A PROBABILISTIC TECHNIQUE FOR FINDING ALMOST-PERIODS OF CONVOLUTIONS

ERNIE CROOT AND OLOF SISASK

Abstract. We introduce a new probabilistic technique for finding 'almost-periods' of convolutions of subsets of groups. This gives results similar to the Bogolyubov-type estimates established by Fourier analysis on abelian groups but without the need for a nice Fourier transform to exist. We also present applications, some of which are new even in the abelian setting. These include a probabilistic proof of Roth's theorem on three-term arithmetic progressions and a proof of a variant of the Bourgain-Green theorem on the existence of long arithmetic progressions in sumsets $A+B$ that works with sparser subsets of $\{1,\dots,N\}$ than previously possible. In the non-abelian setting we exhibit analogues of the Bogolyubov-Freiman-Halberstam-Ruzsa-type results of additive combinatorics, showing that product sets $A_1 \cdot A_2 \cdot A_3$ and $A^2 \cdot A^{-2}$ are rather structured, in the sense that they contain very large iterated product sets. This is particularly so when the sets in question satisfy small-doubling conditions or high multiplicative energy conditions. We also present results on structures in $A \cdot B$.

Our results are 'local' in nature, meaning that it is not necessary for the sets under consideration to be dense in the ambient group. In particular, our results apply to finite subsets of infinite groups provided they 'interact nicely' with some other set.

Contents

2010 *Mathematics Subject Classification.* 11B30.

1. Introduction and statements of results

There are many interesting problems that are concerned with counting various structures in subsets of groups. Many of these can be expressed in terms of the operation of convolution, defined for two functions $f, g : G \to \mathbb{C}$ on a group G to be the function $f * g$ given by

$$f * g(x) := \sum_{y \in G} f(y)g(y^{-1}x),$$

provided this exists for all $x \in G$. For example, many of the central objects of additive combinatorics can be expressed directly in terms of convolutions: the product set $A \cdot B = \{ab : a \in A, b \in B\}$ of two subsets of a group is precisely the support of the function $1_A * 1_B$, where 1_X denotes the indicator function of a set X, and the number of three-term arithmetic progressions in an additive set A, i.e., tuples $(a_1, a_2, a_3) \in A \times A \times A$ with $a_1 + a_3 = 2a_2$, is $1_A * 1_{-2\cdot A} * 1_A(0)$. One may think of a convolution as being a sum of a function weighted by translates of another function and, as such, one may hope that convolutions are somewhat 'smooth'. Indeed there are various senses in which this is true, and having precise notions of what it means can lead to interesting combinatorial consequences. Such results are often proved for abelian groups using the beautiful theory of Fourier analysis, where one uses the fact that convolutions and Fourier transforms interact in a very nice way. In this paper our aim is to demonstrate a new technique for establishing results about convolutions that are similar to those of Fourier analysis but that work on arbitrary groups, as well as to present applications.

1.1. **Notation.** Before we state our results let us introduce some notation—most of which is standard—directing the reader to the book [39] of Tao and Vu or the paper [38] of Tao for more details and interesting information about the concepts we use. Throughout the paper G will denote a group (which may potentially be infinite). For two subsets A and B of G we write $A \cdot B := \{ab : a \in A, b \in B\}$ for the *product set* of A and B, and A^{-1} for the collection of inverses of elements of A. Sometimes we shall omit the $\cdot$ and just juxtapose two sets to indicate the multiplication. For an element t of G we write $tA := \{ta : a \in A\}$ for the left-translate of A by t and similarly for the right-translate At. If k is a positive integer then we write $A^k := A \cdot A \cdots A$ for the k-fold product set of A, and A^{-k} for the k-fold product set of A^{-1}. For abelian groups we write the group operation additively and we give the corresponding definitions to $A + B$, $A - B$, $t + A$, kA, etc. The *multiplicative energy* between two sets A and B is defined to be the quantity

$$E(A, B) := \sum_{x \in G} 1_A * 1_B(x)^2;$$

for abelian groups this is known as the *additive energy*. For a function $f : G \to \mathbb{C}$ and a real number $p \geqslant 1$ we write $\|f\|_p^p = \|f(x)\|_p^p := \sum_{x \in G} |f(x)|^p$ for (the pth power of) the L^p norm of f provided this is finite. Thus $E(A, B) = \|1_A * 1_B\|_2^2$. A final piece of terminology: for finite groups G we say that the *density* of a set $A \subseteq G$ is $|A|/|G|$.

1.2. **The almost-periodicity results.** Our first result, then, is the following almost-periodicity-type theorem.

Proposition 1.1 (L^2-almost-periodicity, local version)**.** *Let G be a group, let $A, B \subseteq G$ be finite subsets, and let $\epsilon \in (0,1)$ be a parameter. Suppose $S \subseteq G$ is such that $|B \cdot S| \leqslant K|B|$. Then there is a set $T \subseteq S$ of size*

$$|T| \geqslant \frac{|S|}{(2K)^{9/\epsilon^2}}$$

such that, for each $t \in TT^{-1}$,

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_2^2 \leqslant \epsilon^2 |A||B|^2.$$

The condition that there should be a set S such that $|B \cdot S| \leqslant K|B|$ is what justifies the terminology 'local': one does not need B to be dense in its ambient group in order to apply the proposition effectively. All one needs is for B to interact nicely with some large set S, a condition that we say more about in §2. If one knows little about the structure of B one can still obtain useful conclusions from the proposition provided B is dense in some structured set. For example, if $G = \mathbb{Z}$ and $B \subseteq [N] := \{1, \dots, N\}$ with $|B| \geqslant \beta N$ (a case of interest in many problems) then one may take $S = [N]$ and $K = 2/\beta$. Similarly, if G is finite then one can always take $S = G$, regardless of B, which immediately gives the following corollary.

Corollary 1.2 (L^2-almost-periodicity, global version)**.** *Let G be a finite group, let $A, B \subseteq G$, and let $\epsilon \in (0,1)$ be a parameter. Suppose B has density β. Then there is a set $T \subseteq G$ of size at least $(\beta/2)^{9/\epsilon^2}|G|$ such that, for each $t \in TT^{-1}$,*

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_2^2 \leqslant \epsilon^2 |A||B|^2.$$

On an informal level these results say that convolutions are somewhat continuous: one may find a large number of translates t such that the function $1_A * 1_B$ does not change by much—in an L^2 sense—when translated by t. Having L^2-almost-periods provides one with good control in many applications, particularly those involving three-fold or higher convolutions, such as when dealing with the number of three-term progressions in a set or with a triple-fold product set $A \cdot B \cdot C$. But for certain applications involving only a single convolution it turns out that having L^p-almost-periods for a somewhat large p is more useful.

Proposition 1.3 (L^p-almost periodicity, local version)**.** *Let G be a group, let $A, B \subseteq G$ be finite subsets, and let $\epsilon \in (0,1)$ and $m \geqslant 1$ be parameters. Suppose $S \subseteq G$ is such that $|B \cdot S| \leqslant K|B|$. Then there is a set $T \subseteq S$ of size*

$$|T| \geqslant \frac{|S|}{(2K)^{50m/\epsilon}}$$

such that

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant \max\left(\epsilon^m |AB||B|^m, \|1_A * 1_B\|_m^m\right) \epsilon^m |B|^m$$

for each $t \in TT^{-1}$.

As before, this has the following 'global' corollary.

Corollary 1.4 (L^p-almost-periodicity, global version)**.** *Let G be a finite group, let $A, B \subseteq G$ be subsets, and let $\epsilon \in (0,1)$ and $m \geqslant 1$ be parameters. Suppose B has density β. Then there is a set $T \subseteq G$ of size at least $(\beta/2)^{50m/\epsilon}|G|$ such that*

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant \max\left(\epsilon^m |AB||B|^m, \|1_A * 1_B\|_m^m\right) \epsilon^m |B|^m$$

for each $t \in TT^{-1}$.

We give some further variants of the above propositions in §3. In particular one can with a slight change to the hypotheses find left-translates instead of right-translates, which may be more useful depending on the application.

Our proofs of the above propositions are of a probabilistic nature, involving a 'random sampling' procedure that finds small subsets of one of the sets that behave similarly to the set itself (in a precise sense). This procedure is the same regardless of whether the group is commutative or not, which places our method in stark contrast to the Fourier-analytic methods that are typically the port of call for dealing with almost-periodicity in abelian groups. We say more about the abelian versions of the above results and the Fourier-analytic methods that lead to them in §9, turning now instead to applications of our results.

1.3. **Applications.** We shall apply the almost-periodicity results in four directions in this paper, namely towards

(i) non-commutative analogues of the Bogolyubov-Freiman-Halberstam-Ruzsa theory that shows that sumsets are structured,
(ii) a low-density version of the Bourgain-Green theorem on long arithmetic progressions in sumsets $A + B$ in $\mathbb{Z}_p$,
(iii) a probabilistic proof of Roth's theorem on arithmetic progressions and
(iv) a new result on the approximate translation-invariance of product sets whose corresponding convolutions are 'discontinuous' in a certain sense.

We discuss each of these in turn.

Structures in product sets. A general objective in additive combinatorics is to show that sumsets in abelian groups are rather structured objects. A rather useful such result due to Bogolyubov [2] that was highlighted by Ruzsa [30] in the additive-combinatorial context shows that sets $2A - 2A$ are highly structured, particularly if A has small doubling. For non-abelian groups an analogue of this was recently proved by Sanders [33]:

Theorem 1.5. *Suppose G is a group, $A \subseteq G$ is a finite set such that $|A^2| \leqslant K|A|$ and $k \in \mathbb{N}$ is a parameter. Then there is a symmetric set S containing the identity such that*

$$S^k \subseteq A^2 \cdot A^{-2} \text{ and } |S| \geqslant \exp\left(-K^{O(k)}\right)|A|.$$

As noted in [33], this is a variant of a result used in Tao's proof [37] of a Freiman-type theorem on the structure of sets with small doubling in solvable groups. Freiman-type

results are ones that characterize subsets of groups that are group-like—more precisely subsets A of a group G that satisfy a small-doubling condition $|A^2| \leqslant K|A|$ or a small-tripling condition $|A^3| \leqslant K|A|$ for some fixed K—and there has been a concerted effort in recent years to try to establish such results in various classes of groups. In the commutative setting a rather precise and useful such characterization is provided by a theorem of Green and Ruzsa [19] that generalizes a fundamental theorem of Freiman [16]. In the non-commutative setting a number of interesting results have appeared recently [6, 7, 8, 15, 22, 27, 37, **?**], though there is not yet a unified theory. Let us remark in the context of this paper, however, that results of the form of Theorem 1.5 can be useful in proving such results: abelian results that find large Bohr sets in $2A - 2A$ form a key step in many proofs of Freiman's theorem, and Theorem 1.5 itself was recently used by Green, Sanders and Tao [20] to provide combinatorial proofs of some Freiman-type results of Hrushovski [22].

