J. Eur. Math. Soc. (Online first) DOI 10.4171/JEMS/1271

Jacob Fox · Huy Tuan Pham · Yufei Zhao

Tower-type bounds for Roth's theorem with popular differences

Received June 5, 2020; revised July 2, 2021

Abstract. Green developed an arithmetic regularity lemma to prove a strengthening of Roth's theorem on arithmetic progressions in dense sets. It states that for every $\epsilon > 0$ there is some $N_0(\epsilon)$ such that for every $N \geq N_0(\epsilon)$ and $A \subset [N]$ with $|A| = \alpha N$, there is some nonzero d such that A contains at least $(\alpha^3 - \epsilon)N$ three-term arithmetic progressions with common difference d.

We prove that the minimum $N_0(\epsilon)$ in Green's theorem is an exponential tower of twos of height on the order of $\log(1/\epsilon)$. Both the lower and upper bounds are new. This shows that the tower-type bounds that arise from the use of a regularity lemma in this application are quantitatively necessary.

Keywords. Roth's theorem, arithmetic regularity lemma, popular difference

1. Introduction

A celebrated theorem of Roth [25] states that for each $\alpha > 0$ there is a least positive integer $N(\alpha)$ such that if $N \geq N(\alpha)$ and $A \subset [N] := \{1, \dots, N\}$ with $|A| \geq \alpha N$, then A contains a three-term arithmetic progression. Over the past six decades, there have been great efforts by many researchers toward understanding the growth of this function, and despite the introduction of important tools, the growth of $N(\alpha)$ is still not well understood. The upper bound was improved by Heath-Brown [21], Szemerédi [31], Bourgain [7, 8], Sanders [27, 28], and most recently Bloom [5] (see also [6]). The lower bound of Behrend [2] was recently improved a bit by Elkin [10] (see also Green and Wolf [20] for a shorter proof). The best known bounds are of the form $\alpha^{-\Omega(\log(\alpha^{-1}))} \leq N(\alpha) \leq 2^{O(\alpha^{-1}(\log \alpha^{-1})^4)}$.

Jacob Fox: Department of Mathematics, Stanford University, Stanford, CA, USA; jacobfox@stanford.edu

Huy Tuan Pham: Department of Mathematics, Stanford University, Stanford, CA, USA; huypham@stanford.edu

Yufei Zhao: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA; yufeiz@mit.edu

Mathematics Subject Classification (2020): Primary 11B25; Secondary 11B30, 05D10

Szemerédi [29] extended Roth's theorem to show that any dense set of integers contains arbitrarily long arithmetic progressions. Szemerédi's proof developed an early version of Szemerédi's regularity lemma [30], which gives a rough structural result for large graphs and is arguably the most powerful tool developed in graph theory. It roughly says that any graph can be equitably partitioned into a bounded number of parts so that between almost all pairs of parts, the graph behaves randomly-like. Szemerédi's proof of the regularity lemma gives an upper bound on the number of parts which is tower-type in an approximation parameter, which gives a seemingly poor bound for the various applications of the regularity lemma. For over two decades there was some hope that a substantially better bound might hold leading to better bounds in the many applications. This hope was shattered by Gowers [16], who proved that the bound on the number of parts in the regularity lemma must grow as a tower-type function. Further results improving on some aspects of the lower bound were obtained in [9, 11, 24].

It has been a major program over the last few decades to find new proofs of the various applications of Szemerédi's regularity lemma and its variants that avoid using the regularity lemma and obtain much better quantitative bounds. This program, popularized by Szemerédi and others, has been quite successful, leading to the development of powerful new methods, such as in Gowers' new proof of Szemerédi's theorem [17] which introduced higher order Fourier analysis [17], and in the resolution of many open problems in extremal combinatorics using the powerful probabilistic technique known as dependent random choice (see the survey [15]). However, until now it was unclear if one could avoid using regularity methods and obtain much better bounds in all known applications of the regularity lemma.

A simple averaging argument of Varnavides [33] shows that not only is there at least one arithmetic progression in a subset of $[N]$ of density α with N sufficiently large, but in fact it must contain a positive constant fraction $c(\alpha)$ of the three-term arithmetic progressions. It is not difficult to show that $c(\alpha)$ is rather small, with $c(\alpha) = \alpha^{\omega(1)}$. In fact, one can show that $c(\alpha)$ is closely related to $N(\alpha)$. This bound on the density of three-term progressions is much smaller than the random bound of α^3 one gets by considering a random set of density α.

Green [18] developed an arithmetic analogue of Szemerédi's regularity lemma and used it to prove the following theorem, which answered a question of Bergelson, Host and Kra [3]. It shows that while the total number of three-term arithmetic progressions can be much smaller than the random bound, there is a nonzero d for which the number of three-term arithmetic progressions with common difference d is at least roughly the random bound.

Theorem 1.1 (Green's popular progression difference theorem [18]). *For each $\epsilon > 0$ there is an integer N_0 such that for any $N \geq N_0$ and subset $A \subset [N]$ with $|A| = \alpha N$, there is a nonzero d such that A contains at least $(\alpha^3 - \epsilon)N$ three-term arithmetic progressions with common difference d.*

Similar to the graph setting, the proof of the arithmetic regularity lemma gives a tower-type upper bound on the size of the partition. Green [18] proved a tower-type lower bound

for the arithmetic regularity lemma in the setting of vector spaces over $\mathbb{F}_2$, and Hosseini, Lovett, Moshkovitz, and Shapira [22] later improved the tower height to $\Omega(\epsilon^{-1})$. Green's proof of Theorem 1.1 uses the arithmetic regularity lemma and consequently shows that N_0 can be taken to be an exponential tower of twos of height $\epsilon^{-O(1)}$. It was unknown if the tower-type bounds that come from any regularity lemma application like Theorem 1.1 are necessary. Our main theorem determines the growth of the minimum N_0 for which Theorem 1.1 holds, showing that it is an exponential tower of twos of height $\Theta(\log(1/\epsilon))$. Let $\text{tower}(m)$ denote an exponential tower of twos of height m.

Theorem 1.2. *Let $N_0(\epsilon)$ denote the smallest choice of N_0 for which Theorem 1.1 holds. There exist absolute constants $c, C > 0$ such that for all $0 < \epsilon < 1/2$,*

$$\text{tower}(c \log(1/\epsilon)) \leq N_0(\epsilon) \leq \text{tower}(C \log(1/\epsilon)).$$

This result, and the considerably easier analogous result in vector spaces over a fixed finite field [12,13], are the first examples of regularity lemma applications that require the tower-type growth.

For investigations of popular differences for other patterns, including recent results on higher-dimensional patterns, see [3,4,14,19,23,26].

1.1. Detailed statement of results

Theorem 1.2 comes in two parts, an upper bound and a lower bound. The upper bound is as follows.

Theorem 1.3 (Upper bound for intervals). *There exists a constant $C > 0$ such that the following is true. Let $\epsilon > 0$ and $N \geq \text{tower}(C \log(1/\epsilon))$. For every $A \subset [N]$, setting $\alpha = |A|/N$, there exists some positive integer d such that $x, x + d, x + 2d \in A$ for at least $(\alpha^3 - \epsilon)N$ integers x.*

The main part of the proof is an analogous result in abelian groups of odd order.

Theorem 1.4 (Upper bound for abelian groups). *There exists a constant $C > 0$ such that the following is true. Let $\epsilon > 0$ and let G be a finite abelian group of odd order with $|G| \geq \text{tower}(C \log(1/\epsilon))$. For every $A \subset G$, setting $\alpha = |A|/|G|$, there exists some $d \in G \setminus \{0\}$ such that $x, x + d, x + 2d \in A$ for at least $(\alpha^3 - \epsilon)|G|$ values of $x \in G$.*

For the lower bound for intervals, we prove the following result, which is somewhat stronger than that lower bound result claimed in Theorem 1.2.

Theorem 1.5 (Lower bound for intervals). *There exist constants $c, \alpha_0 > 0$ such that for every $0 < \alpha \leq \alpha_0$, $0 < \epsilon \leq \alpha^{12}$ and $N \leq \text{tower}(c \log(1/\epsilon))$, there exists $A \subset [N]$ with $|A| \geq \alpha N$ such that for every positive integer $d \leq N/2$, one has $x, x + d, x + 2d \in A$ for at most $(\alpha^3 - \epsilon)(N - 2d)$ integers x.*

Organization. In the next section, we introduce some helpful notation and preliminaries including some basic facts from discrete Fourier analysis. In Section 3, we give an

overview of the proof strategies for our results. In Section 4, we prove Theorems 1.3 and 1.4 giving the upper bound results. In Section 5, we give some auxiliary results for the probabilistic lower bound construction. In Section 6, we give a lower bound construction for groups that can be written as a product of prime cyclic groups with fast growing order. In Section 7, we then use this construction as an important ingredient to obtain the lower bound construction in intervals.

We often omit floor and ceiling signs when they are not crucial for clarity of presentation.

2. Notations and preliminaries

Averaging and expectation. We use E to denote the averaging operator: given a function f on a finite set S, denote

$$\mathsf{E}f = \mathsf{E}_{x\in S}[f(x)] := \frac{1}{|S|}\sum_{x\in S} f(x).$$

We may write E_x instead of $\mathsf{E}_{x\in S}$ if the domain of x is clear from context (usually over a group).

The L^p norms are defined in the usual way:

$$\|f\|_p := (\mathsf{E}[|f|^p])^{1/p}.$$

As our lower bound construction is probabilistic, we will also need to consider expectations of random variables, for which we use the usual notation $\mathbb{E}$ for expectation (note the difference in font compared to the averaging operator E).

Fourier transform and convolutions. Given a finite abelian group G, let $\widehat{G}$ denote its dual group, whose elements are characters of G, i.e., homomorphisms $\chi : G \to S^1 := \{z \in \mathbb{C} : |z| = 1\}$. The Fourier transform of $f : G \to \mathbb{C}$ is a function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by

$$\widehat{f}(\chi) := \mathsf{E}[f\overline{\chi}] = \mathsf{E}_{x\in G}[f(x)\overline{\chi(x)}].$$

We write $\chi^{1/2}$ to denote the character given by $x \mapsto \chi(x/2)$ (we will always work with odd order abelian groups so that $x/2$ makes sense).

It is often convenient to explicitly identify the dual group $\widehat{G}$ with G (they are isomorphic for finite abelian groups). For example, for $f : \mathbb{Z}_N \to \mathbb{C}$ and $r \in \mathbb{Z}_N$, we identify r with the character $\chi_r(x) = e(xr/N)$ where we use the standard notation for the complex exponential

$$e(t) := \exp(2\pi i t), \quad t \in \mathbb{R}.$$

Thus,

$$\widehat{f}(r) := \mathsf{E}_x[f(x)e(-xr/N))].$$

Likewise, for $f : \mathbb{F}_p^n \to \mathbb{C}$ and $r \in \mathbb{F}_p^n$, we identify r with the character $\chi_r(x) = e(x \cdot r/p)$, where $x \cdot r = x_1 r_1 + \cdots + x_n r_n \in \mathbb{F}_p$ is the dot product in $\mathbb{F}_p^n$. Thus,

$$\widehat{f}(r) := \mathsf{E}_x[f(x)e(-x \cdot r/p)].$$

Given two functions f and g on G, their convolution $f * g$ is defined by

$$(f * g)(x) := \mathsf{E}_{y \in G}[f(y)g(x-y)].$$

We recall several useful properties of the Fourier transform:

$$\begin{aligned}
f(x) &= \sum_{\chi \in \widehat{G}} [\widehat{f}(\chi)\chi(x)], && \text{[Fourier inversion formula]} \\
\widehat{f * g} &= \widehat{f} \cdot \widehat{g}, && \text{[Convolution identity]} \\
\mathsf{E}_{x \in G}[f(x)\overline{g(x)}] &= \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\overline{\widehat{g}(\chi)}, && \text{[Plancherel's identity]} \\
\mathsf{E}_{x \in G}[|f(x)|^2] &= \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2. && \text{[Parseval's identity]}
\end{aligned}$$

The Fourier transform is also fundamentally related to the count of 3-APs (or the count of solutions to linear equations in general), as is evident in the following key identity already used in the proof of Roth's theorem [25] (it can be easily shown by substituting the Fourier coefficients and expanding):

$$\mathsf{E}_{x,d \in G}[f(x)f(x+d)f(x+2d)] = \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)^2 \widehat{f}(\chi^{-2}). \tag{1}$$

Densities. For an abelian group G with odd order and a function $f : G \to [0,1]$, we define the *density of* 3*-APs of* f as

$$\mathsf{E}_{x,d \in G}[f(x)f(x+d)f(x+2d)].$$

We define the *density of* 3*-APs with common difference* d *of* f as

$$\mathsf{E}_{x \in G}[f(x)f(x+d)f(x+2d)].$$

For a subset A of G, when we say "density of 3-APs" of A, we mean that of its indicator function 1_A, and likewise for "density of 3-APs with common difference d" of A.

Over the interval $[N]$, we have two possible notions for the density of 3-APs with common difference d of a function $f : [N] \to [0,1]$. One can define the density of 3-APs with common difference d of f as

$$\frac{\sum_{x \in [N-2d]}[f(x)f(x+d)f(x+2d)]}{N},$$

as used in Theorem 1.3. This defines the density of 3-APs with common difference d as the average weight of the 3-APs $(x, x+d, x+2d)$ for $x \in [N]$, setting the value

of f outside $[N]$ to 0. The other possible definition of the density of 3-APs with common difference d of f is

$$\mathsf{E}_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)] = \frac{\sum_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)]}{N-2d},$$

as used in Theorem 1.5. In this case, we take the average only over 3-APs supported in $[N]$. It is easy to see that the density from the second definition is always at least the density from the first definition. In particular, the upper bound using the first definition implies the upper bound using the second definition. Similarly, the lower bound using the second definition implies the lower bound using the first definition. Thus, we give the stronger result in each case.

Constants. We use $c > 0$ and $C > 0$ to denote small and large absolute constants, though their values may differ at every instance. One could imagine attaching a unique subscript to each appearance of c and C.

3. Overview of strategy

In this section we sketch the proof ideas of our main theorems, starting with the upper bound (Theorems 1.4 and 1.3) and then followed by the lower bound (Theorem 1.5).

In both cases, we prove the "functional" versions of the theorems. That is, instead of working with subsets $A \subset G$, we work with functions $f : G \to [0, 1]$, which can also be viewed as subsets with weighted elements. A subset $A \subset G$ can be represented by its indicator function 1_A. Conversely, given a function $f : G \to [0, 1]$, we can produce from it a random subset $A \subset G$ obtained by putting each element $x \in G$ independently into A with probability $f(x)$. The resulting A has similar statistical properties to f due to concentration. Working with functions affords us greater flexibility, which is convenient for both parts of the proofs.

