. (2)* **60**(1) (2016), 133–163.
12. E. Naslund, On improving Roth's theorem in the primes, *Mathematika* **61**(1) (2015), 49–62.
13. K. O'Bryant, Sets of integers that do not contain long arithmetic progressions, *Electron. J. Combin.* **18**(1) Paper 59, 15, (2011).
14. O. Reingold, L. Trevisan, M. Tulsiani and S. Vadhan, New proofs of the Green–Tao–Ziegler dense model theorem: an exposition, preprint (arXiv:0806.0381, 2008).
15. T. Tao and T. Ziegler, The primes contain arbitrarily long polynomial progressions, *Acta Math.* **201**(2) (2008), 213–305.
16. P. Varnavides, On certain sets of positive density, *J. Lond. Math. Soc.* **34** (1959), 358–360.
17. Y. Zhao, An arithmetic transference proof of a relative Szemerédi theorem, *Math. Proc. Cambridge Philos. Soc.* **156**(2) (2014), 255–261.

https://doi.org/10.1017/S0013091518000561 Published online by Cambridge University Press