Journal of Number Theory 281 (2026) 110–138

Contents lists available at ScienceDirect

Journal of Number Theory

journal homepage: www.elsevier.com/locate/jnt

General Section

Some new results on the higher energies

I.D. Shkredov [a,b,*]

[a] *Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907–2067, USA*
[b] *London Institute for Mathematical Sciences, 21 Albermarle St, London W1S 4BS, UK*

A R T I C L E I N F O

Article history:
Received 12 August 2024
Received in revised form 15 September 2025
Accepted 17 September 2025
Available online 21 October 2025
Communicated by L. Smajlovic

MSC:
11B30
11B75

Keywords:
Higher energies
Arithmetic progressions
Gowers norms

A B S T R A C T

We obtain a generalization of the recent Kelley–Meka result on sets avoiding arithmetic progressions of length three. In our proof we develop the theory of the higher energies. Also, we discuss the case of longer arithmetic progressions, as well as a general family of norms, which includes the higher energies norms and Gowers norms.

1. Introduction

The famous Erdős–Turán conjecture [5] asks if it is true that for an arbitrary integer $k \geqslant 3$ any set of positive integers $A = \{n_1 < n_2 < \cdots < n_m < \dots\}$ satisfying

* Correspondence to: Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907–2067, USA.

E-mail address: ishkredo@purdue.edu.

https://doi.org/10.1016/j.jnt.2025.09.018

$$\sum_{j=1}^{\infty} \frac{1}{n_j} = \infty \qquad (1)$$

contains an arithmetic progression of length k (we say A has an APk in this case), that is the sequence of the form $x, x+y, \dots, x+(k-1)y \in A$? This question has a rich history see, e.g., [8], [9] or [23] and is considered as a central one in the area of classical additive combinatorics due to its connection with many adjacent fields as combinatorial ergodic theory and graphs/hypergraphs theory, we just mention some papers [26], [27], [6], [7], [22], [10], [29], [28], [14], [13] etc. If one defines

$$r_k(N) = \frac{1}{N} \max\{|A| \ : \ A \subseteq \{1, \dots, N\}\,, \quad A \text{ has no APk}\}\,,$$

then condition (1) means, roughly, that

$$r_k(N) \ll \frac{1}{\log N \cdot (\log\log N)^{1+\varepsilon}}\,, \qquad N \to \infty \qquad (2)$$

for an arbitrary $\varepsilon > 0$.

The case of arithmetic progressions of length three is considered special thanks to the Fourier approach of Roth [15], the necessary information and references on this topic can be found in [2], [12], as well as in [23]. Bloom and Sisask in [1] proved that $r_3(N) \ll (\log N)^{-1-c_1}$ for a certain $c_1 > 0$ and hence established conjecture (1) in the case of $k = 3$. Recently, Kelley and Meka [12] made significant progress on this issue and proved that

$$r_k(N) \ll \exp(-O((\log N)^{c_1}))\,,$$

where $c_1 > 0$ is an absolute constant. One of the ideas of paper [12] was to use the higher energy E_2^k and the notion of the uniformity relatively to E_2^k (all definitions can be found in Sections 2, 3) with a growing parameter k to control the number of arithmetic progressions in an arbitrary set. Namely, bound (2) is an immediate consequence of the following result (for simplicity we consider the group $\mathbb{F}_p^n$).

Theorem 1. *Let $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$, and $\varepsilon > 0$ be a parameter. Then there is a subspace $V \subseteq \mathbf{G}$ and $x \in \mathbf{G}$ such that $A \cap (V + x)$ is ε–uniform relatively to E_2^k, $\mu_{V+x}(A) \geqslant \delta$, and*

$$\operatorname{codim} V \ll \varepsilon^{-14} k^4 \mathcal{L}^3(\delta) \mathcal{L}^2(\varepsilon\delta)\,. \qquad (3)$$

The aim of this paper is to generalize the Kelley–Meka results to a wider additive–combinatorial family of energies E_l^k see, e.g., [25]. In our regime the parameter l is $l = O(1)$.

Theorem 2. *Let $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$, and $\varepsilon \in (0,1]$ be a parameter. Then there is a subspace $V \subseteq \mathbf{G}$ and $x \in \mathbf{G}$ such that $A \cap (V + x)$ is ε–uniform relatively to E_l^k, $\mu_{V+x}(A) \geqslant \delta$ and*

$$\mathrm{codim} V \ll \varepsilon^{-28l^l} (8l)^{28l^l} k^4 \mathcal{L}^{4l}(\varepsilon) \mathcal{L}^{5l}(\delta) \,. \tag{4}$$

Theorem 2 is interesting in its own right and can be used to solve more general equations and systems than $x + y = 2z$ (the latter corresponds to the case of AP3). The approach develops the strategy of [8], [9], the method of the higher energies (see, e.g., [20], [25]) and of course [12]. Also, we extensively use the brilliant exposition [2], where the Kelley–Meka results were discussed in detail. As an application of the 2 theorem, we consider a well–known two–dimensional generalization of arithmetic progressions of length three, namely the question of the density of sets avoiding *corners*, see, for example, [6] or [9].

In the appendix, we discuss the original Erdős–Turán conjecture, i.e. the case of longer arithmetic progressions and show that there is a number of difficulties at the conceptual and technical levels that make the question of generalizations of the methods from [12] rather hard. The author believes that this part is also interesting in its own right, as it allows us to understand the limitations of the Kelley–Meka approach. In addition, we consider a general family of norms that simultaneously includes the norms E_l^k mentioned above, as well as the classical Gowers norms [9].

We thank the reviewers for their valuable comments and useful remarks.

2. Definitions and preliminaries

Let $\mathbf{G}$ be a finite abelian group and denote by N the cardinality of $\mathbf{G}$. We use the same capital letter to denote a set $A \subseteq \mathbf{G}$ and its characteristic function $A : \mathbf{G} \to \{0,1\}$. Let us define $\mu_A(x) = A(x)/|A|$, and notice that $\sum_{x\in\mathbf{G}} \mu_A(x) = 1$. Finally, let $f_A(x) = A(x) - |A|/N$ be the *balanced function* of A. Given two sets $A, B \subset \mathbf{G}$, define the *sumset* of A and B as

$$A + B := \{a + b \ : \ a \in A, \, b \in B\} \,.$$

In a similar way we define the *difference sets* and the *higher sumsets*, e.g., $2A - A$ is $A + A - A$.

Let f be a function from $\mathbf{G}$ to $\mathbb{C}$. We denote the Fourier transform of f by $\widehat{f}$,

$$\widehat{f}(\xi) = \sum_{x \in \mathbf{G}} f(x) \overline{\chi(x)} \,, \tag{5}$$

where $\chi \in \widehat{\mathbf{G}}$ is a character of $\mathbf{G}$. We rely on the following basic identities

$$\sum_{x \in \mathbf{G}} |f(x)|^2 = \frac{1}{N} \sum_{\chi \in \widehat{\mathbf{G}}} \big|\widehat{f}(\chi)\big|^2 \,, \tag{6}$$

and

$$f(x) = \frac{1}{N} \sum_{\chi \in \widehat{\mathbf{G}}} \widehat{f}(\chi)\chi(x) \,. \tag{7}$$

If

$$(f * g)(x) := \sum_{y \in \mathbf{G}} f(y) g(x-y) \quad \text{and} \quad (f \circ g)(x) := \sum_{y \in \mathbf{G}} f(y) g(y+x) \,,$$

then

$$\widehat{f * g} = \widehat{f}\widehat{g} \tag{8}$$

and similar for $f \circ g$. Clearly, $(f * g)(x) = (g * f)(x)$ and $(f \circ g)(x) = (g \circ f)(-x)$, $x \in \mathbf{G}$. The k–fold convolution, $k \in \mathbb{N}$ we denote by $f^{(k)}$, so $f^{(2)} = f * f$ and $f^{(3)} = f * f * f$ for example.

We need some formalism concerning higher convolutions see, e.g., [25]. Let l be a positive integer. Consider two operators $\mathcal{D}_l, \mathcal{P}_l : \mathbf{G} \to \mathbf{G}^l$ such that for a variable $x \in \mathbf{G}$ one has $\mathcal{D}_l(x) = (x, \dots, x) \in \mathbf{G}^l$ and we formally write $\mathcal{P}_l(x) = (x_1, \dots, x_l) \in \mathbf{G}^l$, that is $\mathcal{P}_l(x) \in \mathbf{G}^l$ is a vector, which runs over $\mathbf{G}^l$. Notice that $\mathcal{P}_1(x) = \mathcal{D}_1(x) = x$. In the same way these operators act on functions $f : \mathbf{G} \to \mathbb{C}$, e.g., $\mathcal{P}_l(f)(x_1, \dots, x_l) = (f(x_1), \dots, f(x_l))$ (more generally, $\mathcal{P}_l(F)(x_1, \dots, x_l) = (f_1(x_1), \dots, f_l(x_l))$ for $F = (f_1, \dots, f_l)$) and $\mathcal{D}_l(f)(x_1, \dots, x_l) = f(x_1)$ if $x_1 = \dots = x_l$ and zero otherwise. Now given a function $f : \mathbf{G} \to \mathbb{C}$ and a positive integer l define the generalized convolution

$$\mathcal{C}_l(f)(x_1, \dots, x_l) = \sum_{z \in \mathbf{G}} f(z + x_1) \dots f(z + x_l) = (\mathcal{D}_l(\mathbf{G}) \circ \mathcal{P}_l(f))(x_1, \dots, x_l) \tag{9}$$

$$:= \sum_{z \in \mathbf{G}} f_{x_1, \dots, x_l}(z) \,. \tag{10}$$

In a similar way we can consider $\mathcal{C}_l(f_1, \dots, f_l)(x_1, \dots, x_l)$ for any functions $f_1, \dots, f_l : \mathbf{G} \to \mathbb{C}$. One has

$$\mathcal{C}_l(f)(x_1, \dots, x_l) = \mathcal{C}_l(f)(x_1 + w, \dots, x_l + w) = \mathcal{C}_l(f)((x_1, \dots, x_l) + \mathcal{D}_l(w)) \tag{11}$$

for any $w \in \mathbf{G}$. Let us emphasize that definitions (9), (10) differ slightly from the usual one, see, e.g., [25] by a linear change of the variables. Namely, it is a little bit more traditional to put

$$f'_{x_1, \dots, x_l}(z) = f_{0, x_1, \dots, x_l}(z) = f(z) f(z + x_1) \dots f(z + x_l) \,, \tag{12}$$

and

$$\mathcal{C}'_{l+1}(f)(x_1,\dots,x_l) = \sum_{z\in\mathbf{G}} f(z)f(z+x_1)\dots f(z+x_l) = \mathcal{C}_l(f)(0,x_1,\dots,x_l)\,. \tag{13}$$

Definitions (12), (13) have an advantage that they allow to consider infinite groups $\mathbf{G}$ as well. To this end we use the dual notation $\|f\|^{kl}_{\bar{\mathsf{E}}^k_l} = \bar{\mathsf{E}}^k_l(f) = N^{-1}\mathsf{E}^k_l(f)$. Now having $k,l \geqslant 2$ and a function $f:\mathbf{G}\to\mathbb{C}$ one can consider

$$\mathsf{E}^k_l(f) = \sum_{x_1,\dots,x_l} \mathcal{C}^k_l(f)(x_1,\dots,x_l) = \sum_{|y|=l} \mathcal{C}^k_l(f)(y) = \mathsf{E}^l_k(f) \tag{14}$$

and it was showed in [25, Proposition 30] (or see Corollary 6 below) that for a real function f and even k,l the formula $(\mathsf{E}^k_l(f))^{1/kl}$ defines a norm of our function f. Here, given a vector $y=(x_1,\dots,x_l)\in\mathbf{G}^l$, the fact that y has l coordinates is expressed as $|y|=l$. The property $\mathsf{E}^k_l(f)=\mathsf{E}^l_k(f)$ of the energies E^k_l we call *duality*, and this equality was proved in [20] (see also [24] and [25]). If one puts $l=1$ in (14), then we formally obtain $\mathsf{E}^k_1(f) = N(\sum_z f(x))^k$ and this is not a norm for any k. Nevertheless, it is convenient to consider the quantities $\mathsf{E}^k_1(f)$ sometimes. Notice that $\mathsf{E}^k_l(f)\geqslant 0$, provided at least one of k,l is even but, nevertheless, it cannot be a norm in this case, see [25, Sections 4,7] (although it is a norm restricted to the family of non–negative functions). A general family of norms, which includes the norms above is considered in the second part of the appendix, where, in particular, one can find the discussed properties of the energies $\mathsf{E}^k_l(f)$. Finally, let us articulate one more formula for the energy $\mathsf{E}^k_l(f)$, namely,

$$\mathsf{E}^k_l(f) = \sum_{x_1,\dots,x_l} (f\circ f_{x_2,\dots,x_l})^k(x_1) = \sum_{y_1,\dots,y_k} (f\circ f_{y_2,\dots,y_k})^l(y_1)\,. \tag{15}$$

For the convenience of the reader, we recall the Croot–Sisask Lemma, see [4], [17].

