ESSENTIAL NUMBER THEORY

The Kelley–Meka bounds for sets free of three-term arithmetic progressions

Thomas F. Bloom and Olof Sisask

2023

vol. 2 no. 1

msp

msp

Essential Number Theory
Vol. 2, No. 1, 2023
https://doi.org/10.2140/ent.2023.2.15

The Kelley–Meka bounds for sets free of three-term arithmetic progressions

Thomas F. Bloom and Olof Sisask

We give a self-contained exposition of the recent remarkable result of Kelley and Meka: if $A \subseteq \{1, \dots, N\}$ has no nontrivial three-term arithmetic progressions then $|A| \leq \exp(-c(\log N)^{1/12})N$, where $c > 0$ is a constant.

Although our proof is identical to that of Kelley and Meka in all of the main ideas, we also incorporate some minor simplifications relating to Bohr sets. This eases some of the technical difficulties tackled by Kelley and Meka and widens the scope of their method. As a consequence, we improve the lower bounds for the problem of finding long arithmetic progressions in $A + A + A$, where $A \subseteq \{1, \dots, N\}$.

How large can a subset of $\{1, \dots, N\}$ be without containing a nontrivial three-term arithmetic progression $\{a, a+d, a+2d\}$? (Nontrivial here means that $d \neq 0$.) This seemingly innocuous question, asked by Erdős and Turán [1936], has led to a wealth of interesting mathematics, and has become one of the central questions in additive combinatorics. A large part of the reason for this is that the tools and techniques that have been developed to tackle it, starting with the argument of Roth [1953], have turned out to be very influential, not only in dealing with other problems in additive number theory, but also in motivating the development of tools in other areas of mathematics — particularly in harmonic analysis.

This note gives an exposition of a recent remarkable breakthrough result of Kelley and Meka [2023] concerning this question: they prove a very strong upper bound for the maximal size of sets free of three-term progressions, far smaller than any previously available.

Let $A \subseteq \{1, \dots, N\}$ be a set which does not contain any (nontrivial) three-term arithmetic progressions. Even establishing that $|A| = o(N)$ is a difficult problem, this being Roth's landmark result [1953]. Since Roth's work there has been a sequence of quantitative improvements to the upper bound, all of the form $|A| \leq N/(\log N)^C$ for some constant $C > 0$, culminating in recent work of the authors [Bloom and Sisask 2020], who showed that $C = 1 + c$ is permissible, where $c > 0$ is some tiny constant (we refer to the introduction of [loc. cit.] for further history).

MSC2020: 11B25, 11B30.
Keywords: additive combinatorics, arithmetic progressions.

Kelley and Meka's new upper bound is of a whole new order of magnitude.

Theorem 1 (Kelley–Meka). *If $A \subseteq \{1, \dots, N\}$ contains no nontrivial three-term arithmetic progressions, then*

$$|A| \leq \frac{N}{\exp(c(\log N)^{1/12})}$$

for some absolute constant $c > 0$.

The Kelley–Meka bound is a huge leap forward, and tantalisingly close to the best possible such bound. Indeed, we know that there are subsets of $\{1, \dots, N\}$ of size at least $\exp(-c(\log N)^{1/2})N$, where $c > 0$ is some universal constant, that contain no nontrivial three-term progressions. This was first proved by Behrend [1946] using a beautiful construction that uses lattice points on high-dimensional spheres; small improvements have also been established by Elkin [2011] and Green and Wolf [2010]. Getting anywhere near Behrend-style bounds for this problem has been a long-standing goal in the area, and the argument of Kelley and Meka achieves this in a beautiful and elegant way.

Our reasons for writing this note are:

(1) To explain the Kelley–Meka approach using terminology and a perspective perhaps more familiar to researchers in additive combinatorics.

(2) To provide some minor technical refinements which allow for a proof of the full Theorem 1 which involves only "classical" Bohr set techniques, rather than the ad hoc method employed in [Kelley and Meka 2023] (in particular we answer [loc. cit., Question 6.2] of whether such an approach is possible in the affirmative). This widens the scope of potential applications over the integers, and allows for integer analogues of the other main results in [loc. cit.].

It must be made clear, however, that our sole contribution is at the technical level, allowing the Bohr set machinery to run a little more smoothly; all of the main ideas are the same as in [loc. cit.].

The finite field case. As usual in this area, a simpler model case is provided by replacing $\{1, \dots, N\}$ with the vector space $\mathbb{F}_q^n$. We will present Kelley and Meka's ideas in this model setting before the general case, since the former is technically much simpler, while still containing all of the important new ideas. Kelley and Meka's argument in $\mathbb{F}_q^n$ establishes the following.

Theorem 2 (Kelley–Meka). *If q is an odd prime and $A \subseteq \mathbb{F}_q^n$ has no nontrivial three-term progressions, then $|A| \leq q^{n-cn^{1/9}}$ for some constant $c > 0$.*

The utility of Theorem 2 is as a demonstration of the proof techniques, since for the result itself a stronger bound of $|A| \leq q^{n-cn}$ is available via the polynomial

method, as shown by Ellenberg and Gijswijt [2017]. Unfortunately, however, there is no known analogue of the polynomial method for the integer problem, so achieving strong bounds for the integer problem via this method is out of reach. Kelley and Meka's proof of Theorem 2 uses no polynomial methods, and instead uses techniques from probability and Fourier analysis, which can be generalised (using classical Bohr set machinery) to the integer setting.

After presenting the Kelley–Meka argument for $\mathbb{F}_q^n$ we will generalise this, using the language of Bohr sets, to prove Theorem 1. Kelley and Meka take a different, more ad hoc route, iterating over high-dimensional progressions. This is an ingenious alternative, that may itself have further applications, but our aim here is to show that more classical existing techniques also suffice. As a consequence, we can also establish the following integer analogue of [Kelley and Meka 2023, Corollary 1.12], which finds large subspaces in $A+A+A$ for $A \subseteq \mathbb{F}_q^n$.

Theorem 3. *If $A \subseteq \{1, \dots, N\}$ has size αN, then $A+A+A$ contains an arithmetic progression of length*

$$\geq \exp(-C \log(2/\alpha)^3) N^{c/\log(2/\alpha)^9},$$

where $C, c > 0$ are constants.

For comparison, the previously best bound known was $N^{\alpha^{1+o(1)}}$, originally due to Sanders [2008]. A construction due to Freiman, Halberstam, and Ruzsa [Freiman et al. 1992] shows that no exponent better than $c/\log(2/\alpha)$ is possible.

Kelley and Meka deduce the above bounds for sets without three-term arithmetic progressions, and more besides, from a more general type of structure result. This says, roughly speaking, that for any reasonably large set A inside a finite abelian group there exists some structured set V (e.g., an affine subspace in the $\mathbb{F}_q^n$ case) such that A restricted to V is very "regular" in its additive behaviour — that is, for "most" x the number of solutions to $x = a + b$ with $a, b \in A \cap V$ is very close to the average number of solutions.

We will first state a precise version of this structural result for $\mathbb{F}_q^n$. It is convenient to introduce the notion of a normalised indicator function: if $A \subseteq G$ is nonempty with size $|A| = \alpha |G|$ then we write $\mu_A = \alpha^{-1} 1_A$. (For a set A, we write 1_A for the indicator function of A, taking the value 1 on A and 0 elsewhere.) If $V \subseteq G$ and $1 \leq p < \infty$, then we denote the $L^p(V)$ norm of $f : G \to \mathbb{C}$ by

$$\|f\|_{p(\mu_V)} = \left(\frac{1}{|V|} \sum_{x \in V} |f(x)|^p \right)^{1/p}.$$

Note, for example, that our choice of normalisation is chosen to ensure $\|\mu_A\|_{1(\mu_G)} = 1$. We will measure the aforementioned regularity of $A \subseteq V$ by bounding the $L^p(V)$

norm of the difference between the convolution

$$\mu_A * \mu_A(x) = \frac{1}{|G|} \sum_{a,b \in G} 1_{a+b=x} \mu_A(a)\mu_A(b) = \frac{|G|}{|A|^2} \sum_{a,b \in A} 1_{a+b=x}$$

and its expected value. An elementary calculation shows that, when $V \leq G$ is a subgroup and $A \subseteq V$,

$$\|\mu_A * \mu_A\|_{1(\mu_V)} = \frac{|G|}{|V|} = \|\mu_V\|_{1(\mu_V)}.$$

That is, the average of $\mu_A * \mu_A$ over V agrees with the average of μ_V itself.

Note carefully that, even though we are taking a *local* L^p norm restricted to V we are keeping μ_A and $*$ normalised relative to the *global* group G. This is a little confusing at first, but when we move from $\mathbb{F}_q^n$ to general groups it is much more convenient to keep as many definitions as possible global, rather than relativising to the local V, since in general (e.g., when we move from $\mathbb{F}_q^n$ to $\mathbb{Z}/N\mathbb{Z}$) the subgroup V will be replaced with a set that is only approximately structured.

Theorem 4 (Kelley–Meka). *Let $\epsilon > 0$. There is a constant $C = C(\epsilon) > 0$ such that the following holds. Let $G = \mathbb{F}_q^n$ for some prime q and $n \geq 1$. Let $p \geq 1$. For any nonempty $A \subseteq G$ with $|A| = \alpha|G|$ there exists a subspace $V \leq G$ with*

$$\operatorname{codim}(V) \leq Cp^4 \log(2/\alpha)^5$$

and $x \in G$ such that, if $A' = (A - x) \cap V$, then

(1) $|A'| \geq (1 - \epsilon)\alpha|V|$ *and*

(2) $\|\mu_{A'} * \mu_{A'} - \mu_V\|_{p(\mu_V)} \leq \epsilon \frac{|G|}{|V|}$.

The factor $\frac{|G|}{|V|}$ here can be thought of as a normalising/scaling factor, corresponding to the fact that the convolution is defined over G rather than over V.

Intuitively, the second item of the conclusion says, as $p \to \infty$, that $\mu_{A'} * \mu_{A'} = (1 + O(\epsilon))\mu_V$ with high probability, as x ranges uniformly over V. Therefore, when studying many problems involving the additive behaviour of A', one can replace A' with a random subset of V of the same density. This is, of course, a very powerful tool. The cost is that we have had to restrict our original set A to an affine subspace $x + V$ to find this regularity, and so having the codimension of V be as small as possible is important for quantitative applications. For example, in the deduction of Theorem 2 from this one may take p to be around $\log(2/\alpha)$ and ϵ a constant. The resulting codimension bound of $C \log(2/\alpha)^9$ then corresponds to the exponent $\frac{1}{9}$ in Theorem 2.

The general group case: Bohr sets. To prove Theorems 1 and 3 we shall use a more general version of Theorem 4, applicable to any finite abelian group (which in applications will be $\mathbb{Z}/N\mathbb{Z}$). One difficulty in even writing down the appropriate statement is working out what should play the role of subspaces for general abelian groups. This is not obvious; fortunately for us, however, this was already done by Bourgain [1999], who showed that a suitable generalisation of a subspace, for these purposes, is a *Bohr set*. A Bohr set is an approximate level set of some characters, or more precisely, a set of the shape

$$B = \mathrm{Bohr}_\nu(\Gamma) = \{x \in G : |1 - \gamma(x)| \leq \nu \text{ for all } \gamma \in \Gamma\}$$

for some $\nu \geq 0$ (known as the width) and some $\Gamma \subseteq \widehat{G}$ (known as the frequency set). Often the most important thing about the frequency set is its size $d = |\Gamma|$, which is called the rank of the Bohr set.