The almost-periodicity results of this paper are particularly well-suited to proving results of the form of Theorem 1.5, and doing so with reasonable bounds. Indeed, the following is a virtually immediate consequence of Proposition 1.1.

Theorem 1.6. *Suppose G is a group, $A \subseteq G$ is a finite set such that $|A^2| \leqslant K|A|$ and $k \in \mathbb{N}$ is a parameter. Then there is a symmetric set $S \subseteq A^{-1}A$ containing the identity such that*

$$S^k \subseteq A^2 \cdot A^{-2} \text{ and } |S| \geqslant \exp\left(-9k^2K \log 2K\right)|A|.$$

Furthermore, each element of S^k has at least $|A|^3/2K$ representations as $a_1a_2a_3^{-1}a_4^{-1}$ with $a_i \in A$.

Four-fold product sets of the above form are particularly pleasant to analyze, but it is not much harder to obtain a result that works with only triple product sets. To state this concisely it is convenient to introduce a small piece of non-standard terminology: for a triple (A, B, C) of finite subsets of G and an element $x \in G$ we shall say that x is *γ-popular* if $1_A * 1_B * 1_C(x) \geqslant \gamma(|A||B|)^{1/2}|C|$. That is, x is γ-popular if it can be written as a product abc with $a \in A$, $b \in B$ and $c \in C$ in at least $\gamma(|A||B|)^{1/2}|C|$ different ways. If $|A \cdot B \cdot C|$ is small then certainly there is a popular element, since

$$|A||B||C| = \sum_{x \in A \cdot B \cdot C} 1_A * 1_B * 1_C(x) \leqslant |A \cdot B \cdot C| \sup_{x \in G} 1_A * 1_B * 1_C(x)$$

(see §2), but there are of course much weaker conditions ensuring this.

Theorem 1.7. *Let G be a group, let $A_1, A_2, A_3 \subseteq G$ be finite, non-empty sets and let $k \in \mathbb{N}$ be a parameter. Suppose x is a $(1/K)$-popular element for (A_1, A_2, A_3) and that there is a set $D \subseteq G$ such that $|A_3 \cdot D| \leqslant K'|A_3|$. Then there is a symmetric set $S \subseteq DD^{-1}$ containing the identity such that*

$$xS^k \subseteq A_1A_2A_3 \text{ and } |S| \geqslant \exp\left(-36k^2K^2 \log 2K'\right)|D|.$$

In the abelian setting the non-local version of this result is in the same vein as a result of Freiman, Halberstam and Ruzsa [17] that finds long arithmetic progressions or Bohr sets in $A + A + A$ (see also [39, Theorem 4.43]); the best bounds currently known in this direction are due to Sanders [32].

For the product of two sets the situation looks rather different, a phenomenon that has been observed in many different contexts. Whereas we cannot ensure that we can find a translate of a large iterated product set in $A \cdot B$, it turns out that we can always find a translate of any small subset of a large iterated product set.

Theorem 1.8. *Let G be a group, let $A, B \subseteq G$ be finite, non-empty subsets and let $k, n \in \mathbb{N}$ be parameters. Suppose $|A \cdot B| \leqslant K|A|$ and $|B \cdot D| \leqslant K'|B|$. Then there is a symmetric set $S \subseteq DD^{-1}$ of size*

$$|S| \geqslant \exp\left(-150k^2K \log 2K' \log 2n\right)|D|$$

such that the product set $A \cdot B$ contains a left-translate of any set $P \subseteq S^k$ of size at most n.

This theorem is a straightforward consequence of the L^p-almost-periodicity of $1_A * 1_B$ given by Proposition 1.3. Our next application restricts this result to subsets of $\{1, \dots, N\}$.

Arithmetic progressions in sumsets $A+B$. Coupled with a 'structure-generation' lemma that finds arithmetic progressions in iterated sumsets kS, Theorem 1.8 quickly yields the following.

Theorem 1.9. *Let N be a positive integer and let $A, B \subseteq [N]$ be non-empty sets of sizes αN, βN. Then $A + B$ contains an arithmetic progression of length at least*

$$\tfrac{1}{2}\exp\left(c\left(\frac{\alpha \log N}{\log 4/\beta}\right)^{1/4}\right),$$

where $c > 0$ is an absolute constant.

Results of this form have a rich history, starting with the paper [4] of Bourgain. There it was shown, using a very insightful and sophisticated manipulation of sets of Fourier coefficients in the group $\mathbb{Z}_p$, that if A and B are subsets of $[N]$ of densities α and β then $A + B$ must contain an arithmetic progression of length at least

$$\exp\left(c\left((\alpha\beta \log N)^{1/3} - \log\log N\right)\right) \tag{1.1}$$

for some absolute constant $c > 0$. This bound was improved by Green [18] using a different Fourier-analytic argument to the best bound that is currently known for high-density sets, increasing the exponent $1/3$ above to $1/2$; a similar bound has since also been established by Sanders [32] using another Fourier-analytic technique. By contrast, our result yields somewhat shorter arithmetic progressions for high-density sets (where α and β are thought of as not depending on N) but is also able to deal with sets that are much smaller than previously possible. Whereas the previous bounds for the length of the arithmetic progressions one can find in $A + B$ are only non-trivial provided $\alpha\beta \geqslant C(\log\log N)^2/\log N$ for some absolute constant C, Theorem 1.9 requires only $\alpha(\log 4/\beta)^{-1} \geqslant C/\log N$. Thus, whereas at least one of the sets had to have density at least $C \log\log N/(\log N)^{1/2}$ with previous bounds, the above theorem allows us to deal with pairs of sets each of which may have density as low as $C \log\log N/\log N$. In fact, one of the sets may have density as low as $\exp(-(\log N)^c)$, which illustrates a significant difference between our results and the Fourier-analytic ones. Our proof also

adds another novelty: we are able to work directly in the group $\mathbb{Z}$, never needing to embed the sets in a group $\mathbb{Z}_p$ (as is typical). We are also able to give a local version of the result; we present this and the proofs in §6.

Roth's theorem. Our next application concerns the quantity $r_3(N)$, the largest size of a subset of the integers $\{1, \dots, N\}$ that is free from non-trivial three-term arithmetic progressions—that is, triples $(x, x+d, x+2d)$ with $d \neq 0$. As a consequence of our probabilistic proof of Proposition 1.1 we are able to establish the following version of Roth's theorem [28] by completely combinatorial means.

Theorem 1.10. *There is a function ω with $\omega(N) \to \infty$ as $N \to \infty$ such that*

$$r_3(N) \leqslant \frac{N}{(\log\log N)^{\omega(N)}}$$

for any positive integer N.

This bound for r_3 is marginally stronger than Roth's original $r_3(N) \ll N/\log\log N$, the beautiful Fourier-analytic proof of which has become a model argument in additive combinatorics. Subsequent Fourier-analytic arguments have demonstrated better bounds for r_3: the best bound currently known is due to Bourgain, who in [5] established that

$$r_3(N) \ll \frac{(\log\log N)^2}{(\log N)^{2/3}} N.$$

Roth's theorem has enjoyed many different proofs, including non-Fourier-analytic ones, and each new proof has typically offered a slightly different perspective on the problem. However, only the Fourier-analytic proofs seem to have given decent bounds: the methods that have not used Fourier analysis have generally been accompanied by tower-type bounds, establishing only that

$$r_3(N) \ll N/\log^* N;$$

see [39, Chapter 10] for references, as well as the more recent work [26]. (The iterated logarithm of N, $\log^* N$, is defined to be the number of times it is necessary to take the logarithm of N in order to get a number less than or equal to 1, and thus grows extremely slowly.) It is therefore perhaps of interest that our method manages to give bounds of a similar quality to the Fourier-analytic proofs despite not using Fourier analysis. We give the proof of Theorem 1.10 in §7.

Discontinuous convolutions. We present one final application of the probabilistic technique: the following result says that product sets of sets whose auto-convolutions are 'discontinuous' in a rather strong sense must have strong almost-periodicity properties.

Proposition 1.11. *Let A be a finite subset of a group and let $\epsilon \in (0,1)$. Suppose A has the property that every $x \in A^2$ has at least $\gamma|A|$ representations as ab with $a, b \in A$. Then there is symmetric set $S \subseteq A^{-1}A$ of size*

$$|S| \geqslant \exp\left(-(\log 8/\epsilon)(\log 2/\gamma)/\gamma^2\right)|A|$$

such that, for each $t \in S$,

$$|tA^2 \triangle A^2| \leqslant \epsilon|A^2|.$$

Clearly, if A is a subgroup of G then $1_A * 1_A$ is 'discontinuous'. But there are more complex examples: let N be a square-free composite and let $H_1 \leqslant \mathbb{Z}_N$ be the subgroup consisting of the multiples of some integer m. Let H_2 be the subgroup of multiples of N/m (which is prime to m); if we find a set $B \subseteq H_2$ such that $1_B * 1_B$ is discontinuous, then the set $A = B + H_1$ has a discontinuous convolution $1_A * 1_A$. Now suppose m is a prime that is congruent to 3 mod 4. Noting that $|H_2| = m$, if we took B to be the set of non-zero squares mod m, interpreted as a subset of H_2, then $1_B * 1_B$ is discontinuous, as $1_B * 1_B(0) = 0$ while $1_B * 1_B(x) \gg m$ for all other $x \in H_2$.

The remainder of this paper is laid out as follows. In the next section we describe some standard background material from the subject of arithmetic combinatorics. In §3 we outline the basic idea behind our method and present the proofs of our almost-periodicity results. The proofs of the results on structures in product sets are very short and we give them immediately afterwards in §4. In §5 we establish a structure-generation lemma that allows us to pass from arbitrary sets of translates to structured sets of translates in $\mathbb{Z}_p$. In §6 we give the proof of Theorem 1.9 on arithmetic progressions in sumsets, and in §7 we present our proof of Roth's theorem. We present the proof of Proposition 1.11 on discontinuous convolutions in §8, and we close in §9 with some further remarks, including a comparison with Fourier-analytic results.

1.4. **Acknowledgements.** We would like to thank Tom Sanders for many interesting and helpful conversations relating to the results of this paper, particularly in connection with Theorem 1.6 and Theorem 1.9.

2. Preliminaries on convolutions and product sets

In this section we record some useful standard results about convolutions and product sets; it may be largely skipped by those familiar with additive combinatorics. We follow Tao [38].