Lemma 3. *Let $\mathbf{G}$ be an abelian group, $\varepsilon\in(0,1)$ and $K\geqslant 1$ be real numbers, q be a positive integer, $A,B\subseteq\mathbf{G}$ be sets such that $|A+B|\leqslant K|A|$, and let $f\in L_l(\mathbf{G})$ be an arbitrary function. Then there exist a $b\in B$ and a set $T\subseteq B$ with $|T|\geqslant |B|(2K)^{-O(\varepsilon^2 q)}$ such that*

$$\|(f*A)(x+t)-(f*A)(x)\|_{L_q(\mathbf{G},x)} \leqslant \varepsilon^q|A|\|f\|^q_{L_q(\mathbf{G})}$$

for all $t\in T-b$.

Also, we recall a special case of Chang's Lemma, see [3]. Recall that for a set $A\subseteq\mathbf{G}$ and $\varepsilon\in(0,1]$ the set

$$\mathrm{Spec}_{\,\varepsilon}(A) = \{z\in\mathbf{G}\ :\ |\widehat{A}(z)|\geqslant\varepsilon|A|\}\,, \tag{16}$$

is called the ε*–spectrum* of A.

Lemma 4. *Let* $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$. *Then*

$$\dim(\mathrm{Spec}_\varepsilon(A)) \ll \varepsilon^{-2} \log(|\mathbf{G}|/|A|)\,.$$

Finally, we need a result on the functions $\mathcal{C}_l(f)(x_1, \dots, x_l)$, see [25, Lemma 29] and [25, Corollary 30].

Lemma 5. *Let* $\mathbf{G}$ *be an abelian group,* $k, l \geqslant 2$ *be even numbers and* $\varphi_1, \dots, \varphi_k : \mathbf{G} \to \mathbb{C}^l$, $\varphi_j = (\varphi_j^{(1)}, \dots, \varphi_j^{(l)})$. *Then*

$$\left| \sum_{x \in \mathbf{G}^l} (\mathcal{D}_l(\mathbf{G}) \circ \mathcal{P}_l(\varphi_1)) \dots (\mathcal{D}_l(\mathbf{G})(x) \circ \mathcal{P}_l(\varphi_k))(x) \right| \leqslant \prod_{j=1}^{k} \prod_{i=1}^{l} \|\varphi_j^{(i)}\|_{\mathsf{E}_l^k}\,. \tag{17}$$

Corollary 6. *Let* $\mathbf{G}$ *be an abelian group,* $k, l \geqslant 2$ *be even numbers. Then for any pair of functions* $f, g : \mathbf{G} \to \mathbb{C}$ *the following holds*

$$\|f + g\|_{\mathsf{E}_k^l} \leqslant \|f\|_{\mathsf{E}_k^l} + \|g\|_{\mathsf{E}_k^l}\,,$$

and $\|\cdot\|_{\mathsf{E}_k^l}$ *is a norm.*

Let $\varepsilon \in (0, 1]$ be a real number. We write $\mathcal{L}(\varepsilon)$ for $\log(2/\varepsilon)$. Let us make a convention that if a product is taken over an empty set, then it equals one. The signs $\ll$ and $\gg$ are the usual Vinogradov symbols. When the constants in the signs depend on a parameter M, we write $\ll_M$ and $\gg_M$. All logarithms are to base 2. By $\mathbb{F}_p$ denote $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for a prime p. We write $V \leqslant \mathbb{F}_p^n$ if V is a subspace of the group $\mathbb{F}_p^n$. Let us denote by $[n]$ the set $\{1, 2, \dots, n\}$.

3. Some results on E_l^k–norms

In this section, we generalize some the Kelley–Meka results which were obtained for the E_2^k–norm to the E_l^k–norm. Also, we discuss some special properties of such norms. Our results naturally fall into two cases: uniform and non–uniform.

3.1. Uniform sets in the sense of E_l^k–norm

Let us give the main definition of this subsection.

Definition 7. Let $\mathbf{G}$ be a finite abelian group, $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$, and $\varepsilon > 0$ be a parameter. Then we say that A is ε–uniform relatively to (the energy) E_l^k if

$$\|f_A\|_{\mathsf{E}_l^k}^{kl} \leqslant \varepsilon^{kl} \delta^{kl} N^{k+l}\,. \tag{18}$$

Usually the number ε belongs to $(0,1]$ but sometimes $\varepsilon > 1$ and hence one can consider the quantity ε as the definition of the energy $\|f_A\|^{kl}_{\mathsf{E}^k_l}$, that is $\|f_A\|^{kl}_{\mathsf{E}^k_l} := \varepsilon^{kl}\delta^{kl}N^{k+l}$. Further by the Hölder inequality, we have

$$(\mathsf{E}^{k-1}_l(f))^k \leqslant (\mathsf{E}^k_l(f))^{k-1}N^l \tag{19}$$

and hence if A is ε–uniform relatively to E^k_l, then A is ε–uniform relatively to $\mathsf{E}^{k'}_{l'}$ for $k' \leqslant k$, $l' \leqslant l$ (we consider just even indices, say). On the other hand, it is easy to see that the smaller norm does not control the higher one.

Example 8. Let $\mathbf{G} = \mathbb{F}_2^n$, $H < \mathbb{F}_2^n$, $\Lambda \subseteq \mathbb{F}_2^n/H$ be a random set such that $|\Lambda| = \delta N/|H|$. Also, suppose that $\delta^2 \gg |H|/N$ and thus with high probability $\Lambda - \Lambda \approx \mathbb{F}_2^n/H$. Let A be the direct sum of H and Λ, then $|A| = \delta N$. It is easy to see that for a random $x \in A - A \approx \mathbf{G}$ one has $|A_x| \sim \delta^2 N$ but for $x, y \in H$ one has $A_x = A$ and $A_{x,y} = A$. Thus for any $k \geqslant 2$ the following holds

$$\mathsf{E}^k_2(A) \sim (\delta^2 N)^k N + (\delta N)^k |H| \sim (\delta^2 N)^k N\,,$$

provided $|H| \ll \delta^k N$ but taking an arbitrary k_*, we see that

$$\mathsf{E}^{k_*}_3(A) \sim (\delta^3 N)^{k_*} N^2 + (\delta N)^{k_*}|H|^2 \gg (\delta N)^{k_*}|H|^2\,,$$

provided $|H| \gg \delta^{k_*} N$. It follows that one can choose any $k_* \geqslant k+1$ and construct a set A such that A is E^k_2–uniform but not $\mathsf{E}^{k_*}_3$–uniform. Of course, one can replace the pair $(2,3)$ by any suitable pair of indices.

Now let us obtain the characteristic property of the energy E^k_l (also, see Remark 4 below).

Lemma 9. *Let $l, k \geqslant 2$ be even numbers, and $A_j \subseteq \mathbf{G}$, $j \in [l]$ be sets. Then for any function $g : \mathbf{G} \to \mathbb{R}$ one has*

$$\sum_x \prod_{j=1}^{l} (f_{A_j} \circ g)(x) \leqslant \|g\|_1^{l(1-1/k)} \|\mathcal{C}_l(g)\|_\infty^{1/k} \cdot \prod_{j=1}^{l} \|f_{A_j}\|_{\overline{\mathsf{E}}^k_l}\,. \tag{20}$$

If all sets A_j are the same, then for any l and an arbitrary even k bound (20) still takes place.

Proof. By the Hölder inequality, [25, Lemma 29] and the duality one has

$$\sum_x \prod_{j=1}^{l} (f_{A_j} \circ g)(x) = N^{-1} \sum_{|z|=l} \mathcal{C}_l(f_{A_1}, \dots, f_{A_l})(z)\mathcal{C}_l(g)(z)$$

$$\leqslant N^{-1+1/k} \prod_{j=1}^{l} \|f_{A_j}\|_{\overline{\mathsf{E}}_l^k} \cdot \left(\sum_{|z|=l} |\mathcal{C}_l^{k/(k-1)}(g)(z)| \right)^{1-1/k}$$

$$\leqslant \prod_{j=1}^{l} \|f_{A_j}\|_{\overline{\mathsf{E}}_l^k} \cdot \|\mathcal{C}_l(g)\|_\infty^{1/k} \|g\|_1^{l(1-1/k)} .$$

This completes the proof. □

Corollary 10. *Let $A, B \subseteq \mathbf{G}$ be sets, $|A| = \delta N$, $|B| = \beta N$, and l be a positive integer. Take $k = 2\lceil 4l \log(1/\beta) \rceil$ and suppose that A is ε–uniform relatively to the energy E_l^k. Then*

$$\sum_x (A \circ B)^l(x) \leqslant \delta^l |B|^l N \cdot \min\{1.25(1+\varepsilon)^l, (1+1.25\varepsilon)^l\} . \qquad (21)$$

Proof. Using the formula $A(x) = f_A(x) + \delta$, combining with Lemma 9, we see that the left–hand side of (21) is

$$\sum_{j=0}^{l} \binom{l}{j} (\delta|B|)^{l-j} \sum_x (f_A \circ B)^j(x) \leqslant \sum_{j=0}^{l} \binom{l}{j} (\delta|B|)^{l-j} \|f_A\|^j_{\overline{\mathsf{E}}_j^k} |B|^{j(1-1/k)+1/k}$$

$$\leqslant \delta^l |B|^l N \sum_{j=0}^{l} \binom{l}{j} \varepsilon^j \beta^{-(j-1)/k} \leqslant \frac{5}{4} \delta^l |B|^l N (1+\varepsilon)^l$$

as required. One can obtain the second bound in a similar way. This completes the proof. □

Let us remark that, of course, the energy E_2^k solely allows us to control sums from (21) but our task is to obtain the correct power of δ and $|B|$ in the right–hand side of this estimate (also, see Remark 4 below).

We need one more result about uniform sets, which is useful for applications.

Lemma 11. *Let $k, l \geqslant 2$ be even numbers, $A_1, \dots, A_l \subseteq \mathbf{G}$ be sets, $|A_j| = \delta_j N$. Suppose that all A_j are ε–uniform relatively to E_l^k and $2l\varepsilon^k \leqslant 1$. Then*

$$\sum_{|x|=l} \left(\mathcal{C}_l(A_1, \dots, A_l)(x) - N \prod_{j=1}^{l} \delta_j \right)^k \leqslant 2^{kl} \varepsilon^k N^{l+k} \left(\prod_{j=1}^{l} \delta_j \right)^k . \qquad (22)$$

Proof. Put $\Pi = \prod_{j=1}^{l} \delta_j$. Then the left–hand side of (22) is

$$\sigma := \sum_{|x|=l} \left(\sum_{\emptyset \neq S \subseteq [l]} \mathcal{C}_l(f_1, \dots, f_l)(x) \right)^k = \sum_{|x|=l} \left(\sum_{\emptyset \neq S \subseteq [l]} F_S(x) \right)^k ,$$

where for $j \in S$ we put $f_j = f_{A_j}$ and if $j \notin S$, then $f_j = \delta_j$. Using ε–uniformity of all sets A_j and estimate (17) of Lemma 5 with $\varphi_j = (f_1, \dots, f_l)$, we get

$$\sum_{|x|=l} F_S^k(x) \leqslant \varepsilon^{|S|k} \Pi^k N^{l+k}\,.$$

Thus by the Hölder inequality one has

$$\sigma \leqslant 2^{(k-1)l} \sum_{\emptyset \neq S \subseteq [l]} \sum_{|x|=l} F_S^k(x) \leqslant 2^{(k-1)l} \Pi^k N^{k+l} ((1+\varepsilon^k)^l - 1) \leqslant 2^{kl} \varepsilon^k \Pi^k N^{k+l}$$

as required. □

Let us consider one more example which shows that it is possible to remove/add a tiny subset from a non–uniform set to get a uniform one. This phenomenon has no place if we consider the classical uniformity in terms of the Fourier transform or in terms of Gowers norms [9], say. The reason is normalization (18), of course.