Background material on Bohr sets can be found in the Appendix, but the unfamiliar reader can for now think of the above Bohr set B of rank $d = |\Gamma|$ as roughly like the embedding in G of the lattice points in a d-dimensional box in $\mathbb{R}^d$ of side-length proportional to ν. Writing B_ρ for the same Bohr set but with the "width" ν replaced by $\rho \cdot \nu$, one then has the approximate closure property that $B + B_\rho \approx B$ provided ρ is small — in particular, $B + B_\rho$ is not much larger than B provided ρ is small compared to $1/d$. It turns out that this approximate additive closure property of Bohr sets of rank d can, for the purposes of this paper, be used in place of the exact rigid structure provided by subspaces of codimension d in $\mathbb{F}_q^n$ (which are indeed Bohr sets themselves, of rank d and width $\nu = 0$).

Since Bohr sets only enjoy an approximate closure property, statements involving them necessarily require more technical conditions and quantitative overhead. This is why the use of the finite field model of $\mathbb{F}_q^n$ is invaluable in this area: an idea can be tested with subspaces in a relatively clean way, and only once the idea has been proven to have real quantitative strength does the work of translating the argument to work over general Bohr sets need to begin.

The following is the generalisation of Theorem 4 required for our applications. Kelley and Meka did not prove such a statement, but speculated that this should be possible in [Kelley and Meka 2023, Footnote 9].

The reader should compare the statement to Theorem 4, and at first reading may wish to pretend that $B = B' = B_\rho$ is the same subspace $V \leq \mathbb{F}_q^n$. For comparison to the conclusion of Theorem 4 it may help to note that, when B is a subspace and $A' \subseteq B$, then

$$\mu_B * \mu_B = \mu_B = \mu_{A'} * \mu_B,$$

and so

$$(\mu_{A'} - \mu_B) * (\mu_{A'} - \mu_B) = \mu_{A'} * \mu_{A'} - \mu_B.$$

Theorem 5. *There is a constant $c > 0$ such that the following holds. Let $\delta, \epsilon \in (0, 1)$, let $p \geq 1$ and let k be a positive integer such that $(k, |G|) = 1$. There is a constant $C = C(\epsilon, \delta, k) > 0$ such that the following holds.*

For any finite abelian group G and any subset $A \subseteq G$ with $|A| = \alpha|G|$ there exists a regular Bohr set B with

$$\mathrm{rk}(B) \leq Cp^4 \log(2/\alpha)^5$$

and

$$|B| \geq \exp(-Cp^5 \log(2p/\alpha) \log(2/\alpha)^6)|G|$$

and $A' \subseteq (A - x) \cap B$ for some $x \in G$ such that

(1) $|A'| \geq (1 - \epsilon)\alpha|B|$,

(2) $|A' \cap B'| \geq (1 - \epsilon)\alpha|B'|$, *where $B' = B_\rho$ is a regular Bohr set with $\rho \in \left(\frac{1}{2}, 1\right) \cdot c\delta\alpha/dk$, and*

$$\|(\mu_{A'} - \mu_B) * (\mu_{A'} - \mu_B)\|_{p(\mu_{k\cdot B'})} \leq \epsilon \frac{|G|}{|B|}. \tag{3}$$

In other words, for any dense set A, we can find a low-rank Bohr set B such that the restriction of (a translate of) A to B has almost the same relative density and has its convolution extremely balanced: it is close to its average, as measured in an L^p-sense over another large Bohr set. This roughly means that, when studying additive problems involving this restriction, we can replace A by a random set of the same density. Part (2) of the conclusion is there for somewhat technical reasons: since we need to work with both the Bohr set B and a narrowed copy B_ρ, as described above, we want to know that A' has large density on B_ρ as well as on B.

To give an idea of the parameters: to deduce Theorem 1, we shall take ϵ and δ some small constants, $k = 2$, and $p \asymp \log(2/\alpha)$. (The dependence on ϵ, δ, k is not hard to track explicitly, but is a distraction for the present applications.) An identical proof gives a statement with the measure $\mu_{k\cdot B'}$ in part (3) replaced by $\mu_{k\cdot B'+t}$ for any $t \in B$, or indeed μ_B or $\mu_B * \mu_B$, the latter two of which may appear more natural. These are, however, slightly weaker, and certainly for the application to three-term arithmetic progressions it is important that the measure over which the L^p norm is taken is supported on some suitably narrowed copy of B.

In Section 1 we provide an informal overview of the main steps of the argument. In Section 2 we prove what are, in our opinion, the most important steps of the argument in sufficient generality for the results over both $\mathbb{F}_q^n$ and $\{1, \dots, N\}$. In Section 3 we make the overview more precise and provide full proofs for the $\mathbb{F}_q^n$ case. Finally, in Sections 4 and 5 we show how this argument can be directly adapted for the integers using Bohr sets. We will proceed by proving Theorem 5 first and then deducing Theorems 1 and 3.

Improvements. Although Kelley and Meka's breakthrough is close to the best possible bound, there is still a gap between the exponent $\frac{1}{12}$ of Theorem 1 and the exponent $\frac{1}{2}$ in the Behrend example, and it is natural to ask whether further progress is possible. Indeed, a small modification of the method as presented in this paper allows for a slight improvement: the $\frac{1}{12}$ of Theorem 1 can be replaced by $\frac{1}{9}$ with a relatively clean argument (the only modification required is to the almost-periodicity part). A further tedious lengthy technical optimisation allows for an exponent of $\frac{5}{41}$. Since the focus of this paper is an exposition of the method and results of Kelley and Meka, we will detail these improvements in a separate forthcoming note. The same modification leads to an improvement of the exponent in Theorem 2 from $\frac{1}{9}$ to $\frac{1}{7}$, and an improvement of the exponent in Theorem 3 from 9 to 7.

We believe that an exponent of $\frac{1}{7}$ (in the statement of Theorem 1) is the natural limit of these methods, in that achieving anything better will require significant new ideas. Of course, the Behrend exponent of $\frac{1}{2}$ would be the final target, but this seems quite far out of reach still; indeed, an exponent of $\frac{1}{3}$ (or perhaps even $\frac{1}{4}$) seems to be the limit of any argument that uses any sort of "density increment" argument with Bohr sets (whether using Kelley–Meka ideas or a more traditional Fourier analytic approach).

Notational conventions. Logarithmic factors will appear often, and so in this paper we use the convenient abbreviation $\mathcal{L}(\alpha)$ to denote $\log(2/\alpha)$. (The 2 here is just a convenient device to make sure that $\mathcal{L}(\alpha) \geq \frac{1}{2}$, say, whenever $\alpha \in (0, 1]$.)

In statements which refer to G, this can be taken to be any finite abelian group (although for the applications this will always be either $\mathbb{F}_q^n$ or $\mathbb{Z}/N\mathbb{Z}$). We use the normalised counting measure on G, so that

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\in G} f(x)\overline{g(x)} \quad \text{and} \quad \|f\|_p = \left(\mathop{\mathbb{E}}_{x\in G} |f(x)|^p\right)^{1/p} \text{ for } 1 \leq p < \infty,$$

where $\mathbb{E}_{x\in G} = \frac{1}{|G|}\sum_{x\in G}$. For any $f, g : G \to \mathbb{C}$ we define the convolution and the difference convolution[1] as

$$f * g(x) = \mathop{\mathbb{E}}_{y} f(y)g(x-y) \quad \text{and} \quad f \circ g(x) = \mathop{\mathbb{E}}_{y} f(x+y)\overline{g(y)}.$$

Note the useful adjoint property

$$\langle f, g * h\rangle = \langle f \circ h, g\rangle.$$

We furthermore write $f^{(p)}$ for the p-fold convolution $f^{(p)} = f * f * \cdots * f$, where there are p copies of f.

[1]We caution that, while convolution is commutative and associative, difference convolution is in general neither.

For some purposes it is conceptually cleaner to work relative to other nonnegative functions on G, so that if $\mu : G \to \mathbb{R}_{\geq 0}$ has $\|\mu\|_1 = 1$ we write

$$\langle f, g\rangle_\mu = \mathop{\mathbb{E}}_{x\in G} \mu(x) f(x)\overline{g(x)} \quad \text{and} \quad \|f\|_{p(\mu)} = \left(\mathop{\mathbb{E}}_{x\in G} \mu(x)|f(x)|^p\right)^{1/p}$$

for $1 \leq p < \infty$. (The special case above is the case when $\mu \equiv 1$.)

We use the slight abuse of notation that if $\mu : G \to \mathbb{R}_{\geq 0}$ with $\|\mu\|_1 = 1$, then $\mu(A) = \|1_A\|_{1(\mu)}$ is the density of A relative to μ. Unless specified otherwise, μ is the uniform measure on G. (So that, for example, $\mu(A) = \alpha = |A|/|G|$ is the density of A within G.) We write $\mu_A = \alpha^{-1}1_A$ for the normalised indicator function of A (so that $\|\mu_A\|_1 = 1$). We will sometimes speak of $A \subseteq B$ with relative density $\alpha = |A|/|B|$.

The Fourier transform of $f : G \to \mathbb{R}$ is $\hat{f} : \widehat{G} \to \mathbb{C}$ defined for $\gamma \in \widehat{G}$ as

$$\hat{f}(\gamma) = \mathop{\mathbb{E}}_{x\in G} f(x)\overline{\gamma(x)},$$

where $\widehat{G} = \{\gamma : G \to \mathbb{C}^\times : \gamma \text{ a homomorphism}\}$ is the dual group of G. We will also use convolution of functions $f, g : \widehat{G} \to \mathbb{C}$, defined as $f * g(\gamma) = \sum_{\chi\in\widehat{G}} f(\chi)g(\gamma - \chi)$, and denote k-fold convolution again by $f^{(k)}$ for such functions.[2] We note the following elementary facts:

- $\widehat{f * g} = \hat{f} \cdot \hat{g}$ and $\widehat{f \circ f} = |\hat{f}|^2$ (so in particular the Fourier transform of $f \circ f$ is a nonnegative function on $\widehat{G}$).
- $\mathbb{E}_x f(x)^k = \hat{f} * \cdots * \hat{f}(0_{\widehat{G}})$, where the convolution is k-fold.
- If $\mu : G \to \mathbb{R}_{\geq 0}$ has $\|\mu\|_1 = 1$, then the Fourier transform of $\mu - 1$ is $\widehat{\mu}1_{\neq 0_{\widehat{G}}}$.

Note that these three facts immediately imply that the Fourier transform of $\mu_A \circ \mu_A - 1$ is nonnegative, that $\mathbb{E}(\mu_A \circ \mu_A - 1)^k \geq 0$ for any integer k, and (coupled with the triangle inequality) that $\|\mu_A * \mu_A - 1\|_p \leq \|\mu_A \circ \mu_A - 1\|_p$ when p is an even integer. Although easily seen via the Fourier transform, the latter two facts also have purely "physical" proofs, as we will see later.

Finally, we use the Vinogradov notation $X \ll Y$ to mean $X = O(Y)$, that is, there exists some constant $C > 0$ such that $|X| \leq CY$. We write $X \asymp Y$ to mean $X \ll Y$ and $Y \ll X$, and $X = \Omega(Y)$ to mean $Y = O(X)$. An expression like $1 + \Omega(1)$ thus means a quantity bounded below by an absolute constant strictly bigger than 1. The appearance of parameters as subscripts indicates that the implied constant may depend on these parameters (in some unspecified fashion).

[2]Note that we are using additive notation for the group operation on $\widehat{G}$.

1. Sketch of the argument

In this section we provide a sketch of the Kelley–Meka proof of the finite field model case, Theorem 2, along with some personal commentary and context. Kelley and Meka also include some fascinating comparisons of this approach with earlier work and speculations about it and alternative approaches, and we encourage the reader to study Appendices A and B of [Kelley and Meka 2023] and consider the interesting questions therein (although note that we answer their Question A.4 in the affirmative below).