3.1. Upper bound

Theorems 1.3 and 1.4 follow from the functional forms below by setting $f = 1_A$.

Theorem 3.1 (Upper bound for intervals, functional version). *There exists a constant $C > 0$ such that the following holds. Let $\epsilon > 0$ and $N \geq \text{tower}(C \log(1/\epsilon))$. Let $f : [N] \to [0, 1]$ with $\mathsf{E} f = \alpha$. Then there exists $d \in G \setminus \{0\}$ such that*

$$\mathsf{E}_{x\in[N]}[f(x)f(x+d)f(x+2d)] \geq \alpha^3 - \epsilon.$$

Theorem 3.2 (Upper bound for abelian groups, functional version). *There exists a constant $C > 0$ such that the following holds. Let $\epsilon > 0$ and let G be a finite abelian group of odd order with $|G| \geq \text{tower}(C \log(1/\epsilon))$. Let $f : G \to [0, 1]$ with $\mathsf{E} f = \alpha$. Then there exists $d \in G \setminus \{0\}$ such that*

$$\mathsf{E}_{x\in G}[f(x)f(x+d)f(x+2d)] \geq \alpha^3 - \epsilon.$$

Green [18] proved the above theorems with a slightly worse bound of tower$(C(1/\epsilon)^C)$ instead of tower$(C\log(1/\epsilon))$. Let us first give a quick sketch of Green's approach. It is easier to first explain it for the finite field vector space setting $G = \mathbb{F}_p^n$ with p fixed.

Green begins by establishing a regularity lemma. Given $f : G \to [0,1]$, one finds a subspace H of $G = \mathbb{F}_p^n$ of codimension at most tower$(C(1/\epsilon)^C)$ such that inside almost all translates of H, f behaves "pseudorandomly" in the sense of having small Fourier coefficients (other than the principal "zeroth" Fourier coefficient that records the density). This subspace H is obtained iteratively, similar to the standard energy-increment proofs of regularity lemmas: starting with $H_0 = G$, at each step one checks if H_i has the desired properties, and if not, then one finds a bounded-codimensional subspace H_{i+1} of H_i witnessing the nonuniformity. Each step increases the "energy", or mean-squared density, by at least $\epsilon^{O(1)}$. As the energy can never exceed 1, the process terminates after at most $(1/\epsilon)^{O(1)}$ steps.

Once we have the above bounded-codimensional subspace H, let g be the function obtained from f by averaging f inside each translate of H. In other words, consider the convolution $g = f * \beta_H$, where β_H denotes the averaging measure on H (normalized so that $\mathsf{E}\beta_H = 1$). The regularity property of H, namely $f - g$ having small Fourier coefficients when restricted to most H-cosets, is enough to deduce that f and g have similar densities of 3-APs *with common differences lying in* H, i.e.,

$$\mathsf{E}_{x\in G, d\in H}[f(x)f(x+d)f(x+2d)] \approx \mathsf{E}_{x\in G, d\in H}[g(x)g(x+d)g(x+2d)].$$

On the other hand, g is constant along H-cosets, so that $g(x) = g(x+d) = g(x+2d)$ for all $d \in H$. Thus the final expression is $\mathsf{E}[g^3] \geq (\mathsf{E}g)^3$ by convexity, and we have $\mathsf{E}g \approx \mathsf{E}f$. Putting everything together, we have

$$\mathsf{E}_{x\in G, d\in H}[f(x)f(x+d)f(x+2d)] \geq (\mathsf{E}f)^3 - \epsilon/2. \tag{2}$$

If G is large enough, so that H is large enough, then the above inequality implies that there is some nonzero common difference $d \in H$ such that

$$\mathsf{E}_{x\in G}[f(x)f(x+d)f(x+2d)] \geq (\mathsf{E}f)^3 - \epsilon,$$

thereby showing that d is a popular common difference.

Let us now sketch how to improve the above bound to tower$(C\log(1/\epsilon))$ in the finite field setting, which had been worked out in [13]. Instead of finding an H that regularizes the function f, we simply seek to satisfy inequality (2). One then shows that if (2) is violated, then we can find a bounded-codimensional subspace of H, via an application of the weak regularity lemma at a "local" level, so that the corresponding *mean-cube density* of f (after averaging along the subspace) nearly doubles at each step (instead of merely increasing by $\epsilon^{-O(1)}$), so that the iteration process must end after $O(\log(1/\epsilon))$ steps (instead of $(1/\epsilon)^{O(1)}$ steps). Once we obtain an H satisfying (2), the rest of the argument is essentially identical.

For general abelian groups G, unlike in the finite field setting $\mathbb{F}_p^n$, the group might not have enough subgroups to run the above arguments. Instead, one uses *Bohr sets*, which play an analogous role to subgroups. Bohr sets are defined in Section 4.1, and their manipulations are much more delicate compared to subspaces, particularly as they do not have as nice closure properties. Green proved his arithmetic regularity lemma for general abelian groups using Bohr sets as basic structural objects. The strategy remains largely similar in spirit to the finite field setting, though more challenging at a technical level (e.g., the group cannot be partitioned into Bohr sets, unlike with subgroups). For instance, we obtain g from f by setting $g(x)$ to be a certain "smooth average" of f around a certain carefully chosen Bohr neighborhood of x. While the values $g(x)$, $g(x+d)$, and $g(x+2d)$ are no longer necessarily identical, they are hopefully approximately the same if d lies inside some Bohr neighborhood of 0. The rest of Green's argument is similar to the finite field vector space case.

To obtain the corresponding result for intervals, one considers embedding $[N]$ in $\mathbb{Z}_N$ and only consider Bohr sets whose elements d are all small in magnitude.

In order to improve the bound from tower$((1/\epsilon)^{O(1)})$ to tower$(O(\log(1/\epsilon)))$ for general groups and for intervals, we carefully execute a combination of the above ideas. New ideas are required to adapt the mean-cube density increment argument from [13] to Bohr sets due to complications that do not arise in the finite field setting. The proof is carried out in full detail in Section 4.

3.2. Lower bound

In this section, we give a brief overview of the proof of Theorem 1.5. We will deduce Theorem 1.5 from its functional analogue given below, where we replace the subset A by a function $f : [N] \to [0,1]$ with density α so that for any nonzero d, the density of 3-APs with common difference d of f is at most $\alpha^3(1-\epsilon)$.

Theorem 3.3 (Lower bound for intervals, functional version). *There are positive absolute constants c, α_0 such that the following holds. If $0 \le \alpha \le \alpha_0$, $0 \le \epsilon \le \alpha^7$, and $N \le$ tower$(c\log(1/\epsilon))$, then there is a function $f : [N] \to [0,1]$ with $\mathsf{E}[f] = \alpha$ such that for any integer $0 < d < N/2$,*

$$\mathsf{E}_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)] \le \alpha^3(1-\epsilon).$$

We remark that we have replaced ϵ by $\epsilon\alpha^3$, which is more convenient to work with in the lower bound construction. Since we treat α as a constant throughout, this has no effect on the behavior of the asymptotic bound we get. The proof of Theorem 1.5 assuming Theorem 3.3 follows from a standard sampling argument, which we defer to Appendix A. There, we also show that it suffices to prove Theorems 1.5 and 3.3 for $N \ge \epsilon^{-15}$.

In the following subsections, we sketch the construction of the function f in Theorem 3.3. This construction utilizes a construction over cyclic groups that can be factored into a product of groups with appropriate growth in size. The construction in this case is inspired by the recursive construction presented in [13] over finite field vector spaces.

However, several important new ideas are needed. The construction of f for such product groups is sketched in Section 3.2.1. Using this construction, we can construct f over intervals, using ideas sketched in Section 3.2.2, thus proving Theorem 3.3.

We remark that in this section we only give proof sketches without complete detail. The details of each construction and full proofs are presented in Sections 6 and 7.

3.2.1. Product groups. We first give the construction for groups that can be written as a product of appropriately growing cyclic groups of prime order.

Theorem 3.4 (Lower bound for product of growing prime cyclic groups). *Let $0 < \alpha \le 1/4$, $0 < \epsilon \le 20^{-9}$, and $G = \mathbb{Z}_n$ where n is a positive integer such that there exist distinct primes $m_1, \dots, m_s$ with $s \le \log_{150}(\epsilon^{-1/4}\alpha^6/8)$ satisfying*

- $n = \prod_{j=1}^{s} m_j$,
- $\epsilon^{-1/3}/2 < m_1 \le \epsilon^{-1/3}$, *and*
- *for $i \ge 2$, $n_{i-1}^6 < m_i < \exp(2^{-1} \cdot 64^{-2} \cdot 150^{i-1}\epsilon^{1/4} n_{i-1})/2$ where $n_i = \prod_{j=1}^{i} m_j$.*

Then there exists a function $f : G \to [0,1]$ with $\mathsf{E}[f] = \alpha$ such that for any $d \in G \setminus \{0\}$,

$$\mathsf{E}_x[f(x) f(x+d) f(x+2d)] \le \alpha^3(1-\epsilon).$$

Furthermore, $\mathsf{E}_x[f(x)^3] \le 3\alpha^3/2$ and there exists $\tilde{\alpha} \in [\alpha, \alpha(1+\epsilon^{1/4})]$ such that $f(x) = \tilde{\alpha}$ for at least a $3/4$ fraction of $x \in G$.

Here, the product structure of G and the bound on the growth of m_i allow us to conduct an iterative construction. The basic framework of the construction builds on the construction in [13], which took place in the setting of $\mathbb{F}_p^n$.

We build functions $f_1, f_2, f_3, \dots$ in this order. Here the domain of f_i is $Q_i = \prod_{j=1}^{i} \mathbb{Z}_{m_j}$. We will maintain that $\mathsf{E} f_i = \alpha$, and that for every $d \in Q_i \setminus \{0\}$, we have

$$\mathsf{E}_{x \in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] \le (1-\epsilon)(\mathsf{E} f_i)^3.$$

The function f_s thus gives the desired construction.

Suppose we have already constructed f_{i-1}. Here is how we construct $f_i : Q_i \to [0,1]$:

(i) Design a family F of functions $\mathbb{Z}_{m_i} \to [0,1]$.

(ii) Choose a subset $M_i \subset Q_{i-1}$ (with some properties).

(iii) For each $x \in M_i$, fill in values of f_i on the coset $x + \mathbb{Z}_{m_i}$ using a random function chosen from F.

(iv) For each $x \notin M_i$, set all values of f_i on the coset $x + \mathbb{Z}_{m_i}$ to be $f_{i-1}(x)$.

We refer to this process as *random modification* (the word "perturbation" was used in [13]). We will show that with positive probability, the random function f_i satisfies the desired properties.

The main difference from the finite vector space case is the choice of the family F. In [13], we first choose $g : \mathbb{F}_p \to [0,1]$ to be a multiple of the indicator of an interval of

length $2p/3$. We then choose the family $\mathcal{F}$ to be functions of the form $g(x \cdot v)$ for some nonzero $v \in \mathbb{F}_p^{m_i}$.

Here, instead, we choose a nice model function $g : \mathbb{Z}_{m_i} \to [0,1]$ satisfying certain properties to be discussed later. We choose F to consist of functions $g_{a,b} : \mathbb{Z}_{m_i} \to [0,1]$ defined by $g_{a,b}(x) = g(ax+b)$, indexed by $a, b \in \mathbb{Z}_{m_i}$ with $a \neq 0$.

We denote elements of Q_i by $x = (x_1, \dots, x_i)$ where $x_j \in \mathbb{Z}_{m_j}$. We write $x_{[i-1]} = (x_1, \dots, x_{i-1}) \in Q_{i-1}$. We can prove that there is a constant $c > 0$ such that for any $x \in Q_i$ with $x_{[i-1]} \in M_i$, and $d \in Q_i \setminus \{0\}$ such that $d_{[i-1]} = 0$,

$$\begin{aligned}\mathbb{E}_{a,b}[g_{a,b}(x)g_{a,b}(x+d)g_{a,b}(x+2d)] &= \mathsf{E}_{y,z\in\mathbb{Z}_{m_i},z\neq 0}[g(y)g(y+z)g(y+2z)] \\ &\leq (1-c)(\mathsf{E}g)^3,\end{aligned}$$

where a, b vary uniformly over all elements of $\mathbb{Z}_{m_i}$ with $a \neq 0$, and y, z vary uniformly over all elements of $\mathbb{Z}_{m_i}$ with $z \neq 0$. The final inequality is due to a property of the model function g. It then follows via concentration that with positive probability,

$$\mathsf{E}_{x\in Q_i}[f_i(x)\, f_i(x+d)\, f_i(x+2d)] \leq (1-\epsilon)(\mathsf{E}f_i)^3$$

for every $d \in Q_i \setminus \{0\}$ with $d_{[i-1]} = 0$.

We are left with the task of bounding the density of 3-APs with common difference d where $d_{[i-1]} \neq 0$. Over $\mathbb{F}_p^n$, this is easy, as we can show, using the structure of vector spaces, that under a mild condition, if $d_{[i-1]} \neq 0$, then

$$\begin{aligned}&\mathsf{E}_{x\in Q_i}[f_i(x)\, f_i(x+d)\, f_i(x+2d)] \\ &\qquad = \mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x)\, f_{i-1}(x+d_{[i-1]})\, f_{i-1}(x+2d_{[i-1]})] \leq \alpha^3(1-\epsilon).\end{aligned}$$

Such equality does not hold in our current setting. Even though we do not need exact equality, obtaining uniform control over all $d \in Q_i$ with $d_{[i-1]} \neq 0$ using standard concentration inequalities does not work because n_1 is small compared to ϵ^{-1}. However, we can indeed guarantee with high probability the equality

$$\begin{aligned}&\mathsf{E}_{x\in Q_i}[f_i(x)\, f_i(x+d)\, f_i(x+2d)] \\ &\qquad = \mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x)\, f_{i-1}(x+d_{[i-1]})\, f_{i-1}(x+2d_{[i-1]})],\end{aligned}$$

assuming that the model function g over $\mathbb{Z}_{m_i}$ satisfies some additional nice properties, which we refer to as *smoothness*.