For functions on abelian groups the operation of convolution is commutative; this is not true in general for non-abelian groups. Convolution is, however, always bilinear and associative. A crucial link between convolutions and products is that the support of $1_{A_1} * \cdots * 1_{A_k}$ is the product set $A_1 \cdots A_k$. More precisely,

$$1_{A_1} * \cdots * 1_{A_k}(x) = |\{(a_1, \ldots, a_k) \in A_1 \times \ldots \times A_k : a_1 \cdots a_k = x\}|; \tag{2.1}$$

convolutions thus count how many representations an element of a product of k sets has a product of elements of the k sets. For pairs of sets one also has the interpretation

$$1_A * 1_B(x) = |A \cap xB^{-1}| = |B \cap A^{-1}x|.$$

For functions this change between left-translates and right-translates is illustrated by the reflection property

$$\widetilde{f * g} = \widetilde{g} * \widetilde{f} \tag{2.2}$$

where $\widetilde{f}(x) := f(x^{-1})$. Note that $\widetilde{1_X} = 1_{X^{-1}}$. Since convolutions are counts, sums of convolutions are also counts, and a full sum counts a particularly simple quantity:

$$\sum_{x \in G} 1_{A_1} * \cdots * 1_{A_k}(x) = |A_1| \cdots |A_k|.$$

Many results in this paper involve conditions on the cardinalities of product sets. For two finite sets A and B in a group G, one always has the inequalities

$$\max(|A|, |B|) \leqslant |A \cdot B| \leqslant |A||B|$$

with equality possible in various scenarios. Of course $|A \cdot B| \leqslant |G|$ as well. Of particular importance to this paper are the cases when the product set $A \cdot B$ is *small*, though precisely what this means will depend on the context. Generally we shall say that $|A \cdot B|$ is small if it is at most $K|A|$ or $K|B|$ for some fixed number K, i.e., if it is within a constant factor of being as small as it could be. One generally thinks of a condition $|A \cdot B| \leqslant K|B|$ as showing that A and B share some structure, particularly if A and B are close in size. In particular this implies that A and B must themselves be somewhat structured, as follows from [38, Lemma 3.2].

Lemma 2.1 (Ruzsa triangle inequality)**.** *Let $A, B, C \subseteq G$ be finite, non-empty subsets of a group. Then*

$$|A \cdot C^{-1}| \leqslant \frac{|A \cdot B^{-1}||B \cdot C^{-1}|}{|B|}.$$

Our almost-periodicity theorems are thus particularly effective when one of the sets A and B is structured in the sense of having *small doubling* $|A^2| \leqslant K|A|$ or $|B^2| \leqslant K|B|$, or *small differencing* $|A \cdot A^{-1}| \leqslant K|A|$ or $|B \cdot B^{-1}| \leqslant K|B|$, for some small, fixed K. In abelian groups the following result is particularly useful for bounding sizes of sumsets; see [39, Chapter 6] for references and a proof.

Theorem 2.2 (Plünnecke-Ruzsa inequality)**.** *Let A and B be finite subsets of an abelian group, and suppose $|A + B| \leqslant K|A|$. Then*

$$|nB - mB| \leqslant K^{n+m}|A|$$

for all integers $m, n \geqslant 1$.

As previously noted, however, one can substitute the above notion of structure for a much weaker one: that of being dense in a structured set (such as the ambient group). There are other ways in which one can weaken the notion of structure used; recall our definition of the multiplicative energy between two sets:

$$E(A, B) = \sum_{x \in G} 1_A * 1_B(x)^2.$$

If the product set $A \cdot B$ is small compared to either A or B, in the sense that it has size at most $K|A|$ or $K|B|$, then $E(A, B)$ is large:

$$E(A, B) \geqslant \frac{1}{|A \cdot B|} \left(\sum_{x \in G} 1_A * 1_B(x) \right)^2 = \frac{|A|^2|B|^2}{|A \cdot B|}, \tag{2.3}$$

where the inequality follows from the Cauchy-Schwarz inequality. On the other hand, the condition $E(A,A) \geqslant |A|^3/K$ need not imply that $|A^2|$ is small, even in the abelian setting. We mention that there is a partial converse, however, that could be used in certain applications to keep the effectiveness of the bounds of this paper in the case when the sets in question have high multiplicative energy instead of small doubling: this is known as the Balog-Szemerédi-Gowers theorem. We point the interested reader to [39, Chapter 2] and [38, Section 5] for more information on this.

Many of the above properties have analogues for functions more general than indicator functions, of course. The distinction between indicator functions and more general functions does not tend to be particularly important in practice; see the comments in §9.

3. Proofs of the main propositions

Each of our propositions on almost-periodicity has to do with finding translates by which the convolution $1_A * 1_B$ is approximately invariant in some norm. There are two basic ideas behind the proofs of these propositions. The first is that if one selects a small random subset $C \subseteq A$, then with high probability the convolution $1_C * 1_B$ will approximate the function $\frac{|C|}{|A|}1_A * 1_B$. This means that the approximation will hold for many subsets C of A; so many, in fact, that there must be some relations amongst the sets: lots of them must in fact be *translates* of one another. These translates will then correspond to translates that leave $1_A * 1_B$ approximately invariant (in the appropriate norm).

Surprisingly little background is needed to prove Proposition 1.1; all we shall assume is some basic familiarity with the probabilistic method—see for example [1] or [39] for more details on this. We shall prove the following equivalent version of Proposition 1.1; the equivalence follows immediately from the reflection identity (2.2).

Proposition 3.1 (L^2-almost-periodicity, left-translates)**.** *Let G be a group, let $A, B \subseteq G$ be finite subsets, and let $\epsilon \in (0,1)$ be a parameter. Suppose $S \subseteq G$ is such that $|S \cdot A| \leqslant K|A|$. Then there is a set $T \subseteq S^{-1}$ of size*

$$|T| \geqslant \frac{|S|}{(2K)^{9/\epsilon^2}}$$

such that, for each $t \in TT^{-1}$,

$$\|1_A * 1_B(tx) - 1_A * 1_B(x)\|_2^2 \leqslant \epsilon^2|A|^2|B|.$$

Proof. Let k be an integer between 1 and $|A|/2$ that we shall fix later and let C be a random subset of A of size k, chosen uniformly out of all such sets. Let us write $\mu_C := 1_C \cdot |A|/k$ for a normalized version of the indicator function of C. It is easy to see that $\mathbb{E}\mu_C * 1_B(x) = 1_A * 1_B(x)$ for each $x \in G$ and that the variance

$$\mathbf{Var}(\mu_C * 1_B(x)) = \mathbb{E}|\mu_C * 1_B(x) - 1_A * 1_B(x)|^2$$

satisfies

$$\mathbf{Var}(\mu_C * 1_B(x)) \leqslant \tfrac{|A|}{k}1_A * 1_B(x).$$

Summing this inequality over all $x \in A \cdot B$, the support of $1_A * 1_B$, we obtain

$$\mathbb{E}\|\mu_C * 1_B(x) - 1_A * 1_B(x)\|_2^2 \leqslant |A|^2|B|/k. \tag{3.1}$$

Let us say that a set $C \in \binom{G}{k}$ *approximates* A if the bound

$$\|\mu_C * 1_B(x) - 1_A * 1_B(x)\|_2^2 \leqslant 2|A|^2|B|/k$$

holds. By (3.1) and Markov's inequality we thus have that

$$\mathbb{P}_{C \in \binom{A}{k}}(C \text{ approximates } A) \geqslant 1/2, \tag{3.2}$$

where $\mathbb{P}_{C \in \binom{X}{k}}$ refers to the uniform distribution on k-sets in a set X.

We now consider k-sets C chosen uniformly at random from $Y := S \cdot A$ instead of A. Let $t \in S^{-1}$. Clearly

$$\mathbb{P}_{C \in \binom{Y}{k}}(tC \text{ approximates } A) = \mathbb{P}_{C \in \binom{tY}{k}}(C \text{ approximates } A),$$

and since $A \subseteq tY$ we see that this is at least

$$\binom{|A|}{k}\binom{|S \cdot A|}{k}^{-1} \mathbb{P}_{C \in \binom{A}{k}}(C \text{ approximates } A).$$

By (3.2) and the hypothesis that $|S \cdot A| \leqslant K|A|$, then, we have that

$$\mathbb{P}_{C \in \binom{Y}{k}}(tC \text{ approximates } A) \geqslant \frac{1}{(2K)^k}.$$

Summing this inequality over all $t \in S^{-1}$ thus gives

$$\mathbb{E}_{C \in \binom{Y}{k}}|\{t \in S^{-1} : tC \text{ approximates } A\}| \geqslant \frac{|S|}{(2K)^k}.$$

In particular there exists a set C for which the set

$$T := \{t \in S^{-1} : tC \text{ approximates } A\}$$

has size at least $|S|/(2K)^k$. For this set C we have

$$\|\mu_C * 1_B(x) - 1_A * 1_B(tx)\|_2^2 \leqslant 2|A|^2|B|/k$$

for each $t \in T$, whence

$$\|1_A * 1_B(tx) - 1_A * 1_B(x)\|_2^2 \leqslant 8|A|^2|B|/k$$

for each $t \in TT^{-1}$ by the triangle inequality. The proposition now follows upon choosing $k := \lceil 8/\epsilon^2 \rceil$. (Note that the conclusion of the proposition is trivial if $k > |A|/2$.) □

We need to argue only a little more subtly in order to establish the analogous estimate for higher L^p norms: we just make use of higher moments than the variance. In order to do this we shall need some more information about random variables of the type $1_C * 1_B(x)$ considered above. Since $1_C * 1_B(x) = |C \cap xB^{-1}|$, a moment's thought reveals that this random variable follows a *hypergeometric distribution*: a random variable X is said to follow a hypergeometric distribution with parameters N, M and k if

$$\mathbb{P}(X = j) = \binom{M}{j}\binom{N-M}{k-j}\Big/\binom{N}{k}$$

for each integer $j \geqslant 0$. Thus one may think of X as counting the number of marked objects one obtains when selecting k objects randomly and without replacement from a population of N objects, a total M of which are marked. The proof of the following bounds on the moments of hypergeometrically distributed random variables is elementary, though somewhat tangential to our main arguments, so we postpone it till Appendix A.

Lemma 3.2. *Let $m \geqslant 1$ and suppose that X follows a hypergeometric distribution with parameters N, M and k as above. Then*

$$\mathbb{E}|X - \tfrac{kM}{N}|^{2m} \leqslant 2\left(3m\tfrac{kM}{N} + m^2\right)^m.$$

With these estimates in hand the proof of Proposition 1.3 is straightforward. Again we prove the following trivially equivalent version.