Let $A \subseteq \mathbb{F}_q^n$ be a set of density α and $C \subseteq \mathbb{F}_q^n$ a set of density γ. (Note that a three-term arithmetic progression is a solution to $x + y = 2z$, and thus for their study we will choose

$$C = 2 \cdot A = \{2a : a \in A\},$$

in which case $\gamma = \alpha$.) How many solutions to $a_1 + a_2 = c$ with $a_1, a_2 \in A$ and $c \in C$ do we expect? If A, C are random sets, then we expect $\approx \alpha^2 \gamma q^{2n}$ many such solutions, which is to say

$$\langle \mu_A * \mu_A, \mu_C \rangle \approx 1.$$

The Kelley–Meka approach begins with some constant discrepancy from this expected count, say $\langle \mu_A * \mu_A, \mu_C \rangle \leq \frac{1}{2}$, and shows that this leads to a large density increment of A on some subspace $V \leq \mathbb{F}_q^n$ with codimension $O(\mathcal{L}(\alpha)^{O(1)})$ (that is, shows that there is such a subspace V and a translate A' of A such that $\mu_V(A') \geq (1 + \Omega(1))\alpha$, meaning that A' has significantly larger density in V than A has in $\mathbb{F}_q^n$). This is done using mostly physical-based methods, rather than the Fourier-based methods that have dominated the study of three-term progressions thus far.

Our presentation of the Kelley–Meka strategy breaks it down into five key steps.

Step 1 (Hölder lifting). If $\langle \mu_A * \mu_A, \mu_C \rangle \leq \frac{1}{2}$, then $\|\mu_A \circ \mu_A - 1\|_p \geq \frac{1}{4}$ for some $p \ll \mathcal{L}(\gamma)$.

This is essentially a one-line application of Hölder's inequality:

$$\tfrac{1}{2} \leq |\langle \mu_A * \mu_A - 1, \mu_C \rangle| \leq \|\mu_A * \mu_A - 1\|_p \|\mu_C\|_{p/(p-1)} \leq 2\|\mu_A * \mu_A - 1\|_p$$

for sufficiently large p, since $\|\mu_C\|_{p/(p-1)} = \gamma^{-1/p}$, and noting that $\|\mu_A * \mu_A - 1\|_p \leq \|\mu_A \circ \mu_A - 1\|_p$ when p is an even integer. Although trivial, the passage from few three-term arithmetic progressions to large L^p norm of $\mu_A \circ \mu_A - 1$ was rarely used in previous work, most of which begins with the Fourier deduction that $\sum_{\gamma \neq 0} |\widehat{\mu_A}(\gamma)|^2 |\widehat{\mu_C}(\gamma)| \gg 1$.

This physical Hölder step was used (along with almost-periodicity) in [Bloom and Sisask 2019] to achieve density bounds of $\alpha \leq (\log N)^{-1+o(1)}$. Indeed, in a sense the approach of [loc. cit.] corresponds to carrying out Steps 1, 4, and 5 of the

present sketch. The advancement of [Bloom and Sisask 2020] past the logarithmic density barrier couples this with delicate structural information on the Fourier side (following seminal ideas of Bateman and Katz [2012]).

It is incredible that the following two steps, which are simple enough to prove in only a couple of pages, perform far better quantitatively than this elaborate Fourier-side approach.

Step 2 (unbalancing). For any $f : G \to \mathbb{R}$ such that $\hat{f} \geq 0$, if $\|f\|_p \geq \frac{1}{4}$, then $\|f + 1\|_{p'} \geq 1 + \frac{1}{8}$ for some $p' \ll p$. In particular, if $\|\mu_A \circ \mu_A - 1\|_p \geq \frac{1}{4}$, then $\|\mu_A \circ \mu_A\|_{p'} \geq 1 + \frac{1}{8}$.

This step is essential to the success of the Kelley–Meka argument, and rests on the fact that the Fourier transform of f is nonnegative. (Note that it is not true for an arbitrary function.) While the spectral nonnegativity of $\mu_A \circ \mu_A - 1$ has played a role in some arguments before (e.g., in the spectral boosting aspect of [Bloom and Sisask 2020]), to our knowledge it has not before been so cleanly expressed, and its potential had not been fully appreciated within additive combinatorics.

We remark that, as pointed out to the authors by Shkredov, an entirely physical proof of this step is possible, with no mention of the Fourier transform at all, if we replace the assumption $\hat{f} \geq 0$ with $f = g \circ g$ for some function $g : \mathbb{R} \to \mathbb{C}$ (note that this is indeed satisfied in our application since $\mu_A \circ \mu_A - 1 = (\mu_A - 1) \circ (\mu_A - 1)$). We refer to the proof of Lemma 7 for details.

At this point we digress to note that, instead of following Steps 1 and 2 as Kelley and Meka do, one could obtain the conclusion $\|\mu_A \circ \mu_A\|_p \geq 1 + \Omega(1)$ for some $p \ll \mathcal{L}(\gamma)$ from the assumption that $\langle \mu_A * \mu_A, \mu_C \rangle \leq \frac{1}{2}$ using more classical Fourier-based methods. (In particular this observation answers [Kelley and Meka 2023, Question A.4] in the affirmative.) By converting the inner product to Fourier space and applying the triangle inequality we observe that

$$\tfrac{1}{2} \leq \sum_{\lambda \neq 0} |\widehat{\mu_A}(\lambda)|^2 |\widehat{\mu_C}(\lambda)|.$$

It follows that, for some choice of signs $c_\lambda \in \mathbb{C}$, we have

$$1 + \tfrac{1}{2} \leq \sum_{\lambda} |\widehat{\mu_A}(\lambda)|^2 |\widehat{\mu_C}(\lambda)| = \mathop{\mathbb{E}}_{x \in C} \sum_{\lambda} c_\lambda |\widehat{\mu_A}(\lambda)|^2 \lambda(-x).$$

Applying Hölder's inequality to the left-hand side and using orthogonality of characters yields, for any even integer p,

$$1 + \tfrac{1}{2} \leq \gamma^{-1/p} \Bigg(\sum_{\lambda_1, \dots, \lambda_p} c_{\lambda_1} \cdots \overline{c_{\lambda_p}} |\widehat{\mu_A}(\lambda_1)|^2 \cdots |\widehat{\mu_A}(\lambda_p)|^2 1_{\lambda_1 + \cdots - \lambda_p = 0} \Bigg)^{1/p}.$$

We can discard the signs c_λ by the triangle inequality, and then by orthogonality the sum here is in fact equal to $\|\mu_A \circ \mu_A\|_p$, and we are done choosing p suitably large. This sort of step has already played a major role in previous Fourier-based approaches to Roth's theorem, although generally with the $\widehat{\mu_A}$ restricted to some "large spectrum", where it then yields information about the additive relations within this large spectrum. A striking feature of the work of Kelley and Meka is that this information is far more useful on the physical side.

We end our digression here and return to the sketch.

Step 3 (dependent random choice). If $\|\mu_A \circ \mu_A\|_p \geq 1+\frac{1}{8}$, then there are $A_1, A_2 \subseteq A$ of density at least $\alpha^{O(p)}$ such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle \geq 1 - \tfrac{1}{32}$$

where $S = \{x : \mu_A \circ \mu_A(x) > 1 + \frac{1}{16}\}$.

This is, in our opinion, the most important step (although of course every step is necessary, and in particular the previous unbalancing step is crucial to obtaining the information required for this step). In combination with the previous two steps, we have now converted the original three variable deficiency information of $\langle \mu_A * \mu_A, \mu_C \rangle \leq \frac{1}{2}$ into four variable abundancy information $\langle \mu_{A_1} \circ \mu_{A_2}, \mu_A \circ \mu_A \rangle \geq 1 + \Omega(1)$. This is very promising, since Schoen and the second author [Schoen and Sisask 2016] have already shown that almost-periodicity can prove quasipolynomial bounds (that is, of the same shape as the Kelley–Meka bound) for 4 variable equations. Another indication of the strength of the conclusion obtained in Step 3 is that Sanders [2012b] has shown (also using almost-periodicity) that S contains a large proportion of a large subspace with codimension $O(\mathcal{L}(\alpha)^{O(1)})$.

The proof of this step, called sifting in [Kelley and Meka 2023], is completely elementary, and uses dependent random choice — one takes A_i to be the intersection of p randomly chosen translates of A. A simple expectation calculation combined with the L^p information then verifies that there must exist some choice of translates which satisfy both the density conditions and the inner product condition.

Arguments of this kind have appeared before in additive combinatorics, dating back in some form to Gowers' proof [1998] of Szemerédi's theorem. For example, when $p = 2$ this method was used by Schoen [2015] to prove strong bounds for the Balog–Szemerédi–Gowers theorem, and similar manipulations for larger p have played an extensive role in work of Schoen and Shkredov; see for example [Schoen and Shkredov 2013; Shkredov 2013]. A very similar statement also appears in work of Sanders [2010, Lemma 1.9], itself a generalisation of an argument of Gowers [1998, Lemma 11].

Despite this previous work, the true potential of this method (when coupled with the powerful technique of almost-periodicity) in applications to the study

of three-term progressions and related problems had been overlooked before the breakthrough of Kelley and Meka.

Step 4 (almost periodicity). For any sets $A_1, A_2, S \subseteq \mathbb{F}_q^n$, if A_1, A_2 have density at least α, then there is a subspace V with codimension $O(\mathcal{L}(\alpha)^4)$ such that

$$|\langle \mu_V * \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle - \langle \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle| \leq \tfrac{1}{100}.$$

Almost-periodicity statements of this type have played an important role in additive combinatorics since their introduction by Croot and the second author [Croot and Sisask 2010], most notably in the work of Sanders achieving quasipolynomial bounds for inverse sumset theorems [Sanders 2012b; 2013]. The conventional wisdom was that, despite their success in inverse sumset problems and for translation invariant equations in four or more variables (see [Schoen and Sisask 2016] and the earlier [Schoen and Shkredov 2014] for six or more variables), they were not able to achieve significant results for three-term arithmetic progressions. (Although the argument of [Bloom and Sisask 2020] did make fundamental use of almost-periodicity, most of the work in that paper was Fourier-analytic.) Kelley and Meka have dispelled this illusion completely.

It should be noted, however, that there is no novelty in the actual form of almost-periodicity used by Kelley and Meka — the new strength is a result of the context in which they use it.

Step 5 (density increment). If $\langle \mu_A * \mu_A, \mu_C\rangle \leq \frac{1}{2}$, then there is an affine subspace V of codimension $O(\mathcal{L}(\alpha)^4\mathcal{L}(\gamma)^4)$ on which A has density at least $\big(1 + \frac{1}{100}\big)\alpha$.

This is just a trivial combination of the previous 4 steps (noting that the density increment condition can also be phrased as $\|\mu_A * \mu_V\|_\infty \geq 1 + \frac{1}{100}$). This density increment condition can now be iteratively applied to eventually obtain a lower bound for $\langle \mu_{A'} * \mu_{A'}, \mu_C\rangle$ (with A' now perhaps some subset of a translate of A).

For example, the deduction of Theorem 2 is routine with $C = 2 \cdot A$. Indeed, a density increment such as $\alpha \mapsto (1 + \Omega(1))\alpha$ can occur at most $O(\mathcal{L}(\alpha))$ many times, after which we must halt with many three-term arithmetic progressions found in the intersection of A with some affine subspace, the codimension of which is bounded above by $O(\mathcal{L}(\alpha)^9)$.

Alternatively, carrying through these steps with C being the complement of $A + A$ leads to the following.