Example 12. Let $\mathbf{G} = \mathbb{F}_2^n$, $H < \mathbb{F}_2^n$, $|H| = \beta N$, $\Lambda \subseteq \mathbf{G}$ be a random set, $|\Lambda| = \delta N$, $\beta \leqslant \delta \leqslant 1/2$ and put $A = \tilde{H} \bigsqcup \Lambda$, where $\tilde{H} = H \setminus \Lambda$. Then with high probability $|\tilde{H}| \sim \beta(1-\delta)N := \varepsilon|A|$, the set Λ is uniform in all possible senses but A is a non η–uniform set with rather large η. Indeed, by the Kelley–Meka method [12] or just see Lemma 9, we know that

$$\begin{aligned}\sigma(A,A) &:= \sum_x (A \circ A)(x) H(x) = |H|^{-1} \sum_x (A \circ A)(x)(H \circ H)(x) \\ &= (\delta + (1-\delta)\beta)^2 \beta N^2 + \theta \eta^2 \delta^2 \beta N^2 \\ &= \beta \delta^2 (1+\varepsilon)^2 N^2 + \theta \eta^2 \delta^2 \beta N^2\,, \end{aligned} \tag{23}$$

where $|\theta| \leqslant 4$, say, is a certain number and A is supposed to be E_2^k–uniform with $k \sim \mathcal{L}(\beta)$. On the other hand, direct calculation shows that

$$\begin{aligned}\sigma(A,A) = \sigma(\tilde{H},\tilde{H}) + 2\sigma(\tilde{H},\Lambda) + \sigma(\Lambda,\Lambda) &= |\tilde{H}|^2 + 2\delta\beta^2(1-\delta)N^2 + \delta^2 |H| N \\ &= (\varepsilon^2\delta^2 + 2\delta^2\beta\varepsilon + \delta^2\beta)N^2 \end{aligned} \tag{24}$$

plus a negligible error term. Comparing (23) and (24), we obtain

$$\eta^2 \delta^3 \varepsilon \gg \varepsilon^2 \delta^2 (1-\beta) \gg \varepsilon^2 \delta^2$$

and thus $\eta^2 \gg \varepsilon/\delta$ which is much larger than ε for small δ.

Similarly, one can show that removing a subspace H from a random set Λ, $|H| = \varepsilon|A|$, $A = \Lambda \setminus H$, H lives on the first coordinates, say, we obtain a non η–uniform set with $\eta \gg 1$ thanks to the equality $\sum_x H(x)(A \circ A)(x) = 0$.

Remark 1. Example 12 shows us that E^k_l–norm is a rather delicate object and addition/removal of a small set can change the norm dramatically. Nevertheless, notice that if $A = \bigsqcup_{j=1}^s A_j$, $\delta = \mu_{\mathbf{G}}(A)$, $\delta_j = \mu_{\mathbf{G}}(A_j)$ and all sets A_j are ε–uniform relatively to the norm E^k_l, where k, $l \geqslant 2$ are even numbers, then by the triangle inequality for the norm E^k_l the set A is also ε–uniform. If k is an even number and l is an arbitrary positive integer, then still the characteristic property (21) of the norm E^k_l takes place. Indeed, by the Hölder inequality one has

$$\sigma := \sum_x (\sum_{j=1}^s f_{A_j} \circ B)^l (x) = \sum_{i_1+\dots+i_s=l} \binom{l}{i_1,\dots,i_s} \sum_x (f_{A_1} \circ B)^{i_1}(x) \dots (f_{A_s} \circ B)^{i_s}(x)$$
$$\leqslant \sum_{i_1+\dots+i_s=l} \binom{l}{i_1,\dots,i_s} \prod_{j=1}^s \left(\sum_x |(f_{A_j} \circ B)(x)|^l \right)^{i_j/l} .$$

Now consider $\sigma_0 := \sum_x |(f_S \circ B)(x)|^l$ for an arbitrary ε–uniform relatively to the norm E^k_l set S, $\varepsilon \leqslant 1/4$, say, and let $\mu_{\mathbf{G}}(S) = \sigma$. Of course, one has $\sigma_0 = \sum_x (f_S \circ B)(x)^l$ for even l but for odd l, we have a similar bound, namely,

$$\sigma_0 \leqslant \sum_x (f_S \circ B)(x)^{l-1} ((f_S \circ B)(x) + 2\delta|B|) = \sum_x (f_S \circ B)(x)^l + 2\sigma|B| \sum_x (f_A \circ B)(x)^{l-1}$$
$$\leqslant 5/4 \cdot (2+\varepsilon)\varepsilon^{l-1}\sigma^l |B|^l N \leqslant 3\varepsilon^{l-1}\sigma^l |B|^l N$$

thanks to Lemma 9 (also, see Corollary 10). Here we have assumed that $k \gg l\mathcal{L}(\mu_{\mathbf{G}}(B))$, of course. Thus

$$\sigma \leqslant 3\varepsilon^{l-1}|B|^l N \sum_{i_1+\dots+i_s=l} \binom{l}{i_1,\dots,i_s} \prod_{j=1}^s \delta_j^{i_j} = 3\varepsilon^{l-1}|B|^l N \left(\sum_{j=1}^s \delta_j \right)^l = 3\varepsilon^{l-1}\delta^l |B|^l N$$

and we see that the difference between two cases is almost absent.

3.2. Non–uniformity and almost periodicity

The aim of this subsection is to obtain Sanders' almost periodicity result for *higher convolutions*, see Lemma 15 below.

At the beginning we want to transfer a lower bound for the energy $\mathsf{E}^k_l(f_A)$ to the largeness of the energy $\mathsf{E}^k_l(A)$. We follow a more simple method from [2] which differs from the approach of [12] by some logarithms. The dependence on l in the first multiple in (26) is, probably, can be improved significantly (also, see Remark 6 from the appendix) but in our regime $l = O(1)$ and thus it is not so critical for us.

Lemma 13. *Let $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$ and $\varepsilon > 0$ be a parameter, $\varepsilon_* := \min\{\varepsilon, 1\}$. Suppose that for an odd $k \geqslant 5$ one has*

$$\mathsf{E}^k_l(f_A) = \varepsilon^{lk}\delta^{lk}N^{l+k}\,, \tag{25}$$

and that for $k_* = O(kl\varepsilon_*^{-l}\mathcal{L}(\varepsilon_*))$ *the set* A *is* $\frac{\varepsilon\varepsilon_*^{l-1}}{8l}$*–uniform relatively to* $\mathsf{E}^{k_*}_{l-1}$*. Then there is an even* $k_1 \leqslant k_*$ *such that*

$$\mathsf{E}^{k_1}_l(A) \geqslant \left(1+\frac{\varepsilon\varepsilon_*^{l-1}}{8l}\right)^{lk_1}\delta^{lk_1}N^{l+k_1}\,. \tag{26}$$

Proof. Write $f(x) = f_A(x)$ and put $P = \{x \,:\, \mathcal{C}_l(f)(x) \geqslant 0\}$. Since k is an odd number, we have

$$\sum_{x\in P}\mathcal{C}^k_l(f)(x) \geqslant \varepsilon^{lk}\delta^{lk}N^{l+k}\,. \tag{27}$$

Now let us consider the subset of the set P, namely,

$$P_\varepsilon := \{x \,:\, \mathcal{C}_l(f)(x) \geqslant \frac{3}{4}\varepsilon^l\delta^l N\}\,.$$

Then we have

$$\sum_{x\notin P_\varepsilon}\mathcal{C}^k_l(f)(x) \leqslant \left(\frac{3}{4}\varepsilon^l\delta^l N\right)^k N^l \leqslant 2^{-2}\varepsilon^{lk}\delta^{lk}N^{l+k}\,. \tag{28}$$

Combining (27), (28) and using the Hölder inequality, we obtain

$$|P_\varepsilon|\mathsf{E}^{2k}_l(f) \geqslant 2^{-4}\varepsilon^{2lk}\delta^{2lk}N^{2l+2k}\,. \tag{29}$$

By the norm property of $\mathsf{E}^{2k}_l(f)$ for positive functions (see Corollary 6) one has

$$\mathsf{E}^{2k}_l(f) \leqslant \left(\|A\|_{\mathsf{E}^{2k}_l} + \|\delta\|_{\mathsf{E}^{2k}_l}\right)^{2kl} \leqslant (2+\varepsilon/8)^{2kl}\delta^{2kl}N^{l+2k}$$

otherwise there is nothing to prove with $k_1 = 2k$ and much larger ε. Thus we derive from (29) that $|P_\varepsilon| \geqslant (2\varepsilon_*/5)^{2kl}N^l$. Now

$$\mathcal{C}_l(A)(x) = \mathcal{C}_l(f+\delta)(x) = \delta^l N + \mathcal{C}_l(f) + \sum_{S\subseteq[l] \,:\, 1\leqslant|S|<l}\delta^{l-|S|}\mathcal{C}_{|S|}(f)(x_S)$$
$$= \delta^l N + \mathcal{C}_l(f) + \mathcal{E}(x)\,,$$

where for a set $S \subseteq [l]$ the vector x_S has coordinates x_j, $j \in S$. By the triangle inequality for L_{k_1}–norm, we have

$$(\mathsf{E}^{k_1}_l(A))^{1/k_1} = \|\mathcal{C}_l(A)\|_{k_1} = \|\mathcal{C}_l(f+\delta)\|_{k_1} \geqslant \|\delta^l N + \mathcal{C}_l(f)\|_{k_1} - \|\mathcal{E}\|_{k_1}\,. \tag{30}$$

Using our bound for the cardinality of the set P_ε, we get

$$\|\delta^l N + \mathcal{C}_l(f)\|_{k_1}^{k_1} \geqslant \sum_{x\in P_\varepsilon} (\delta^l N + \mathcal{C}_l(f))^{k_1}(x) \geqslant (2\varepsilon_*/5)^{2kl} N^l \cdot (1+3\varepsilon^l/4)^{k_1} \delta^{lk_1} N^{k_1}$$
$$\geqslant (1+\varepsilon^l/2)^{k_1} \delta^{lk_1} N^{k_1+l}\,, \qquad (31)$$

provided $k_1 \geqslant 20kl\varepsilon_*^{-l}\mathcal{L}(\varepsilon_*)$. On the other hand, by our assumption the set A is $\zeta := \frac{\varepsilon^l}{8l}$–uniform relatively to $\mathsf{E}_j^{k_*}$ for all $j < l$ and $k_* = k_1$. It follows that

$$\|\mathcal{E}\|_{k_1} \leqslant \sum_{S\subseteq[l]\,:\,1\leqslant|S|<l} \delta^{l-|S|} N^{\frac{l-|S|}{k_1}} \|f\|^{|S|}_{\mathsf{E}^{k_1}_{|S|}} \leqslant \delta^l N^{\frac{l+k_1}{k_1}} ((1+\zeta)^l - 1)\,. \qquad (32)$$

Combining (30), (31) and (32), we obtain

$$\mathsf{E}_l^{k_1}(A) \geqslant \delta^{lk_1} N^{k_1+l} \left(2+\varepsilon^l/2 - (1+\zeta)^l\right)^{k_1} \geqslant \delta^{lk_1} N^{k_1+l} \left(1+\frac{\varepsilon\varepsilon_*^{l-1}}{8l}\right)^{lk_1}$$

as required. □

Now we use duality (14) to obtain the appropriate version of multi–dimensional version of the Balog–Szemerédi–Gowers theorem as was done in [19] (also, see [24, Theorem 17]). Thanks to duality (65) one can show that a similar result takes place for more general energies $\mathcal{E}^k_{s,t}$, see the appendix. Of course, in this case one needs to replace $\mathcal{C}_{|x|}(\cdot)(x)$ to $\mathcal{C}_{|x||z|}(\cdot)(x\oplus z)$ or $\mathcal{C}_{|y||z|}(\cdot)(y\oplus z)$ and use symmetries (67) instead of the symmetry (11) below.