Proposition 3.3 (L^p-almost-periodicity, left-translates)**.** *Let G be a group, let $A, B \subseteq G$ be finite subsets, and let $\epsilon \in (0,1)$ and $m \geqslant 1$ be parameters. Suppose $S \subseteq G$ is such that $|S \cdot A| \leqslant K|A|$. Then there is a set $T \subseteq S^{-1}$ of size*

$$|T| \geqslant \frac{|S|}{(2K)^{50m/\epsilon}}$$

such that

$$\|1_A * 1_B(tx) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant \max\left(\epsilon^m |A \cdot B||A|^m, \|1_A * 1_B\|_m^m\right) \epsilon^m |A|^m$$

for each $t \in TT^{-1}$.

Proof. We follow the proof of Proposition 3.1, letting C be a random subset of A of size k for some k that is to be fixed. Fix an element $x \in G$. As alluded to above, the random variable $1_C * 1_B(x)$ follows a hypergeometric distribution:

$$\mathbb{P}(1_C * 1_B(x) = j) = \binom{M}{j}\binom{|A| - M}{k - j}\Big/\binom{|A|}{k}$$

where $M := 1_A * 1_B(x) = |A \cap xB^{-1}|$, the probability being nothing but the proportion of k-sets C in A that contain precisely j elements from $A \cap xB^{-1}$. Applying Lemma 3.2 therefore tells us that

$$\mathbb{E}\big|1_C * 1_B(x) - \tfrac{k}{|A|}1_A * 1_B(x)\big|^{2m} \leqslant 2\left(3mk \cdot 1_A * 1_B(x)/|A| + m^2\right)^m$$

or, using the notation $\mu_C := \frac{|A|}{k}1_C$,

$$\mathbb{E}|\mu_C * 1_B(x) - 1_A * 1_B(x)|^{2m} \leqslant 2(m|A|/k)^m \left(3 \cdot 1_A * 1_B(x) + m|A|/k\right)^m.$$

Summing over all $x \in A \cdot B$ then yields

$$\mathbb{E}\|\mu_C * 1_B(x) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant 2(m|A|/k)^m \sum_{x \in A \cdot B} \left(3 \cdot 1_A * 1_B(x) + m|A|/k\right)^m,$$

the right-hand side of which we denote by λ. From this it follows by Markov's inequality that

$$\mathbb{P}\big(\|\mu_C * 1_B(x) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant 2\lambda\big) \geqslant 1/2.$$

We may now argue exactly as in the proof of Proposition 3.1, replacing the L^2-version of approximation there with this L^{2m}-version. We thus obtain a set $C \subseteq S \cdot A$ of size k such that the set

$$T := \{t \in S^{-1} : \|\mu_C * 1_B(x) - 1_A * 1_B(tx)\|_{2m}^{2m} \leqslant 2\lambda\}$$

has size at least $|S|/(2K)^k$. The result now follows from the triangle inequality upon noting the bound

$$\lambda \leqslant 2(m|A|/k)^m 3.05^m \max\left(\|1_A * 1_B\|_m, 20m|A \cdot B||A|/k)\right)^m$$

and choosing $k := \lceil 49m/\epsilon \rceil$. □

Remark 3.4. We have not attempted to optimize the constant 50 that appears in the exponent of the density of the set T in this proposition; one can certainly reduce it, though any such reduction would be largely irrelevant for our applications.

4. Structures in product sets

In this section we provide proofs of the applications discussed in the first part of §1.3. These results were all versions of the statement that product sets are structured objects, with various meanings. Theorem 1.6 said that sets $A^2 \cdot A^{-2}$ are structured in the sense that they contain large iterated product sets; this is perhaps the most straightforward consequence of Proposition 3.1:

Proof of Theorem 1.6. Set $\epsilon := 1/k\sqrt{K}$ and apply Proposition 3.1 to A with $B = S = A$ to obtain a set $T \subseteq A$ of size at least $|A|/(2K)^k$ such that

$$\|1_A * 1_A(tx) - 1_A * 1_A(x)\|_2^2 \leqslant \epsilon^2 |A|^3$$

for each $t \in TT^{-1}$. Write $S := TT^{-1}$. By the triangle inequality we then have

$$\|1_A * 1_A(tx) - 1_A * 1_A(x)\|_2^2 \leqslant |A|^3/K$$

for each $t \in S^k$. The left-hand side of this inequality can be expanded as

$$\begin{aligned} &2\sum_{x \in G} 1_A * 1_A(x)^2 - 2\sum_{x \in G} 1_A * 1_A(tx) 1_A * 1_A(x) \\ &\quad = 2\left(E(A,A) - 1_A * 1_A * 1_{A^{-1}} * 1_{A^{-1}}(t)\right). \end{aligned}$$

Since A has small doubling, it also has large multiplicative energy by (2.3): $E(A,A) \geqslant |A|^3/K$. Hence

$$1_A * 1_A * 1_{A^{-1}} * 1_{A^{-1}}(t) \geqslant |A|^3 (1/K - 1/2K) \geqslant |A|^3/2K.$$

Since $1_A * 1_A * 1_{A^{-1}} * 1_{A^{-1}}$ has support $A^2 \cdot A^{-2}$, we thus have that $S^k \subseteq A^2 \cdot A^{-2}$ as desired. Furthermore, each element $t \in S^k$ has many representations as products in the way claimed, as follows from (2.1). □

We record the following more general version of Theorem 1.6; the proof is the same except we do not specialize all the parameters when applying Proposition 3.1.

Theorem 4.1. *Let G be a group, let $A, B \subseteq G$ be finite, non-empty subsets and let $k \in \mathbb{N}$ be a parameter. Suppose $E(A,B) \geqslant |A|^2|B|/K$ and that $|D \cdot A| \leqslant K'|A|$ for some set $D \subseteq G$. Then there is a symmetric set $S \subseteq D^{-1}D$ containing the identity such that*

$$S^k \subseteq A \cdot B \cdot B^{-1} \cdot A^{-1} \text{ and } |S| \geqslant \exp\left(-9k^2K \log 2K'\right)|D|.$$

Furthermore, each element of S^k has at least $|A|^2|B|/2K$ representations as $a_1b_1b_2^{-1}a_2^{-1}$ with $a_i \in A$, $b_i \in B$.

Note that this really does generalize Theorem 1.6 by (2.3).

Theorem 1.7 dealt with the product of three sets under the assumption of the existence of a 'popular element'. Note that there are various conditions that will ensure the existence of a popular element for a triple of sets (A,B,C): $A \cdot B \cdot C$ being small will certainly do, as will $\|1_A * 1_B * 1_C\|_2$ being large. The condition $E(A,B) \geqslant |A|^2|B|/K$ is also a popularity-type condition, $E(A,B)$ equalling $1_A * 1_B * 1_{B^{-1}} * 1_{A^{-1}}(1)$, and the pigeonhole principle shows that if the multiplicative energy $E(A,B)$ is large then there is a popular element for the triple (B, B^{-1}, A^{-1}).

Proof of Theorem 1.7. Recall that we are given three finite sets A_1, A_2 and A_3, a 'popular' element x such that

$$1_{A_1} * 1_{A_2} * 1_{A_3}(x) \geqslant (|A_1||A_2|)^{1/2}|A_3|/K,$$

and a set D such that $|A_3 \cdot D| \leqslant K'|A_3|$. Apply Proposition 1.1 to the sets $A = A_2$ and $B = A_3$ with $\epsilon := 1/2kK$ to obtain a set T of size at least $|D|/(2K')^{36k^2K^2}$ such that

$$\|1_{A_2} * 1_{A_3}(yt) - 1_{A_2} * 1_{A_3}(y)\|_2^2 \leqslant \epsilon^2|A_2||A_3|^2$$

for each $t \in S := TT^{-1}$. Thus for each $t \in S^k$ we have

$$\|1_{A_2} * 1_{A_3}(yt) - 1_{A_2} * 1_{A_3}(y)\|_2^2 \leqslant |A_2||A_3|^2/4K^2.$$

Let $t \in S^k$. Then

$$\begin{aligned}
&|1_{A_1} * 1_{A_2} * 1_{A_3}(xt) - 1_{A_1} * 1_{A_2} * 1_{A_3}(x)| \\
&\quad = \left|\sum_{y \in G} 1_{A_1}(y)\left(1_{A_2} * 1_{A_3}(y^{-1}xt) - 1_{A_2} * 1_{A_3}(y^{-1}x)\right)\right| \\
&\quad \leqslant |A_1|^{1/2}\|1_{A_2} * 1_{A_3}(yt) - 1_{A_2} * 1_{A_3}(y)\|_2,
\end{aligned}$$

the inequality being an application of the Cauchy-Schwarz inequality. Thus

$$|1_{A_1} * 1_{A_2} * 1_{A_3}(xt) - 1_{A_1} * 1_{A_2} * 1_{A_3}(x)| \leqslant (|A_1||A_2|)^{1/2}|A_3|/2K.$$

Since x is a $(1/K)$-popular element and $t \in S^k$ was arbitrary, this completes the proof. □

We turn now to the case of two sets. Theorem 1.8 is a special case of the following result, which has the advantage of giving stronger results in the situation when $|A \cdot B|$ is not small but the multiplicative energy $E(A,B)$ is still large.

Theorem 4.2. *Let G be a group, let $A, B \subseteq G$ be finite, non-empty subsets and let $k, n \in \mathbb{N}$ be parameters. Suppose that*

(i) $E(A,B) \geqslant |A||B|^2/K_1$,
(ii) $|A \cdot B| \leqslant K_2|A|$ *and*
(iii) $|B \cdot S| \leqslant K_3|B|$.

Then there is a set $T \subseteq S$ *of size*

$$|T| \geqslant \exp\left(-150k^2(K_1K_2)^{1/2}(\log 2K_3)(\log 2n)\right)|S|$$

such that the product set $A \cdot B$ *contains a left-translate of any set* $P \subseteq (TT^{-1})^k$ *of size at most* n.

Proof. We may assume that $n \geqslant 2$. Set $m := \log 2n$, define γ by requiring $\|1_A * 1_B\|_m^m = \gamma^m|AB||B|^m$ and set $\epsilon := \gamma/ek^2$. Applying Proposition 1.3 to A and B with these parameters gives us a set $T \subseteq D$ with

$$|T| \geqslant \frac{|S|}{(2K_3)^{50ek^2(\log 2n)/\gamma}}$$

such that

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_{2m}^{2m} \leqslant \epsilon^m|B|^m\|1_A * 1_B\|_m^m$$

for each $t \in TT^{-1}$. Let $P \subseteq (TT^{-1})^k$ be a set of size at most n. Suppose for a contradiction that $A \cdot B$ does not contain a left-translate of P. Then for every $x \in G$ there must be an element $t \in P$ for which $xt \notin A \cdot B$, i.e., for which $1_A * 1_B(xt) = 0$. Hence

$$\begin{aligned} nk^{2m}\epsilon^m|B|^m\|1_A * 1_B\|_m^m &\geqslant \sum_{t\in P}\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_{2m}^{2m} \\ &\geqslant \sum_{x\in G} 1_A * 1_B(x)^{2m}. \end{aligned}$$

By the Cauchy-Schwarz inequality this is at least $\|1_A * 1_B\|_m^{2m}/|AB|$. Recalling the definition of ϵ and m then gives the required contradiction; hence there must be some element x for which $xP \subseteq A \cdot B$. The result now follows upon noting that $\gamma \geqslant 1/(K_1K_2)^{1/2}$; this follows from the Cauchy-Schwarz inequality. □

Remark 4.3. The constant 150 in the conclusion should not be taken seriously; it can obviously be improved.