Theorem 6 (Kelley–Meka). *If $A \subseteq \mathbb{F}_q^n$ has density α and $\gamma \in (0, 1]$, then there is some affine subspace $V \leq \mathbb{F}_q^n$ of codimension $O(\mathcal{L}(\alpha)^5\mathcal{L}(\gamma)^4)$ such that*

$$|(A + A) \cap V| \geq (1 - \gamma)|V|.$$

For comparison, the best bound previously available, due to Sanders [2012b], had codimension $O(\mathcal{L}(\alpha)^4\gamma^{-2})$. The latter is slightly better when γ is constant (as

was the case of interest in [loc. cit.]) but much weaker when $\gamma \approx \alpha$, which is the regime of interest for three-term arithmetic progressions.

2. The key new lemmas

In this section we prove general forms of Steps 2 and 3 that will be used for both the $\mathbb{F}_q^n$ and integer case.

2.1. ***Unbalancing of spectrally nonnegative functions.*** The following precise form of Step 2 will suffice for all our applications.

Lemma 7. *Let $\epsilon \in (0,1)$ and $\nu : G \to \mathbb{R}_{\geq 0}$ satisfy $\|\nu\|_1 = 1$ and $\widehat{\nu} \geq 0$. Let $f : G \to \mathbb{R}$ be such that $\hat{f} \geq 0$. (Or, alternatively, assume that $f = g \circ g$ and $\nu = h \circ h$ for some $g, h : G \to \mathbb{C}$.)*

If $\|f\|_{p(\nu)} \geq \epsilon$ for some $p \geq 1$, then

$$\|f + 1\|_{p'(\nu)} \geq 1 + \tfrac{1}{2}\epsilon$$

for some $p' \ll_\epsilon p$.

We have left the dependence on ϵ unspecified since our applications only use $\epsilon \gg 1$. The proof below delivers $p' \ll \epsilon^{-1}\log(\epsilon^{-1})p$. Kelley and Meka use a different method (see [Kelley and Meka 2023, Appendix D]) which is more efficient, giving $p' \ll \epsilon^{-1}p$, but requires use of an external (though simple) fact about the binomial distribution.

Proof. We first establish the important fact that, for any integer $k \geq 1$,

$$\langle \nu, f^k \rangle \geq 0. \tag{1}$$

We will give two proofs of (1), depending on whether the Fourier assumption $\hat{f}, \widehat{\nu} \geq 0$ or the physical assumption $f = g \circ g$ and $\nu = h \circ h$ is used. Given (1) no further use of the Fourier transform is required, and the proofs converge.

The Fourier proof of (1), which is used by Kelley and Meka, is immediate from Parseval's identity:

$$\langle \nu, f^k \rangle = \langle \widehat{\nu}, \hat{f}^{(k)} \rangle,$$

where we recall that $f^{(k)} = f * f * \cdots * f$ denotes the k-fold convolution, and the right-hand side is nonnegative since $\hat{f}, \widehat{\nu} \geq 0$.

We now present the alternative physical proof of (1), assuming $f = g \circ g$ and $\nu = h \circ h$. This argument was pointed out to the authors by Shkredov, who has made extensive use of the following kind of manipulations; for example in [Shkredov

2013]. We observe that

$$\begin{aligned}\langle \nu, f^k\rangle &= \mathop{\mathbb{E}}_x h\circ h(x) g\circ g(x)^k\\ &= \mathop{\mathbb{E}}_{y_1,y_2} h(y_1)\overline{h(y_2)}\left(\mathop{\mathbb{E}}_z g(y_1+z)\overline{g(y_2+z)}\right)^k\\ &= \mathop{\mathbb{E}}_{z_1,\dots,z_k}\mathop{\mathbb{E}}_{y_1,y_2} h(y_1)\overline{h(y_2)}g(y_1+z_1)\cdots\overline{g(y_2+z_k)}\\ &= \mathop{\mathbb{E}}_{z_1,\dots,z_k}\left|\mathop{\mathbb{E}}_y h(y)g(y+z_1)\cdots g(y+z_k)\right|^2,\end{aligned}$$

which is clearly nonnegative as each summand is.

We now show how (1) implies the conclusion. Without loss of generality we can assume that $p\geq 5$ is an odd integer. Using (1), since $2\max(x,0)=x+|x|$ for $x\in\mathbb{R}$ and $f^{p-1}=|f|^{p-1}$, we have

$$2\langle \max(f,0), f^{p-1}\rangle_\nu = \langle \nu, f^p\rangle + \langle |f|, f^{p-1}\rangle_\nu \geq \|f\|_{p(\nu)}^p \geq \epsilon^p.$$

Therefore, if $P=\{x : f(x)\geq 0\}$, then $\langle 1_P, f^p\rangle_\nu \geq \frac{1}{2}\epsilon^p$. Furthermore, if $T=\left\{x\in P : f(x)\geq \frac{3}{4}\epsilon\right\}$, then $\langle 1_{P\setminus T}, f^p\rangle_\nu < \left(\frac{3}{4}\epsilon\right)^p \leq \frac{1}{4}\epsilon^p$, and hence by the Cauchy–Schwarz inequality

$$\nu(T)^{1/2}\|f\|_{2p(\nu)}^p \geq \langle 1_T, f^p\rangle_\nu \geq \tfrac{1}{4}\epsilon^p.$$

On the other hand, by the triangle inequality

$$\|f\|_{2p(\nu)} \leq 1 + \|f+1\|_{2p(\nu)} \leq 3,$$

or else we are done, with $p'=2p$. Hence $\nu(T)\geq(\epsilon/10)^{2p}$. It follows that, for any $p'\geq 1$,

$$\|f+1\|_{p'(\nu)} \geq \langle 1_T, |f+1|^{p'}\rangle_\nu^{1/p'} \geq \left(1+\tfrac{3}{4}\epsilon\right)(\epsilon/10)^{2p/p'}.$$

The desired bound now follows if we choose p' a sufficiently large multiple (depending on ϵ) of p. □

2.2. ***An application of dependent random choice.*** We now use dependent random choice (or what Kelley and Meka call "sifting") to prove a general form of Step 3. This makes use of (a generalisation of) the fact that

$$\|1_A\circ 1_A\|_p^p = \mathop{\mathbb{E}}_{s_1,\dots,s_p\in G}\mu((A+s_1)\cap\cdots\cap(A+s_p))^2$$

to convert L^p-information about a convolution to information about the nested intersections appearing in the right-hand side. This identity features extensively in some works of Shkredov (see [Shkredov 2013] for example) and Schoen and Shkredov [2014], and is already implicitly used in work of Sanders [2010, Lemma 1.9], but its

strength and utility in the current context was far from apparent before the work of Kelley and Meka. At a first reading of the following result the reader may wish to take $B_1 = B_2 = G$, in which case $\mu = \mu_{B_1} \circ \mu_{B_2}$ is just the usual uniform measure on G.

Lemma 8. *Let $p \geq 1$ be an integer and $\epsilon, \delta > 0$. Let $B_1, B_2 \subseteq G$, and let $\mu = \mu_{B_1} \circ \mu_{B_2}$. For any finite set $A \subseteq G$ with density α, if*

$$S = \{x \in G : \mu_A \circ \mu_A(x) > (1-\epsilon)\|\mu_A \circ \mu_A\|_{p(\mu)}\},$$

then there are $A_1 \subseteq B_1$ and $A_2 \subseteq B_2$ such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle \geq 1 - \delta$$

and

$$\min\left(\frac{|A_1|}{|B_1|}, \frac{|A_2|}{|B_2|}\right) \gg (\alpha \|\mu_A \circ \mu_A\|_{p(\mu)})^{2p+O_{\epsilon,\delta}(1)}.$$

Again, since we will apply this only in the case $\epsilon, \delta \gg 1$, we are not concerned with the behaviour of the $O_{\epsilon,\delta}(1)$ term, although we record here that the proof (which is identical to that in [Kelley and Meka 2023]) in fact allows for $O(\epsilon^{-1}\log(\delta^{-1}))$. We furthermore note that A_i will take the form $B_i \cap (A + s_1) \cap \cdots \cap (A + s_p)$ for some randomly chosen shifts $s_j \in G$.

We shall prove Lemma 8 after Lemma 10 below, but first we note the following immediate special case, which is all we use when studying $\mathbb{F}_q^n$.

Corollary 9. *Let $p \geq 1$ be an integer and $\epsilon > 0$. If $A \subseteq G$ is such that $\|\mu_A \circ \mu_A\|_p \geq 1 + \epsilon$ and $S = \{x : \mu_A \circ \mu_A(x) > 1 + \epsilon/2\}$, then there are $A_1, A_2 \subseteq G$, both of density*

$$\gg \alpha^{2p+O_\epsilon(1)},$$

such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle \geq 1 - \epsilon/8.$$

We encode the dependent random choice argument underpinning Lemma 8 as the following general lemma.

Lemma 10. *Let $p \geq 2$ be an even integer. Let $B_1, B_2 \subseteq G$ and $\mu = \mu_{B_1} \circ \mu_{B_2}$. For any finite set $A \subseteq G$ with density α and function $f : G \to \mathbb{R}_{\geq 0}$ there exist $A_1 \subseteq B_1$ and $A_2 \subseteq B_2$ such that*

$$\langle \mu_{A_1} \circ \mu_{A_2}, f \rangle \leq 2\frac{\langle (\mu_A \circ \mu_A)^p, f \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p(\mu)}^p}$$

and

$$\min\left(\frac{|A_1|}{|B_1|}, \frac{|A_2|}{|B_2|}\right) \geq \tfrac{1}{4}\alpha^{2p}\|\mu_A \circ \mu_A\|_{p(\mu)}^{2p}.$$

Proof. For $s \in G^p$ let $A_1(s) = B_1 \cap (A+s_1) \cap \cdots \cap (A+s_p)$, and similarly for $A_2(s)$. Note that

$$\begin{aligned}\langle (\mu_A \circ \mu_A)^p, f\rangle_\mu &= \mathop{\mathbb{E}}_{\substack{b_1\in B_1\\ b_2\in B_2}} \mu_A\circ\mu_A(b_1-b_2)^p f(b_1-b_2)\\ &= \mathop{\mathbb{E}}_{\substack{b_1\in B_1\\ b_2\in B_2}} \left(\alpha^{-2}\mathop{\mathbb{E}}_{t\in G} 1_{A+t}(b_1)1_{A+t}(b_2)\right)^p f(b_1-b_2)\\ &= \alpha^{-2p}\mathop{\mathbb{E}}_{\substack{b_1\in B_1\\ b_2\in B_2}}\mathop{\mathbb{E}}_{s\in G^p} 1_{A_1(s)}(b_1)1_{A_2(s)}(b_2) f(b_1-b_2)\\ &= \tfrac{\alpha^{-2p}|G|^2}{|B_1||B_2|}\mathop{\mathbb{E}}_{s\in G^p}\langle 1_{A_1(s)}\circ 1_{A_2(s)}, f\rangle.\end{aligned}$$