Roughly speaking, $g : \mathbb{Z}_{m_i} \to [0,1]$ is smooth if, for random $a_1, \dots, a_h \in \mathbb{Z}_{m_i} \setminus \{0\}$, with high probability, one has

$$\mathsf{E}_y\Big[\prod_{j=1}^{h} g(a_j y + b_j)\Big] = \mathsf{E}[g]^h \quad \text{for all } b_1, \dots, b_h \in \mathbb{Z}_{m_i}.$$

To see how this smoothness property helps, assume that we are given $x' \in Q_{i-1}$ and $d' \in Q_{i-1} \setminus \{0\}$ such that

$$\begin{aligned} f_i(x) &= g(a_1 x_i + b_1) \quad \text{for all } x \text{ such that } x_{[i-1]} = x', \\ f_i(x) &= g(a_2 x_i + b_2) \quad \text{for all } x \text{ such that } x_{[i-1]} = x' + d', \\ f_i(x) &= g(a_3 x_i + b_3) \quad \text{for all } x \text{ such that } x_{[i-1]} = x' + 2d'. \end{aligned}$$

Then, for all d with $d_{[i-1]} = d'$,

$$\begin{aligned} &\mathsf{E}_{x \in Q_i : x_{[i-1]} = x'} [f_i(x) f_i(x+d) f_i(x+2d)] \\ &\qquad = \mathsf{E}_{y \in \mathbb{Z}_{m_i}} [g(a_1 y + c_1) g(a_2 y + c_2) g(a_3 y + c_3)], \end{aligned}$$

where c_1, c_2, c_3 depend on x' and d. Moreover, if g is smooth, then with high probability over random a_1, a_2, a_3, one has

$$\mathsf{E}_{y \in \mathbb{Z}_{m_i}} [g(a_1 y + c_1) g(a_2 y + c_2) g(a_3 y + c_3)] = \mathsf{E}[g]^3 \tag{3}$$

for all $c_1, c_2, c_3 \in \mathbb{Z}_{m_i}$. In that case, for all d with $d_{[i-1]} = d'$,

$$\mathsf{E}_{x \in Q_i : x_{[i-1]} = x'} [f_i(x) f_i(x+d) f_i(x+2d)] = \mathsf{E}[g]^3.$$

Thus, as long as the parameters a_1, a_2, a_3 corresponding to each $x' \in Q_{i-1}$ and $d' \in Q_{i-1} \setminus \{0\}$ satisfy (3), we obtain the desired equality

$$\begin{aligned} &\mathsf{E}_{x \in Q_i} [f_i(x) f_i(x+d) f_i(x+2d)] \\ &\qquad = \mathsf{E}_{x \in Q_{i-1}} [f_{i-1}(x) f_{i-1}(x + d_{[i-1]}) f_{i-1}(x + 2d_{[i-1]})]. \end{aligned}$$

We guarantee this property by applying the union bound over possible values of d' and x' in Q_{i-1}. Note that it is crucial here that the smoothness property allows us to avoid the union bound over d_i, which takes $m_i \approx \exp(O(|Q_{i-1}|))$ possible values.

The model function g over $\mathbb{Z}_{m_i}$ is constructed in Section 5. The essential idea behind the construction is that g should be supported on only a few Fourier characters. The details of the construction over product groups are included in Section 6.

3.2.2. Intervals. Using Theorem 3.4, we can prove Theorem 3.3, giving the desired lower bound over intervals. The construction of the function f in Theorem 3.3 over intervals consists of three steps.

In the first step, we construct a function f_1 with density α which is 0 in the interval $[N' + 1, N]$ for N' slightly smaller than N. This sets the density of 3-APs with common difference d close to $N/2$ to 0.

In the second step, we let f_2 be the function obtained from the following procedure applied to f_1:

(i) Partition $[N']$ into N'/q intervals $I_1, \ldots, I_{N'/q}$ of length q where q can be written as a product of prime numbers as required in Theorem 3.4.

(ii) Using Theorem 3.4, construct a function $g : \mathbb{Z}_q \to [0, 1]$ which satisfies $\mathsf{E}[g] = \alpha$ and $\mathsf{E}_{x \in \mathbb{Z}_q} [g(x) g(x+d) g(x+2d)] \le \alpha^3 (1 - \epsilon)$ for any $d \in \mathbb{Z}_q \setminus \{0\}$.

(iii) For each $j = 1, \dots, N'/q$, identify each interval I_j with $\mathbb{Z}_q$ and place a copy of g on each of them.

For any d with $0 < d < N/2$ and $q \nmid d$, one can show that the density of 3-APs with common difference d of f_2 is at most $\alpha^3(1-\epsilon)$. However, for d divisible by q, the density of 3-APs with common difference d of f_2 is larger than α^3.

Note that the function f_2 constructed in the second step is constant on each mod q residue class in $[N']$. In the third step, we construct the function f_3 as follows:

(i) Construct a subset X of $\mathbb{Z}_{N'/q}$ with much fewer 3-APs compared to the random bound using a variant of the Behrend construction.

(ii) Let $P_t = \{x \in [N'] : x \equiv t \pmod q\}$. With some appropriate $T \subseteq \mathbb{Z}_q$, for each $t \in T$, take a random linear transformation of X inside set $\mathbb{Z}_{N'/q}$, and set f_3 on P_t to be the indicator function of this randomly transformed X.

(iii) On $[N'] \setminus \bigcup_{t\in T} P_t$, set f_3 to be equal to f_2.

The function f_3 has the property that in expectation, for a nonzero d divisible by q, the density of 3-APs with common difference d of f_3 is at most $\alpha^3(1-\epsilon)$.

We let $f = f_3$. Using concentration inequalities, we can show that with positive probability (over the randomness in the third step), for any $d \in [N/2]$,

$$\mathsf{E}_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)] \le \alpha^3(1-\epsilon),$$

proving Theorem 3.3. The details of this construction are contained in Section 7.

4. Upper bound

In this section, we prove Theorems 1.4 and 1.3, showing the existence of a popular difference for 3-APs when $|G| \ge \text{tower}(C\log(1/\epsilon))$ or $N \ge \text{tower}(C\log(1/\epsilon))$. Here G always denotes a finite abelian group of odd order. For $x \in G$, we write $x/2$ to mean the inverse of the isomorphism $x \mapsto 2x$. In Section 4.1, we give some preliminaries on Bohr sets, which is an important tool to make Fourier analysis work over general abelian groups. In Section 4.2, we give the complete proofs of Theorems 1.4 and 1.3.

4.1. Bohr sets

Denote the distance from $x \in \mathbb{R}$ to the nearest integer by $\|x\|_{\mathbb{R}/\mathbb{Z}} := \min_{n\in\mathbb{Z}} |x-n|$. Let $\arg(z)$ denote the argument of $z \in \mathbb{C}$, so that $\arg(e^{it}) \in [0, 2\pi]$ and $e^{it} = e^{i\arg(e^{it})}$.

Definition 4.1. Let G be an abelian group of odd order. For a subset $S \subseteq \widehat{G}$ and a parameter $\rho \in [0,1]$, define the *Bohr set* $B(S,\rho) = \{x \in G : \|\arg(\chi(x))/(2\pi)\|_{\mathbb{R}/\mathbb{Z}} \le \rho \ \forall \chi \in S\}$. We call S the *frequency set* of the Bohr set $B(S,\rho)$ and ρ the *radius*. The *codimension* of the Bohr set is $|S|$.

We often drop S and ρ from the notation and denote the Bohr set by B if it is clear from context. Given a Bohr set B, we write $S(B)$ to denote the frequency set of B. For

a real number $\nu \geq 0$, we denote by $(B)_\nu$ the Bohr set with the same frequency set and scaled radius $B(S, \nu\rho)$.

Define the normalized indicator function of Bohr sets by

$$\beta_B(x) = \frac{|G|B(x)}{|B|},$$

where we chose the normalization so that $\mathsf{E}_x[\beta_B(x)] = 1$. Define

$$\phi_B(x) = \beta_B * \beta_B(x).$$

Then we also have $\mathsf{E}_x[\phi_B(x)] = 1$. The functions β_B can be thought of as the density with respect to the uniform distribution on G of the uniform distribution on B, and ϕ_B is the density of a smoothened version of the uniform distribution on B. In general, a function $\tau : G \to [0, \infty)$ with $\mathsf{E}\tau = 1$ can be thought of as the density of a distribution with respect to the uniform distribution on G. Note also that $\beta_B(x) = \beta_B(-x)$ and $\phi_B(x) = \phi_B(-x)$.

Conventions. For simplicity of notation, we often omit the subscript B and use consistent subscripts throughout. For example, $\phi = \phi_B$, $\beta = \beta_B$, $\phi_1 = \phi_{B_1}$, $\beta_1 = \beta_{B_1}$.

As introduced by Bourgain [7], it is often useful to work with *regular Bohr sets*, those for which a small change to the radius does not significantly change the size of the Bohr set.

Definition 4.2. A Bohr set $B = B(S, \rho)$ of codimension d is *regular* if for all $\delta \leq 1/(80d)$,

$$|(B)_{1+\delta} \setminus (B)_{1-\delta}| \leq 160\delta d|B|.$$

In the next proposition, we state some basic properties of Bohr sets, whose proofs can be found in [32, Section 4.4]. Denote by $2 \cdot X = \{2x : x \in X\}$ the dilation of X by a factor 2. Recall that for a character χ, we denote by $\chi^{1/2}$ the character given by $x \mapsto \chi(x/2)$.

Proposition 4.3. *The following properties hold:*

(i) $|B(S, \rho)| \geq |G|\rho^{|S|}$.

(ii) $B(S, \rho) + B(S, \rho') \subseteq B(S, \rho + \rho')$.

(iii) *For every Bohr set $B = B(S, \rho)$, there exists $\nu \in [1/2, 1]$ such that $(B)_\nu$ is regular.*

(iv) *For every Bohr set $B = B(S, \rho)$, $2 \cdot B$ is also a Bohr set with frequency set $\{\chi^{1/2} : \chi \in S\}$ and radius ρ. Furthermore, $2 \cdot B \subseteq (B)_2$.*

(v) *For every Bohr set $B = B(S, \rho)$, $(2 \cdot B)_\nu = 2 \cdot (B)_\nu$. Hence, if $(B)_\nu$ is regular then $(2 \cdot B)_\nu$ is also regular.*

Definition 4.4. Let ϕ be a function on G with $\mathsf{E}\phi = 1$. Define

$$f_\phi(x) = (f * \phi)(x) = \mathsf{E}_y[f(x - y)\phi(y)].$$

The next estimate shows that regular Bohr sets are essentially invariant under convolutions with a distribution whose support is contained in a Bohr set with the same frequency set and smaller radius. This is analogous to the additive closure property of subgroups.

Proposition 4.5. *Let B be a regular Bohr set of codimension d, $\beta = \beta_B$ and $\phi = \phi_B$. Let $\nu \le 1/(80d)$. Let τ be a function supported in B_ν with $\mathsf{E}\tau = 1$. Then*

$$\mathsf{E}_x[|(\beta * \tau)(x) - \beta(x)|] \le 160\nu d, \tag{4}$$

$$\mathsf{E}_x[|(\phi * \tau)(x) - \phi(x)|] \le 160\nu d. \tag{5}$$

Furthermore, for any $f : G \to [0,1]$, letting κ be either β or ϕ,

$$\mathsf{E}_x[|(f_\tau * \kappa)(x) - (f * \kappa)(x)|] \le 160\nu d. \tag{6}$$

Proof. We have

$$(\beta * \tau)(x) = \mathsf{E}_y[\beta(x-y)\tau(y)] \in [0, |G|/|B|].$$

The support of $\beta * \tau$ is a subset of $\mathrm{supp}(\beta) + \mathrm{supp}(\tau) \subseteq B + (B)_\nu \subseteq (B)_{1+\nu}$. Thus, if $x \notin (B)_{1+\nu}$, then $(\beta * \tau)(x) = 0 = \beta(x)$. Furthermore, if $x \in (B)_{1-\nu}$, then for all $y \in \mathrm{supp}(\tau)$, we have $x - y \in B$, so $(\beta * \tau)(x) = |G|/|B| = \beta(x)$. Hence,

$$\mathsf{E}_x[|(\beta * \tau)(x) - \beta(x)|] \le \frac{1}{|G|} \sum_{x \in (B)_{1+\nu} \setminus (B)_{1-\nu}} \frac{|G|}{|B|} = \frac{|(B)_{1+\nu} \setminus (B)_{1-\nu}|}{|B|} \le 160\nu d,$$

giving (4).

For (5), note that

$$\begin{aligned}
&\mathsf{E}_x[|(\phi * \tau)(x) - \phi(x)|] \\
&\qquad = \mathsf{E}_x[|(\beta * \beta * \tau)(x) - (\beta * \beta)(x)|] \\
&\qquad = \mathsf{E}_x[|\mathsf{E}_y[\beta(y)((\beta * \tau)(x-y) - \beta(x-y))]|] \\
&\qquad \le \mathsf{E}_y[\beta(y)\mathsf{E}_x[|(\beta * \tau)(x-y) - \beta(x-y)|]] && \text{(by the triangle inequality)} \\
&\qquad \le 160\nu d\,\mathsf{E}_y[\beta(y)] && \text{(by (4))} \\
&\qquad = 160\nu d.
\end{aligned}$$

For (6), we have

$$\begin{aligned}
\mathsf{E}_x[|f_\tau * \kappa - f_\kappa|] &= \mathsf{E}_x[|\mathsf{E}_y[f(y)(\kappa * \tau)(x-y) - f(y)\kappa(x-y)]|] \\
&\le \mathsf{E}_{x,y}[|(\kappa * \tau)(x-y) - \kappa(x-y)|] && \text{(by the triangle inequality)} \\
&\le 160\nu d. && \text{(by (4), (5))}
\end{aligned}$$

∎

The next lemma says that if $B_2 \subseteq (B_1)_{\nu/2}$ and $\tau = \beta_{B_2}$ or $\tau = \phi_{B_2}$, then the k-th moment of f_τ is at least the k-th moment of $f_{\phi_{B_1}}$ up to a small error term.