Remark 4.4. If $|A \cdot B|$ is not small compared to $|A|$ then the conclusion of the theorem becomes much less effective. If one still has the energy condition $E(A,B) \geqslant |A||B|^2/K$ and A and B are of a similar size then one can use the Balog-Szemerédi-Gowers theorem [38, Theorem 5.2] to obtain large subsets $A' \subseteq A$ and $B' \subseteq B$ that one can then apply the theorem to effectively; this would yield better bounds than using a large value of K_2 directly. We omit the details.

5. Obtaining structured sets of translates

While Propositions 1.1 and 1.3 yield very large sets of translates for which $1_A * 1_B$ is approximately translation-invariant, one often needs these sets to be structured as well. Indeed, for the abelian applications in this paper we shall need to find *arithmetic*

progressions of such translates. With Fourier-analytic methods the existence of an arithmetic progression of almost-periods is usually easy to obtain since one usually gets a Bohr set of translates, but we do not have this convenience. Instead we shall generate the structure by repeated set-addition.

We say that an arithmetic progression P in an abelian group has length k if P can be written as

$$P = \{a, a+d, \dots, a+(k-1)d\}$$

for some non-zero element d. Note that this notion may be somewhat degenerate in some groups.

Lemma 5.1. *Let G be an abelian group, let $S \subseteq G$ be a finite subset that satisfies $|S+S| \leqslant K|S|$ or $|S-S| \leqslant K|S|$ and let $k > 2$ be an integer. Suppose $A \subseteq S$ satisfies $|A| \geqslant \delta|S|$. If*

$$\delta > K^{3k/2}/|S|^{1/(k+1)}$$

then the set $kA-kA$ contains a symmetric arithmetic progression of length at least 2^{k+2} passing through 0, with non-zero step $d \in A-A$.

To prove this we require a simple preliminary result. (This is not required if $K = 1$, as would be the case if X is a group.)

Lemma 5.2. *Let G be an abelian group, and let $A \subseteq G$ be a finite subset satisfying $|A+A| \leqslant K|A|$ or $|A-A| \leqslant K|A|$. Let $k \in \mathbb{N}$. Then*

$$|A - 2^k \cdot A| \leqslant K^{3k}|A|.$$

Here we write $\lambda \cdot A$ for the dilate $\{\lambda a : a \in A\}$. This result is Theorem 15 of Bukh [9] specialized to the case $\lambda = -2$. We include the short proof for completeness.

Proof. By the Ruzsa triangle inequality, Lemma 2.1, we have that

$$|A - 2^k \cdot A| \leqslant \frac{|A - 2\cdot A||2\cdot A - 2^k \cdot A|}{|A|} \leqslant \frac{|A - 2\cdot A||A - 2^{k-1} \cdot A|}{|A|}.$$

Hence

$$\frac{|A - 2^k \cdot A|}{|A|} \leqslant \left(\frac{|A - 2\cdot A|}{|A|}\right)^k.$$

The lemma then follows from the instance $|A - 2\cdot A| \leqslant K^3|A|$ of the Plünnecke-Ruzsa inequality, Theorem 2.2. □

Proof of Lemma 5.1. It suffices to show that there are distinct elements $a, b \in A$ for which

$$2^j(a-b) \in A - A \text{ for each } j = 1, \dots, k,$$

for then $[-2^{k+1}, 2^{k+1}] \cdot (a-b) \subseteq kA - kA$. Upon rearranging, this is equivalent to there being distinct $a, b \in A$ and elements $x_j, y_j \in A$ for which

$$\begin{aligned} x_1 - 2a &= y_1 - 2b \\ x_2 - 4a &= y_2 - 4b \\ &\vdots \\ x_k - 2^k a &= y_k - 2^k b. \end{aligned}$$

We claim that this system of equations must have a solution with $a \neq b$. Indeed, for each $(k+1)$-tuple of elements $(a, x_1, \dots, x_k) \in A^{k+1}$, set

$$f(a, x_1, \dots, x_k) := (x_1 - 2a, x_2 - 4a, \dots, x_k - 2^k a).$$

The image of this function is a subset of $(S - 2 \cdot S) \times \cdots \times (S - 2^k \cdot S)$, which by Lemma 5.2 has size at most $K^{3k(k+1)/2}|S|^k$. So if $|A|^{k+1} > K^{3k(k+1)/2}|S|^k$, which is the case given our bound on δ, then there must be two distinct tuples $\mathbf{a} = (a, x_1, \dots, x_k)$ and $\mathbf{b} = (b, y_1, \dots, y_k)$ in A^{k+1} for which $f(\mathbf{a}) = f(\mathbf{b})$. Clearly such tuples must have $a \neq b$ and so provide a non-trivial solution to our system. □

Remark 5.3. One may wish to generate different types of structure depending on the group; for example, for problems in $\mathbb{F}_3^n$ it is more natural to generate subspaces rather than arithmetic progressions. Establishing such a result in $\mathbb{F}_3^n$ is relatively straightforward: it is easy to see that adding a symmetric subset of $\mathbb{F}_3^n$ to itself generates a subspace of dimension equal to the number of summands provided the set has enough linearly independent vectors.

The proof of Lemma 5.1 should be compared with an argument of the first-named author, Ruzsa and Schoen [12] that finds arithmetic progressions in single sumsets $A + B$, even when A and B are very sparse (much sparser than the sets considered in this paper).

Next we record a combination of Lemma 5.1 and Proposition 1.1 that will be useful to us in our proof of Roth's theorem. Recall that $[N] := \{1, \dots, N\}$.

Corollary 5.4. *Let $\delta \in (0,1)$ be a parameter and suppose that $A \subseteq [N]$ has size αN, where $\alpha > 4N^{-\delta^2/2304}$. Then there is a symmetric arithmetic progression $P \subseteq [-N/2, N/2]$ of length*

$$|P| \geqslant \exp\left(\tfrac{1}{30}\left(\frac{\delta^2 \log N}{\log 4/\alpha}\right)^{1/3}\right)$$

such that $0 \in P$ and, for each $t \in P$,

$$\|1_A * 1_A(x+t) - 1_A * 1_A(x)\|_2^2 \leqslant \delta^2 |A|^3.$$

Proof. Set

$$\begin{aligned} k &:= \left\lceil \left(\frac{\delta^2 \log N}{36 \log 4/\alpha}\right)^{1/3} \right\rceil - 2, \\ \epsilon &:= \delta/k, \end{aligned}$$

and apply Proposition 1.1 to A with $B = A$ and $S = [N]$. Note that we may certainly take $K = 2/\alpha$ since $A + [N] \subseteq [2N]$. Thus we get a set $T \subseteq [N]$ of size at least $(\alpha/4)^{9/\epsilon^2}N$ that has

$$\|1_A * 1_A(x+t) - 1_A * 1_A(x)\|_2^2 \leqslant \epsilon^2|A|^3$$

for each $t \in T - T$. Now apply Lemma 5.1 to the set T to get a symmetric arithmetic progression $P \subseteq kT - kT$ of length at least 2^{k+2}. By the triangle inequality we then have that any $t \in P$ gives

$$\|1_A * 1_A(x+t) - 1_A * 1_A(x)\|_2^2 \leqslant \delta^2|A|^3;$$

this progression would thus satisfy the conclusion of the corollary were it not for the fact that it may not be contained in $[-N/2, N/2]$. It is however contained in $[-kN, kN]$, and so we may simply select a symmetric subprogression $P' \subseteq [-N/2, N/2]$ of P of length at least $2^k/k$; this progression will then do. Note that the condition on α comes from the requirement that k be bigger than 2. □

6. Arithmetic progressions in sumsets

In this section we shall prove Theorem 1.9. Our task is thus to exhibit, for two sets A and B in $[N]$, the existence of a long arithmetic progression in the sumset $A+B$. We do this by combining Theorem 4.2—a consequence of Proposition 1.3—with Lemma 5.1.

Proof of Theorem 1.9. Set

$$k := \left\lceil \tfrac{1}{10}\left(\frac{\alpha \log N}{\log 4/\beta}\right)^{1/4} \right\rceil - 2$$

and

$$n := 2^{k+2}.$$

Apply Theorem 4.2 to A and B with these parameters and $S = [N]$. Since

$$A + B \subseteq B + [N] \subseteq [2N]$$

we may certainly take $K_2 = 2/\alpha$ and $K_3 = 2/\beta$, and we may take $K_1 = K_2$ by (2.3). This gives us a set $T \subseteq [N]$ of size δN, where

$$\delta \geqslant \exp\left(-300k^2(\log 4/\beta)(\log 2n)/\alpha\right),$$

such that $A + B$ contains a translate of any subset P of $kT - kT$ of size at most n. By Lemma 5.1 we can find an arithmetic progression P of length n in $kT - kT$ provided

$$\delta \geqslant 2^{3k/2}/N^{1/(k+1)}$$

and $k > 2$. A short calculation shows that the first condition holds, and if $k \leqslant 2$ then the conclusion of the theorem is trivial. □

Remark 6.1. In contrast to previous proofs of results of the form of Theorem 1.9, there was no need for us to embed the sets A and B in a finite group $\mathbb{Z}_p$ for some prime p larger than N in order for us to carry out our analysis. Had we performed this embedding into $\mathbb{Z}_p$, however, we would have been able to use a slight simplification of Lemma 5.1, since we would only need to use it for $K = 1$.

Remark 6.2. By a minor modification of the proof of Proposition 3.3—using the triangle inequality to get rid of the terms $1_A * 1_B(x)$ instead of $\mu_C * 1_B(x)$—one can deduce that $P \subseteq A + C$ for a very small set $C \subseteq B$. One thus needs to translate A by very few elements of B in order to generate long arithmetic progressions.

We similarly get the following local version of Theorem 1.9.