In particular, applying this with $f\equiv 1$ we see that, if $\alpha_i(s) = |A_i(s)|/|G|$, then

$$\|\mu_A\circ\mu_A\|_{p(\mu)}^p = \tfrac{\alpha^{-2p}|G|^2}{|B_1||B_2|}\mathop{\mathbb{E}}_s \alpha_1(s)\alpha_2(s)$$

and

$$\frac{\langle (\mu_A\circ\mu_A)^p, f\rangle_\mu}{\|\mu_A\circ\mu_A\|_{p(\mu)}^p} = \frac{\mathbb{E}_s\langle 1_{A_1(s)}\circ 1_{A_2(s)}, f\rangle}{\mathbb{E}_s\,\alpha_1(s)\alpha_2(s)} = \eta,$$

say. Note that, if

$$M = \tfrac{1}{2}\alpha^p(|B_1||B_2|/|G|^2)^{1/2}\|\mu_A\circ\mu_A\|_{p(\mu)}^p,$$

then

$$\begin{aligned}\mathop{\mathbb{E}}_s 1_{\alpha_1(s)\alpha_2(s)<M^2}\alpha_1(s)\alpha_2(s) &< M\left(\mathop{\mathbb{E}}_s\mathop{\mathbb{E}}_{x\in G} 1_{A_1(s)}(x)\right)^{1/2}\left(\mathop{\mathbb{E}}_s\mathop{\mathbb{E}}_{x\in G} 1_{A_2(s)}(x)\right)^{1/2}\\ &= M\alpha^p(|B_1||B_2|/|G|^2)^{1/2}\\ &= \tfrac{1}{2}\mathop{\mathbb{E}}_s\alpha_1(s)\alpha_2(s)\end{aligned}$$

and so

$$\mathop{\mathbb{E}}_s\langle 1_{A_1(s)}\circ 1_{A_2(s)}, f\rangle = \eta\mathop{\mathbb{E}}_s\alpha_1(s)\alpha_2(s) < 2\eta\mathop{\mathbb{E}}_s\alpha_1(s)\alpha_2(s)1_{\alpha_1(s)\alpha_2(s)\geq M^2}.$$

In particular there must exist some s such that

$$\langle 1_{A_1(s)}\circ 1_{A_2(s)}, f\rangle < 2\eta\alpha_1(s)\alpha_2(s)1_{\alpha_1(s)\alpha_2(s)\geq M^2},$$

and the claim follows (note that the left-hand side is trivially ≥ 0 and hence such an s must satisfy $\alpha_1(s)\alpha_2(s)\geq M^2$). □

The deduction of Lemma 8 is immediate from Lemma 10 with $f = 1_{G\setminus S}$. Indeed, by nesting of L^p norms we can assume that p is sufficiently large in terms of ϵ and δ (this is where the $O_{\epsilon,\delta}(1)$ term arises in the exponent), and that p is an even integer. It then suffices to note that

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle = 1 - \langle \mu_{A_1} \circ \mu_{A_2}, 1_{G\setminus S} \rangle$$

and by definition of S we have

$$\frac{\langle (\mu_A \circ \mu_A)^p, 1_{G\setminus S} \rangle}{\|\mu_A \circ \mu_A\|_p^p} \leq (1-\epsilon)^p$$

which is $\leq \delta/2$ if p is large enough.

3. The finite field case

In this section we prove Theorem 2, following the sketch of Section 1. Theorem 4 can proved in a very similar way. The following straightforward lemma is a form of Step 1.

Lemma 11. *Let $\epsilon > 0$. If $A, C \subseteq G$, where C has density at least γ, then either*

(1) $|\langle \mu_A * \mu_A, \mu_C \rangle - 1| \leq \epsilon$ *or*

(2) $\|\mu_A \circ \mu_A - 1\|_p \geq \epsilon/2$ *for some* $p \ll \mathcal{L}(\gamma)$.

Proof. If the first alternative fails, then by Hölder's inequality, for any $p \geq 1$

$$\epsilon < |\langle \mu_A * \mu_A - 1, \mu_C \rangle| \leq \|\mu_A * \mu_A - 1\|_p \gamma^{-1/p}.$$

In particular, if we choose $p = 2\lceil K\mathcal{L}(\gamma)\rceil$ for some large constant K, then we deduce that

$$\|\mu_A * \mu_A - 1\|_p \geq \tfrac{1}{2}\epsilon.$$

It remains to note that, assuming without loss of generality that p is an even integer,

$$\|\mu_A * \mu_A - 1\|_p^p = (\widehat{\mu_A}^2 1_{\neq 0_{\widehat{G}}})^{(p)}(0_{\widehat{G}}) \leq (|\widehat{\mu_A}|^2 1_{\neq 0_{\widehat{G}}})^{(p)}(0_{\widehat{G}}) = \|\mu_A \circ \mu_A - 1\|_p^p.$$

Again, although we have used a one-line Fourier proof here, this can also be seen using an entirely physical argument, which we sketch here. Note that $\mu_A * \mu_A - 1 = (\mu_A - 1) * (\mu_A - 1)$ and similarly for $\mu_A \circ \mu_A - 1$. It suffices therefore to show that, for any function $f : G \to \mathbb{C}$, we have

$$\|f * f\|_p^p \leq \|f \circ f\|_p^p.$$

We can write the left-hand side as

$$\|f*f\|_p^p = \mathop{\mathbb{E}}_{x,y}\Big(\mathop{\mathbb{E}}_{u} f(x+u)f(y-u)\Big)^p$$
$$= \mathop{\mathbb{E}}_{u_1,\dots,u_p}\Big(\mathop{\mathbb{E}}_{x} f(x+u_1)\cdots f(x+u_p)\Big)\Big(\mathop{\mathbb{E}}_{y} f(y-u_1)\cdots f(y-u_p)\Big)$$

and the right-hand side as

$$\|f\circ f\|_p^p = \mathop{\mathbb{E}}_{x,y}\Big(\mathop{\mathbb{E}}_{u} f(x+u)\overline{f(y+u)}\Big)^p = \mathop{\mathbb{E}}_{u_1,\dots,u_p}\Big|\mathop{\mathbb{E}}_{x} f(x+u_1)\cdots f(x+u_p)\Big|^2,$$

and the desired inequality now follows from the Cauchy–Schwarz inequality. □

Steps 2 and 3 have already been proved as Lemma 7 and Corollary 9. For Step 4 we can use the following almost-periodicity result, which is [Schoen and Sisask 2016, Theorem 3.2], as a black box.

Theorem 12 (almost-periodicity). *If $A_1, A_2, S \subseteq \mathbb{F}_q^n$ are such that A_1 and A_2 both have density at least α, then there is a subspace V of codimension*

$$\operatorname{codim}(V) \ll_\epsilon \mathcal{L}(\alpha)^4$$

such that

$$|\langle \mu_V * \mu_{A_1} * \mu_{A_2}, 1_S\rangle - \langle \mu_{A_1} * \mu_{A_2}, 1_S\rangle| \le \epsilon.$$

Importantly, note that no assumption is made on S, and there is no dependency on the density of S. We now complete the final step by combining everything thus far into a single density increment statement, which suffices for Theorem 2 as discussed in Section 1.

Proposition 13. *Let $\epsilon \in (0,1)$. If $A, C \subseteq \mathbb{F}_q^n$, where C has density at least γ, then either*

(1) $|\langle \mu_A * \mu_A, \mu_C\rangle - 1| \le \epsilon$ *or*

(2) *there is a subspace V of codimension*

$$\ll_\epsilon \mathcal{L}(\gamma)^4\mathcal{L}(\alpha)^4$$

such that $\|1_A * \mu_V\|_\infty \ge (1+\Omega(\epsilon))\alpha$.

Proof. By Lemma 11, if the first alternative fails, then $\|\mu_A \circ \mu_A - 1\|_p \ge \epsilon/2$ for some $p \ll \mathcal{L}(\gamma)$. By Lemma 7 (applied to $f = \mu_A \circ \mu_A - 1$) we deduce that $\|\mu_A \circ \mu_A\|_p \ge 1 + \epsilon/4$ for some $p \ll_\epsilon \mathcal{L}(\gamma)$. Hence, by Corollary 9, there are A_1, A_2, both of density

$$\ge \alpha^{O_\epsilon(\mathcal{L}(\gamma))},$$

such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle \geq 1 - \epsilon/32,$$

where $S = \{x : \mu_A \circ \mu_A(x) \geq 1 + \epsilon/8\}$. By Theorem 12 there is a subspace V of the required codimension such that

$$\langle \mu_V * \mu_{A_1} \circ \mu_{A_2}, 1_S \rangle \geq 1 - \tfrac{1}{16}\epsilon.$$

By definition of S, it follows that

$$\begin{aligned} 1 + \Omega(\epsilon) &\leq (1 + \epsilon/8)(1 - \epsilon/16) \\ &\leq \langle \mu_V * \mu_{A_1} \circ \mu_{A_2}, \mu_A \circ \mu_A \rangle \\ &\leq \|\mu_V * \mu_A\|_\infty \|\mu_A * \mu_{A_2} \circ \mu_{A_1}\|_1 \\ &= \|\mu_V * 1_A\|_\infty \alpha^{-1}, \end{aligned}$$

and the proof is complete. □

4. The integer case

A useful strategy in additive combinatorics is to first prove a result of interest over $\mathbb{F}_q^n$, making use of the abundancy of subspaces, and then translate this to a result over the integers, replacing various equalities with approximate equalities and subspaces with Bohr sets.

Bohr sets of low rank are analogues of subspaces of low codimension, and have played a central role in additive combinatorics since the work of Bourgain [1999], with much of the theory further developed by Sanders [2011; 2012a]. We have recalled the relevant definitions and properties in the Appendix.

In this section we will employ Steps 3, 4, and 5 of the Kelley–Meka approach, performed relative to Bohr sets, to prove the following technical statement, whose statement is convenient for the iterative proof. In the following section we will show how it, together with unbalancing and the regularity of Bohr sets, implies Theorem 5, and thence Theorems 1 and 3.

We apologise for the daunting appearance and technicality of the statements in this and the next section; a certain overhead of notation and caveats is a sad fact of life when working with Bohr sets. The reader should be reassured, however, that all the essential ideas are as in the $\mathbb{F}_q^n$ case.

The approach taken here should be compared with that in [Kelley and Meka 2023, Section 8] — there Kelley and Meka also follow the $\mathbb{F}_q^n$ model, but instead use mixed analogues over multidimensional progressions and Bohr sets, instead of just Bohr sets as we do here. The two objects are, in a heuristic sense, identical, but the need to pass between them adds some complexity to the argument of Kelley and Meka.

Theorem 14. *There is a constant $c>0$ such that the following holds. Let $\epsilon,\delta\in(0,1)$ and $p,k\geq 1$ be integers such that $(k,|G|)=1$. For any $A\subseteq G$ with density α there is a regular Bohr set B with*

$$d=\mathrm{rk}(B)=O_\epsilon(\mathcal{L}(\alpha)^5p^4)\quad\text{and}\quad |B|\geq\exp(-O_{\epsilon,\delta}(\mathcal{L}(\alpha)^6p^5\mathcal{L}(\alpha/p)))|G|$$

and some $A'\subseteq(A-x)\cap B$ for some $x\in G$ such that

(1) $|A'|\geq(1-\epsilon)\alpha|B|$,

(2) $|A'\cap B'|\geq(1-\epsilon)\alpha|B'|$, *where $B'=B_\rho$ is a regular Bohr set with $\rho\in\left(\frac{1}{2},1\right)\cdot c\delta\alpha/d$, and*

(3) $\|\mu_{A'}\circ\mu_{A'}\|_{p(\mu_{k\cdot B''}\circ\mu_{k\cdot B''}*\mu_{k\cdot B'''}\circ\mu_{k\cdot B'''})}<(1+\epsilon)\mu(B)^{-1}$, *for any regular Bohr sets $B''=B'_{\rho'}$ and $B'''=B''_{\rho''}$ satisfying $\rho',\rho''\in\left(\frac{1}{2},1\right)\cdot c\delta\alpha/d$.*

The proof of Theorem 14 proceeds by repeated application of the following statement, which is suitable for iteration.