Lemma 4.6. *Let $f : G \to [0,1]$. Let B_1, B_2 be regular Bohr sets such that B_1 has codimension d_1. Let $\phi_1 = \phi_{B_1}$, $\beta_2 = \beta_{B_2}$ and $\phi_2 = \phi_{B_2}$. Let $k \geq 1$ be an integer and $\nu \leq 1/(80d_1)$. If $B_2 \subseteq (B_1)_{\nu/2}$, then*

$$\mathsf{E}_x[f_{\phi_2}(x)^k] \geq \mathsf{E}_x[f_{\phi_1}(x)^k] - 160\nu d_1 k. \tag{7}$$

If $B_2 \subseteq (B_1)_\nu$, then

$$\mathsf{E}_x[f_{\beta_2}(x)^k] \geq \mathsf{E}_x[f_{\phi_1}(x)^k] - 160\nu d_1 k. \tag{8}$$

Proof. By (6) of Proposition 4.5, applied with $B = B_1$, $\tau = \phi_2$ (noting that $\operatorname{supp}(\phi_2) \subseteq B_2 + B_2 \subseteq (B_1)_\nu$) and $\kappa = \phi_1$,

$$\mathsf{E}_x[|f_{\phi_1}(x) - \mathsf{E}_d[f_{\phi_2}(x-d)\phi_1(d)]|] \leq 160\nu d_1.$$

Thus,

$$\mathsf{E}_x[|f_{\phi_1}(x)^k - (\mathsf{E}_d[f_{\phi_2}(x-d)\phi_1(d)])^k|] \leq 160\nu d_1 k.$$

By Jensen's inequality applied to the convex function $t \mapsto t^k$, we obtain

$$\begin{aligned}\mathsf{E}_x[f_{\phi_1}(x)^k] &\leq \mathsf{E}_x[(\mathsf{E}_d[f_{\phi_2}(x-d)\phi_1(d)])^k] + 160\nu d_1 k\\ &\leq \mathsf{E}_{x,d}[f_{\phi_2}(x-d)^k\phi_1(d)] + 160\nu d_1 k\\ &= \mathsf{E}_x[f_{\phi_2}(x)^k] + 160\nu d_1 k.\end{aligned}$$

The proof of (8) is similar. ∎

4.2. *Proofs of Theorems 1.4 and 1.3*

In the following, we prove two results that are used in the proof of Theorem 1.4, the counting lemma (Lemma 4.7) and the mean-cube density increment (Lemma 4.9).

For a function ϕ on G with $\mathsf{E}\phi = 1$ and a function $f : G \to [0,1]$, we denote

$$\Lambda_\phi(f) = \mathsf{E}_{x,d}[f(x)f(x+d)f(x+2d)\phi(d)].$$

Lemma 4.7 (Counting lemma). *Let $B_1 = B(S_1, \rho_1)$ and $B_2 = B(S_2, \rho_2)$ be two Bohr sets. Let $\phi_1 = \phi_{B_1}$ and $\phi_2 = \phi_{B_2}$. Then*

$$\Lambda_{\phi_1}(f_{\phi_2}) \geq \Lambda_{\phi_1}(f) - 3\sup_{\chi\in\widehat{G}} |\widehat{f}(\chi) - \widehat{f_{\phi_2}}(\chi)|\mathbb{E}[f(x)^2]\frac{|G|}{|B_1|}.$$

Proof. By expanding in the Fourier basis,

$$\Lambda_{\phi_1}(f) = \sum_{\chi_1\chi_2\chi_3=1} \widehat{f}(\chi_1)\widehat{f}(\chi_2)\widehat{f}(\chi_3)\widehat{\phi_1}(\chi_2^{-1}\chi_3^{-2}).$$

By similar expansion for $\Lambda_{\phi_1}(f_{\phi_2})$, we can write $\Lambda_{\phi_1}(f) - \Lambda_{\phi_1}(f_{\phi_2})$ as

$$\begin{aligned}\sum_{\chi_1\chi_2\chi_3=1} &(\widehat{f}(\chi_1) - \widehat{f_{\phi_2}}(\chi_1))\widehat{f}(\chi_2)\widehat{f}(\chi_3)\widehat{\phi_1}(\chi_2^{-1}\chi_3^{-2}) \\ &+ \sum_{\chi_1\chi_2\chi_3=1} \widehat{f_{\phi_2}}(\chi_1)(\widehat{f}(\chi_2) - \widehat{f_{\phi_2}}(\chi_2))\widehat{f}(\chi_3)\widehat{\phi_1}(\chi_2^{-1}\chi_3^{-2}) \\ &+ \sum_{\chi_1\chi_2\chi_3=1} \widehat{f_{\phi_2}}(\chi_1)\widehat{f_{\phi_2}}(\chi_2)(\widehat{f}(\chi_3) - \widehat{f_{\phi_2}}(\chi_3))\widehat{\phi_1}(\chi_2^{-1}\chi_3^{-2}).\end{aligned}$$

Note that

$$\begin{aligned}\Big|\sum_{\chi_2\chi_3^2=\chi^{-1}} \widehat{f}(\chi_2)\widehat{f}(\chi_3)\Big| &\le \Big(\sum_{\chi_2}|\widehat{f}(\chi_2)|^2\Big)^{1/2}\Big(\sum_{\chi_3}|\widehat{f}(\chi_3)|^2\Big)^{1/2} \quad \text{(Cauchy–Schwarz)} \\ &= \sum_{\chi}|\widehat{f}(\chi)|^2 \\ &= \mathsf{E}[f(x)^2], \quad \text{(Parseval)} \qquad (9)\end{aligned}$$

and

$$\begin{aligned}\sum_{\chi}|\widehat{\phi_1}(\chi)| &= \sum_{\chi}|\widehat{\beta_1}(\chi)|^2 \\ &= \mathsf{E}_x[\beta_1(x)^2]. \quad \text{(Parseval)} \qquad (10)\end{aligned}$$

We can now bound the first term as

$$\begin{aligned}\Big|\sum_{\chi_1\chi_2\chi_3=1} &(\widehat{f}(\chi_1) - \widehat{f_{\phi_2}}(\chi_1))\widehat{f}(\chi_2)\widehat{f}(\chi_3)\widehat{\phi_1}(\chi_2^{-1}\chi_3^{-2})\Big| \\ &\le \sup_{\chi}|\widehat{f}(\chi) - \widehat{f_{\phi_2}}(\chi)| \cdot \sum_{\chi}|\widehat{\phi_1}(\chi)|\Big|\sum_{\chi_2\chi_3^2=\chi^{-1}} \widehat{f}(\chi_2)\widehat{f}(\chi_3)\Big| \\ &\le \sup_{\chi}|\widehat{f}(\chi) - \widehat{f_{\phi_2}}(\chi)| \cdot \sum_{\chi}|\widehat{\phi_1}(\chi)|\mathsf{E}[f(x)^2] \text{ (by (9))} \\ &\le \sup_{\chi}|\widehat{f}(\chi) - \widehat{f_{\phi_2}}(\chi)|\mathsf{E}[f(x)^2]\mathsf{E}_x[\beta_1(x)^2] \quad \text{(by (10))} \\ &\le \sup_{\chi}|\widehat{f}(\chi) - \widehat{f_{\phi_2}}(\chi)|\mathbb{E}[f(x)^2]\frac{|G|}{|B_1|},\end{aligned}$$

where in the last inequality we use the fact that $\sup_x \beta_1(x) \le |G|/|B_1|$ and $\mathsf{E}\beta_1 = 1$.

The remaining two terms are bounded similarly. ∎

We next state and prove the mean-cube density increment lemma. We make use of the following classical inequality in the proof of the lemma:

Theorem 4.8 (Schur's inequality). *For real numbers $a, b, c \ge 0$, one has*

$$a^3 + b^3 + c^3 + 3abc \ge a^2b + b^2a + a^2c + c^2a + b^2c + c^2b. \qquad (11)$$

Lemma 4.9 (Mean-cube density increment). *Let $f : G \to [0,1]$. Let $B_1 = B(S_1, \rho_1)$ and $B_2 = B(S_2, \rho_2)$ be two regular Bohr sets with codimension d_1, d_2 respectively. Let $\phi_1 = \phi_{B_1}$ and $\phi_2 = \phi_{B_2}$. Let $\nu \le 1/(1000d_1)$. Assume that $B_2 \subseteq (B_1)_{\nu^2/8} \cap (2\cdot B_1)_{\nu^2/8}$. For every regular Bohr set $B = (B_1)_{\delta\nu/2}$ with $\delta \in [1/2, 1]$ (Proposition 4.3 guarantees the existence of such δ that makes B regular), we have, setting $\phi = \phi_B$,*

$$\Lambda_\phi(f_{\phi_2}) \ge 2\mathsf{E}[f_{\phi_1}(x)^3] - \mathsf{E}[f_{\phi_2}(x)^3] - 1920\nu d_1.$$

Proof. Let $\beta = \beta_B$ and $\phi = \phi_B$. We denote by $\tilde\beta$ the normalized measure associated with the Bohr set $2\cdot B$, and denote $\tilde\phi = \tilde\beta * \tilde\beta$.

Applying Schur's inequality (Theorem 4.8) with $a = f_{\phi_2}(x), b = f_{\phi_2}(x+d), c = f_{\phi_2}(x+2d)$ for each x and d, and using linearity of expectation, we have

$$\begin{aligned}\Lambda_\phi(f_{\phi_2}) &= \mathsf{E}_{x,d}[f_{\phi_2}(x) f_{\phi_2}(x+d) f_{\phi_2}(x+2d)\phi(d)] \\ &\ge \frac{4\mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+d)\phi(d)] + 2\mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+2d)\phi(d)]}{3} \\ &\quad - \mathsf{E}_x[f_{\phi_2}(x)^3].\end{aligned}$$

By Proposition 4.3 (v), since B is a regular Bohr set, $2\cdot B$ is also a regular Bohr set. The codimensions of $2\cdot B$ and B are d_1. We have $B_2 \subseteq (B_1)_{\nu^2/8} \cap (2\cdot B_1)_{\nu^2/8} \subseteq (B)_{\nu/2} \cap (2\cdot B)_{\nu/2}$. By (6) in Proposition 4.5, applied with the Bohr set $B, \kappa = \phi, \tau = \phi_2$,

$$\mathsf{E}_x\big[|\mathsf{E}_d[f_{\phi_2}(x+d)\phi(d)] - f_\phi(x)|\big] = \mathsf{E}_x\big[|\mathsf{E}_d[f_{\phi_2}(x-d)\phi(d)] - f_\phi(x)|\big] \le 160\nu d_1,$$

where we have used $\phi(d) = \phi(-d)$. Hence,

$$\mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+d)\phi(d)] \ge \mathsf{E}_x[f_{\phi_2}(x)^2 f_\phi(x)] - 160\nu d_1.$$

Similarly,

$$\mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+2d)\phi(d)] \ge \mathsf{E}_x[f_{\phi_2}(x)^2 f_{\tilde\phi}(x)] - 160\nu d_1.$$

We have

$$\begin{aligned}\mathsf{E}_x[f_{\phi_2}(x)^2 f_\phi(x)] &= \mathsf{E}_{x,y}[f_{\phi_2}(x)^2\beta(y) f_\beta(x-y)] \\ &= \mathsf{E}_{x,y}[f_{\phi_2}(x)^2\beta(y) f_\beta(x+y)] \qquad (\text{using } \beta(-y) = \beta(y)) \\ &= \mathsf{E}_x\big[f_\beta(x)\mathsf{E}_y[f_{\phi_2}(x-y)^2\beta(y)]\big] \\ &\ge \mathsf{E}_x\big[f_\beta(x)\mathsf{E}_y[f_{\phi_2}(x-y)\beta(y)]^2\big]. \qquad (\text{Cauchy–Schwarz})\end{aligned}$$

By (6) in Proposition 4.5, applied with the Bohr set $B, \kappa = \beta, \tau = \phi_2$,

$$\mathsf{E}_x\big[|\mathsf{E}_y[f_{\phi_2}(x-y)\beta(y)] - f_\beta(x)|\big] \le 160\nu d_1.$$

Thus,

$$\mathsf{E}_x\big[f_\beta(x)\mathsf{E}_y[f_{\phi_2}(x-y)\beta(y)]^2\big] \ge \mathsf{E}_x[f_\beta(x)^3] - 320\nu d_1.$$

Hence,

$$\mathsf{E}_x[f_{\phi_2}(x)^2 f_{\phi_2}(x+d)\phi(d)] \geq \mathsf{E}_x[f_\beta(x)^3] - 480\nu d_1.$$

Similarly, noting that $B_2 \subseteq (2\cdot B)_{\nu/2}$, we have

$$\begin{aligned}\mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+2d)\phi(d)] &= \mathsf{E}_{x,d}[f_{\phi_2}(x)^2 f_{\phi_2}(x+d)\tilde{\phi}(d)]\\ &\geq \mathsf{E}_x[f_{\tilde{\beta}}(x)^3] - 480\nu d_1.\end{aligned}$$

Since $B \subseteq (B_1)_{\delta\nu}$, by (8) of Lemma 4.6 applied to the Bohr sets B_1 and B,

$$\mathsf{E}_x[f_\beta(x)^3] \geq \mathsf{E}_x[f_{\phi_1}(x)^3] - 480\nu d_1.$$

Thus,

$$\mathsf{E}_x[f_{\phi_2}(x)^2 f_{\phi_2}(x+d)\phi(d)] \geq \mathsf{E}_x[f_{\phi_1}(x)^3] - 960\nu d_1.$$

Similarly, since $2\cdot B \subseteq (B)_2 \subseteq (B_1)_{\delta\nu}$, by (8) of Lemma 4.6 applied with the Bohr sets B_1 and $2\cdot B$,

$$\mathsf{E}_x[f_{\tilde{\beta}}(x)^3] \geq \mathsf{E}_x[f_{\phi_1}(x)^3] - 480\nu d_1.$$

Thus,

$$\mathsf{E}_x[f_{\phi_2}(x)^2 f_{\phi_2}(x+2d)\phi(d)] \geq \mathsf{E}_x[f_{\phi_1}(x)^3] - 960\nu d_1.$$

Combining, we get

$$\Lambda_\phi(f_{\phi_2}) \geq 2\mathsf{E}[f_{\phi_1}^3] - \mathsf{E}[f_{\phi_2}^3] - 1920\nu d_1. \qquad \blacksquare$$

Lemma 4.10. *Let $\alpha, \epsilon > 0$. Let $a_1, a_2, \ldots$ be a sequence of positive real numbers such that $\alpha^3 \leq a_i \leq 1$ for all i. Then for some $i \leq 2\log_2(2/\epsilon)$, $2a_i - a_{i+1} > \alpha^3 - \epsilon/2$.*

Proof. Assume for the sake of contradiction that for all $i \leq 2\log_2(2/\epsilon)$,

$$a_{i+1} \geq 2a_i - \alpha^3 + \epsilon/2.$$

Then $a_2 \geq \alpha^3 + \epsilon/2$ since $a_1 \geq \alpha^3$. For $2 \leq i \leq 2\log_2(2/\epsilon)$, $a_{i+1} - \alpha^3 \geq 2(a_i - \alpha^3)$, so $a_{i+1} \geq \alpha^3 + 2^i\epsilon/2$. Since $a_{i+1} \leq 1$ for all i, we arrive at a contradiction since $2^{2\log_2(2/\epsilon)}\epsilon/2 > 1$. $\blacksquare$

Proof of Theorem 1.4. We inductively define parameters ρ_i such that $\rho_1 = \epsilon^{10}$, and for $i \geq 2$, $\rho_i = \exp(-\rho_{i-1}^{-5})$. Let $\nu_i = 10^{-5}\epsilon\rho_i^2$.