Lemma 14. *Let $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$ and $\varepsilon > 0$, $\eta \in (0, 1/2)$ be parameters. Suppose that for some integers $k, l \geqslant 2$ with $kl \geqslant 4\varepsilon_*^{-1}\mathcal{L}(\eta)$ one has*

$$\mathsf{E}_l^k(A) \geqslant (1+\varepsilon)^{lk} \delta^{lk} N^{l+k}\,. \qquad (33)$$

Define the set

$$S = \{|x| = l \;:\; \mathcal{C}_l(A)(x) \geqslant (1+\varepsilon/4)^l \delta^l N\}\,. \qquad (34)$$

Then there is a set B such that

$$N^{-1} \sum_{|x|=l} S(x)\mathcal{C}_l(B)(x) \geqslant (1-2\eta)|B|^l\,, \qquad (35)$$

and $|B| \geqslant 2^{-1/(l-1)}(1+\varepsilon)^k\delta^k N$.

Proof. For *any* set $S \subseteq \mathbf{G}^l$ with the property $S(x+\mathcal{D}_l(t)) = S(x)$, $t \in \mathbf{G}$, $x \in \mathbf{G}^l$, we get

$$\sum_{|x|=l} S(x)\mathcal{C}_l^k(A)(x) = \sum_{|x|=l} S(x) \sum_{|z|=k} A^k(z+\mathcal{D}_k(x_1))\dots A^k(z+\mathcal{D}_k(x_l))$$

$$= \sum_{|z|=k} \sum_{|x|=l} S(x) A_z(x_1) \dots A_z(x_l) = N^{-1} \sum_{|z|=k} \sum_{|x|=l} S(x) \mathcal{C}_l(A_z)(x) \,, \tag{36}$$

where we have made the change of the variables $x_j \to x_j + t$ in the last formula. Clearly, we have from identity (11) and definition (34) that $S(x + \mathcal{D}_l(t)) = S(x)$ for $t \in \mathbf{G}$ and $x \in \mathbf{G}^l$ and thus the argument above can be applied for the set S as well. Thus using the definition of the set S, as well as conditions (33) and $kl \geqslant 4\varepsilon_*^{-1} \mathcal{L}(\eta)$, we get

$$\sum_{x \notin S} \mathcal{C}_l^k(A)(x) \leqslant (1 + \varepsilon/4)^{lk} \delta^{lk} N^{k+l} \leqslant 2^{-2} \eta (1+\varepsilon)^{lk} \delta^{lk} N^{k+l} \leqslant 2^{-2} \eta \mathsf{E}_l^k(A) \,. \tag{37}$$

Now let us define the set

$$\Omega = \left\{ |z| = k \ : \ |A_z| \geqslant 2^{-1/(l-1)} (1+\varepsilon)^k \delta^k N \right\} \,.$$

By the definition of the set Ω, we have

$$\sum_{z \notin \Omega} |A_z|^l \leqslant (\max_{z \notin \Omega} |A_z|)^{l-1} |A|^k N \leqslant 2^{-1} (1+\varepsilon)^{lk} \delta^{lk} N^{l+k} \leqslant 2^{-1} \mathsf{E}_l^k(A) \,. \tag{38}$$

In view of bounds (37), (38) one has

$$N^{-1} \sum_{z \in \Omega} \left(\sum_{x \in S} \mathcal{C}_l(A_z)(x) - \eta^{-1} \sum_{x \notin S} \mathcal{C}_l(A_z)(x) \right) \geqslant 2^{-1} \sum_{|z|=k} |A_z|^l - 2^{-1} \mathsf{E}_l^k(A) = 0 \,. \tag{39}$$

Hence there is $z \in \Omega$ such that inequality (35) holds for $B = A_z$ and $|B| \geqslant 2^{-1/(l-1)}(1+\varepsilon)^k \delta^k N$ as required. $\square$

Now we need an analogue of the almost periodicity result [4] (also, see [16], [17], [18] and, especially, [21, Theorem 3.2]) for the higher convolutions. This theme is rather well–known and thus we give just a scheme of the proof, emphasizing the necessary distinctions we need to make. For the convolution $\mathcal{C}_{|x||z|}(x \oplus z)$ a similar result takes place, see Lemma 21 from the appendix.

Lemma 15. *Let $\mathbf{G} = \mathbb{F}_p^n$, l be an integer and $\epsilon \in (0, 1]$ be a real parameter. Also, let $B \subseteq \mathbf{G}$ be a set, $|B| = \beta N$, and $f : \mathbf{G}^l \to [-1, 1]$ be a function. Then there is a subspace $V \leqslant \mathbf{G}$ with*

$$\mathrm{codim} V \ll \epsilon^{-2} l \mathcal{L}^2(\beta) \mathcal{L}^2(\epsilon \beta^l) \tag{40}$$

and such that

$$\left| \sum_{|x|=l} f(x) (B^l \circ \mathcal{D}_l(B * \mu_V))(x) - \sum_{|x|=l} f(x) (B^l \circ \mathcal{D}_l(B))(x) \right| \leqslant \epsilon |B|^{l+1} \,. \tag{41}$$

Proof. We begin with a rather general argument that takes place in any abelian group $\mathbf{G}$. Let $k \geqslant 2$ be an integer parameter and $q \geqslant 2$ be a real parameter. Applying the Croot–Sisask Lemma 3 with $\varepsilon = \varepsilon/(4k)$, $A = \mathcal{D}_l(B)$, $B = \mathcal{D}_l(\mathbf{G})$ (clearly, one has $|\mathcal{D}_l(B) + \mathcal{D}_l(\mathbf{G})| \leqslant \beta^{-1}|\mathcal{D}_l(B)|$), we find a set $T \subseteq \mathbf{G}$, $|T| \geqslant |B| \exp(-O(\epsilon^{-2} q k^2 \log(1/\beta)))$ and such that for any $t \in kT$ the following holds

$$\sum_{|x|=l} \left| (f \circ \mathcal{D}_l(B))(x + \mathcal{D}_l(t)) - \sum_{|x|=l} (f \circ \mathcal{D}_l(B))(x) \right|^q \leqslant \left(\frac{\epsilon}{4}\right)^q \|f\|_q^q |B|^q \leqslant \left(\frac{\epsilon}{4}\right)^q |B|^q N^l . \tag{42}$$

Fixing $t \in kT$ and using the Hölder inequality, combining with estimate (42), we get

$$\left| \sum_{|x|=l} f(x)(B^l \circ \mathcal{D}_l(B))(x + \mathcal{D}_l(t)) - \sum_{|x|=l} f(x)(B^l \circ \mathcal{D}_l(B))(x) \right| \leqslant \frac{\epsilon}{4} |B| N^{l/q} |B|^{l(1-1/q)}$$
$$= \frac{\epsilon}{4} \beta^{-l/q} |B|^{l+1} \leqslant \frac{\epsilon}{2} |B|^{l+1} , \tag{43}$$

where we have taken $q = Cl \log(1/\beta)$ for a sufficiently large constant $C > 0$. It follows that

$$\left| \sum_{|x|=l} f(x)(B^l \circ \mathcal{D}_l(B * \mu_T^{(k)}))(x) - \sum_{|x|=l} f(x)(B^l \circ \mathcal{D}_l(B))(x) \right| \leqslant 2^{-1} \epsilon |B|^{l+1} . \tag{44}$$

Let us analyze the sum $\sigma := | \sum_{|x|=l} f(x)(B^l \circ \mathcal{D}_l(B * \mu_T^{(k)}))(x)|$ from (44). Clearly, one has $\widehat{\mu}_{\mathcal{D}_l(T)}(r_1, \dots, r_l) = |T|^{-1} \widehat{T}(r_1 + \dots + r_l)$ and thus

$$\sigma \leqslant \frac{|B|}{|T|^k N^l} \sum_z |\widehat{T}(z)|^k \sum_{r_1 + \dots + r_l = z} |\widehat{f}(r_1, \dots, r_l)| |\widehat{B}(r_1)| \dots |\widehat{B}(r_l)| . \tag{45}$$

As usual let us estimate the last sum over $z \in \mathrm{Spec}_c(T)$, where $c \in (0, 1]$ is a parameter and over $z \notin \mathrm{Spec}_c(T)$, see formula (16). By the definition of the set $\mathrm{Spec}_c(T)$, the Hölder inequality and the Parseval identity, we have

$$\sigma_1 := \frac{|B|}{|T|^k N^l} \sum_{z \notin \mathrm{Spec}_c(T)} |\widehat{T}(z)|^k \sum_{r_1 + \dots + r_l = z} |\widehat{f}(r_1, \dots, r_l)| |\widehat{B}(r_1)| \dots |\widehat{B}(r_l)|$$
$$\leqslant \frac{c^k |B|}{N^l} \sum_{r_2, \dots, r_l} |\widehat{B}(r_2)| \dots |\widehat{B}(r_l)| \sum_z |\widehat{f}(z - r_2 + \dots + r_l, \dots, r_l)| |\widehat{B}(z - r_2 + \dots + r_l)|$$
$$\leqslant \frac{c^k |B|^{3/2}}{N^{l-1}} \sum_{r_2, \dots, r_l} |\widehat{B}(r_2)| \dots |\widehat{B}(r_l)| \left(\sum_a |\widehat{f}_a(r_2, \dots, r_l)|^2 \right)^{1/2} ,$$

where $f_a(x_2,\dots,x_l) = f(a,x_2,\dots,x_l)$. Using the Hölder inequality and the Parseval formula one more time, we derive

$$\sigma_1 \leqslant \frac{c^k|B|^{(l+2)/2}}{N^{(l-1)/2}} \left(\sum_{r_2,\dots,r_l} \sum_a |\widehat{f}_a(r_2,\dots,r_l)|^2 \right)^{1/2} \leqslant c^k |B|^{(l+2)/2} N^{l/2} \leqslant 2^{-1}\epsilon |B|^{l+1} \,.$$

Here we have taken $c = 1/2$ and $k = \lceil 2\mathcal{L}(\epsilon\beta^l)\rceil$, say. For the sum over $z \in \mathrm{Spec}_c(T)$ we use Chang's Lemma 4 with the parameters $\varepsilon = c$ and $A = T$, and find a subspace V such that (41) takes place and

$$\mathrm{codim} V \ll \log(N/|T|) \ll \epsilon^{-2} l \mathcal{L}^2(\beta) k^2 \ll \epsilon^{-2} l \mathcal{L}^2(\beta)\mathcal{L}^2(\epsilon\beta^l)\,,$$

see details in [4], [17], [18] or in [24, Section 5]. This completes the proof. □

4. Some generalizations of the Kelley–Meka results

Using the density increment, Kelley–Meka [12] (or just repeat the calculations of the previous section, combining forthcoming Proposition 18 in the case $l = 2$) obtained the following result.

Theorem 16. *Let $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$, and $\varepsilon > 0$ be a parameter. Then there is a subspace $V \subseteq \mathbf{G}$ and $x \in \mathbf{G}$ such that $A \cap (V + x)$ is ε–uniform relatively to E^k_2, $\mu_{V+x}(A) \geqslant \delta$, and*

$$\mathrm{codim} V \ll \varepsilon^{-14} k^4 \mathcal{L}^3(\delta)\mathcal{L}^2(\varepsilon\delta) \cdot \mathcal{L}^4(\varepsilon)\,. \tag{46}$$

Remark 2. Actually, we formulate Theorem 16 in the form of Bloom–Sisask [2]. Kelley–Meka [12] obtained this result without $\mathcal{L}^4(\varepsilon)$ in (46).

We generalize Theorem 16 for the higher energies E^k_l.

Theorem 17. *Let $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$ be a set, $|A| = \delta N$, and $\varepsilon \in (0, 1]$ be a parameter. Then there is a subspace $V \subseteq \mathbf{G}$ and $x \in \mathbf{G}$ such that $A \cap (V + x)$ is ε–uniform relatively to E^k_l, $\mu_{V+x}(A) \geqslant \delta$ and*

$$\mathrm{codim} V \ll \varepsilon^{-28 l^l} (8l)^{28 l^l} k^4 \mathcal{L}^{4l}(\varepsilon)\mathcal{L}^{5l}(\delta)\,. \tag{47}$$

Remark 3. Using the second part of Example 12 one can easily see that the quantity k is necessary in estimates (46), (47). Indeed, just delete from $\mathbb{F}_2^n$ a subspace H of density 2^{-k}. Then the obtained set is not E^k_2–uniform and to find a uniform piece we need k dimensions.