Theorem 6.3 (Arithmetic progressions in small sumsets)**.** *Suppose A and B are finite subsets of an abelian group such that*

$$|A + B| \leqslant K_1|A| \text{ and } |A + B| \leqslant K_2|B|.$$

Then $A + B$ contains an arithmetic progression of length at least

$$\tfrac{1}{2}\exp\left(c\left(\frac{\log|A|}{K_1 \log 2K_2}\right)^{1/4}\right),$$

where $c > 0$ is an absolute constant.

Proof. This proof is virtually the same as that above. Set

$$k := \left\lceil \tfrac{1}{10}\left(\frac{\log|A|}{K_1 \log 2K_2}\right)^{1/4} \right\rceil - 2$$

and $n := 2^{k+2}$. As before we apply Theorem 4.2 to A and B with these parameters, but this time with $S = A$. Thus we get a set $T \subseteq A$ of size

$$|T| \geqslant \exp\left(-150k^2K_1(\log 2K_2)(\log 2n)\right)|A|$$

such that $A + B$ contains any subset of $kT - kT$ of size at most n. By the Plünnecke-Ruzsa inequality, Theorem 2.2, we have that $|A + A| \leqslant K_1K_2^2|A|$. Another calculation now shows that we can apply Lemma 5.1 to $T \subseteq A$ to find an arithmetic progression of length n in $kT - kT$, which yields the result. (Note again that the theorem is trivial if $k \leqslant 2$.) □

Remark 6.4. Recall that arithmetic progressions may be degenerate in some groups; consider for example the group $\mathbb{F}_2^n$.

Remark 6.5. Other local versions of this result are possible: we could for example work relative to a set S of small doubling such that $|B + S| \leqslant K|B|$; this would yield slightly better bounds.

We cannot mention this topic without drawing the reader's attention to a delightful construction [29] of Ruzsa that places a limit on the potential strength of results of the above form.

Theorem 6.6. *Let $\epsilon > 0$. For every prime $p > p_0(\epsilon)$ there is a symmetric set $A \subseteq \mathbb{Z}_p$ of size at least $(1/2 - \epsilon)p$ such that $A + A$ contains no arithmetic progression of length*

$$\exp\left((\log p)^{2/3+\epsilon}\right).$$

Let us also mention that if one only wishes to find arithmetic progressions of length about $\log N$ in $A + B$ then much better results are available: one can work with much sparser sets than those considered in this paper by using the results in [12].

7. Roth's theorem

In this section we give our proof of Theorem 1.10. We shall employ a density-increment strategy, showing that if $A \subseteq \{1, \dots, N\}$ is large and contains no three-term progressions then we can find a long arithmetic progression on which A has significantly increased density. We can then iterate this argument in order to obtain a contradiction.

Let us introduce some notation before we begin. We denote the sum of a function $f : \mathbb{Z} \to \mathbb{C}$ with finite support over the three-term progressions in $\mathbb{Z}$ by $T_3(f)$; thus

$$T_3(f) := \sum_{x,y\in\mathbb{Z}} f(x)f(y)f(2y-x) = \sum_y f(y)(f*f)(2y).$$

Note that we may drop parts of subscripts when the meaning is clear. If $f = 1_A$ is the indicator function of a set then $T_3(f)$ is simply the number of three-term progressions in A. Note that this includes trivial (constant) three-term progressions and that it counts $(x, x+d, x+2d)$ separately from $(x+2d, x+d, x)$. We shall use the notation μ_X to denote the normalized indicator function $1_X/|X|$ of a finite set X. For a subset A of X we shall say that A has *density α relative to X* if $|A| = \alpha|X|$; when X is clear from the context we shall refer to α simply as the *density* of A. Finally, we write $\mathbb{E}_{x\in X} f(x) = \frac{1}{|X|}\sum_{x\in X} f(x)$ for the average of f over X.

The core of our proof of Roth's theorem lies in the following proposition.

Proposition 7.1. *Let $\epsilon > 0$ and suppose that $A \subseteq [N]$ has size αN. Then there is a symmetric arithmetic progression $P \subseteq [-N/8, N/8]$ of length at least*

$$|P| \geqslant c\exp\left(c\left(\frac{\epsilon^2\log N}{\log 4/\alpha}\right)^{1/3}\right),$$

where $c > 0$ is an absolute constant, such that

$$|T_3(1_A * \mu_P) - T_3(1_A)| \leqslant \epsilon|A|^2.$$

Proof. Let Q be the arithmetic progression obtained from Corollary 5.4 applied to A with parameter ϵ^2; thus Q is large, $Q = -Q$ and $Q \subseteq [-N/2, N/2]$. Let P be a symmetric subprogression of Q of length at least $|Q|/8$ such that $4P \subseteq Q$; thus $P \subseteq [-N/8, N/8]$. We claim that this P satisfies the conclusion of the proposition. Indeed,

$$T_3(1_A * \mu_P) = \mathbb{E}_{(y,z,w)\in P^3}\sum_x 1_A(x)1_A * 1_A(2x+2y-z-w)$$

and so

$$\begin{aligned}|T_3(1_A * \mu_P) - T_3(1_A)| &= \big|\mathbb{E}_{y,z,w\in P}\textstyle\sum_x 1_A(x)\big(1_A * 1_A(2x+2y-z-w) - 1_A * 1_A(2x)\big)\big| \\ &\leqslant |A|^{1/2}\,\mathbb{E}_{\substack{z,w\in P\\ y\in 2\cdot P}}\|1_A * 1_A(x-y-z-w) - 1_A * 1_A(x)\|_2 \\ &\leqslant \epsilon|A|^2,\end{aligned}$$

these inequalities being instances of the triangle and Cauchy-Schwarz inequalities and the fact that $P + P + 2\cdot P \subseteq Q$. □

We also require a preliminary lemma about T_3. The following lemma gives a lower bound for the minimal number of three-term progressions that a set (or a function)

can contain given upper bounds on the function r_3; it is a quantitative version of an averaging argument of Varnavides [40].

Lemma 7.2 (Varnavides' theorem)**.** *Let N be a positive integer and suppose that $f : [N] \to [0,1]$ is a function with average $\mathbb{E}_{x\in[N]}f(x) = \alpha$. Then, for any positive integer $M \leqslant N^{1/18}$,*

$$T_3(f) \geqslant \left(\alpha - \tfrac{r_3(M)+2}{M}\right) M^{-4}N^2.$$

The proof of this lemma proceeds via a double-counting argument and can be found in [13] for the case when f is the indicator function of a set. In order to pass from a result about sets, like the lemma stated in [13], to a result about a function f one can employ a standard probabilistic trick of defining a random set A in $[N]$ by letting $x \in A$ with probability $f(x)$ independently for each x. See [39, Exercise 10.1.7] for more details.

We are now ready to proceed with the main body of the proof. We shall prove Theorem 1.10 in the following equivalent form.

Theorem 7.3. *For any $c > 0$ there are positive numbers C and N_0 such that*

$$r_3(N) \leqslant CN/(\log\log N)^c$$

for all $N \geqslant N_0$.

Proof. We begin by establishing the theorem for some $c > 0$; we shall then be able to bootstrap this to establish the full result. Various inequalities in the argument will hold by the assumption that N is large enough; we shall not state this assumption explicitly each time it is used.

Let A be a subset of $\{1,\dots,N\}$ of size $\alpha N = r_3(N)$ that does not contain any non-trivial three-term progressions, and let $\epsilon > 0$ be a parameter that is to be fixed later. Applying Proposition 7.1 to A we obtain a long arithmetic progression P such that

$$|T_3(1_A * \mu_P) - T_3(1_A)| \leqslant \epsilon|A|^2. \tag{7.1}$$

Our argument will be centred around the function

$$1_A * \mu_P(x) = |A \cap (x - P)|/|P|;$$

we shall show that if $0 < \delta < 1$ is chosen appropriately then there must be an x for which

$$|A \cap (x - P)| > \delta^{-1}\alpha|P|.$$

This will form the base of our density increment argument.

Suppose, then, that $1_A * \mu_P(x) \leqslant \delta^{-1}\alpha$ for all $x \in \mathbb{Z}$. Let $f(x) := (\delta/\alpha)1_A * \mu_P(x)$, so that $0 \leqslant f(x) \leqslant 1$ for all x, $\sum_x f(x) = \delta N$, and

$$T_3(f) = (\delta/\alpha)^3 T_3(1_A * \mu_P).$$

Note also that f is supported on $A + P \subseteq [1 - N/8, 9N/8] \cap \mathbb{Z}$, an interval of size at most $5N/4$. Now, A contains only trivial three-term progressions and so $T_3(1_A) = |A|$. Thus (7.1) implies that

$$T_3(f) \leqslant 2\delta^3\epsilon N^2/\alpha \tag{7.2}$$

provided $\epsilon \geqslant 1/|A|$. On the other hand, Lemma 7.2 tells us that

$$T_3(f) \geqslant \left(\tfrac{4}{5}\delta - \tfrac{r_3(M)+2}{M}\right) M^{-4}N^2$$

provided $M \leqslant N^{1/18}$.

Let us initially pick $\delta = 9/10$. One may check by hand that $r_3(10) = 5$; by picking $M = 10$ we therefore see that $T_3(f) \geqslant c_0 N^2$ for some positive absolute constant c_0. Comparing this to (7.2) we see that we obtain a contradiction provided we pick $\epsilon = c_1\alpha$ for some small constant $c_1 > 0$. (This is permissible provided $\alpha \geqslant 1/\sqrt{c_1 N}$, which we assume.) Hence we must have that

$$|A \cap (x - P)| \geqslant \tfrac{10}{9}\alpha|P|$$

for some integer x, where P is a rather long progression. Let us assume that $\alpha \geqslant (\log N)^{-1/6}$. Then

$$|P| \geqslant \exp\left((\log N)^{1/8}\right);$$

we have thus shown that A has density at least $\frac{10}{9}\left(\frac{r_3(N)}{N}\right)$ on an arithmetic progression of length $N_1 := |P|$. We may thus rescale to obtain a set $A_1 \subseteq \{1, \dots, N_1\}$ that is also free of arithmetic progressions, but that is now much denser than the original set A.

We may now iterate this argument, obtaining a sequence of integers N_j with

$$N_j \geqslant \exp\left((\log N_{j-1})^{1/8}\right)$$

and a sequence of densities δ_j such that

$$\delta_j \geqslant \left(\tfrac{10}{9}\right)^j \left(\tfrac{r_3(N)}{N}\right),$$

the only requirements for proceeding to the next stage of the iteration being that $\delta_j \geqslant (\log N_j)^{-1/6}$ and $N_j \geqslant C$ for some absolute constant C. Since no δ_j can exceed 1, this iteration must stop at some stage K with $K \leqslant \frac{\log(N/r_3(N))}{\log 10/9}$, at which point one of these requirements must fail. From this we may deduce that

$$r_3(N) \leqslant \frac{CN}{(\log\log N)^{\frac{\log 10/9}{\log 8}}}$$

for some absolute constant C.