Let $S_1 = \{\chi \in \widehat{G} : |\widehat{f}(\chi)| \geq \rho_1/2\}$, and for $i \geq 2$, $S_i = \{\chi \in \widehat{G} : |\widehat{f}(\chi)| \geq \rho_i/2\} \cup \{\chi^{1/2} : \chi \in S_{i-1}\}$. Note that $S_1 \subseteq S_2$ as $\rho_1 \geq \rho_2$, and inductively, if $S_{i-1} \subseteq S_i$ for some $i \geq 2$, then

$$\begin{aligned}S_{i+1} &= \{\chi \in \widehat{G} : |\widehat{f}(\chi)| \geq \rho_{i+1}/2\} \cup \{\chi^{1/2} : \chi \in S_i\}\\ &\supseteq \{\chi \in \widehat{G} : |\widehat{f}(\chi)| \geq \rho_i/2\} \cup \{\chi^{1/2} : \chi \in S_{i-1}\} = S_i.\end{aligned}$$

Thus $S_i \subseteq S_{i+1}$ for all $i \geq 1$.

By Parseval's identity, $|\{\chi \in \widehat{G} : |\widehat{f}(\chi)| \geq \rho_i/2\}| \leq 4\rho_i^{-2}$, so $|S_i| \leq \sum_{j=1}^{i} 4\rho_j^{-2} < 5\rho_i^{-2}$. Let $B_i = B(S_i, \rho_i/(4\pi))$. We first note that for $\chi \in S_i$,

$$|1 - \widehat{\beta_i}(\chi)| = |1 - \mathsf{E}_x[\beta_i(x)\chi(x)]| \leq \mathsf{E}_x[\beta_i(x)|\chi(x) - 1|] \leq \rho_i/2,$$

since $\|\arg(\chi(x))/(2\pi)\|_{\mathbb{R}/\mathbb{Z}} \leq \rho_i/(4\pi)$ for all x such that $\beta_i(x) \neq 0$. Hence,

$$|1 - \widehat{\phi_i}(\chi)| = |1 - \widehat{\beta_i}(\chi)^2| \leq \rho_i.$$

Thus, for $\chi \in S_i$,

$$|\widehat{(f - f_{\phi_i})}(\chi)| \leq \rho_i,$$

and for $\chi \notin S_i$, $|\widehat{f}(\chi)| \leq \rho_i/2$ so

$$|\widehat{(f - f_{\phi_i})}(\chi)| \leq \rho_i.$$

Observe that $\mathsf{E}[f_{\phi_i}^3] \geq \alpha^3$ by convexity, and $\mathsf{E}[f_{\phi_i}^3] \leq 1$ for all i. By Lemma 4.10, there exists $i \leq 2\log_2(2/\epsilon)$ such that

$$2\mathsf{E}[f_{\phi_i}(x)^3] - \mathsf{E}[f_{\phi_{i+1}}(x)^3] \geq \alpha^3 - \epsilon/2.$$

Fix such an i. We have $B_{i+1} \subseteq (B_i)_{\nu_i^2/8} \cap (2 \cdot B_i)_{\nu_i^2/8}$ since for any $\chi \in S_i \subseteq S_{i+1}$ and $x \in B_{i+1}$,

$$\|\arg(\chi(x))/(2\pi)\|_{\mathbb{R}/\mathbb{Z}} \leq \rho_{i+1} \leq \nu_i^2\rho_i/8,$$

and furthermore $\chi^{1/2} \in S_{i+1}$ so

$$\|\arg(\chi^{1/2}(x))/(2\pi)\|_{\mathbb{R}/\mathbb{Z}} \leq \rho_{i+1} \leq \nu_i^2\rho_i/8.$$

By Lemma 4.9, there exists a regular Bohr set $B = (B_i)_{\delta\nu_i/2}$ for $\delta \in [1/2, 1]$ such that

$$\Lambda_\phi(f_{\phi_{i+1}}) \geq 2\mathsf{E}[f_{\phi_i}(x)^3] - \mathsf{E}[f_{\phi_{i+1}}(x)^3] - 1920\nu_i|S_i| \geq \alpha^3 - \epsilon/2 - 1920\nu_i|S_i|.$$

By Lemma 4.7,

$$\begin{aligned}\Lambda_\phi(f_{\phi_{i+1}}) &\leq \Lambda_\phi(f) + \sup_\chi |\widehat{f}(\chi) - \widehat{f_{\phi_{i+1}}}(\chi)|\mathsf{E}[f(x)^2]\frac{|G|}{|B|} \\ &\leq \Lambda_\phi(f) + \rho_{i+1}(16\pi\nu_i^{-1}\rho_i^{-1})^{5\rho_i^{-2}}.\end{aligned}$$

By our choice of ρ_i, ν_i, we have

$$\rho_{i+1}(16\pi\nu_i^{-1}\rho_i^{-1})^{5\rho_i^{-2}} \leq \exp(5\rho_i^{-2} \cdot \log(10^8\epsilon^{-1}\rho_i^{-3}))\exp(-\rho_i^{-5}) < \epsilon/8,$$

noting that $\rho_i \leq \rho_1 \leq \epsilon^{10}$. Hence,

$$\begin{aligned}\Lambda_\phi(f) &\geq \alpha^3 - \epsilon/2 - 1920\nu_i d_i - \rho_{i+1}(16\pi\nu_i^{-1}\rho_i^{-1})^{5\rho_i^{-2}} > \alpha^3 - \epsilon/2 - \epsilon/4 - \epsilon/8 \\ &= \alpha^3 - 7\epsilon/8.\end{aligned}$$

Observe that there exists an absolute constant $C' > 0$ such that $\rho_i \geq 1/\text{tower}(C'i)$. Furthermore, the codimension of B_i is bounded above by $5\rho_i^{-2}$ and the radius of B_i is $\rho_i/(4\pi)$. Hence, we obtain a Bohr set B with size at least $|G|/\text{tower}(10C'\log(1/\epsilon))$ such that $\Lambda_\phi(f) \geq \alpha^3 - 7\epsilon/8$. Hence, for a sufficiently large constant $C > 0$, assuming that $|G| \geq \text{tower}(C\log(1/\epsilon))$, we have

$$\mathsf{E}_{x,d}[f(x)f(x+d)f(x+2d)\phi(d)I(d \neq 0)] \geq \Lambda_\phi(f) - \frac{1}{|B|} \geq \alpha^3 - \epsilon.$$

Here $I(d \neq 0)$ is the indicator function which evaluates to 1 if $d \neq 0$ and evaluates to 0 otherwise. Thus, there exists $d \neq 0$ such that

$$\mathsf{E}_x[f(x)f(x+d)f(x+2d)] \geq \alpha^3 - \epsilon. \qquad \blacksquare$$

Proof of Theorem 1.3. We can assume without loss of generality that N is odd by possibly increasing N by 1. Let $G = \mathbb{Z}_N$. We repeat the proof of Theorem 1.4 with the inclusion of the character $\chi_0(x) = e^{2\pi i x/N}$ in the sets S_i. We then obtain a Bohr set B whose frequency set contains χ_0 such that B has size at least $|G|/\text{tower}(C'\log(1/\epsilon))$ and $\Lambda_\phi(f) \geq \alpha^3 - 7\epsilon/8$. Assuming that $N \geq \text{tower}(C\log(1/\epsilon))$ for sufficiently large C, following the last step in the proof of Theorem 1.4, we obtain a positive integer $d < N/2$ such that $d \in \text{supp}(\phi)$ when viewed as an element in $\mathbb{Z}_N$ and

$$\mathsf{E}_x[f(x)f(x+d)f(x+2d)] \geq \alpha^3 - 15\epsilon/16.$$

Since $d \in \text{supp}(\phi) \subseteq B + B$ and χ_0 is in the frequency set defining B, $\|\arg(\chi_0(d))/(2\pi)\|_{\mathbb{R}/\mathbb{Z}} \leq 2\rho \leq 2\epsilon^{10}$. Thus, as d is a positive integer less than $N/2$, we have $d < 2\epsilon^{10}N$. Thus, restricting to $x \in [N-2d]$ in the above expectation, we have

$$\sum_{x\in[N-2d]} f(x)f(x+d)f(x+2d) \geq N(\alpha^3 - 15\epsilon/16) - 2d \geq N(\alpha^3 - \epsilon). \qquad \blacksquare$$

5. Lower bound construction: preparations

We assume throughout this section that N is an odd prime number. As a building block in our construction, we will make use of a function g which has relatively low 3-AP density (considerably smaller than the random bound given the density of g), but behaves randomly-like in the following way. If $a_1, \ldots, a_h$ are chosen independently and uniformly at random from the nonzero elements of $\mathbb{Z}_N$, then with high probability, for all $b_1, \ldots, b_h \in \mathbb{Z}_N$,

$$\mathsf{E}_x\Big[\prod_{j=1}^{h} g(a_j x + b_j)\Big] = \mathsf{E}[g]^h. \tag{12}$$

In the following, we identify $\widehat{\mathbb{Z}_N}$ with $\mathbb{Z}_N$, so that we write

$$\widehat{g}(r) = \mathsf{E}_{x\in\mathbb{Z}_N}[g(x)e(-rx/N)].$$

Lemma 5.1. *Suppose $g : \mathbb{Z}_N \to [0,1]$ and $a_1, \dots, a_h \in \mathbb{Z}_N \setminus \{0\}$ satisfy the following properties:*

(i) *The support of $\widehat{g}$ has size at most ℓ.*

(ii) *For all $r_1, \dots, r_h \in \mathbb{Z}_N$ such that $\sum_{j=1}^h r_j a_j = 0$ and $(r_1, \dots, r_h) \neq (0, \dots, 0)$, there is some $j \in [h]$ such that r_j is not contained in the support of $\widehat{g}$.*

Then for all $b_1, \dots, b_h \in \mathbb{Z}_N$,

$$\mathsf{E}_x\Big[\prod_{j=1}^h g(a_j x + b_j)\Big] = \mathsf{E}[g]^h.$$

Furthermore, if $a_1, \dots, a_h$ are chosen from $\mathbb{Z}_N \setminus \{0\}$ uniformly and independently at random, then property (ii) is satisfied with probability at least $1 - \ell^h/(N-1)$.

Proof. By the Fourier inversion formula,

$$\begin{aligned}
\mathsf{E}_x\Big[\prod_{j=1}^h g(a_j x + b_j)\Big] &= \mathsf{E}_x\Big[\prod_{j=1}^h \Big(\sum_{r_j \in \mathbb{Z}_N} \widehat{g}(r_j) e\Big(\frac{r_j a_j x + r_j b_j}{N}\Big)\Big)\Big] \\
&= \sum_{r_1, \dots, r_h \in \mathbb{Z}_N} \prod_{j=1}^h e\Big(\frac{r_j b_j}{N}\Big)\widehat{g}(r_j) \cdot \mathsf{E}_x\Big[e\Big(\frac{\sum_{j=1}^h r_j a_j x}{N}\Big)\Big] \\
&= \sum_{\substack{r_1, \dots, r_h \in \mathbb{Z}_N \\ \sum_{j=1}^h r_j a_j = 0}} \prod_{j=1}^h e\Big(\frac{r_j b_j}{N}\Big)\widehat{g}(r_j).
\end{aligned}$$

Note that $\prod_{j=1}^h \widehat{g}(0) = \mathsf{E}[g]^h$. Consider $(r_1, \dots, r_h) \neq (0, \dots, 0)$ where $\sum_{j=1}^h r_j a_j = 0$. Property (ii) guarantees that $\widehat{g}(r_j) = 0$ for some $j \in [h]$, so $\prod_{j=1}^h \widehat{g}(r_j) = 0$. Hence, if $a_1, \dots, a_h$ satisfy (ii), then

$$\mathsf{E}_x\Big[\prod_{j=1}^h g(a_j x + b_j)\Big] = \widehat{g}(0)^h = \mathsf{E}[g]^h.$$

Next, we show that if $a_1, \dots, a_h$ are chosen uniformly and independently at random from $\mathbb{Z}_N \setminus \{0\}$, then (ii) is satisfied with probability at least $1 - \ell^h/(N-1)$. Indeed, consider a fixed $(r_1, \dots, r_h) \neq (0, \dots, 0)$ such that r_j is in the support of $\widehat{g}$ for each $j \in [h]$. There exists $i \in [h]$ such that $r_i \neq 0$. For each fixed choice of a_j for $j \in [h] \setminus \{i\}$, there is a unique choice of a_i such that $\sum_{j=1}^h r_j a_j = 0$. Hence, the probability that $\sum_{j=1}^h r_j a_j = 0$ is at most $1/(N-1)$. By the union bound over the choice of r_j in the support of $\widehat{g}$, we conclude that (ii) is violated with probability at most $\ell^h/(N-1)$. ∎

Next, for each $\alpha \leq 1/2$, we construct a function g_α with mean α and prove that g_α has the desired properties in Lemma 5.1. We recall that the 3-AP density of a function g is denoted by $\Lambda(g) = \mathsf{E}_{x,d}[g(x)g(x+d)g(x+2d)]$.

Lemma 5.2. *For $\alpha \le 1/2$, define a function $g_\alpha : \mathbb{Z}_N \to [0,1]$ by*

$$g_\alpha(x) = \alpha - \frac{\alpha\cos(2\pi x/N)}{2} - \frac{\alpha\cos(4\pi x/N)}{2}.$$

Then g_α satisfies the following properties:

(i) $\mathsf{E}[g_\alpha] = \alpha \le 1/2$ *and* $g_\alpha(x) \in [0, 2\alpha]$ *for all* $x \in \mathbb{Z}_N$.

(ii) $\Lambda(g_\alpha) = \mathsf{E}_{x,d}[g_\alpha(x)g_\alpha(x+d)g_\alpha(x+2d)] = (1-\frac{1}{32})\alpha^3$, *and* $\mathsf{E}_x[g_\alpha(x)^3] \le \frac{3}{2}\alpha^3$.

(iii) *For h a positive integer, if we choose $a_1, \dots, a_h$ uniformly and independently at random from $\mathbb{Z}_N \setminus \{0\}$, then with probability at least $1 - 5^h/(N-1)$, for all choices of $b_1, \dots, b_h \in \mathbb{Z}_N$,*

$$\mathsf{E}_x\Big[\prod_{j=1}^{h} g_\alpha(a_j x + b_j)\Big] = \mathsf{E}[g_\alpha]^h.$$

(iv) $\mathsf{E}_{x\neq y}[g_\alpha(x)g_\alpha(y)] \le \alpha^2$.