Proposition 15. *There is a constant $c>0$ such that the following holds. Let $\epsilon>0$ and $p,k\geq 1$ be integers such that $(k,|G|)=1$. Let B, B', $B''\subseteq G$ be regular Bohr sets of rank d such that $B''\subseteq B'_{c/d}$ and $A\subseteq B$ with relative density α. If*

$$\|\mu_A\circ\mu_A\|_{p(\mu_{k\cdot B'}\circ\mu_{k\cdot B'}*\mu_{k\cdot B''}\circ\mu_{k\cdot B''})}\geq(1+\epsilon)\mu(B)^{-1},$$

then there is a regular Bohr set $B'''\subseteq B''$ of rank at most

$$\mathrm{rk}(B''')\leq d+O_\epsilon(\mathcal{L}(\alpha)^4p^4)$$

and size

$$|B'''|\geq\exp(-O_\epsilon(dp\mathcal{L}(\alpha/d)+\mathcal{L}(\alpha)^5p^5))|B''|$$

such that

$$\|\mu_{B'''}*\mu_A\|_\infty\geq(1+c\epsilon)\mu(B)^{-1}.$$

We first explain how Theorem 14 follows by iteration. In doing so we shall require a regularity "narrowing" trick originally due to Bourgain. The following form is [Bloom and Sisask 2020, Lemma 12.1].

Lemma 16. *There is a constant $c>0$ such that the following holds. Let B be a regular Bohr set of rank d, suppose $A\subseteq B$ has density α, let $\epsilon>0$, and suppose B', $B''\subseteq B_\rho$ where $\rho\leq c\alpha\epsilon/d$. Then either*

(1) *there is some translate A' of A such that $|A'\cap B'|\geq(1-\epsilon)\alpha|B'|$ and $|A'\cap B''|\geq(1-\epsilon)\alpha|B''|$, or*

(2) $\|1_A*\mu_{B'}\|_\infty\geq(1+\epsilon/2)\alpha$, *or*

(3) $\|1_A*\mu_{B''}\|_\infty\geq(1+\epsilon/2)\alpha$.

Proof of Theorem 14 assuming Proposition 15. Let C_ϵ, $D_{\epsilon,\delta} \geq 1$ be parameters to be specified later, and let c be the smaller of $\frac{1}{2}$ and the constant c in Proposition 15. Let $t \geq 0$ be maximal such that there is a sequence of regular Bohr sets, say $B^{(0)}, \dots, B^{(t)}$, and subsets of translates of A, say $A_0, \dots, A_t$, such that the following holds:

(1) $B^{(0)} = G$ and $A_0 = A$.

(2) Each $B^{(i)}$ is a regular Bohr set of rank d_i and

$$d_{i+1} \leq d_i + C_\epsilon \mathcal{L}(\alpha)^4 p^4$$

and

$$|B^{(i+1)}| \geq \exp(-D_{\epsilon,\delta}(dp\mathcal{L}(\alpha/d) + \mathcal{L}(\alpha)^5 p^5))|B^{(i)}|.$$

(3) Each A_i is a subset of $B^{(i)}$ with density α_i such that $\alpha_{i+1} \geq (1 + c\epsilon/4)\alpha_i$ for $0 \leq i < t$.

Observe from point (3), and the trivial fact that $\alpha_i \leq 1$, that $t \ll_\epsilon \mathcal{L}(\alpha)$. Note that this implies $d_t \ll_\epsilon \mathcal{L}(\alpha)^5 p^4$.

We apply Lemma 16 with $c\epsilon/2$ in place of ϵ, and $B = B^{(t)}$, $B' = B_{c'\alpha\epsilon/d_t}$ and $B'' = B'_{c''\delta\alpha/d_t}$, where the constants are in particular chosen to ensure that B', B'' are both regular. Provided we pick $D_{\epsilon,\delta}$ large enough in terms of C_ϵ, ϵ, and δ, Lemma A.4 and the maximality of t ensure that we must be in the first alternative of Lemma 16's conclusion: there exists a translate $A_t - x$ such that $|(A_t - x) \cap B'| \geq (1 - c\epsilon/2)\alpha|B'|$ and $|(A_t - x) \cap B''| \geq (1 - c\epsilon/2)\alpha|B''|$.

We claim that $A' = (A_t - x) \cap B'$, with the B' and B'' above playing the role of B and B' respectively, satisfies the conclusions of Theorem 14. Indeed, the bounds on the rank and size of B', and the density conditions on A', are clearly satisfied.

Suppose for a contradiction that

$$\|\mu_{A'} \circ \mu_{A'}\|_{p(\mu_{k\cdot B'''} \circ \mu_{k\cdot B'''} * \mu_{k\cdot B''''} \circ \mu_{k\cdot B''''})} \geq (1+\epsilon)\mu(B')^{-1},$$

for some regular Bohr sets $B''' = B''_\rho$ and $B'''' = B'''_{\rho'}$ satisfying $\rho, \rho' \in \left(\frac{1}{2}, 1\right) \cdot c\delta\alpha/d_t$.

The conditions of Proposition 15 are met, and hence we deduce there is some $\tilde{B} \subseteq B''''$ of rank

$$\mathrm{rk}(\tilde{B}) \leq \mathrm{rk}(B) + O_\epsilon(\mathcal{L}(\alpha)^4 p^4)$$

and

$$|\tilde{B}| \geq \exp(-O_\epsilon(d_t p\mathcal{L}(\alpha/d_t) + \mathcal{L}(\alpha)^5 p^5))|B''|$$

and there is a translate of A_t, say $A_t - y$, such that

$$\mu_{\tilde{B}}(A_t - y) \geq (1 + c\epsilon)(1 - c\epsilon/2)\alpha \geq (1 + c\epsilon/4)\alpha,$$

say. This a contradiction to the maximality of t, provided C_ϵ matches the implicit constant in the first O_ϵ-term, since we can take $B^{(t+1)} = \tilde{B}$ and $A_{t+1} = (A_t - y) \cap \tilde{B}$,

noting that by Lemma A.4

$$|B''''| \geq \exp(-O_{\epsilon,\delta}(d\mathcal{L}(\alpha/d_t)))|B^{(t)}|. \qquad \square$$

Proposition 15 is a consequence of Steps 3 and 4 (dependent random choice and almost-periodicity) of the Kelley–Meka approach. We will use the following version of almost-periodicity, which is essentially [Schoen and Sisask 2016, Theorem 5.4].

Theorem 17 (almost-periodicity). *There is a constant $c > 0$ such that the following holds. Let $\epsilon > 0$ and $B, B' \subseteq G$ be regular Bohr sets of rank d. Suppose that $A_1 \subseteq B$ with density α_1 and A_2 is such that there exists x with $A_2 \subseteq B' - x$ with density α_2. Let S be any set with $|S| \leq 2|B|$. There is a regular Bohr set $B'' \subseteq B'$ of rank at most*

$$d + O_\epsilon(\mathcal{L}(\alpha_1)^3\mathcal{L}(\alpha_2))$$

and size

$$|B''| \geq \exp(-O_\epsilon(d\mathcal{L}(\alpha_1\alpha_2/d) + \mathcal{L}(\alpha_1)^3\mathcal{L}(\alpha_2)\mathcal{L}(\alpha_1\alpha_2/d)))|B'|$$

such that

$$|\langle \mu_{B'} * \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle - \langle \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle| \leq \epsilon.$$

Proof. We apply [Schoen and Sisask 2016, Theorem 5.4] with the choices (with apologies for the unfortunate clash in variable naming between papers)

$$A \to -A_2, \quad M \to A_1, \quad L \to -S, \quad S \to B'_{c/d}, \quad B \to B'_{c/d}$$

and note that

$$|-A_2 + B'_{c/d}| \leq |B' + B'_{c/d}| \leq 2|B'| \leq 2\alpha_2^{-1}|A_2|$$

by regularity of B'. The statement now almost follows from [loc. cit., Theorem 5.4] after observing that (in that theorem's language) we have $K \leq 2\alpha_2^{-1}$, $\sigma = 1$, and $\eta \geq \alpha_1/2$, except that there the statement has a constraint on the rank and the width of B'', rather than the rank and the size. Nonetheless the width condition can be immediately converted into a lower bound for the size of B'' with Lemma A.4. $\square$

We will now use this almost-periodicity together with Lemma 8 to deduce the iterative step.

Proof of Proposition 15. By averaging there exists some $x \in k \cdot B' + k \cdot B''$ such that

$$\|\mu_A \circ \mu_A\|_{p(\mu_{k\cdot B'} * \mu_{k\cdot B''-x})} \geq (1+\epsilon)\mu(B)^{-1}.$$

We now apply Lemma 8 with $B_1 = k \cdot B'$ and $B_2 = k \cdot B'' + x$. This produces some $A_1 \subseteq k\cdot B'$ and $A_2 \subseteq k\cdot B'' - x$ such that, with $S = \{x : \mu_A \circ \mu_A(x) \geq (1+\epsilon/2)\mu(B)^{-1}\}$,

$$\langle \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle \geq 1 - \epsilon/4$$

and

$$\min\left(\frac{|A_1|}{|B'|}, \frac{|A_2|}{|B''|}\right) \gg \alpha^{2p+O_\epsilon(1)}.$$

We now apply Theorem 17 (with $k \cdot B'$ and $k \cdot B''$ playing the roles of B and B' respectively), noting that we can, without loss of generality, take S to be supported in $A_1 - A_2 \subseteq k \cdot B' + k \cdot B'' - x$ and so by regularity of B'

$$|S| \leq |B' + B''| \leq 2|B'|.$$

This produces some $B''' \subseteq k \cdot B''$ of the required rank and size such that

$$\langle \mu_{B'''} * \mu_{A_1} \circ \mu_{A_2}, \mu_A \circ \mu_A \rangle \geq (1+\epsilon/2)(1-\epsilon/4)(1-\epsilon/8)\mu(B)^{-1} \geq (1+c\epsilon)\mu(B)^{-1}$$

for some absolute constant $c > 0$. The result now follows from averaging, since

$$\begin{aligned}\langle \mu_{B'''} * \mu_{A_1} \circ \mu_{A_2}, \mu_A \circ \mu_A \rangle &\leq \|\mu_{B'''} * \mu_A\|_\infty \|\mu_{A_2} \circ \mu_{A_1} * \mu_A\|_1 \\ &= \|\mu_{B'''} * \mu_A\|_\infty. \qquad \square\end{aligned}$$

5. Deduction of Theorems 5, 1 and 3

Finally, in this section we deduce Theorem 5 from Theorem 14 (using Step 2) and show (using Step 1) how it implies Theorems 1 and 3. The following form of unbalancing relative to Bohr sets is sufficient.