Now we are ready to obtain our driving result about the density increment. As always we will apply Proposition 18 in an iterative way and we see that estimate (50) allows us to do it in at most $O(\varepsilon^{-1}\mathcal{L}(\delta))$ times.

Proposition 18. *Let $\mathbf{G} = \mathbb{F}_p^n$, $A \subseteq \mathbf{G}$, $|A| = \delta N$, $\varepsilon > 0$ be a real number and $k, l \geqslant 2$ be positive integers, $kl \gg \varepsilon^{-1}\mathcal{L}(\varepsilon)$. Suppose that*

$$\mathsf{E}_l^k(A) \geqslant (1+\varepsilon)^{kl}\delta^{kl}N^{k+l}\,, \tag{48}$$

and that A is $\varepsilon/5$–uniform relatively to $\mathsf{E}_{l-1}^{k_*}$ for an even $k_* = O(kl\mathcal{L}(\delta))$. Then there is a subspace $V \subseteq \mathbf{G}$ such that

$$\mathrm{codim} V \ll \varepsilon^{-2} l^3 k^4 \mathcal{L}^2(\delta)\mathcal{L}^2(\varepsilon\delta)\,, \tag{49}$$

and for a certain $x \in \mathbf{G}$ one has

$$|A \cap (V + x)| \geqslant (1 + \varepsilon/8)\delta|V|\,. \tag{50}$$

Proof. Applying Lemma 14 for the energy $\mathsf{E}_l^k(A)$ with the parameters $\epsilon = \varepsilon$, $\eta = \varepsilon/30$, we construct the set

$$S = \{|x| = l \ : \ \mathcal{C}_l(A)(x) \geqslant (1+\varepsilon/4)^l\delta^l N\}$$

and such that for a certain set $B \subseteq \mathbf{G}$, $|B| > 2^{-1/(l-1)}(1+\varepsilon)^k\delta^k N := \beta N$ the following holds

$$N^{-1}\sum_{|x|=l} S(x)\mathcal{C}_l(B)(x) \geqslant (1-2\eta)|B|^l\,. \tag{51}$$

We have $kl \geqslant 4\varepsilon^{-1}\mathcal{L}(\eta)$ and thus Lemma 14 can be applied indeed. Using formulae (11) and making the required change of the variables (in (9) we put, consequently, $z \to z - x_1$) one can see that (51) is equivalent to

$$\sum_{|x|=l-1} \bar{S}(x)(B^{l-1} \circ \mathcal{D}_{l-1}(B))(x) \geqslant (1-2\eta)|B|^l\,, \tag{52}$$

where $\bar{S} \subseteq \mathbf{G}^{l-1}$ is a certain set which is constructed via the set S, see formulae (11), (13). Now we apply Lemma 15 with $f = \bar{S}$, $B = B$, $\epsilon = \eta$, and $l = l-1$. By this result and inequality (52) we find a subspace $V \subseteq \mathbf{G}$ such that the co–dimension of V is controlled by estimate (40) and

$$\sum_{|x|=l-1} \bar{S}(x)(B^{l-1} \circ \mathcal{D}_{l-1}(B * \mu_V))(x) \geqslant (1-3\eta)|B|^l\,. \tag{53}$$

By the definition of the set S (and hence $\bar{S}$), we have $\mathcal{C}_l(A)(x) \geqslant (1+\varepsilon/4)^l \delta^l N$ for any $x \in \bar{S}$ and hence inequality (53) gives us

$$(1-3\eta)(1+\varepsilon/4)^l \delta^l |B|^l N \leqslant \sum_{|x|=l-1} (A^{l-1} \circ \mathcal{D}_{l-1}(A))(x)(B^{l-1} \circ \mathcal{D}_{l-1}(B * \mu_V))(x)$$

$$= \sum_\alpha (A*B*\mu_V)(\alpha) \sum_{|x|=l-1} A^{l-1}(x) B^{l-1}(x+\mathcal{D}_{l-1}(\alpha)) = \sum_\alpha (A*B*\mu_V)(\alpha)(A \circ B)^{l-1}(\alpha)$$

$$\leqslant |B| \|A * \mu_V\|_\infty \sum_\alpha (A \circ B)^{l-1}(\alpha)\,. \qquad (54)$$

Now by our assumption A is $\varepsilon/5$–uniform relatively to $\mathsf{E}^{k_*}_{l-1}$ and a certain even $k_* = O(kl\mathcal{L}(\delta))$. Using Corollary 10, we derive

$$\|A * \mu_V\|_\infty \geqslant \delta(1-3\eta)(1+\varepsilon/4) \geqslant \delta(1+\varepsilon/8)\,.$$

Finally, thanks to (40), we get

$$\mathrm{codim} V \ll \varepsilon^{-2} l \mathcal{L}^2(\beta) \mathcal{L}^2(\varepsilon\beta^l) \ll \varepsilon^{-2} l^3 k^4 \mathcal{L}^2(\delta) \mathcal{L}^2(\varepsilon\delta)\,. \qquad (55)$$

This completes the proof. □

Remark 4. Formula (52) of Proposition 18 shows that Corollary 10 is a **criterium** and we can say, roughly, that for sufficiently large even k (depending on l and on $|N|/|B|$) one has

$$\|f_A\|^{kl}_{\mathsf{E}^k_l} = o(\delta^{kl} N^{k+l}) \qquad \text{iff} \qquad \forall m \in [l],\ \forall B \ :\ \sum_x (f_A \circ B)^m(x) = o(\delta^m |B|^m N)\,. \ (56)$$

Alternatively, one can see this criterium in a slightly more direct way, namely, from the second formula of (15). Let us emphasize one more time that crucial point of the second formula in (56) is the correct dependence on $|B|$ and on δ.

Now we can prove our new Theorem 17.

As always the proof follows the density increment scheme and our aim is to construct a shift of a subspace $V_\varepsilon(l,k)$ where the set A is ε–uniform relatively to E^k_l. Also, to obtain Theorem 17 we use induction on parameter $l \geqslant 2$ and the first step of the induction for $l=2$ and an arbitrary k one can use either Kelley–Meka Theorem 16 or the arguments of Section 3 (one can check that we do not need any uniformity conditions in this case), combining with Proposition 18. Now let $l \geqslant 3$ and suppose that the set A is not ε–uniform relatively to E^k_l for a certain k because otherwise there is nothing to prove. Of course (see, e.g., inequality (19)), one can take k to be a sufficiently large number and we choose k such that $k \gg l\varepsilon^{-l}\mathcal{L}(\varepsilon)$. Put $\varepsilon_l = \frac{\varepsilon^l}{8l}$. We can freely assume that our set A is $\varepsilon_l/5$–uniform relatively to $\mathsf{E}^{k_*}_{l-1}$ with $k_* = k_*(l) = O(kl^2\varepsilon^{-l}\mathcal{L}(\varepsilon)\mathcal{L}(\delta))$ in a shift of a subspace $V_{\varepsilon_l/5}(l-1, k_*(l))$ thanks to Theorem 16 in the case $l = 3$ or by the induction assumption for larger l. Now we apply Lemma 13 and find $k_1 = O(kl\varepsilon^{-l}\mathcal{L}(\varepsilon))$ such that

$$\mathsf{E}_l^{k_1}(A) \geqslant (1+\varepsilon_l)^{lk_1} \delta^{lk_1} N^{l+k_1} .$$

After that we use Proposition 18 with $\varepsilon = \varepsilon_l$. One can check that $kl \gg \varepsilon_l^{-1}\mathcal{L}(\varepsilon_l)$ and that A is sufficiently uniform set (to apply our proposition) relatively to $\mathsf{E}_{l-1}^{k_*}$ thanks to our choice of k_*. Estimate (50) implies that the procedure must stop after $O(\varepsilon_l^{-1}\mathcal{L}(\delta))$ number of steps and thus the final co–dimension is

$$\begin{aligned}\mathrm{codim} V_\varepsilon(l,k) &\ll \varepsilon_l^{-1}\mathcal{L}(\delta)\mathrm{codim} V_{\varepsilon_l/5}(l-1, k_*(l)) \\ &\ll l\varepsilon^{-l}\mathcal{L}(\delta)\mathrm{codim} V_{\varepsilon_l/5}(l-1, kl^2\varepsilon^{-l}\mathcal{L}(\varepsilon)\mathcal{L}(\delta)) .\end{aligned}$$

Put $L = l!$. Solving the functional inequality above and using (46), we get

$$\begin{aligned}\mathrm{codim} V_\varepsilon(l,k) &\ll L\varepsilon^{-l(l-1)/2}\mathcal{L}^{l-2}(\delta)\mathrm{codim} V_{(\varepsilon/8l)^{2L}}(2, k(l!)^2\varepsilon^{-l(l-1)/2}(\mathcal{L}(\varepsilon)\mathcal{L}(\delta))^{l-2}) \\ &\ll \varepsilon^{-28l^l}(8l)^{28l^l}k^4\mathcal{L}^{4l}(\varepsilon)\mathcal{L}^{5l}(\delta) \end{aligned} \tag{57}$$

as required. Actually, one can see that the number of steps of our algorithm is at most $O(\varepsilon_l^{-1}\mathcal{L}(\delta))$ (due to every time we increase the density δ to $\delta(1+\varepsilon_l/8)$ and hence we do not need the first multiple in (57)). Nevertheless, it gives us a bound of the same sort. This concludes the proof of the theorem. □

5. An application to two–dimensional corners

In this section we consider the simplest two–dimensional generalization of arithmetic progressions of length three. Namely, let $\mathbf{G}$ be an abelian group. Any triple of the form

$$\{(x,y), (x+d,y), (x,y+d)\} \in (\mathbf{G}\times\mathbf{G})^3$$

is called a *corner*, see [6], [9]. If $d \neq 0$, then we say that our corner is a *non–trivial* one. Using projection, one can see that if a two–dimensional set contains a corner, then its projection contains an AP3 (see [6] or [23]). Thus, this problem does indeed generalize the question of the upper bound of $r_3(N)$.

Further let $A \subseteq S_1 \times S_2 \subseteq \mathbf{G}\times\mathbf{G}$, $|A| = \delta|S_1||S_2|$, $|S_1| = \sigma_1 N$, $|S_2| = \sigma_2 N$. We say that A is *ε–uniform relatively to the rectangular norm* if

$$\|f_A\|_\square^4 := \sum_{x,y}\left|\sum_y f_A(x,y)f_A(x',y)\right|^2 \leqslant \varepsilon^4\delta^4|S_1|^2|S_2|^2 .$$

Let us obtain a counting result on the number of corners in uniform sets, see [11], [23]. Namely, we show that E_4^k–norm controls the number of corners in any uniform set A.

Theorem 19. *Let $\mathbf{G} = \mathbb{F}_2^n$ and $A \subseteq S_1 \times S_2 \subseteq \mathbf{G} \times \mathbf{G}$, $|A| = \delta|S_1||S_2|$, $|S_1| = \sigma_1 N$, $|S_2| = \sigma_2 N$. Suppose that A is η–uniform relatively to the rectangular norm, $\eta \leqslant \delta^{3/2}/4$, and S_1, S_2 are ε–uniform relatively to E_4^k with $\varepsilon = 2^{-12}\eta^4\delta^2$ and $k = O(\mathcal{L}(\delta\sigma_1\sigma_2))$. Then A contains at least $2^{-1}\delta^3\sigma_1^2\sigma_2^2N^3$ corners. In particular, if $N > 2\delta^{-2}\sigma_1^{-1}\sigma_2^{-1}$, then A has a non–trivial corner.*

Proof. Let f_1, f_2, f_3 be three arbitrary functions on $\mathbf{G}\times\mathbf{G} := \mathcal{P}$. Consider the functional

$$T(f_1, f_2, f_3) = \sum_{x,y,z} f_1(x,y)f_2(y+z,y)f_3(x,x+z)\,.$$

It is clear that T is linear in each of the arguments. Moreover, the value $T(A,A,A)$ is equal to the number of triples $\{(x,y),(x+d,y),(x,y+d)\}$ in A (here and below we use the fact that our group is $\mathbb{F}_2^n$). We have $T(A,A,A) = \delta T(\mathcal{P},A,A) + T(f_A,A,A)$ and $\|f_A\|_\square^4 \leqslant \eta^4\delta^4\sigma_1^2\sigma_2^2N^4$. Below we write $f = f_A$ for brevity.