(v) $\mathsf{E}_x[g_\alpha(x)^2] = \frac{5}{4}\alpha^2$.

Proof. We first observe that

$$\widehat{g_\alpha}(0) = \alpha, \quad \widehat{g_\alpha}(1) = \widehat{g_\alpha}(N-1) = \widehat{g_\alpha}(2) = \widehat{g_\alpha}(N-2) = -\alpha/4,$$

and for all $r \notin \{0, 1, 2, N-1, N-2\}$,

$$\widehat{g_\alpha}(r) = 0.$$

Furthermore, it is clear from the definition of g_α that for all $x \in \mathbb{Z}_N$, $g_\alpha(x) \in [0, 2\alpha]$. Moreover, $\mathsf{E}[g_\alpha] = \widehat{g_\alpha}(0) = \alpha$. This proves (i).

From (1),

$$\Lambda(g_\alpha) = \sum_r \widehat{g_\alpha}(r)^2\widehat{g_\alpha}(-2r) = \alpha^3 - 2\left(\frac{\alpha}{4}\right)^3 = \left(1 - \frac{1}{32}\right)\alpha^3,$$

and

$$\begin{aligned}
\mathsf{E}_x[g_\alpha(x)^3] &= \mathsf{E}_x\left[\left(\sum_r \hat{g}_\alpha(r)e\left(\frac{rx}{N}\right)\right)^3\right] \\
&= \sum_{r_1,r_2,r_3\in\mathbb{Z}_N} \widehat{g_\alpha}(r_1)\widehat{g_\alpha}(r_2)\widehat{g_\alpha}(r_3)\mathsf{E}_x\left[e\left(\frac{r_1x + r_2x + r_3x}{N}\right)\right] \\
&= \sum_{\substack{r_1,r_2,r_3\in\mathbb{Z}_N \\ r_1+r_2+r_3=0}} \widehat{g_\alpha}(r_1)\widehat{g_\alpha}(r_2)\widehat{g_\alpha}(r_3) \\
&= \alpha^3 + 6\alpha\frac{\alpha^2}{16} - 6\frac{\alpha^3}{64} < \frac{3}{2}\alpha^3.
\end{aligned}$$

This proves (ii).

Property (iii) follows directly from Lemma 5.1 applied to the function g_α and $\ell = 5$. To prove (iv), notice that $\sum_x g_\alpha(x)^2 \geq \frac{1}{N}(\sum_x g_\alpha(x))^2 = \alpha^2 N$ so

$$\mathsf{E}_{x\neq y}[g_\alpha(x)g_\alpha(y)] \leq \frac{\alpha^2 N^2 - \alpha^2 N}{N(N-1)} = \alpha^2.$$

Finally, (v) follows from Parseval's identity,

$$\mathsf{E}_x[g_\alpha(x)^2] = \sum_r |\widehat{g_\alpha}(r)|^2 = \frac{5\alpha^2}{4}.$$

■

6. Lower bound construction for product groups

In this section, we prove Theorem 3.4. For convenience, we recall the theorem statement here.

Theorem. *Let $0 < \alpha \leq 1/4$, $0 < \epsilon \leq 20^{-9}$, and $G = \mathbb{Z}_n$ where n is a positive integer such that there exist distinct primes $m_1, \ldots, m_s$ with $s \leq \log_{150}(\epsilon^{-1/4}\alpha^6/8)$ satisfying*

- $n = \prod_{j=1}^s m_j$,
- $\epsilon^{-1/3}/2 < m_1 \leq \epsilon^{-1/3}$, *and*
- *for $i \geq 2$, $n_{i-1}^6 < m_i < \exp(2^{-1}\cdot 64^{-2}\cdot 150^{i-1}\epsilon^{1/4}n_{i-1})/2$ where $n_i = \prod_{j=1}^i m_j$.*

Then there exists a function $f : G \to [0,1]$ with $\mathsf{E}[f] = \alpha$ such that for any $d \in G \setminus \{0\}$,

$$\mathsf{E}_x[f(x)f(x+d)f(x+2d)] \leq \alpha^3(1-\epsilon).$$

Furthermore, $\mathsf{E}_x[f(x)^3] \leq 3\alpha^3/2$ and there exists $\tilde{\alpha} \in [\alpha, \alpha(1+\epsilon^{1/4})]$ such that $f(x) = \tilde{\alpha}$ for at least a $3/4$ fraction of $x \in G$.

We first make a few notation conventions. Note that if $n = \prod_{i=1}^s m_i$ for distinct primes m_i, then

$$G = \mathbb{Z}_n \cong \prod_{i=1}^s \mathbb{Z}_{m_i}.$$

Each element of G can be represented by an s-tuple $(x_1, \ldots, x_s)$ where $x_i \in \mathbb{Z}_{m_i}$. Let $Q_i = \prod_{j=1}^i \mathbb{Z}_{m_i}$. We can think of Q_i as a quotient of G by the subgroup $H_i = \{x \in G : x_j = 0 \text{ for all } j \leq i\}$. We identify $\mathbb{Z}_{m_i}$ as the subgroup of Q_i consisting of elements with $x_j = 0$ for $j < i$, and we identify the quotient $Q_i/\mathbb{Z}_{m_i}$ with Q_{i-1}. We hence use elements of Q_{i-1} to index $\mathbb{Z}_{m_i}$-cosets in Q_i. For an element $x \in G$ or $x \in Q_j$ with $j \geq i$, we denote $x_{[i]} = (x_1, \ldots, x_i)$. For $j < i$, we say that an element x of Q_i is a *lift* of an element y in Q_j if $x_{[j]} = y$. In the following discussion, when the level i is clear from context, if not specified otherwise, the 3-APs would refer to 3-APs in Q_i.

6.1. The construction

Let $s = \lceil \log_{150}(\epsilon^{-1/4}\alpha^6/8)\rceil$. In each level i, for $i \in [s]$, we construct a function $f_i : Q_i \to [0,1]$. Finally, we let $f = f_s : G \to [0,1]$.

We introduce parameters $\mu_1 = \epsilon^{1/4}$ and $\mu_i = 150^{i-1}\tilde{\alpha}^{-6}\epsilon^{1/4}$ for $i \geq 2$, where $\tilde{\alpha} = \alpha(1 + \frac{1}{m_1-1})$.

In the first level, define $f_1 : Q_1 \to [0,1]$ by $f_1(0) = 0$ and $f_1(x) = \alpha(1 + \frac{1}{m_1-1})$ for each $x \in Q_1 \setminus \{0\}$.

For $i \geq 2$, let M_{i-1} be any set of $\mu_i n_{i-1}$ elements of Q_{i-1} so that $f_{i-1}(x) = \tilde{\alpha}$ for any $x \in M_{i-1}$. In level i, we define f_i to be a random function as follows.

For each $x \in M_{i-1}$, we choose $a_x \in \mathbb{Z}_{m_i} \setminus \{0\}$ and $b_x \in \mathbb{Z}_{m_i}$ uniformly and independently at random. For each $y \in Q_i$ such that $y_{[i-1]} = x$, we define

$$f_i(y) = g_{\tilde{\alpha}}(a_x y_i + b_x) = \tilde{\alpha} - \frac{\tilde{\alpha}\cos(2\pi(a_x y_i + b_x)/m_i)}{2} - \frac{\tilde{\alpha}\cos(4\pi(a_x y_i + b_x)/m_i)}{2},$$

where $g_{\tilde{\alpha}}$ is the function with density $\tilde{\alpha}$ and with low 3-AP density defined earlier in Lemma 5.2. Otherwise, for $x \notin M_{i-1}$ and $y \in Q_i$ such that $y_{[i-1]} = x$, we define

$$f_i(y) = f_{i-1}(x).$$

We refer to this as the *random modification* in level i. This defines (random) $f_i : Q_i \to [0,1]$. Finally, we let $f = f_s : G \to [0,1]$. We will show that with positive probability, for each level i, we can pick f_i such that the function f has the desired properties claimed in Theorem 3.4.

6.2. *Proof of Theorem 3.4*

We first claim that the construction is feasible with the above choice of parameters. Note that $\mu_1 \geq 1/m_1$, so $f_1(x) = \tilde{\alpha}$ for all but a μ_1 fraction of elements $x \in Q_1$. For $i \geq 2$, observe that if $f_i(y) \neq \tilde{\alpha}$, then we must have $y_{[1]} = 0$ or $y_{[j]} \in M_j$ for some $j < i$. Thus, the fraction of $y \in Q_i$ for which $f_i(y) \neq \tilde{\alpha}$ is at most $\sum_{j=1}^{i} \mu_j$. Since

$$\sum_{j=1}^{s} \mu_j < 2\mu_s = 2 \cdot 150^{s-1}\epsilon^{1/4}\tilde{\alpha}^{-6} < 1/4 \tag{13}$$

as $s - 1 \leq \log_{150}(\epsilon^{-1/4}\alpha^6/8) < \log_{150}(\epsilon^{-1/4}\tilde{\alpha}^6/8)$, it is possible to choose M_i for each $i \leq s-1$ such that $f_i(x) = \tilde{\alpha}$ for all $x \in M_i$.

We next prove that the function f_i has density α and f_i maps Q_i to $[0,1]$. This is true for $i = 1$. Assume that f_{i-1} has density α and takes values in $[0,1]$, we show that f_i also has these properties. For $x \in Q_i$ such that $x_{[i-1]} \notin M_{i-1}$, $f_i(x) = f_{i-1}(x_{[i-1]}) \in [0,1]$. If $x_{[i-1]} \in M_{i-1}$ then $f_{i-1}(x_{[i-1]}) = \tilde{\alpha} \leq 1/2$. Hence, if $x_{[i-1]} \in M_{i-1}$ then $f_i(x) = g_{\tilde{\alpha}}(ax_i + b)$ for some $a \in \mathbb{Z}_{m_i} \setminus \{0\}$ and $b \in \mathbb{Z}_{m_i}$. Since $g_{\tilde{\alpha}}$ also has density $\tilde{\alpha}$ and takes values in $[0,1]$, we have $f_i(x) = g_{\tilde{\alpha}}(ax_i + b) \in [0,1]$ and the density of f_i over the $\mathbb{Z}_{m_i}$-coset $x_{[i-1]}$ is $f_{i-1}(x_{[i-1]}) = \tilde{\alpha}$. Hence, the density of f_i is the same as the density of f_{i-1}, and f_i takes values in $[0,1]$. By induction, the density of f_i is α and the values of f_i are in $[0,1]$ for all $i \in [s]$.

We denote by $\mathbb{E}_{f_i}$ the expectation over the randomness of f_i (the local modifications in level i), conditioned on a fixed choice of f_{i-1}. Furthermore, all of the probability we

consider will be conditioned on this fixed choice of f_{i-1}, hence in level i we only consider the randomness of the random modification in level i.

The random modification in level i has the following key property. For any $x = (x_1, \dots, x_i) \in Q_i$ such that $x_{[i-1]} = (x_1, \dots, x_{i-1}) \in M_{i-1}$ and $d \in Q_i \setminus \{0\}$ such that $d_{[i-1]} = 0 \in Q_{i-1}$, we have

$$\begin{aligned}&\mathbb{E}_{f_i}[f_i(x) f_i(x+d) f_i(x+2d)]\\ &\quad = \mathbb{E}_{a\in\mathbb{Z}_{m_i}\setminus\{0\}, b\in\mathbb{Z}_{m_i}}[g_{\tilde{\alpha}}(ax_i+b) g_{\tilde{\alpha}}(ax_i+ad_i+b) g_{\tilde{\alpha}}(ax_i+2ad_i+b)]\\ &\quad \le \Lambda(g_{\tilde{\alpha}}) = \tfrac{31}{32}\tilde{\alpha}^3. \end{aligned} \tag{14}$$

This is since when a is chosen uniformly at random from $\mathbb{Z}_{m_i} \setminus \{0\}$ and b is chosen uniformly at random from $\mathbb{Z}_{m_i}$, then for any fixed x_i and nonzero d_i, $(ax_i + b, ax_i + ad_i + b, ax_i + 2ad_i + b)$ is distributed uniformly among all 3-APs with nonzero common difference in $\mathbb{Z}_{m_i}$.

We now proceed to prove that there exists a choice of the modification in each level so that for any $d \in G \setminus \{0\}$,

$$\mathsf{E}_x[f(x) f(x+d) f(x+2d)] \le \alpha^3(1-\epsilon).$$

The main idea is to maintain by induction that for any $i \in [s]$, we can choose f_i which is a random modification of f_{i-1} so that for any $d \in Q_i \setminus \{0\}$,

$$\mathsf{E}_{x\in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] \le \alpha^3(1-\epsilon).$$

For all d such that $d_{[i-1]} = 0$, the above property follows from observation (14) and concentration inequalities. On the other hand, if $d_{[i-1]} \neq 0 \in Q_{i-1}$, then $\mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x) f_{i-1}(x + d_{[i-1]}) f_{i-1}(x + 2d_{[i-1]})]$ is small by the induction hypothesis. We guarantee that with large probability,

$$\mathsf{E}_{x\in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] = \mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x) f_{i-1}(x+d_{[i-1]}) f_{i-1}(x+2d_{[i-1]})]$$

for all d such that $d_{[i-1]} \neq 0$, so that $\mathsf{E}_{x\in Q_i}[f_i(x) f_i(x + d) f_i(x + 2d)]$ is small. Combining these two cases, we obtain a modification f_i of f_{i-1} whose density of 3-APs with common difference d is small for all nonzero $d \in Q_i$.

We now give the proof of Theorem 3.4.

Proof of Theorem 3.4. It is easy to see that

$$\begin{aligned}\mathsf{E}_{x\in Q_1}[f_1(x)^3] &= \alpha^3 \frac{\left(1 + \frac{1}{m_1-1}\right)^3 (m_1 - 1)}{m_1}\\ &< \alpha^3\left(1 + \frac{3}{m_1}\right) \le \alpha^3(1 + 6\epsilon^{1/3}) \le \alpha^3(1 + 2\epsilon^{1/4})\\ &< \tilde{\alpha}^3(1 + 2\mu_1)\end{aligned}$$

for $\epsilon \le 20^{-9}$. Inductively, if

$$\mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x)^3] \le \tilde{\alpha}^3(1 + 2\mu_{i-1}),$$

then

$$\mathsf{E}_{x\in Q_i}[f_i(x)^3] \le \tilde{\alpha}^3(1+2\mu_{i-1}) + \tfrac{1}{2}\mu_i\tilde{\alpha}^3 < \tilde{\alpha}^3(1+2\mu_i), \tag{15}$$

where the first inequality is by (ii) in Lemma 5.2 as we apply the local modification to a μ_i fraction of the $\mathbb{Z}_{m_i}$-cosets, getting at most a $\frac{1}{2}\tilde{\alpha}^3$ increment in the mean-cube density over each of them, and the second inequality follows from our choice of parameters $\mu_i \ge 150\mu_{i-1}$ for all $i \ge 2$.