Proposition 18. *There is a constant $c > 0$ such that the following holds. Let $\epsilon > 0$ and $p \geq 2$ be an integer. Let $B \subseteq G$ be a regular Bohr set and $A \subseteq B$ with relative density α. Let $\nu : G \to \mathbb{R}_{\geq 0}$ be supported on B_ρ, where $\rho \leq c\epsilon\alpha/\operatorname{rk}(B)$, such that $\|\nu\|_1 = 1$ and $\widehat{\nu} \geq 0$. If*

$$\|(\mu_A - \mu_B) \circ (\mu_A - \mu_B)\|_{p(\nu)} \geq \epsilon\mu(B)^{-1},$$

then there exists $p' \ll_\epsilon p$ such that

$$\|\mu_A \circ \mu_A\|_{p'(\nu)} \geq \left(1 + \tfrac{1}{4}\epsilon\right)\mu(B)^{-1}.$$

Proof. Let us write

$$g = \mu_A \circ \mu_B + \mu_B \circ \mu_A - \mu_B \circ \mu_B$$

and note that for any $p' \geq 1$

$$\begin{aligned}\|\mu_A \circ \mu_A\|_{p'(\nu)} &= \|(\mu_A - \mu_B) \circ (\mu_A - \mu_B) + \mu(B)^{-1} + g - \mu(B)^{-1}\|_{p'(\nu)} \\ &\geq \|(\mu_A - \mu_B) \circ (\mu_A - \mu_B) + \mu(B)^{-1}\|_{p'(\nu)} - \|g - \mu(B)^{-1}\|_{p'(\nu)}.\end{aligned}$$

To bound the error term, note that $\mu_A \circ \mu_B(0) = \mu(B)^{-1}$ and that by regularity (for example with Lemma A.5), for $x \in \mathrm{supp}(\nu)$,

$$\begin{aligned}|\mu_A \circ \mu_B(x) - \mu_A \circ \mu_B(0)| &\leq \alpha^{-1}\mu(B)^{-1}\|\mu_B(\cdot + x) - \mu_B\|_1 \\ &\ll \rho\,\mathrm{rk}(B)\alpha^{-1}\mu(B)^{-1} \\ &\leq \tfrac{1}{12}\epsilon\mu(B)^{-1},\end{aligned}$$

say, assuming ρ is sufficiently small, and similarly for the other terms constituting g. Hence

$$\|g - \mu(B)^{-1}\|_{p'(\nu)} \leq \|g - \mu(B)^{-1}\|_{L^\infty(\mathrm{supp}(\nu))} \leq \tfrac{1}{4}\epsilon\mu(B)^{-1}.$$

It therefore suffices to find some $p' \ll_\epsilon p$ such that

$$\|(\mu_A - \mu_B) \circ (\mu_A - \mu_B) + \mu(B)^{-1}\|_{p'(\nu)} \geq (1 + \tfrac{1}{2}\epsilon)\mu(B)^{-1}.$$

This is immediate from Lemma 7 applied to $f = \mu(B)(\mu_A - \mu_B) \circ (\mu_A - \mu_B)$. □

To deduce Theorem 5 from Theorem 14 we will need to pass from the measure μ_B to $\mu_{B'} \circ \mu_{B'} * \mu_{B''} \circ \mu_{B''}$. This is a technical issue with Bohr sets that does not arise in the $\mathbb{F}_q^n$ model case (note that these measures are identical when $B = B' = B'' = G$).

Proposition 19. *There is a constant $c > 0$ such that the following holds. Let $p \geq 2$ be an even integer. Let $f : G \to \mathbb{R}$, let $B \subseteq G$ and $B', B'' \subseteq B_{c/\mathrm{rk}(B)}$ all be regular Bohr sets. Then*

$$\|f \circ f\|_{p(\mu_{B'} \circ \mu_{B'} * \mu_{B''} \circ \mu_{B''})} \geq \tfrac{1}{2}\|f * f\|_{p(\mu_B)}.$$

Proof. By an application of Lemma A.6 with $L = 4$, $\rho = c/4\,\mathrm{rk}(B)$ and $\nu = \mu_{B'} * \mu_{B'} * \mu_{B''} * \mu_{B''}$, we have

$$\mu_B \leq 2\mu_{B_{1+4\rho}} * \nu.$$

Hence

$$\begin{aligned}\|f * f\|_{p(\mu_B)}^p &= \mathbb{E}_{x \in G}\mu_B(x) f * f(x)^p \\ &\leq 2\mathbb{E}_{x \in G}(\mu_{B_{1+4\rho}} * \nu)(x) f * f(x)^p \\ &= 2\mathbb{E}_{t \in B_{1+4\rho}}\mathbb{E}_{x \in G}\nu(x - t) f * f(x)^p.\end{aligned}$$

By averaging there exists some t such that

$$\begin{aligned}\|f * f\|_{p(\mu_B)}^p &\leq 2\mathbb{E}_{x \in G}\nu(x - t) f * f(x)^p \\ &= 2\sum_{\gamma \in \widehat{G}} \widehat{\nu}(\gamma)\gamma(-t)(\hat{f}^2)^{(p)}(\gamma) \\ &\leq 2\sum_{\gamma \in \widehat{G}} \widehat{\nu}(\gamma)(|\hat{f}|^2)^{(p)}(\gamma) \\ &= 2\|f \circ f\|_{p(\nu)}^p,\end{aligned}$$

where we have used the fact that $\widehat{\nu} \geq 0$. □

Proof of Theorem 5. We may, without loss of generality, assume that δ is sufficiently small in terms of ϵ and k. Let $p' \ll_\epsilon p$ satisfy the condition in Proposition 18. Let A' and B be the regular Bohr sets provided by Theorem 14 applied with p replaced by p' and ϵ replaced by $\epsilon/8$. We claim that this choice satisfies the conclusion of Theorem 5. It suffices to prove

$$\|(\mu_{A'} - \mu_B) * (\mu_{A'} - \mu_B)\|_{p(\mu_{k\cdot B'})} \leq \epsilon \mu(B)^{-1}.$$

Suppose not. Let $B'' = B'_\rho$ and $B''' = B''_{\rho'}$ be regular Bohr sets with $\rho = c_1/d$ and $\rho' = c_2/d$ for some sufficiently small constants $c_1, c_2 > 0$. By Proposition 19 we have, if we let $f = \mu_{A'} - \mu_B$ for brevity,

$$\|f * f\|_{p(\mu_{k\cdot B'})} \leq 2\|f \circ f\|_{p(\nu)}$$

where $\nu = \mu_{k\cdot B''} \circ \mu_{k\cdot B''} * \mu_{k\cdot B'''} \circ \mu_{k\cdot B'''}$. In particular, $\|f \circ f\|_{p(\nu)} > \frac{1}{2}\epsilon\mu(B)^{-1}$, and so by Proposition 18 (noting that ν is supported on $k(B'' + B'' + B''' + B''') \subseteq B'_{4k\rho} \subseteq B_{c/d}$) we deduce that

$$\|\mu_{A'} \circ \mu_{A'}\|_{p'(\nu)} \geq (1 + \epsilon/8)\mu(B)^{-1},$$

which contradicts the conclusion of Theorem 14, and we are done. □

Finally, to apply Theorem 5 to three-term progressions and finding long arithmetic progressions in $A + A + A$, we record the following version of the Hölder lifting Step 1.

Proposition 20. *There is a constant $c > 0$ such that the following holds. Let $\epsilon > 0$. Let $B \subseteq G$ be a regular Bohr set and $A \subseteq B$ with relative density α, and let $B' \subseteq B_{c\epsilon\alpha/\operatorname{rk}(B)}$ be a regular Bohr set and $C \subseteq B'$ with relative density γ. Either*

(1) $|\langle \mu_A * \mu_A, \mu_C\rangle - \mu(B)^{-1}| \leq \epsilon\mu(B)^{-1}$ *or*

(2) *there is some $p \ll \mathcal{L}(\gamma)$ such that $\|(\mu_A - \mu_B) * (\mu_A - \mu_B)\|_{p(\mu_{B'})} \geq \frac{1}{2}\epsilon\mu(B)^{-1}$.*

Proof. We first note that

$$\langle \mu_A * \mu_A, \mu_C\rangle = \langle (\mu_A - \mu_B) * (\mu_A - \mu_B), \mu_C\rangle + 2\langle \mu_A * \mu_B, \mu_C\rangle - \langle \mu_B * \mu_B, \mu_C\rangle.$$

By the regularity of B (more specifically Lemma A.5) and the fact that $C \subseteq B_{c\epsilon/\operatorname{rk}(B)}$, we have

$$|\langle \mu_B * \mu_B, \mu_C\rangle - \mu(B)^{-1}| \leq \|\mu_B\|_\infty \|\mu_B * \mu_C - \mu_B\|_1 \leq \tfrac{1}{8}\epsilon\mu(B)^{-1}.$$

Similarly, the fact that $C \subseteq B_{c\epsilon\alpha/\operatorname{rk}(B)}$ implies that

$$|\langle \mu_A * \mu_B, \mu_C\rangle - \mu(B)^{-1}| \leq \|\mu_A\|_\infty \|\mu_B * \mu_C - \mu_B\|_1 \leq \tfrac{1}{16}\epsilon\mu(B)^{-1}.$$

It follows that, with $f = (\mu_A - \mu_B) * (\mu_A - \mu_B)$, we have

$$|\langle \mu_A * \mu_A, \mu_C\rangle - \mu(B)^{-1} - \langle f, \mu_C\rangle| \leq \tfrac{1}{4}\epsilon\mu(B)^{-1}.$$

Therefore, if the first possibility fails, then

$$\gamma^{-1}|\langle f, 1_C\rangle_{\mu_{B'}}| = |\langle f, \mu_C\rangle| \geq \tfrac{3}{4}\epsilon\mu(B)^{-1}.$$

By Hölder's inequality, for any $p \geq 1$,

$$\|f\|_{p(\mu_{B'})}\gamma^{1-1/p} \geq |\langle f, 1_C\rangle_{\mu_{B'}}|.$$

We can choose some $p \ll \mathcal{L}(\gamma)$ such that $\gamma^{-1/p} \leq \frac{3}{2}$, and the proof is complete. □

5.1. ***Three-term arithmetic progressions.*** Theorem 1 is an immediate consequence of the following result that gives a lower bound for the number of three-term arithmetic progressions in an arbitrary set, coupled with the observation that if A contains only trivial three-term arithmetic progressions then this count is at most N.

Theorem 21 (Kelley–Meka). *If $A \subseteq \{1, \dots, N\}$ has size $|A| = \alpha N$, then A contains at least*

$$\exp(-O(\mathcal{L}(\alpha)^{12}))N^2$$

many three-term arithmetic progressions.

Proof. As usual, we begin by considering $A \subseteq \{1, \dots, N\}$ as a subset of $G = \mathbb{Z}/(2N+1)\mathbb{Z}$ — the density of this set within G is still $\asymp \alpha$, and any three-term arithmetic progression in $A \subseteq G$ yields one in $A \subseteq \{1, \dots, N\}$.

We apply Theorem 5 with $\epsilon = \frac{1}{4}$, $k = 2$, and $p = \lceil K\mathcal{L}(\alpha)\rceil$ for some large constant K. Let $A'' = A' \cap B'$. If

$$\langle \mu_{A'} * \mu_{A'}, \mu_{2\cdot A''}\rangle \geq \tfrac{1}{2}\mu(B)^{-1}$$

we are done, since the left-hand side is at most $\ll \alpha^{-3}\mu(B)^{-2}\mu(B')^{-1}$ times the number of three-term arithmetic progressions in A, and by Lemma A.4 we have

$$\mu(B)\mu(B') \geq \exp(-O_\epsilon(\mathcal{L}(\alpha)^{12})).$$

Otherwise, we are in the second case of Proposition 20, which contradicts the conclusion of Theorem 5, and we are done. □

5.2. ***Arithmetic progressions in $A+A+A$.*** As in the previous section, Theorem 3 follows immediately from the following statement for general groups, by embedding $\{1, \dots, N\}$ in a cyclic group $\mathbb{Z}/M\mathbb{Z}$ for a prime M between $2N$ and $4N$.

Theorem 22. *If $A \subseteq G$ has size αN, then $A + A + A$ contains a translate of a Bohr set B with*

$$\mathrm{rk}(B) \ll \mathcal{L}(\alpha)^9 \quad \textit{and} \quad \mu(B) \geq \exp(-O(\mathcal{L}(\alpha)^{12}).$$

In particular, if $G = \mathbb{Z}/N\mathbb{Z}$ for a prime N, then $A + A + A$ contains an arithmetic progression of length

$$\geq \exp(-O(\mathcal{L}(\alpha)^3))N^{\Omega(1/\mathcal{L}(\alpha)^9)}.$$

Proof. We apply Theorem 5 with $\epsilon = \frac{1}{4}$, $k = 1$, and $p = \lceil K\mathcal{L}(\alpha)\rceil$ for some large constant K. Let B, B' be the Bohr sets produced by that conclusion, and $A' = (A - x) \cap B$ the corresponding restricted translate of A.