Let $g(z) = \sum_x A(x,x+z)$. Then $T(\mathcal{P},A,A) = \sum_z g(z)^2$. We have $\sum_z g(z) = \delta\sigma_1\sigma_2N^2$ and thus by the Hölder inequality,

$$T(A,A,A) \geqslant \delta^3\sigma_1^2\sigma_2^2N^3 + T(f,A,A)\,. \tag{58}$$

Let us estimate the second term on the right-hand side of (58). Using once again the Hölder inequality, we see that

$$T(f,A,A) = \sum_{y,z} A(y+z,y)\sum_x S_1(y+z)f(x,y)A(x,x+z)$$

$$\leqslant \left(\sum_{y,z} A(y+z,y)\right)^{1/2} \times \left(\sum_{x,x',y,z} S_1(y+z)A(x,x+z)A(x',x'+z)f(x,y)f(x',y)\right)^{1/2}. \tag{59}$$

We have $\sum_{y,z} A(y+z,y) = \delta\sigma_1\sigma_2N^2$. Further,

$$\begin{aligned}\sigma &:= \sum_{x,x',y,z} S_1(y+z)A(x,x+z)A\left(x',x'+z\right)f(x,y)f\left(x',y\right)\\ &= \sum_{x,x',z} A(x,x+z)A\left(x',x'+z\right)\sum_y S_1(y+z)S_2(x+z)S_2\left(x'+z\right)f(x,y)f\left(x',y\right)\end{aligned}$$

Let

$$\begin{aligned}\omega\left(x,x',y,y'\right) &= \sum_z S_1(y+z)S_1\left(y'+z\right)S_2(x+z)S_2\left(x'+z\right)\\ &= \mathcal{C}_4(S_2,S_2,S_1,S_1)(x,x',y,y')\,.\end{aligned}$$

A third use of the Hölder inequality gives us

$$\sigma \leqslant \left(\sum_{x,x',z} S_1(x) S_1(x') S_2(x+z) S_2(x'+z) \right)^{1/2}$$

$$\times \left(\sum_{x,x',y,y'} \omega(x,x',y,y') f(x,y) f(x',y) f(x,y') f(x',y') \right)^{1/2} .$$

Applying Corollary 10 with $l = 2$, $A = S_1$, $B = S_2$, we get

$$\sum_{x,x',z} S_1(x) S_1(x') S_2(x+z) S_2(x'+z) = \sum_z (S_1 \circ S_2)^2(z) \leqslant 2\sigma_1^2 \sigma_2^2 N^3$$

due to $k \gg \mathcal{L}(\sigma_2)$ and $\varepsilon \leqslant 1/4$. Further applying the obtained inequalities, we derive

$$T(f,A,A)^4 \leqslant 2\delta^2 \sigma_1^4 \sigma_2^4 N^7 \sum_{x,x',y,y'} \omega(x,x',y,y') f(x,y) f(x',y) f(x,y') f(x',y')$$

$$= 2\delta^2 \sigma_1^6 \sigma_2^6 N^8 \|f\|_\square^4$$
$$+ 2\delta^2 \sigma_1^4 \sigma_2^4 N^7 \sum_{x,x',y,y'} (\omega(x,x',y,y') - \sigma_1^2\sigma_2^2 N) f(x,y) f(x',y) f(x,y') f(x',y')$$

$$\leqslant 2\eta^4 \delta^6 \sigma_1^8 \sigma_2^8 N^{12} + 2\delta^2 \sigma_1^4 \sigma_2^4 N^7 S_* \,. \tag{60}$$

Here we have used the assumption that the set A is η–uniform relatively to the rectangular norm. To estimate the sum S_* we apply Lemma 11 with $l = 4$, combining with the Hölder inequality to obtain

$$S_* \leqslant 2^{12} \varepsilon (\sigma_1^{2k} \sigma_2^{2k} N^{4+k})^{1/k} (4\delta^2 \sigma_1^2 \sigma_2^2 N^4)^{1-1/k} \leqslant 2^{15} \varepsilon \delta^2 \sigma_1^4 \sigma_2^4 N^5$$

due to $k = C\mathcal{L}(\delta\sigma_1\sigma_2)$, where $C > 1$ is a sufficiently large absolute constant. Returning to (60) and recalling that $\varepsilon = 2^{-12}\eta^4\delta^2$, we get

$$T(f,A,A)^4 \leqslant 2\eta^4 \delta^6 \sigma_1^8 \sigma_2^8 N^{12} + 2^{15} \varepsilon \delta^4 \sigma_1^8 \sigma_2^8 N^{12} \leqslant 16 \eta^4 \delta^6 \sigma_1^8 \sigma_2^8 N^{12} \,.$$

The last bound and (58) give us

$$T(A,A,A) \geqslant \sigma_1^2 \sigma_2^2 N^3 (\delta^3 - 2\eta\delta^{3/2}) \geqslant 2^{-1} \delta^3 \sigma_1^2 \sigma_2^2 N^3$$

as required. $\square$

6. Appendix

6.1. Possible generalizations of the Kelley–Meka method for longer progressions

In this section we will show why the Kelly-Meka method does not provide anything new for arithmetic progressions of length four and longer. To have deal with arithmetic progressions of length more than three we need a generalization of norms (14). Having vectors $x = (x_1, \dots, x_s) \in \mathbf{G}^s$ and $y = (y_1, \dots, y_t) \in \mathbf{G}^t$ (recall that we code this fact as $|x| = s$ and $|y| = t$) define its "Minkowski" sum as $x \oplus y \in \mathbf{G}^{st}$, where the components of $x \oplus y$ are $x_i + y_j$, $i \in [s]$, $j \in [t]$ (and similarly for higher sums). Given a function $f : \mathbf{G} \to \mathbb{C}$ put

$$\mathcal{E}^k_{s,t}(f) = \sum_{|x|=s} \sum_{|y|=t} \mathcal{C}^k_{st}(f)(x \oplus y) = \sum_{|x|=s} \sum_{|y|=t} \sum_{|z|=k} \mathcal{P}_{stk}(f)(x \oplus y \oplus z)\,. \tag{61}$$

In these terms

$$\mathsf{E}^k_l(f) = \sum_{|x|=l} \sum_{|y|=k} \mathcal{P}_{kl}(f)(x \oplus y)\,. \tag{62}$$

For even k, s, t and a real function f one has $\mathcal{E}^k_{s,t}(f) \geqslant \mathsf{E}^k_s(f), \mathsf{E}^k_t(f), \mathsf{E}^s_t(f) \geqslant 0$ and the triangle inequality for $\mathcal{E}^k_{s,t}$ can be obtained exactly as in [25, Appendix] or just see Subsection 6.2 (of course one needs an additional application of the Hölder inequality due to we have the longer sum in (61) than in (62)). Thus $(\mathcal{E}^k_{s,t}(f))^{1/kst}$ defines a norm of $f : \mathbf{G} \to \mathbb{R}$ in the case of even k, s, t. Notice that similar to $\mathsf{E}^k_l(f)$ the quantity $\mathcal{E}^k_{s,t}(f)$ is non–negative, provided at least one of the numbers k, s, t is even but, nevertheless, it is not always a norm in this case, see [25, Section 4]. By some symmetricity reasons (see, e.g., formulae (67) below) we make a normalization and put

$$\|f\|_{\mathcal{E}^k_{s,t}} := (|\mathbf{G}|^{-2} \mathcal{E}^k_{s,t}(f))^{1/kst} := (\bar{\mathcal{E}}^k_{s,t}(f))^{1/kst}$$

for $f : \mathbf{G} \to \mathbb{R}$. Clearly, one has

$$\mathcal{E}^k_{s,t}(f) = \sum_{|x|=s} \sum_{|z|=k} \mathcal{C}^t_s(f_z)(x) = \sum_{|x|=s} \sum_{|z|=k} \mathcal{C}^t_{sk}(f)(x \oplus z)\,, \tag{63}$$

and

$$\mathcal{E}^k_{s,t}(f) = \sum_{|y|=t} \sum_{|z|=k} \mathcal{C}^s_t(f_z)(y) = \sum_{|y|=t} \sum_{|z|=k} \mathcal{C}^s_{tk}(f)(y \oplus z)\,. \tag{64}$$

Thus we have the duality relation similar to formula (14)

$$\mathcal{E}^k_{t,s}(f) = \mathcal{E}^k_{s,t}(f) = \mathcal{E}^t_{s,k}(f) = \mathcal{E}^s_{t,k}(f)\,. \tag{65}$$

Also, let us remark that the expectations over $x \oplus y$ of the generalized convolution of any real function $f : \mathbf{G} \to \mathbb{R}$ is connected with the higher energies

$$\sum_{|x|=s} \sum_{|y|=t} \mathcal{C}_{st}(f)(x \oplus y) = N\mathsf{E}^t_s(f)\,. \tag{66}$$

In particular, the expectation above is always non–negative if s or t is an even number and we see immediately that the duality (14) takes place. Formula (66) can be proved directly or it follows from (63), (64) and the fact that $\mathcal{E}^k_{t,1}(f) = N\mathsf{E}^k_t(f)$. Finally, notice that in contrast to $\mathcal{C}_l(x)$ the function $\mathcal{C}_{st}(x \oplus y)$ enjoys even two symmetries, namely,

$$\mathcal{C}_{st}(f)(x \oplus y) = \mathcal{C}_{st}(f)((x + \mathcal{D}_s(w_1)) \oplus y) = \mathcal{C}_{st}(f)(x \oplus (y + \mathcal{D}_t(w_2))) \tag{67}$$

for any $w_1, w_2 \in \mathbf{G}$. It gives, in particular,

$$\mathcal{C}_{st}(f)(x \oplus y) = \mathcal{C}_{st}(f)((x - \mathcal{D}_s(x_1)) \oplus (y - \mathcal{D}_t(y_1))) = N^2\mathcal{C}'_{st}(f)(w)\,, \tag{68}$$

where $|w| = st - 1$ and, more concretely, $w_{ij} = (x_i - x_1) + (y_j - y_1)$, $i \in [s]$, $j \in [t]$ and $(i,j) \neq (1,1)$.

Now we are ready to obtain our counting lemma. Let us write $L(x,y) = \alpha x + \beta y + \gamma$ for a non–trivial linear form. We say that two forms are non–proportional if their coefficients are not proportional. Given a real number $q > 1$ put $q^* = \frac{q}{q-1}$.

Theorem 20. *Let N be a prime number and $k = 4$, $l_1, l_2 \geqslant 2$ be positive integers. Also, let $f_1, \dots, f_k : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be functions and $L_1, \dots, L_k$ be non–proportional linear forms such that $L_2, \dots, L_k$ depend on both variables. Then*

$$\left| \sum_{x,y} f_1(L_1(x,y)) \dots f_k(L_k(x,y)) \right| \leqslant \|f_1\|_{l_1^*} \|f_2\|_{l_2^*} \|f_3\|_{\mathcal{E}^2_{l_1,l_2}} \|f_4\|_{\mathcal{E}^2_{l_1,l_2}}\,. \tag{69}$$

Proof. Let σ be the left–hand side of (69). Without loss of generality one can assume that $L_j(x,y) = \alpha_j x + \beta_j y$, $j \in [k]$. Consider the nonzero form L_1 and suppose for concreteness that $\alpha_1 \neq 0$. Changing the variables $\alpha_1 x + \beta_1 y \to x$, we obtain

$$\sigma = \sum_{x,y} f_1(x) f_2(\tilde{L}_2(x,y)) \dots f_k(\tilde{L}_k(x,y))\,, \tag{70}$$

where here and below we write $\tilde{L}_j(x,y) = L_j(x,y) = \alpha_j x + \beta_j y$ and the coefficients α_j, β_j may change from line to line. Anyway one can check that all new forms $\tilde{L}_2, \dots, \tilde{L}_k$ in (70) are nonzero and non–proportional. Moreover, by assumption the initial forms $L_2, \dots, L_k$ depend on both variables and we see that the new forms in (70) depend on both variables as well. Now we use the Hölder inequality and get

$$(\sigma/\|f_1\|_{l_1^*})^{l_1} \leqslant \sum_x \left(\sum_y f_2(L_2(x,y)) \dots f_k(L_k(x,y)) \right)^{l_1}$$
$$= \sum_{x,y} \mathcal{P}_{l_1}(f_2)(L_2(\mathcal{D}_{l_1}(x), \mathcal{P}_{l_1}(y))) \dots \mathcal{P}_{l_1}(f_k)(L_k(\mathcal{D}_{l_1}(x), \mathcal{P}_{l_1}(y))) \,.$$