This proves the theorem for a fixed exponent c of $\log\log N$. We may now use this to run the argument again, except that we do not now need to rely on numerical data in order to apply Lemma 7.2 effectively. That is, we may now pick δ arbitrarily small and then find a fixed value M for which $\frac{4}{5}\delta - \frac{r_3(M)+2}{M} \geqslant \delta/2$. This means that, instead of obtaining a density increment of a factor of $\frac{10}{9}$, we may obtain an increment of an arbitrarily large factor δ^{-1}, still on a progression of length at least $\exp\left((\log N)^{1/8}\right)$ (though we now need N to be large enough in terms of δ). Following the above argument through again, this shows that

$$r_3(N) \leqslant \frac{CN}{(\log\log N)^{\frac{\log 1/\delta}{\log 8}}}$$

for $N \geqslant N_0(\delta)$ and some constant C depending on δ. □

8. Discontinuous convolutions

Finally we prove Proposition 1.11, the result about sets whose auto-convolutions are either zero or very large. This proposition does not follow directly from the almost-periodicity results; instead it uses the ideas in the proofs of those results in a slightly different way.

Proof of Proposition 1.11. We shall show that if $C \subseteq A$ is chosen at random then $CA \approx A^2$ with good probability. Indeed, let us start by picking a random set $C \subseteq A$ of size k. By the hypothesis on A, any $x \in A^2$ that satisfies $|\mu_C * 1_A(x) - 1_A * 1_A(x)| < \gamma|A|$ lies in CA, whence

$$\mathbb{P}(x \notin CA) \leqslant \mathbb{P}\left(|\mu_C * 1_A(x) - 1_A * 1_A(x)| \geqslant \gamma|A|\right) \leqslant 2e^{-2\gamma^2 k},$$

the latter inequality being a standard distributional inequality for hypergeometric distributions; see, for example, [11] (and cf. Proposition A.3). Summing this over all $x \in A^2$ we obtain the estimate

$$\mathbb{E}|\{x \in A^2 : x \notin CA\}| \leqslant 2e^{-2\gamma^2 k}|A^2|.$$

Markov's inequality therefore yields

$$\mathbb{P}\left(|A^2 \triangle CA| \leqslant \lambda|A^2|\right) \geqslant 1 - 2e^{-2\gamma^2 k}/\lambda;$$

let us pick $\lambda := 4e^{-2\gamma^2 k}$ to make this probability be at least $1/2$.

Now note that $|A^2| \leqslant |A|/\gamma$; this follows from the inequality $1_A * 1_A(x) \geqslant \gamma|A|1_{A^2}(x)$ holding for all x. As in the proof of Proposition 3.1, this means that there is a set C and a set $T \subseteq A^{-1}$ of size at least $(\gamma/2)^k|A|$ such that

$$|A^2 \triangle tCA| \leqslant \lambda|A^2|$$

for any $t \in T$. For any two elements $t_1, t_2 \in T$ we therefore have

$$|t_2 t_1^{-1} A^2 \triangle A^2| \leqslant 2\lambda|A^2|$$

by the triangle inequality. Thus we may take $S := TT^{-1}$ after choosing $k := \left\lceil \frac{\log 8/\epsilon}{2\gamma^2} \right\rceil$. □

9. Further remarks

We conclude with some remarks.

9.1. **Convolutions of functions.** Although we have focused on convolutions of sets in this paper, it is relatively easy to deduce results for convolutions of functions. Indeed, let f and $g : G \to [0,1]$ be two functions with finite supports S_f and S_g. Define random sets $A, B \subseteq G$ by stipulating that $x \in A$ with probability $f(x)$ and $x \in B$ with probability $g(x)$, all independently. One may then use a concentration inequality such as Chernoff's inequality [39, Theorem 1.8] to deduce that there is a choice of of sets $A \subseteq S_f$ and $B \subseteq S_g$ for which A has size very close to $\sum f$, B has size very close to

$\sum g$ and $|1_A * 1_B(x) - f * g(x)|$ is small for every $x \in S_1 + S_2$. An almost-periodicity result saying that

$$\|f * g(tx) - f * g(x)\|_2^2 \leqslant \epsilon^2 \left(\sum f\right) \left(\sum g\right)^2$$

for every $t \in TT^{-1}$ for a large set T then follows from the corresponding result for sets, and similarly for L^p almost-periodicity. One may then deal with arbitrary real-valued functions with finite support by rescaling. It is also likely that one can prove the almost-periodicity results directly for functions, though the statements will look slightly different. We leave the details to a future paper.

9.2. **Comparisons with Fourier-analytic results.** Our proofs of the almost-periodicity results in this paper have been combinatorial, which meant that there was no need for us to distinguish between abelian and non-abelian groups. When dealing with finite abelian groups, however, it is possible to derive results similar to Corollaries 1.2 and 1.4 using Fourier analysis. Indeed, in the abelian setting Corollary 1.2 is essentially a result of Bogolyubov [2] coupled with a result of Chang [10] on the large spectra of subsets of abelian groups; see Lemma 4.36 and (the proof of) Proposition 4.39 in [39]. An important difference between the two approaches is that Fourier analysis provides one with more information about the set T: it is a so-called Bohr set (an approximate annihilator of a set of characters in the Pontryagin dual of G), and it is well-known that Bohr sets are arithmetically structured sets. For instance, Bohr sets contain long arithmetic progressions, which means that one does not need to appeal to structure-generation results like Lemma 5.1. If one uses this as the base for the arguments of §7 (set in $\mathbb{Z}_N$ rather than $[N]$) then one can obtain a bound for $r_3(N)$ similar to that of a recently published proof of Roth's theorem due to Szemerédi [36]; indeed, our argument is in some ways quite similar to Szemerédi's. We present further details of this argument in the note [14].

It is much less clear that one can obtain an L^p-almost-periodicity result of a type similar to Corollary 1.4 for abelian groups using Fourier analysis. One may extract such a result from the paper [4] of Bourgain that exhibits the existence of long arithmetic progressions in $A+B$; indeed, the main thrust of the paper was to establish the estimates required to prove such an almost-periodicity result. Specifically one obtains a result of the following type.

Proposition 9.1. *Let G be a finite abelian group and let $\epsilon > 0$ and $m \in \mathbb{N}$ be two parameters. Suppose that $f, g : G \to [0,1]$ have averages $\mathbb{E}_{x\in G} f(x) = \alpha$ and $\mathbb{E}_{x\in G} g(x) = \beta$. Then there is a Bohr set $B = B(\Gamma, \rho)$ of rank $|\Gamma| \ll m^2 \log(1/\epsilon)/\epsilon^2$ and radius $\rho = c\epsilon^3/m$ such that*

$$\|f * g(x + t) - f * g(x)\|_{2m} \leqslant \epsilon(\alpha\beta)^{1/2}|G|^{1+1/2m}$$

for each $t \in B$.

Bourgain's argument is very elegant though also somewhat complex, relying on some quite sophisticated manipulations of sets of Fourier coefficients. We shall not say more about this here, save for making two comments. First, the set B produced by the above proposition will in general be somewhat smaller than the set T given by Corollary 1.4, but is also guaranteed to contain more structure, which is ultimately what yields

Bourgain's superior exponent of 1/3 in place of our 1/4 in the length of the arithmetic progressions one finds in $A+B$. Second, if one wishes to compare the L^p norm to $\alpha\beta$, say, then Corollary 1.4 is useful even if one of the sets A and B is rather sparse whereas Proposition 9.1 requires both sets to be quite large. More details about Proposition 9.1 may be found in the note [35].

Obtaining the local versions of our results using Fourier analysis seems harder. We note that there are tools that get around this to some extent; notably there is the 'modelling' lemma of Green and Ruzsa [19] that allows one to 'isomorphically embed' a set $A \subseteq G$ with small doubling $|A+A| \leqslant K|A|$ as a dense set $A' \subseteq G'$, where $|A'| \geqslant f(K)|G'|$. See for example the paper [34] of Sanders for an efficient proof of a local version of Roth's theorem that makes use of this lemma.

9.3. **Roth's theorem in other settings.** In this paper we proved Roth's theorem in the setting of the integers $\{1,\dots,N\}$. The Fourier-analytic proofs of Roth's theorem generally become simpler when studied in the vector space $\mathbb{F}_3^n$ over the finite field $\mathbb{F}_3$, and this holds true for our argument as well. There are two main reasons for this. One is that it is very easy to establish a result similar to Lemma 5.1 in $\mathbb{F}_3^n$, as remarked in §5. The other is that it becomes easier to run through the density increment strategy itself, since one can induct on subspaces rather than on arithmetic progressions. In particular one does not really need a result corresponding to Lemma 7.2 (Varnavides' theorem). The bounds one obtains for $r_3(\mathbb{F}_3^n)$ are not significantly better than the corresponding ones for $r_3(N)$ with $N \approx 3^n$, however.

We should mention in this context that Seva Lev has recently produced a proof [24] of the $\mathbb{F}_3^n$-version of Roth's theorem that removes the use of characters from the general framework of Meshulam's proof [25]. Lev's proof involves very different ideas to those of this paper, however.

9.4. **Extensions.** There are many possible potential extensions of the methods presented in this paper. For example, we plan to tackle locally compact groups in a future paper; in this paper we have only considered groups without any associated topological structure or, equivalently, topological groups endowed with the discrete topology. Such an extension would be natural given that Fourier analysis on locally compact abelian groups has proved such a useful theory.

Appendix A. The moments of the binomial and hypergeometric distributions

As noted in the proof of Proposition 3.3, if one selects a random k-element subset C from a set A in an ambient group G then, for any fixed element $x \in G$, the random variable $|G|1_C * 1_B(x)$ follows a hypergeometric distribution. In this appendix we prove the bounds of Proposition 3.2 on the moments of such a distribution.

Recall that X follows a hypergeometric distribution with parameters N, M and k if

$$\mathbb{P}(X = j) = \binom{M}{j}\binom{N-M}{k-j}\Big/\binom{N}{k},$$

so that X can be thought of as counting the number of marked objects selected when k objects are picked without replacement from a population of N objects, M of which are marked. If the k objects are selected *with* replacement then the number of marked objects selected follows a binomial distribution with parameters $n = k$ and $p = M/N$, and the two distributions are closely related. We have found certain estimates for the binomial distribution to be more readily available in print than the corresponding estimates for the hypergeometric distribution; the following corollary of a result of Hoeffding [21, Theorem 4] allows us to make use of these results.