We first argue that $|(A' + A') \cap B'| \geq (1 - \alpha/4)|B'|$. Indeed, otherwise if we let $C = B'\backslash(A' + A')$, then the first case of Proposition 20 is violated and the conclusion of Theorem 5 means the second also cannot hold.

Let $B'' = B'_{c\alpha/d}$, where $c > 0$ is some small constant. We argue that $B'' \subseteq A' + A' + A'$. If not, there is some $x \in B''$ such that $(A' + A') \cap (x - A') = \varnothing$, and so

$$|A' \cap (B' - x)| = |(x - A') \cap B'| \leq |B'\backslash(A' + A')| \leq \tfrac{\alpha}{4}|B'|.$$

By regularity, however, the left-hand side is at least

$$|A' \cap B'| - |B\backslash(B' - x)| \geq |A' \cap B'| - \tfrac{1}{4}\alpha|B'| \geq \tfrac{\alpha}{2}|B'|,$$

which is a contradiction.

We have found some Bohr set B'' of rank $O(\mathcal{L}(\alpha)^9)$ and density

$$\mu(B'') \geq \exp(-O(\mathcal{L}(\alpha)^{12}))$$

such that $B'' \subseteq A' + A' + A'$. It remains to note that $A' + A' + A'$ is contained in a translate of $A + A + A$ and to appeal to Lemma A.7 to find an arithmetic progression in B'' of length

$$\geq \exp(-O(\mathcal{L}(\alpha)^3))N^{\Omega(1/\mathcal{L}(\alpha)^9)}. \qquad \square$$

Appendix: Bohr sets

In abelian groups more general than $\mathbb{F}_q^n$, a useful substitute for genuine subgroups is the class of Bohr sets, introduced to additive combinatorics by Bourgain [1999]. Below we collect some standard facts about Bohr sets.

Definition A.1 (Bohr sets). For a nonempty $\Gamma \subseteq \widehat{G}$ and $\nu \in [0, 2]$ we define the Bohr set $B = \mathrm{Bohr}_\nu(\Gamma)$ as

$$\mathrm{Bohr}_\nu(\Gamma) = \{x \in G : |1 - \gamma(x)| \leq \nu \text{ for all } \gamma \in \Gamma\}.$$

We call Γ the *frequency set* of B and ν the *width*, and define the *rank* of B to be the size of Γ, denoted by $\mathrm{rk}(B)$. We note here that all Bohr sets are symmetric and contain 0.

In fact, when we speak of a Bohr set we implicitly refer to the triple

$$(\Gamma, \nu, \mathrm{Bohr}_\nu(\Gamma)),$$

since the set $\mathrm{Bohr}_\nu(\Gamma)$ alone does not uniquely determine the frequency set nor the width. When we use subset notation, such as $B' \subseteq B$, this refers only to the

set inclusion (and does not, in particular, imply any particular relation between the associated frequency sets or width functions). Furthermore, if $B = \mathrm{Bohr}_\nu(\Gamma)$ and $\rho \in (0, 1]$, then we write B_ρ for the same Bohr set with the width dilated by ρ, i.e., $\mathrm{Bohr}_{\rho\nu}(\Gamma)$, which is known as a *dilate* of B.

Bohr sets are, in general, not even approximately group-like, and may grow exponentially under addition. Bourgain [1999] observed that certain Bohr sets are approximately closed under addition in a weak sense which is suitable for our applications.

Definition A.2 (regularity[3]). A Bohr set B of rank d is regular if for all $|\kappa| \leq \frac{1}{100d}$ we have

$$(1 - 100d|\kappa|)|B| \leq |B_{1+\kappa}| \leq (1 + 100d|\kappa|)|B|.$$

We record here the useful observation, frequently used in this paper, that if $(k, |G|) = 1$ and B is a regular Bohr set of rank d then $k \cdot B$ is also a regular Bohr set of rank d (and of course the same density), simply by replacing each character in the frequency set by an appropriate dilate.

For further introductory discussion of Bohr sets see, for example, [Tao and Vu 2006, Chapter 4], in which the following basic lemmas are established.

Lemma A.3. *For any Bohr set B there exists $\rho \in \left[\frac{1}{2}, 1\right]$ such that B_ρ is regular.*

Lemma A.4. *If $\rho \in (0, 1)$ and B is a Bohr set of rank d, then $|B_\rho| \geq (\rho/4)^d |B|$.*

The following standard lemmas indicate how regularity of Bohr sets will be exploited. The following is proved as, for example, [Bloom and Sisask 2020, Lemma 4.5].

Lemma A.5. *If B is a regular Bohr set of rank d and $\mu : G \to \mathbb{R}_{\geq 0}$ is supported on B_ρ, with $\rho \in (0, 1)$, then*

$$\|\mu_B * \mu - \mu_B\|_1 \ll \rho d \|\mu\|_1.$$

The following is a minor generalisation of, for example, [Bloom and Sisask 2020, Lemma 4.7], which is stated with $\nu = \mu_{B'}^{(L)}$ for a subset $B' \subseteq B_\rho$; the proof is identical.

Lemma A.6. *There is a constant $c > 0$ such that the following holds. Let B be a regular Bohr set of rank d and $L \geq 1$ be any integer. If $\nu : G \to \mathbb{R}_{\geq 0}$ is supported on LB_ρ, where $\rho \leq c/Ld$, and $\|\nu\|_1 = 1$, then*

$$\mu_B \leq 2\mu_{B_{1+L\rho}} * \nu.$$

Finally, we note the following simple lemma, which is useful for finding arithmetic progressions.

[3]The constant 100 here is fairly arbitrary. Smaller constants are permissible.

Lemma A.7. *If N is a prime and $B \subseteq \mathbb{Z}/N\mathbb{Z}$ is a Bohr set of rank d, then B contains an arithmetic progression of length*

$$\gg |B|^{1/d}.$$

Proof. Let $\rho = 4(2/|B|)^{1/d}$, and note that by Lemma A.4 we have

$$|B_\rho| \geq (\rho/4)^d |B| = 2.$$

In particular there exists some $x \in B_\rho \backslash \{0\}$. By the triangle inequality it is clear that $\{x, \dots, \lfloor \rho^{-1} \rfloor x\} \subseteq B$, whence B contains an arithmetic progression of length $\gg \rho^{-1}$. □

Acknowledgements

Bloom is supported by a Royal Society University Research Fellowship. We would like to thank Zander Kelley and Raghu Meka for generously sharing a copy of their preprint with us, and their encouragement in writing this exposition. We would also like to thank Ben Green for helpful conversations, Ilya Shkredov for showing us an alternative argument for the proof of Lemma 7, and an anonymous referee for useful suggestions.

References

[Bateman and Katz 2012] M. Bateman and N. H. Katz, "New bounds on cap sets", *J. Amer. Math. Soc.* **25**:2 (2012), 585–613. MR Zbl

[Behrend 1946] F. A. Behrend, "On sets of integers which contain no three terms in arithmetical progression", *Proc. Nat. Acad. Sci. U.S.A.* **32** (1946), 331–332. MR Zbl

[Bloom and Sisask 2019] T. F. Bloom and O. Sisask, "Logarithmic bounds for Roth's theorem via almost-periodicity", *Discrete Anal.* (2019), art. id. 4. MR Zbl

[Bloom and Sisask 2020] T. F. Bloom and O. Sisask, "Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions", preprint, 2020. arXiv 2007.03528

[Bourgain 1999] J. Bourgain, "On triples in arithmetic progression", *Geom. Funct. Anal.* **9**:5 (1999), 968–984. MR Zbl

[Croot and Sisask 2010] E. Croot and O. Sisask, "A probabilistic technique for finding almost-periods of convolutions", *Geom. Funct. Anal.* **20**:6 (2010), 1367–1396. MR Zbl

[Elkin 2011] M. Elkin, "An improved construction of progression-free sets", *Israel J. Math.* **184** (2011), 93–128. MR Zbl

[Ellenberg and Gijswijt 2017] J. S. Ellenberg and D. Gijswijt, "On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression", *Ann. of Math.* (2) **185**:1 (2017), 339–343. MR Zbl

[Erdös and Turán 1936] P. Erdös and P. Turán, "On Some Sequences of Integers", *J. London Math. Soc.* **11**:4 (1936), 261–264. MR Zbl

[Freiman et al. 1992] G. A. Freiman, H. Halberstam, and I. Z. Ruzsa, "Integer sum sets containing long arithmetic progressions", *J. London Math. Soc.* (2) **46**:2 (1992), 193–201. MR Zbl

[Gowers 1998] W. T. Gowers, "A new proof of Szemerédi's theorem for arithmetic progressions of length four", *Geom. Funct. Anal.* **8**:3 (1998), 529–551. MR Zbl

[Green and Wolf 2010] B. Green and J. Wolf, "A note on Elkin's improvement of Behrend's construction", pp. 141–144 in *Additive number theory*, edited by D. Chudnovsky and G. Chudnovsky, Springer, 2010. MR Zbl

[Kelley and Meka 2023] Z. Kelley and R. Meka, "Strong Bounds for 3-Progressions", preprint, 2023. arXiv 2302.05537

[Roth 1953] K. F. Roth, "On certain sets of integers", *J. London Math. Soc.* **28** (1953), 104–109. MR Zbl

[Sanders 2008] T. Sanders, "Additive structures in sumsets", *Math. Proc. Cambridge Philos. Soc.* **144**:2 (2008), 289–316. MR

[Sanders 2010] T. Sanders, "Popular difference sets", *Online J. Anal. Comb.* 5 (2010), 4. MR

[Sanders 2011] T. Sanders, "On Roth's theorem on progressions", *Ann. of Math.* (2) **174**:1 (2011), 619–636. MR

[Sanders 2012a] T. Sanders, "On certain other sets of integers", *J. Anal. Math.* **116** (2012), 53–82. MR

[Sanders 2012b] T. Sanders, "On the Bogolyubov–Ruzsa lemma", *Anal. PDE* **5**:3 (2012), 627–655. MR

[Sanders 2013] T. Sanders, "The structure theory of set addition revisited", *Bull. Amer. Math. Soc.* (*N.S.*) **50**:1 (2013), 93–127. MR

[Schoen 2015] T. Schoen, "New bounds in Balog–Szemerédi–Gowers theorem", *Combinatorica* **35**:6 (2015), 695–701. MR Zbl

[Schoen and Shkredov 2013] T. Schoen and I. D. Shkredov, "Higher moments of convolutions", *J. Number Theory* **133**:5 (2013), 1693–1737. MR Zbl

[Schoen and Shkredov 2014] T. Schoen and I. D. Shkredov, "Roth's theorem in many variables", *Israel J. Math.* **199**:1 (2014), 287–308. MR Zbl

[Schoen and Sisask 2016] T. Schoen and O. Sisask, "Roth's theorem for four variables and additive structures in sums of sparse sets", *Forum Math. Sigma* **4** (2016), art. id. e5. MR Zbl

[Shkredov 2013] I. D. Shkredov, "Some new results on higher energies", *Trans. Moscow Math. Soc.* (2013), 31–63. MR Zbl

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, 2006. MR Zbl

Received 20 Feb 2023. Revised 21 Nov 2023.

THOMAS F. BLOOM:

bloom@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

OLOF SISASK:

olof.sisask@math.su.se
Department of Mathematics, Stockholm University, Stockholm, Sweden

ESSENTIAL NUMBER THEORY

msp.org/ent

See inside back cover or msp.org/ent for submission instructions.

Essential Number Theory (ISSN 2834-4634 electronic, 2834-4626 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ENT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

ESSENTIAL NUMBER THEORY

2023 vol. 2 no. 1