Notice that we have decreased the number of our linear forms (but increased the number of variables). Now let us make the changing of the variables similar to above, namely, $\alpha_1 \mathcal{D}_{l_1}(x) + \beta_1 \mathcal{P}_{l_1}(y) \to \mathcal{P}_{l_1}(y)$ and again one can easily check that we preserve all conditions on our linear forms $L_3, \dots, L_k$. Thus one has

$$(\sigma/\|f_1\|_{l_1^*})^{l_1}$$
$$\leqslant \sum_y \mathcal{P}_{l_1}(f_2)(\mathcal{P}_{l_1}(y)) \sum_x \mathcal{P}_{l_1}(f_3)(L_3(\mathcal{D}_{l_1}(x), \mathcal{P}_{l_1}(y))) \dots \mathcal{P}_{l_1}(f_k)(L_k(\mathcal{D}_{l_1}(x), \mathcal{P}_{l_1}(y)))$$

and using the Hölder inequality one more time, as well as the obvious identity

$$\left(\sum_y \mathcal{P}_{l_1}^{l_2^*}(f_2)(\mathcal{P}_{l_1}(y)) \right)^{l_2 - 1} = \|f_2\|_{l_2^*}^{l_1 l_2} \,, \qquad (71)$$

we derive recalling that $k = 4$

$$(\sigma/\|f_1\|_{l_1^*} \|f_2\|_{l_2^*})^{l_1 l_2}$$
$$\leqslant \sum_{x,y} \mathcal{P}_{l_1 l_2}(f_3)(L_3(\mathcal{P}_{l_2}\mathcal{D}_{l_1}(x), \mathcal{D}_{l_2}\mathcal{P}_{l_1}(y))) \mathcal{P}_{l_1 l_2}(f_4)(L_4(\mathcal{P}_{l_2}\mathcal{D}_{l_1}(x), \mathcal{D}_{l_2}\mathcal{P}_{l_1}(y))) \,. \qquad (72)$$

Now let us analyze the right–hand side of formula (72). First of all, it is easy to see that there are l_2 different variables x_i and l_1 different variables y_j in (72). Secondly, take the form L_{k-1} (for L_k the argument is the same) and notice that it depends on $\alpha_{k-1} x_i + \beta_{k-1} y_j$, $i \in [l_2]$, $j \in [l_1]$ and that every such expression appears exactly once. Now introducing two more variables z, w such that $x_i \to x_i + z$, $y_j \to y_j + w$ and then replacing z, w to other variables Z, W, where $Z = \alpha_{k-1} z + \beta_{k-1} w$, $W = \alpha_k z + \beta_k w$ (this change of the variables is allowable because the forms L_{k-1}, L_k are not proportional), we arrive to the quantities $\mathcal{C}_{l_1 l_2}(f_{k-1})$, $\mathcal{C}_{l_1 l_2}(f_k)$ in (72). Writing $x = (x_1, \dots, x_{l_2})$, $y = (y_1, \dots, y_{l_1})$, we have finally

$$(\sigma/\|f_1\|_{l_1^*} \|f_2\|_{l_2^*})^{l_1 l_2} \leqslant N^{-2} \sum_{\vec{x}, \vec{y}} \mathcal{C}_{l_1 l_2}(f_{k-1})(\alpha_{k-1} \cdot x \oplus \beta_{k-1} \cdot y) \mathcal{C}_{l_1 l_2}(f_k)(\alpha_k \cdot x \oplus \beta_k \cdot y) \,.$$

Using the Hölder inequality the last time, as well as the fact that $\alpha_{k-1}, \alpha_k, \beta_{k-1}, \beta_k \neq 0$, we obtain

$$(\sigma/\|f_1\|_{l_1^*} \|f_2\|_{l_2^*})^{l_1 l_2} \leqslant \left(N^{-2} \sum_{|x| = l_2,\, |y| = l_1} \mathcal{C}_{l_1 l_2}^2(f_{k-1})(x \oplus y) \right)^{1/2}$$

$$\times \left(N^{-2} \sum_{|x|=l_2,\, |y|=l_1} \mathcal{C}^2_{l_1 l_2}(f_k)(x \oplus y) \right)^{1/2} = \|f_{k-1}\|^{l_1 l_2}_{\mathcal{E}^2_{l_1,l_2}} \cdot \|f_k\|^{l_1 l_2}_{\mathcal{E}^2_{l_1,l_2}}$$

as required. □

Remark 5. One can check that for any l_1, l_2 one has

$$\frac{1}{l_1^*} + \frac{1}{l_2^*} + \frac{1}{l_1} + \frac{1}{l_2} = 2$$

and hence the right–hand side of bound (69) has the correct order in N. Similarly, taking $f_j(x) = f_A(x)$, $j \in [4]$ as the balanced function of a set A and $l_1 \sim l_2 \sim \mathcal{L}(\delta)$ we see that the dependence on δ is also correct.

Remark 6. As we have said in the previous remark the optimal dependence on the parameters l_1, l_2 in Theorem 20 is $l_1 \sim l_2 \sim \mathcal{L}(\delta)$. Suppose that the dependence on ε in Lemma 13 and in all statements below is almost optimal, say, $c\varepsilon$ for a constant $c \in (0,1)$. Thanks to the induction scheme of the proof, it gives us the multiple $c^{\mathcal{L}(\delta)} = \delta^{-C}$ for a certain $C > 0$ in codimension of the subspace V, where our set A is uniform. But δ^{-C} is more or less that usual Gowers' method gives to us and hence we have no special gain. Thus, even at a technical level, the extension of the Kelley-Meka method to more complex objects than arithmetic progressions of length three is fraught with significant difficulties.

We conclude this part of the appendix showing that the convolutions $\mathcal{C}_{st}(f)(x \oplus y)$, $|x| = s$, $|y| = t$ enjoy the almost periodicity properties similar to the ordinary convolutions $\mathcal{C}_s(f)(x)$. Given a vector $x = (x_1, \dots, x_r)$ let us write for convenience $\bar{x} = (0, x_1, \dots, x_r)$.

Lemma 21. *Let $\varepsilon \in (0,1]$ be a real number, $s,t,q \geqslant 2$ be positive integers, $l := st - 1$, $B \subseteq \mathbf{G}$, $|B| = \beta N$ and $F : \mathbf{G}^l \to \mathbb{R}$. Then there is a set $T \subseteq \mathbf{G}$, $|T| \geqslant |B| \exp(-O(\varepsilon^{-2} q \log(1/\beta)))$ and such that for any $t \in T$ one has*

$$\sum_{|x|=s-1} \sum_{|y|=t-1} |(F \circ \mathcal{D}_l(B+t))(\bar{x} \oplus \bar{y}) - (F \circ \mathcal{D}_l(B))(\bar{x} \oplus \bar{y})|^q$$
$$\leqslant \varepsilon^q |B|^{q-1} \sum_{|y|=t-1} \mathcal{C}'_t(|F|^q)^{s-1}(y) \cdot \mathcal{C}'_t(B, |F|^q, \dots, |F|^q)(y)\,. \tag{73}$$

Proof. We choose $k = O(\varepsilon^{-2} q)$ random points $b_1, \dots, b_k \in B$ uniformly and independently and let $Z_j((\bar{x} \oplus \bar{y}) = F((\bar{x} \oplus \bar{y}) + \mathcal{D}_l(b_j)) - (F \circ \mathcal{D}_l(\mu_B))(\bar{x} \oplus \bar{y})$. Clearly, the random variables Z_j are independent, have zero expectation and their variances do not exceed $(|F|^2 \circ \mathcal{D}_l(\mu_B))(\bar{x} \oplus \bar{y})$. By the Khintchine inequality for sums of independent random variables,

$$\|\sum_{j=1}^{k} Z_j(\bar{x}\oplus\bar{y})\|_{L_p(\mu_B^k)} \ll (|F|^2\circ\mathcal{D}_l(\mu_B))(\bar{x}\oplus\bar{y})^{1/2}.$$

Raising the last inequality to the power q, dividing by k^q, summing over $\bar{x}\oplus\bar{y}$, and using the Hölder inequality, which gives $(|F|^2\circ\mathcal{D}_l(\mu_B))(\bar{x}\oplus\bar{y})^{q/2} \leqslant (|F|^q\circ\mathcal{D}_l(\mu_B))(\bar{x}\oplus\bar{y})$, we get that

$$\sum_{|x|=s-1}\sum_{|y|=t-1}\int\left|\frac{1}{k}\sum_{j=1}^{k} F((\bar{x}\oplus\bar{y})+\mathcal{D}_l(b_j)) - (F\circ\mathcal{D}_l(\mu_B))(\bar{x}\oplus\bar{y})\right|^q d\mu_B^k(x_1,\dots,x_k)$$
$$\ll (qk^{-1})^{q/2}\sum_{|x|=s-1}\sum_{|y|=t-1}(|F|^q\circ\mathcal{D}_l(\mu_B))(\bar{x}\oplus\bar{y})$$
$$= (qk^{-1})^{q/2}|B|^{-1}\sum_{|y|=t-1}\mathcal{C}'_t(|F|^q)^{s-1}(y)\cdot\mathcal{C}'_t(B,|F|^q,\dots,|F|^q)(y).$$

After that we repeat the argument from [4], [17], [18] and [24, Theorem 15]. This completes the proof. □

6.2. On a family of norms

In this section we define a very general family of norms, which includes the norms E^k_l, $\mathcal{E}^k_{s,t}$ above, as well as the classical Gowers norms [9]. As the reader can see we do not use the Fourier transform in our proofs below.

Let $\mathbf{G}$ be an abelian group, $r, k_1,\dots,k_r \geqslant 2$ be integers and $f:\mathbf{G}\to\mathbb{R}$ be an arbitrary function. Let $K=\prod_{j=1}^r k_j$, $B=[k_1]\times\dots\times[k_r]$ and write $x_1=(x_1^{(1)},\dots,x_1^{(k_1)}),\dots,x_r=(x_r^{(1)},\dots,x_r^{(k_r)})$. Also, for $\omega\in B$ we write $\omega=(\omega_1,\dots,\omega_r)$. Define

$$\|f\|^K_{E_{k_1,\dots,k_r}} = \sum_{|x_1|=k_1}\cdots\sum_{|x_r|=k_r}\mathcal{P}_K(f)(x_1\oplus\dots\oplus x_r)$$
$$= \sum_{|x_1|=k_1}\cdots\sum_{|x_r|=k_r}\prod_{\omega\in B} f(x_1^{(\omega_1)}+\dots+x_r^{(\omega_r)}). \qquad (74)$$

The case $r=2$ corresponds to E^k_l–norm, $r=3$ is just $\mathcal{E}^k_{s,t}$–norms and for $k_1=\dots=k_r=2$, we obtain Gowers' U^k norms (up to some normalizations). If we choose f in (74) as the characteristic function of a set $A\subseteq\mathbf{G}$, then $\|A\|^K_{E_{k_1,\dots,k_r}}$ equals the number of complete subgraphs $\mathcal{K}_{k_1,\dots,k_r}$ in addition Cayley graphs (i.e., one has an edge between $x,y\in\mathbf{G}$ iff $x+y\in A$), in the case $r=2$ we obtain the ordinary Cayley graph. In a similar way one can define the multi–scalar product for the quantity $\|\cdot\|_{E_{k_1,\dots,k_r}}$ as was done in [9], namely, having any functions $(f^\omega)_{\omega\in B}$, we write

$$\langle f^\omega\rangle_{E_{k_1,\dots,k_r}} = \sum_{|x_1|=k_1}\cdots\sum_{|x_r|=k_r}\prod_{\omega\in B} f^\omega(x_1^{(\omega_1)}+\dots+x_r^{(\omega_r)}). \qquad (75)$$

It is easy to see that if K is an even number, then

$$\|f\|^K_{E_{k_1,\dots,k_r}} \geqslant 0\,. \tag{76}$$

Indeed, let, say, k_r be an even number, then we can write x_r as $x_r = (x'_r, x''_r)$, where $|x'_r| = |x''_r| = k_r/2$ and whence

$$\|f\|^K_{E_{k_1,\dots,k_r}} = \sum_{|x_1|=k_1} \dots \sum_{|x_{r-1}|=k_{r-1}} \left(\sum_{|x'_r|=k_r/2} \mathcal{P}_{K/2}(f)(x_1 \oplus \dots \oplus x_{r-1} \oplus x'_r) \right)^2 \geqslant 0\,.$$

Also, let us remark the inductive property of the norm $E_{k_1,\dots,k_r}$. For concreteness, we take the rth coordinate and obtain from definition (74) that

$$\|f\|^K_{E_{k_1,\dots,k_r}} = \sum_{|z|=k_r} \|f_z\|^{K/k_r}_{E_{k_1,\dots,k_{r-1}}}\,. \tag{77}$$

Let us make a simple remark concerning $E_{k_1,\dots,k_r}$–norm.