Let $\mathcal{P}(i)$ be the property that for all $d \in Q_i \setminus \{0\}$,

$$\mathsf{E}_{x\in Q_i}[f_i(x)f_i(x+d)f_i(x+2d)] \le \alpha^3(1-\epsilon).$$

We will prove by induction that in level i, the modifications can be chosen so that $\mathcal{P}(i)$ holds.

Consider the base case $i = 1$. Recall that $\epsilon^{-1/3}/2 \le m_1 \le \epsilon^{-1/3}$. For any $d \in Q_1 \setminus \{0\}$,

$$\begin{aligned}\mathsf{E}_{x\in Q_1}[f_1(x)f_1(x+d)f_1(x+2d)] &= \alpha^3\frac{\left(1+\frac{1}{m_1-1}\right)^3(m_1-3)}{m_1}\\ &= \alpha^3\frac{m_1^2(m_1-3)}{(m_1-1)^3} = \alpha^3\left(1-\frac{3m_1-1}{(m_1-1)^3}\right)\\ &\le \alpha^3\left(1-\frac{1}{m_1^2}\right) \le \alpha^3(1-\epsilon).\end{aligned}$$

This establishes $\mathcal{P}(1)$. Next, we continue with the inductive step. Assume that $\mathcal{P}(i-1)$ holds. We prove that we can choose the modification in level i so that $\mathcal{P}(i)$ also holds. This follows from the following two claims.

Claim 6.1. *With probability larger than* $1/2$ *(over the randomness of* f_i*), conditioned on a fixed choice of* f_{i-1} *satisfying* $\mathcal{P}(i-1)$*, for all* $d \in Q_i \setminus \{0\}$ *with* $d_{[i-1]} = 0$*,*

$$\mathsf{E}_{x\in Q_i}[f_i(x)f_i(x+d)f_i(x+2d)] \le \alpha^3(1-\epsilon).$$

Claim 6.2. *With probability larger than* $1/2$ *(over the randomness of* f_i*), conditioned on a fixed choice of* f_{i-1} *satisfying* $\mathcal{P}(i-1)$*, for all* $d \in Q_i \setminus \{0\}$ *with* $d_{[i-1]} \ne 0$*,*

$$\begin{aligned}&\mathsf{E}_{x\in Q_i}[f_i(x)f_i(x+d)f_i(x+2d)]\\ &\qquad = \mathsf{E}_{x\in Q_{i-1}}[f_{i-1}(x)f_{i-1}(x+d_{[i-1]})f_{i-1}(x+2d_{[i-1]})].\end{aligned}$$

Combining Claims 6.1 and 6.2, by the union bound, the modification in level i fails to satisfy $\mathcal{P}(i)$ with probability strictly less than 1. Thus we can choose a modification satisfying $\mathcal{P}(i)$ in level i. This completes the induction. Thus, there exists $f = f_s$ which satisfies $\mathcal{P}(s)$, so for any nonzero d in G,

$$\mathsf{E}_{x\in G}[f(x)f(x+d)f(x+2d)] = \mathsf{E}_{x\in G}[f_s(x)f_s(x+d)f_s(x+2d)] \le \alpha^3(1-\epsilon).$$

This completes the proof of Theorem 3.4.

Now we turn to the proofs of Claims 6.1 and 6.2.

Proof of Claim 6.1. Let $d \in Q_i \setminus \{0\}$ be such that $d_{[i-1]} = 0$. By (14), for any $x \in Q_i$ with $x_{[i-1]} \in M_{i-1}$,

$$\mathbb{E}_{f_i}[f_i(x) f_i(x+d) f_i(x+2d)] \le \tfrac{31}{32}\tilde{\alpha}^3.$$

Hence, for $y \in M_{i-1}$,

$$\mathbb{E}_{f_i} \mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)] \le \tfrac{31}{32}\tilde{\alpha}^3.$$

Note that the random variables $\mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)]$, for $y \in M_{i-1}$, are independent (under the randomness of the modification in level i, conditioned on a fixed choice of f_{i-1}). Thus the probability that

$$\mathsf{E}_{y \in M_{i-1}} \mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)] \ge \tfrac{63}{64}\tilde{\alpha}^3$$

is at most $\exp(-2^{-1} \cdot 64^{-2} \mu_i n_{i-1} \tilde{\alpha}^6)$ by Hoeffding's inequality.

For $d \in Q_i$ such that $d_{[i-1]} = 0$, we have

$$\begin{aligned}
&\mathsf{E}_{x \in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] \\
&= \mathsf{E}_{y \in Q_{i-1}}[f_{i-1}(y)^3] \\
&\quad + \frac{|M_{i-1}|}{|Q_{i-1}|}\left(\mathsf{E}_{y \in M_{i-1}} \mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)] - \mathsf{E}_{y \in M_{i-1}}[f_{i-1}(y)^3]\right) \\
&\le \tilde{\alpha}^3(1 + 2\mu_{i-1}) + \mu_i \cdot \left(\mathsf{E}_{y \in M_{i-1}} \mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)] - \tilde{\alpha}^3\right)
\end{aligned}$$

where the equality follows from $f_i(x) = f_i(x+d) = f_i(x+2d) = f_{i-1}(y)$ if $d_{[i-1]} = 0$ and $x_{[i-1]} = y \notin M_{i-1}$, and the inequality follows from (15) and $f_{i-1}(y) = \tilde{\alpha}$ for $y \in M_{i-1}$. Thus, if

$$\mathsf{E}_{y \in M_{i-1}} \mathsf{E}_{x \in Q_i, x_{[i-1]}=y}[f_i(x) f_i(x+d) f_i(x+2d)] \le \tfrac{63}{64}\tilde{\alpha}^3,$$

then

$$\mathsf{E}_{x \in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] \le \tilde{\alpha}^3(1 + 2\mu_{i-1}) - \mu_i \tilde{\alpha}^3/64.$$

Since

$$\alpha^3(1-\epsilon) \ge \tilde{\alpha}^3(1-\epsilon^{1/4}) > \tilde{\alpha}^3(1 + 2\mu_{i-1}) - \mu_i \tilde{\alpha}^3/64,$$

by the union bound, the probability that there exists $d \in Q_i \setminus \{0\}$ with $d_{[i-1]} = 0$ and

$$\mathsf{E}_{x \in Q_i}[f_i(x) f_i(x+d) f_i(x+2d)] \ge \alpha^3(1-\epsilon)$$

is at most

$$m_i \exp(-2^{-1} \cdot 64^{-2} \mu_i n_{i-1} \tilde{\alpha}^6) < 1/2,$$

where we have used the upper bound on m_i in the theorem statement. ∎

Proof of Claim 6.2. Recall that for a $\mathbb{Z}_{m_i}$-coset represented by $w \in Q_{i-1}$, f_i is either a constant function on w if $w \notin M_{i-1}$, or otherwise $f_i(x) = g_{\tilde{\alpha}}(a_w x_i + b_w)$ where

$$g_{\tilde{\alpha}}(x) = \tilde{\alpha} - \frac{\tilde{\alpha}\cos(2x/m_i)}{2} - \frac{\tilde{\alpha}\cos(4x/m_i)}{2}$$

as defined in Lemma 5.2 and $a_w \in \mathbb{Z}_{m_i} \setminus \{0\}$ and $b_w \in \mathbb{Z}_{m_i}$ are chosen uniformly and independently for each $w \in M_{i-1}$. For each 3-AP $(w, w+d', w+2d')$ with common difference $d' \in Q_{i-1} \setminus \{0\}$, and for any lift d of d', we have

$$\begin{aligned}
&\mathsf{E}_{x\in Q_i, x_{[i-1]}=w}[f_i(x)f_i(x+d)f_i(x+2d)] \\
&\quad = \mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(1)}(a_1y+b_1)g_{(2)}(a_2y+a_2d_i+b_2)g_{(3)}(a_3y+2a_3d_i+b_3)] \\
&\quad = \mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(1)}(a_1y+c_1)g_{(2)}(a_2y+c_2)g_{(3)}(a_3y+c_3)], \qquad (16)
\end{aligned}$$

where $g_{(j)} : \mathbb{Z}_{m_i} \to [0,1]$ can be either the function $g_{\tilde{\alpha}}$ or a constant function, $a_j \in \mathbb{Z}_{m_i} \setminus \{0\}$, $b_j \in \mathbb{Z}_{m_i}$ are chosen uniformly and independently at random, and $c_1 = b_1$, $c_2 = a_2d_i + b_2$, $c_3 = 2a_3d_i + b_3$. Note that if we fix the modification (i.e., fixing each a_j and b_j), changing d to a different lift of d' would only change c_j in (16), and would not change the coefficients of y in $g_{(1)}, g_{(2)}, g_{(3)}$ in the last line of (16). Let $J \subseteq [3]$ be the set of indices such that $g_{(j)} = g_{\tilde{\alpha}}$. By Lemma 5.1 applied to the function $g_{\tilde{\alpha}}$ and $h = |J| \le 3$, with probability at least $1 - 125/(m_i - 1)$,

$$\mathsf{E}_{y\in\mathbb{Z}_{m_i}} \prod_{j\in J} g_{(j)}(a_jy+u_j) = \prod_{j\in J} \mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(j)}(y)]$$

for all $u_j \in \mathbb{Z}_{m_i}$. Since $g_{(j)}$ is a constant function for $j \notin J$, we find that with probability at least $1 - 125/(m_i - 1)$,

$$\begin{aligned}
&\mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(1)}(a_1y+c_1)g_{(2)}(a_2y+c_2)g_{(3)}(a_3y+c_3)] \\
&\qquad = \mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(1)}(y)]\mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(2)}(y)]\mathsf{E}_{y\in\mathbb{Z}_{m_i}}[g_{(3)}(y)] \\
&\qquad = f_{i-1}(w)f_{i-1}(w+d')f_{i-1}(w+2d').
\end{aligned}$$

Thus, by the union bound, with probability at least $1 - 125n_{i-1}^2/(m_i - 1)$, for every 3-AP $(w, w+d', w+2d')$ in Q_{i-1} with nonzero common difference d', and for all lifts d of d' in Q_i,

$$\begin{aligned}
&\mathsf{E}_{x\in Q_i, x_{[i-1]}=w}[f_i(x)f_i(x+d)f_i(x+2d)] \\
&\qquad = f_{i-1}(w)f_{i-1}(w+d')f_{i-1}(w+2d').
\end{aligned}$$

For $i \ge 2$, $m_i \ge n_{i-1}^6$, and $m_i \ge m_2 \ge \epsilon^{-2}/64 > 10^6$, so $125n_{i-1}^2/(m_i - 1) < 1/2$. Hence with probability larger than $1/2$, for all $d \in Q_i$ such that $d_{[i-1]} \neq 0$,

$$\begin{aligned}
&\mathsf{E}_{x\in Q_i}[f_i(x)f_i(x+d)f_i(x+2d)] \\
&\qquad = \mathsf{E}_{x\in Q_{i-1}} f_{i-1}(x)f_{i-1}(x+d_{[i-1]})f_{i-1}(x+2d_{[i-1]}). \qquad \blacksquare
\end{aligned}$$

Thus, assuming that $\mathcal{P}(i-1)$ holds, we can choose the modification in level i so that $\mathcal{P}(i)$ holds. By induction, we can find a function f_s which satisfies $\mathcal{P}(s)$. Notice that $f_s(x) = \tilde{\alpha}$ for at least a $3/4$ fraction of $x \in G$ by (13), and $\mathbb{E}_x[f_s(x)^3] \le 3\alpha^3/2$ by (15) with $i = s$. The function $f = f_s$ then satisfies the conclusion of Theorem 3.4. $\blacksquare$

7. Lower bound construction for intervals

In this section we prove Theorem 3.3, restated below for convenience.

Theorem. *There are positive absolute constants c, α_0 such that the following holds. If $0 \le \alpha \le \alpha_0$, $0 < \epsilon \le \alpha^7$, and $N \le \text{tower}(c \log(1/\epsilon))$, then there is a function $f : [N] \to [0,1]$ with $\mathsf{E}[f] = \alpha$ such that for any $0 < d < N/2$,*

$$\mathsf{E}_{x \in [N-2d]}[f(x) f(x+d) f(x+2d)] \le \alpha^3(1-\epsilon).$$

By Appendix A, in order to prove Theorem 3.3, we can (and will) assume that $N \ge \epsilon^{-15}$.

Before proving Theorem 3.3, we first need an auxiliary construction of a set with relatively low 3-AP density given its density. Recall from the introduction that $N(\alpha)$ is the least positive integer such that if $N \ge N(\alpha)$ and $A \subset [N]$ with $|A| \ge \alpha N$, then A contains a 3-AP.

Lemma 7.1. *For $\alpha > 0$ sufficiently small, there is a subset $T \subset \mathbb{Z}_n$ with $|T| \ge \alpha n$ and with 3-AP density at most* $\max(1/n, 2\alpha/N(6\alpha))$.

Proof. Let $N = N(6\alpha) - 1$, so there is $A \subset [N]$ with $|A| = \lceil 6\alpha N \rceil$ which has no nontrivial 3-AP.

First assume $n \le 4N$. Partition $[N]$ into at most $2N/n + 1 \le 6N/n$ intervals of length at most $\lceil n/2 \rceil$. The set A contains at least $|A|/(6N/n) \ge \alpha n$ elements in one of these intervals. Viewed as a subset of $\mathbb{Z}_n$, we have a subset of $\mathbb{Z}_n$ with density at least α and with no nontrivial 3-AP, and hence 3-AP density at most $1/n$.