Proposition A.1. *Let X follow a hypergeometric distribution as above and let Y follow a binomial distribution with parameters $n = k$ and $p = M/N$. Then for any convex, continuous function f we have*

$$\mathbb{E}f(X) \leqslant \mathbb{E}f(Y).$$

In particular, for $m \geqslant 1/2$ we have

$$\mathbb{E}|X - \tfrac{kM}{N}|^{2m} \leqslant \mathbb{E}|Y - np|^{2m}.$$

Lemma 3.2 therefore follows immediately from the following proposition.

Proposition A.2. *Let $m \geqslant 1$ and suppose that X follows a binomial distribution with parameters n and p. Then*

$$\mathbb{E}|X - np|^{2m} \leqslant 2(3mnp + m^2)^m. \tag{A.1}$$

In order to prove this we shall make use of the following deviation estimates, the type of which is often associated with the names of Bennett, Bernstein, Chernoff and Hoeffding.

Proposition A.3. *Let X follow a binomial distribution with parameters n and p. Then*

$$\mathbb{P}(X \leqslant np - t) \leqslant \exp\left(-\frac{t^2}{2np}\right) \tag{A.2}$$

$$\text{and} \quad \mathbb{P}(X \geqslant np + t) \leqslant \exp\left(-\frac{t^2}{2(np + t/3)}\right) \tag{A.3}$$

for any $t \geqslant 0$.

Proofs of these estimates may be found in [23]; see also [3] and [1]. They can be derived from an application of Markov's inequality to the random variable $e^{\lambda(X-np)}$ using the fact that the moment generating function $\mathbb{E}e^{\lambda(X-np)}$ is $e^{-\lambda pn}(pe^{\lambda} + 1 - p)^n$.

Proof of Proposition A.2. We may write

$$\mathbb{E}|X - np|^{2m} = \int_0^{\infty} \mathbb{P}(|X - np|^{2m} > t)\,\mathrm{d}t. \tag{A.4}$$

Since $\mathbb{P}(|X - np| > t) = \mathbb{P}(X < np - t) + \mathbb{P}(X > np + t)$ we may decompose the right-hand side of (A.4) as a sum of two integrals I^- and I^+ in an obvious way. The deviation estimates (A.2) and (A.3) then give

$$I^- \leqslant \int_0^\infty \exp\left(-\frac{t^{1/m}}{2np}\right) \mathrm{d}t = (2np)^m \Gamma(m+1)$$

and

$$I^+ \leqslant \int_0^\infty \exp\left(\frac{-t^{1/m}}{2(np + t^{1/2m}/3)}\right) \mathrm{d}t.$$

We split the range of integration of this latter integral into two parts I_1 and I_2 defined as follows. Let $\lambda := \frac{1}{3} + \frac{1}{3}\sqrt{1 + 6np/m}$, so that $9(\lambda m)^2/2(np + \lambda m) = 3m$; I_1 is then the integral over the range $0 \leqslant t \leqslant (3\lambda m)^{2m}$ and I_2 the integral over the remaining range. Thus

$$I_1 \leqslant \int_0^\infty \exp\left(-\frac{t^{1/m}}{2(np + \lambda m)}\right) \mathrm{d}t = (2np + 2\lambda m)^m \, \Gamma(m+1).$$

We need to take a little more care with I_2. Let us write $w := \frac{9(\lambda m)^2}{2(np+\lambda m)} = 3m$. Then

$$I_2 \leqslant \int_{(3\lambda m)^{2m}}^\infty \exp\left(-\frac{3t^{1/2m}}{2(1 + np/\lambda m)}\right) \mathrm{d}t = 2m \left(\frac{2(np + \lambda m)}{3\lambda m}\right)^{2m} \int_w^\infty z^{2m-1} e^{-z} \, \mathrm{d}z.$$

Making the change of variables $u = z - w$, this last integral becomes

$$w^{2m-1} e^{-w} \int_0^\infty \left(1 + \tfrac{u}{w}\right)^{2m-1} e^{-u} \, \mathrm{d}u \leqslant w^{2m-1} e^{-w} \int_0^\infty e^{-u(1-2m/w)} \, \mathrm{d}u,$$

the inequality holding since $1 + x \leqslant e^x$ for all x, and this expression equals $w^{2m} e^{-w}/m$. Thus

$$I_2 \leqslant 2(3\lambda m)^{2m} e^{-3m}.$$

Combining these estimates for I^- and $I^+ = I_1 + I_2$ we obtain

$$\mathbb{E}|X - np|^{2m} \leqslant (2np)^m \Gamma(m+1) + (2np + 2\lambda m)^m \Gamma(m+1) + 2(9\lambda^2 m^2/e^3)^m.$$

Using the easily-verifiable bound $\Gamma(m+1) \leqslant 2(3m/5)^m$ and the definition of λ then yields (A.1) after some routine but technical calculations. □

Remark A.4. By being a bit more careful in the above proof one could obtain somewhat smaller values for the constants appearing in the proposition, though this is not particularly important for our applications. We should also remark that, although we only required it for binomial random variables, Proposition A.2 holds even when X is a sum of more general independent Bernoulli random variables that are not necessarily identically distributed. In that setting n is the number of summands and p is $\mathbb{E}X/n$, and one may prove the result exactly as above since Proposition A.3 holds for such random variables.

References

[1] N. Alon and J. H. Spencer, *The probabilistic method,* third edition (John Wiley & Sons, 2008).
[2] N. Bogolioùboff, *Sur quelques propriétés arithmétiques des presque-périodes,* Ann. Chaire Phys. Math. Kiev **4** (1939), 185–205.
[3] B. Bollobás, *Random graphs,* second edition (CUP, 2001).
[4] J. Bourgain, *On arithmetic progressions in sums of sets of integers,* A tribute to Paul Erdős, 105–109 (CUP, 1990).
[5] J. Bourgain, *Roth's theorem on progressions revisited,* J. Anal. Math. **104** (2008), 155–192.
[6] E. Breuillard and B. Green, *Approximate groups, I: the torsion-free nilpotent case,* arXiv:0906.3598.
[7] E. Breuillard and B. Green, *Approximate groups, II: the solvable linear case,* arXiv:0907.0927.
[8] E. Breuillard, B. Green and T. Tao, *Linear Approximate Groups,* arXiv:1001.4570.
[9] B. Bukh, *Sums of dilates,* Combin. Probab. Comput. **17** (2008), no. 5, 627–639.
[10] M.-C. Chang, *A polynomial bound in Freiman's theorem,* Duke Math. J. **113** (2002), no. 3, 399–419.
[11] V. Chvátal, *The tail of the hypergeometric distribution,* Discrete Math. **25** (1979), no. 3, 285–287.
[12] E. Croot, I. Z. Ruzsa and T. Schoen, *Arithmetic progressions in sparse sumsets,* Combinatorial number theory, 157–164 (de Gruyter, Berlin, 2007).
[13] E. Croot and O. Sisask, *A new proof of Roth's theorem on arithmetic progressions,* Proc. Amer. Math. Soc. **137** (2009), no. 3, 805–809.
[14] E. Croot and O. Sisask, *A note on proving Roth's theorem using Bogolyubov's method,* notes available at `http://people.math.gatech.edu/~ecroot/expository.html`.
[15] D. Fisher, N. H. Katz and I. Peng, *On Freiman's Theorem in Nilpotent Groups,* arXiv:0901.1409.
[16] G. A. Freiman, *Foundations of a structural theory of set addition,* Translations of Mathematical Monographs, Vol 37 (AMS, 1973).
[17] G. A. Freiman, H. Halberstam and I. Z. Ruzsa, *Integer sum sets containing long arithmetic progressions,* J. London Math. Soc. (2) **46** (1992), no. 2, 193–201.
[18] B. Green, *Arithmetic progressions in sumsets,* Geom. Funct. Anal. **12** (2002), no. 3, 584–597.
[19] B. Green and I. Z. Ruzsa, *Freiman's theorem in an arbitrary abelian group,* J. Lond. Math. Soc. (2) **75** (2007), no. 1, 163–175.
[20] B. Green, T. Sanders and T. Tao, personal communication.
[21] W. Hoeffding, *Probability inequalities for sums of bounded random variables,* J. Amer. Statist. Assoc. **58** 1963 13–30.
[22] E. Hrushovski, *Stable group theory and approximate subgroups,* arXiv:0909.2190.
[23] S. Janson, *Large deviation inequalities for sums of indicator variables,* Tech. Report 1994:34, Uppsala, available at `http://www.math.uu.se/~svante/papers/index.html`.
[24] V. F. Lev, *Character-free approach to progression-free sets,* arXiv:0911.0513.
[25] R. Meshulam, *On subsets of finite abelian groups with no* 3*-term arithmetic progressions,* J. Combin. Theory Ser. A **71** (1995), no. 1, 168–172.
[26] D. H. J. Polymath, *A new proof of the density Hales-Jewett theorem,* arXiv:0910.3926.
[27] L. Pyber and E. Szabó, *Growth in finite simple groups of Lie type,* arXiv:1001.4556.
[28] K. F. Roth, *On certain sets of integers,* J. London Math. Soc. **28** (1953), 104–109.
[29] I. Z. Ruzsa, *Arithmetic progressions in sumsets,* Acta Arith. **60** (1991), no. 2, 191–202.
[30] I. Z. Ruzsa, *Generalized arithmetical progressions and sumsets,* Acta Math. Hungar. **65** (1994), no. 4, 379–388.
[31] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles,* Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, pp. 939–945, Colloq. Math. Soc. János Bolyai, 18, North-Holland, Amsterdam-New York, 1978.
[32] T. Sanders, *Additive structures in sumsets,* Math. Proc. Cambridge Philos. Soc. **144** (2008), no. 2, 289–316.
[33] T. Sanders, *On a non-abelian Balog-Szemerédi-type lemma,* arXiv:0912.0306.
[34] T. Sanders, *Three-term arithmetic progressions and sumsets,* Proc. Edinb. Math. Soc. (2) **52** (2009), no. 1, 211–233.
[35] O. Sisask, *Bourgain's proof of the existence of long arithmetic progressions in* $A+B$*,* note available at `http://www.maths.qmul.ac.uk/~olof/`.

[36] E. Szemerédi, *An old new proof of Roth's theorem,* Additive combinatorics, 51–54, CRM Proc. Lecture Notes, 43, Amer. Math. Soc., Providence, RI, 2007.
[37] T. Tao, *Freiman's theorem for solvable groups,* arXiv:0906.3535.
[38] T. Tao, *Product set estimates for non-commutative groups,* Combinatorica **28** (2008), no. 5, 547–594.
[39] T. Tao and V. H. Vu, *Additive Combinatorics* (CUP, 2006).
[40] P. Varnavides, *On certain sets of positive density,* J. London Math. Soc. **34** 1959 358–360.

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail address: ecroot@math.gatech.edu

School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, United Kingdom

E-mail address: O.Sisask@qmul.ac.uk