Lemma 22. *Let $f : \mathbf{G} \to \mathbb{R}$ be a function. Suppose that there is $j \in [r]$ such that k_j is even and K/k_j is also even. Then $\|f\|_{E_{k_1,\dots,k_r}} = 0$ iff $f \equiv 0$.*

Proof. Without loosing of the generality assume that $j = r$. Write

$$\|f\|^K_{E_{k_1,\dots,k_r}} = \sum_{|x_1|=k_1} \dots \sum_{|x_{r-1}|=k_{r-1}} \left(\sum_z \mathcal{P}_{K/k_r}(f)(x_1 \oplus \dots \oplus x_{r-1} \oplus z) \right)^{k_r} = 0\,.$$

Since k_r is an even number, it follows that, in particular, $\sum_z f^{K/k_r}(z) = 0$ (we have taken $x_1 = \dots = x_{r-1} = 0$ in the last formula) and hence $f \equiv 0$. This completes the proof. $\square$

Now let us show that the multi–scalar product is controlled via $E_{k_1,\dots,k_r}$–norm.

Lemma 23. *Let $r \geqslant 2$ be a positive integer, $k_1, \dots, k_r \geqslant 2$ be even integers and $f^\omega : \mathbf{G} \to \mathbb{R}$, $\omega \in B$ be any functions. Then*

$$|\langle f^\omega \rangle_{E_{k_1,\dots,k_r}}| \leqslant \prod_{\omega \in B} \|f^\omega\|_{E_{k_1,\dots,k_r}}\,. \tag{78}$$

Proof. We write

$$\langle f^\omega \rangle_{E_{k_1,\dots,k_r}} = \sum_{|x_1|=k_1} \dots \sum_{|x_{r-1}|=k_{r-1}} \left(\sum_{x_r^{(1)}} \prod_{\omega \in B,\, \omega_r=1} f^\omega(x_1^{(\omega_1)} + \dots + x_{r-1}^{(\omega_{r-1})} + x_r^{(1)}) \right)$$

$$\dots \left(\sum_{x_r^{(k_1)}} \prod_{\omega \in B,\, \omega_r = k_r} f^{\omega}(x_1^{(\omega_1)} + \dots + x_{r-1}^{(\omega_{r-1})} + x_r^{(k_r)}) \right) .$$

After that apply the Hölder inequality (here we have used the fact that k_r is an even number) and we arrive to the new k_r families of functions. Take any of them, say, $(\tilde{f}^{\omega})$, $\omega \in B$ and notice that

$$\tilde{f}^{\omega} = \tilde{f}^{\omega'} = f^{\omega}$$

for all $\omega = (\omega_1, \dots, \omega_r)$, $\omega' = (\omega'_1, \dots, \omega'_r)$ with $(\omega_1, \dots, \omega_{r-1}) = (\omega'_1, \dots, \omega'_{r-1})$. In particular, the family $(\tilde{f}^{\omega})$, $\omega \in B$ has K/k_r different functions. Now we use the same argument for all remaining variables $x_1, \dots, x_{r-1}$ subsequently changing the families (f^{ω}), $\omega \in B$. One can easily see that after all these r steps we arrive to K families consisting of single functions f^{ω}, $\omega \in B$ (just thanks to the fact that any two points of our box B can be reached by a path in the directions of the coordinate axes). This is equivalent to inequality (78) and we complete the proof. □

Finally, we are ready to obtain the main result of this section. Let us write $(k_1, \dots, k_r) \leqslant (m_1, \dots, m_t)$ if the first vector is lexigraphically smaller than the second one (i.e., $r \leqslant t$ and $k_j \leqslant m_j$, $j \in [r]$). Also, put

$$\|f\|^K_{\bar{E}_{k_1,\dots,k_r}} = N^{-(k_1+\dots+k_r)} \|f\|^K_{E_{k_1,\dots,k_r}} . \tag{79}$$

Thus for any $f : \mathbf{G} \to [-1, 1]$ one has $\|f\|_{\bar{E}_{k_1,\dots,k_r}} \leqslant 1$.

Theorem 24. *Let $r \geqslant 2$ be a positive integer, $k_1, \dots, k_r \geqslant 2$ be even integers and $f : \mathbf{G} \to \mathbb{R}$ be a function. Then formula (74) defines a norm of f. Further if $(k_1, \dots, k_r) \leqslant (m_1, \dots, m_t)$, then*

$$\|f\|_{\bar{E}_{k_1,\dots,k_r}} \leqslant \|f\|_{\bar{E}_{m_1,\dots,m_t}} . \tag{80}$$

Proof. Take two functions $f, g : \mathbf{G} \to \mathbb{R}$. In view of Lemma 23, we have

$$\|f + g\|^K_{E_{k_1,\dots,k_r}} = \langle f + g \rangle_{E_{k_1,\dots,k_r}} \leqslant \sum_{j=1}^{k} \binom{K}{j} \|f\|^j_{E_{k_1,\dots,k_r}} \|g\|^{K-j}_{E_{k_1,\dots,k_r}}$$
$$= (\|f\|_{E_{k_1,\dots,k_r}} + \|g\|_{E_{k_1,\dots,k_r}})^K$$

and we have obtained the triangle inequality for $E_{k_1,\dots,k_r}$. By estimate (76) we know that our quantity $\|f\|_{\bar{E}_{k_1,\dots,k_r}}$ is non–negative. Also, Lemma 22 guaranties that $\|f\|_{E_{k_1,\dots,k_r}} = 0$ iff $f \equiv 0$. Thus indeed formula (74) defines a norm of f.

It remains to obtain (80). Let $M = \prod_{j=1}^t m_j$, $B' = [m_1] \times \cdots \times [m_t]$, $S = \sum_{j=1}^r k_j$ and $S' = \sum_{j=1}^t m_j$. Consider the family of functions $(g^\omega)_{\omega \in B'}$ such that for $\omega \in B$ one has $g^\omega = f^\omega$ and let $g^\omega \equiv 1$ otherwise. It is easy to see that

$$\|f\|^K_{\bar{E}_{k_1,\dots,k_r}} N^S = \|f\|^K_{E_{k_1,\dots,k_r}} = N^{S-S'} \cdot \langle g^\omega \rangle_{E_{m_1,\dots,m_t}}$$

Using the last formula, definition (79), as well as Lemma 23, we obtain

$$\|f\|^K_{\bar{E}_{k_1,\dots,k_r}} N^S \leqslant N^{S-S'} \|f\|^K_{E_{m_1,\dots,m_t}} (N^{S'/M})^{M-K} = N^{S-S'K/M} \cdot N^{S'K/M} \|f\|^K_{\bar{E}_{m_1,\dots,m_t}}$$
$$= N^S \|f\|^K_{\bar{E}_{m_1,\dots,m_t}}$$

as required. □

Remark 7. Inspecting the proofs of Lemma 23 and Theorem 24 one can check that formula (74) defines a norm of f for **any** numbers $r \geqslant 2$, $k_1, \dots, k_r \geqslant 2$, provided f is taken from the family of non–negative functions. In the case $r = 2$ it was obtained before in [25, Propositions 16, 30].

Data availability

No data was used for the research described in the article.

References

[1] T.F. Bloom, O. Sisask, Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions, arXiv preprint, arXiv:2007.03528, 2020.
[2] T.F. Bloom, O. Sisask, The Kelley–Meka bounds for sets free of three-term arithmetic progressions, arXiv preprint, arXiv:2302.07211, 2023.
[3] M.-C. Chang, A polynomial bound in Freiman's theorem, Duke Math. J. 113 (3) (2002) 399–419.
[4] E. Croot, O. Sisask, A probabilistic technique for finding almost-periods of convolutions, Geom. Funct. Anal. 20 (2010) 1367–1396.
[5] P. Erdős, P. Turán, On some sequences of integers, J. Lond. Math. Soc. 1 (4) (1936) 261–264.
[6] H. Furstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory, vol. 10, Princeton University Press, 2014.
[7] H. Furstenberg, Y. Katznelson, D. Ornstein, The ergodic theoretical proof of Szemerédi's theorem, Bull. Am. Math. Soc. 7 (1982) 527–552.
[8] W.T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, Geom. Funct. Anal. 8 (3) (1998) 529–551, https://doi.org/10.1007/s000390050065, https://proxy.library.spbu.ru:2060/10.1007/s000390050065.
[9] W.T. Gowers, A new proof of Szemerédi's theorem, Geom. Funct. Anal. 11 (3) (2001) 465–588, https://doi.org/10.1007/s00039-001-0332-9, https://proxy.library.spbu.ru:2060/10.1007/s00039-001-0332-9.
[10] W.T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, Ann. Math. (2007) 897–946.
[11] B. Green, Finite field models in additive combinatorics, in: Bridget S. Webb (Ed.), Surveys in Combinatorics 2005, 2005, pp. 1–27.
[12] Z. Kelley, R. Meka, Strong Bounds for 3-Progressions, 2023, pp. 933–973.
[13] B. Nagle, V. Rödl, M. Schacht, The counting lemma for regular k-uniform hypergraphs, Random Struct. Algorithms 28 (2) (2006) 113–179.

[14] V. Rödl, J. Skokan, Regularity lemma for k-uniform hypergraphs, Random Struct. Algorithms 25 (1) (2004) 1–42.
[15] K.F. Roth, On certain sets of integers, J. Lond. Math. Soc. 28 (1) (1953) 104–109.
[16] T. Sanders, On Roth's theorem on progressions, Ann. Math. (2011) 619–636.
[17] T. Sanders, On the Bogolyubov–Ruzsa lemma, Anal. PDE 5 (3) (2012) 627–655.
[18] T. Sanders, The structure theory of set addition revisited, Bull. Am. Math. Soc. 50 (1) (2013) 93–127.
[19] T. Schoen, New bounds in Balog-Szemerédi-Gowers theorem, Combinatorica 35 (6) (2015) 695–701.
[20] T. Schoen, I.D. Shkredov, Higher moments of convolutions, J. Number Theory 133 (5) (2013) 1693–1737, https://doi.org/10.1016/j.jnt.2012.10.010, https://proxy.library.spbu.ru:2060/10.1016/j.jnt.2012.10.010.
[21] T. Schoen, O. Sisask, Roth's theorem for four variables and additive structures in sums of sparse sets, Forum Math. Sigma 4 (2016) e5.
[22] S. Shelah, Primitive recursive bounds for van der Waerden numbers, J. Am. Math. Soc. 1 (3) (1988) 683–697.
[23] I.D. Shkredov, Szemerédi's theorem and problems of arithmetic progressions, Usp. Mat. Nauk 61 (6(372)) (2006) 111–178, https://doi.org/10.1070/RM2006v061n06ABEH004370, https://proxy.library.spbu.ru:2060/10.1070/RM2006v061n06ABEH004370.
[24] I.D. Shkredov, Structure theorems in additive combinatorics, Usp. Mat. Nauk 70 (1(421)) (2015) 123–178, https://doi.org/10.4213/rm9647, https://proxy.library.spbu.ru:2060/10.4213/rm9647.
[25] I.D. Shkredov, Some remarks on the Balog-Wooley decomposition theorem and quantities D^+, $D^\times$, Proc. Steklov Inst. Math. 298 (1) (2017) 74–90.
[26] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, Acta Math. Hung. 20 (1–2) (1969) 89–104.
[27] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (299–345) (1975) 21.
[28] T. Tao, A variant of the hypergraph removal lemma, J. Comb. Theory, Ser. A 113 (7) (2006) 1257–1280.
[29] B.L. Van der Waerden, Beweis einer Baudetschen Vermutung, Nieuw Arch. Wiskd. 15 (1927) 212–216.