So we may assume $n > 4N$. Integers x, y, z form an *approximate 3-AP* if $|2y - x - z| \le 1$. Let $S := \{2a : a \in A\}$, so S has no approximate 3-AP. Let $t = \lfloor \frac{n}{4N} \rfloor$. Consider the set $I_i := \{(i-1)t + 1, (i-1)t + 2, \ldots, (i-1)t + t\}$ of t consecutive integers. Let T be the union of the sets I_i with $i \in S$. The set T has size $|T| = |A|t \ge \alpha n$. Also, every element of T is a positive integer at most $(2N-1)t + t \le n/2$. So if $x, y, z \in T$ are such that $(x, y, z) \pmod n$ form a 3-AP in $\mathbb{Z}_n$, then (x, y, z) is also a 3-AP of integers. Since S has no approximate 3-AP, it follows that the only 3-APs in T are those where the three terms are in the same interval I_i. In each interval I_i, which has size t, the number of 3-APs (with any integer difference allowed) is $t + 2\lfloor \frac{t^2-1}{4} \rfloor$. There are $|A|$ intervals I_i whose union is T. The number of 3-APs in $\mathbb{Z}_n$ is n^2. Hence, the 3-AP density of T as a subset of $\mathbb{Z}_n$ is

$$\left(t + 2\left\lfloor \frac{t^2-1}{4} \right\rfloor\right)|A|/n^2 \le \frac{2\alpha}{N(6\alpha)}. \qquad \blacksquare$$

The Behrend construction [2] implies that if $\alpha > 0$ is sufficiently small, then $N(6\alpha) \ge 2^{\frac{1}{9}(\log_2(1/\alpha))^2}$. Together with the previous lemma, we have the following immediate corollary.

Lemma 7.2. *If $\alpha > 0$ is sufficiently small, then for any positive integer n, there is a subset of $\mathbb{Z}_n$ with density at least α and 3-AP density at most* $\max(1/n, 2^{-\frac{1}{9}(\log_2(1/\alpha))^2})$.

7.1. The construction and proof of Theorem 3.3

We next construct a function $f : [N] \to [0,1]$ with $\mathsf{E}_x[f(x)] = \alpha$ such that for any $0 < d < N/2$, $\mathsf{E}_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)] \le \alpha^3(1-\epsilon)$. The construction is done in three steps.

Step 1: Choose β so that $\beta \le \epsilon^2$ and $N' = N(1-\beta)$ has a divisor q such that $N^{1/5} < q < \sqrt{\beta\alpha^3(1-\epsilon)N}$, N'/q is prime, and q satisfies the condition in Theorem 3.4 with parameters $\alpha_{3.4} = \alpha$ and $\epsilon_{3.4} = 4\epsilon$. Here, for clarity, we include the theorem index in the subscript of the parameters in the theorem. Note that if q satisfies the condition in Theorem 3.4 with parameters $\alpha_{3.4} = \alpha$ and $\epsilon_{3.4} = 4\epsilon$, then for any $\alpha_1 \in [\alpha, 1/4]$, q also satisfies the condition in Theorem 3.4 with parameters $\alpha_{3.4} = \alpha_1$ and $\epsilon_{3.4} = 4\epsilon$. The existence of β, N', q is guaranteed by Lemma 7.4, which is deferred to the end of the section. Let

$$\tilde{\alpha} := (1-\beta)^{-1}\alpha.$$

Step 2: Since $\tilde{\alpha} \in [\alpha, 1/4]$, we can apply Theorem 3.4 with $G = \mathbb{Z}_q$, $\alpha_{4.4} = \tilde{\alpha}$, $\epsilon_{4.4} = 4\epsilon$. We obtain $g : \mathbb{Z}_q \to [0,1]$ with density $\tilde{\alpha}$, and mean-cube density at most $\frac{3}{2}\tilde{\alpha}^3$, such that for each $d \in \mathbb{Z}_q \setminus \{0\}$,

$$\mathsf{E}_x[g(x)g(x+d)g(x+2d)] \le \tilde{\alpha}^3(1-4\epsilon) \le \alpha^3(1-3\epsilon),$$

and there exists $\alpha_* \in [\tilde{\alpha}, \tilde{\alpha}(1+\epsilon^{1/4})]$ such that $|\{x \in \mathbb{Z}_q : g(x) = \alpha_*\}| \ge 3q/4$. For an integer x, denote $\bar{x}_q = x \bmod q \in \mathbb{Z}_q$. Define $f_2 : [N] \to [0,1]$ by $f_2(x) = g(\bar{x}_q)$ for $x \in [N']$ and $f_2(x) = 0$ for $x > N'$.

Step 3: Let $n = N'/q$, which is prime. Apply Lemma 7.2 to find $X \subset \mathbb{Z}_n$ with density at least α_* and 3-AP density at most $\max(1/n, 2^{-\log_2(1/\alpha_*)^2/9})$. We use ξ to denote the characteristic function of X scaled by $\alpha_* n/|X|$, so $\xi(x) = \alpha_* n X(x)/|X|$. Then $\mathsf{E}_{x\in\mathbb{Z}_n}[\xi(x)] = \alpha_*$ and since $|X|/n \ge \alpha_*$,

$$\begin{aligned}\mathsf{E}_{x,d\in\mathbb{Z}_n}[\xi(x)\xi(x+d)\xi(x+2d)] &\le \mathsf{E}_{x,d\in\mathbb{Z}_n}[X(x)X(x+d)X(x+2d)] \\ &\le \max(1/n, 2^{-\log_2(1/\alpha_*)^2/9}).\end{aligned}$$

For each $t \in \mathbb{Z}_q$, we define $P_t = \{x \in [N'] : \bar{x}_q = t\}$, which forms an arithmetic progression of length n. Let $x_1 < \ldots < x_n$ be the elements of P_t in increasing order. We define a bijection $\phi_t : P_t \to \mathbb{Z}_n$ such that $\phi_t(x_i) = i \bmod n$. Observe that if (x, y, z) is a 3-AP in P_t, then $(\phi(x), \phi(y), \phi(z))$ is a 3-AP in $\mathbb{Z}_n$. For each $t \in \mathbb{Z}_q$, we choose independently and uniformly at random $a_t \in \mathbb{Z}_n \setminus \{0\}, b_t \in \mathbb{Z}_n$, with independent choices for different t. Define f_3 by $f_3(x) = \xi(a_t\phi_t(x) + b_t)$ for x such that $\bar{x}_q = t$ and $g(t) = \alpha_*$, and $f_3(x) = f_2(x)$ otherwise.

We let $f = f_3$. It is easy to see that $\mathsf{E}_x[f_3(x)] = \alpha$. We now prove that there exists a choice of randomness (in Step 3) such that for each positive integer $d < N/2$,

$$\mathsf{E}_{x\in[N-2d]}[f(x)f(x+d)f(x+2d)] \le \alpha^3(1-\epsilon).$$

Proof of Theorem 3.3. We reuse the notations from the description of the construction. For a 3-AP $(x, x+d, x+2d)$ in $[N]$, we refer to $f_3(x)f_3(x+d)f_3(x+2d)$ as its *weight*. For any common difference $d \geq N'/2$, all the 3-APs $(x, x+d, x+2d)$ in $[N]$ have zero weight since $x+2d > N'$. Hence the density of 3-APs with common difference d of f_3 is 0. For any common difference $d \geq \frac{N'-\beta N\alpha^3(1-\epsilon)}{2}$, let $u = N'-2d$. Then $u \leq \beta N\alpha^3(1-\epsilon)$. The number of 3-APs with common difference d in $[N']$ is at most u, and hence the number of 3-APs in $[N]$ with nonzero weight is at most u. The number of 3-APs with common difference d in $[N]$ is $N-2d = \beta N + u$, so the density of 3-APs with common difference d in $[N]$ is at most $\frac{u}{\beta N+u} < \alpha^3(1-\epsilon)$ since $u \leq \frac{\beta N\alpha^3(1-\epsilon)}{1-\alpha^3(1-\epsilon)}$.

For d such that $0 < d < \frac{N'-\beta N\alpha^3(1-\epsilon)}{2}$, the number of 3-APs of common difference d in $[N']$ is at least $\beta N\alpha^3(1-\epsilon)$. Partition the 3-APs with common difference d in $[N']$ into different classes according to the congruence class modulo q of the 3-AP (so the class a 3-AP belongs to is determined by the congruence class modulo q of the first element of the 3-AP). Since $\beta N\alpha^3(1-\epsilon) > q^2$, all classes of 3-APs modulo q with common difference $\bar{d}_q$ appear, each class with at least q elements, and any two classes differ in size by at most 1. Hence,

$$\left|\mathsf{E}_{x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)] - \mathsf{E}_{y\in\mathbb{Z}_q}[g(y)g(y+\bar{d}_q)g(y+2\bar{d}_q)]\right| \leq 1/q.$$

By the construction, if $\bar{d}_q \neq 0$ then

$$\mathsf{E}_{y\in\mathbb{Z}_q}[g(y)g(y+\bar{d}_q)g(y+2\bar{d}_q)] \leq \alpha^3(1-3\epsilon),$$

and if $\bar{d}_q = 0$ then

$$\mathsf{E}_{y\in\mathbb{Z}_q}[g(y)g(y+\bar{d}_q)g(y+2\bar{d}_q)] \leq \tfrac{3}{2}\tilde{\alpha}^3.$$

Thus, for d nonzero modulo q,

$$\mathsf{E}_{x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)] \leq \alpha^3(1-3\epsilon) + 1/q < \alpha^3(1-2\epsilon),$$

and for d divisible by q,

$$\mathsf{E}_{x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)] \leq \tfrac{3}{2}\tilde{\alpha}^3 + \tfrac{1}{q}. \tag{17}$$

In the third step, suppose d is nonzero and divisible by q and let $t \in \mathbb{Z}_q$ with $g(t) = \alpha_*$. For $x \in [N'-2d]$ with $\bar{x}_q = t$, one has $(f_3(x), f_3(x+d), f_3(x+2d)) = (\xi(a_t\phi_t(x)+b_t), \xi(a_t\phi_t(x+d)+b_t), \xi(a_t\phi_t(x+2d)+b_t))$. Recall that a_t is uniformly distributed over $\mathbb{Z}_n \setminus \{0\}$ and b_t is uniformly distributed over $\mathbb{Z}_n$, so $(a_t\phi_t(x)+b_t, a_t\phi_t(x+d)+b_t, a_t\phi_t(x+2d)+b_t)$ is uniformly distributed over the 3-APs in $\mathbb{Z}_n$ with nonzero common difference. Thus,

$$\mathbb{E}_{f_3}\mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] = \lambda(\xi) \leq \Lambda(\xi)$$
$$\leq \max(1/n, 2^{-\log_2(1/\alpha_*)^2/9}) \leq \alpha_*^3/10,$$

where $\lambda(\xi)$ is the density of 3-APs with nonzero common difference of ξ, $\Lambda(\xi)$ is the density of 3-APs of ξ, and we have used that $n \geq \sqrt{N}/2 \geq \epsilon^{-15/2}/2 \geq \alpha^{-10}$ and $\alpha \leq \alpha_0$ is sufficiently small. Thus, for each $t \in \mathbb{Z}_q$ such that $g(t) = \alpha_*$,

$$\mathbb{E}_{f_3}\mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] \leq \mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)] - 9\alpha_*^3/10. \quad (18)$$

For each $t \in \mathbb{Z}_q$ such that $g(t) \neq \alpha_*$,

$$\mathbb{E}_{f_3}\mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] = \mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)]. \quad (19)$$

Hence,

$$\mathbb{E}_{f_3}\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] \leq \frac{3\alpha_*^3}{2} + \frac{1}{q} - \frac{3}{4}\cdot\frac{9\alpha_*^3}{10} \leq \frac{5\alpha_*^3}{6},$$

where in the first inequality we used (17)–(19) together with the fact that $g(t) = \alpha_*$ for at least a $3/4$ fraction of $t \in \mathbb{Z}_q$, and in the second inequality we used $q > N^{1/5} \geq \epsilon^{-3} \geq \alpha^{-21} > 120/\alpha_*^3$. Notice that for fixed nonzero d divisible by q, the random variables $\mathsf{E}_{\bar{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)]$, for $t \in \mathbb{Z}_q$, are independent. By Hoeffding's inequality, the probability that

$$\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] \geq \alpha_*^3 - \alpha_*^3/12$$

is at most $\exp(-2\cdot(12^{-1}\alpha_*^3)^2 q) = \exp(-72^{-1}\alpha_*^6 q)$. Noting that $\alpha_*^3 - \alpha_*^3/12 \leq \alpha^3(1-\epsilon)$, by the union bound, the probability that there exists a nonzero common difference d which is divisible by q such that

$$\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] \geq \alpha^3(1-\epsilon)$$

is at most $(N/q)\exp(-72^{-1}\alpha_*^6 q) < q^4\exp(-72^{-1}\alpha_*^6 q) < 1/2$, as $N < q^5$, $q > \epsilon^{-3}$ and $\epsilon \leq \alpha^7$.

If d is not divisible by q, each 3-AP with common difference d occupies three different modulo q classes, and hence the weights of the elements in the 3-AP are independent random variables. By construction, for each $x \in [N']$, $\mathbb{E}_{f_3}[f_3(x)] = f_2(x)$. Hence, by independence, if $(x, x+d, x+2d) \in [N']^3$,

$$\mathbb{E}_{f_3}[f_3(x)f_3(x+d)f_3(x+2d)] = f_2(x)f_2(x+d)f_2(x+2d).$$

Thus,

$$\mathbb{E}_{f_3}\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] = \mathsf{E}_{x\in[N'-2d]}[f_2(x)f_2(x+d)f_2(x+2d)] \leq \alpha^3(1-2\epsilon).$$

For this fixed d, we can partition $\mathbb{Z}_q$ into five sets $S_1, \ldots, S_5$ such that for each $1 \leq i \leq 5$, the 3-APs $(t, t+\bar{d}, t+2\bar{d})$, $t \in S_i$, are disjoint, and $|S_i| \geq q/10$. For each set S_i, the ran-

dom variables $\mathsf{E}_{\tilde{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)]$, for $t\in S_i$, are independent. By Hoeffding's inequality, the probability that

$$\mathsf{E}_{t\in S_i}\mathsf{E}_{\tilde{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)] \geq \mathbb{E}_{f_3}\mathsf{E}_{t\in S_i}\mathsf{E}_{\tilde{x}_q=t,x\in[N'-2d]}[f_3(x)f_3(x+d)f_3(x+2d)]+\epsilon\alpha^3$$

is at most $\exp(-2(\epsilon\alpha^3)^2q/10)=\exp(-5^{-1}\cdot\epsilon^2\alpha^6q)$. By the union bound, the probability that there exists a common difference d not divisible by q with

$$\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+2d)f_3(x+2d)]\geq\alpha^3(1-2\epsilon)+\epsilon\alpha^3=\alpha^3(1-\epsilon)$$

is at most $5N\exp(-5^{-1}\epsilon^2\alpha^6q)<1/2$, where we use $N<q^5$, $q>\epsilon^{-3}$ and $\epsilon\leq\alpha^7$.

Since $f_3(x)=0$ for all $x\notin[N']$,

$$\mathsf{E}_{x\in[N-2d]}[f_3(x)f_3(x+2d)f_3(x+2d)]\leq\mathsf{E}_{x\in[N'-2d]}[f_3(x)f_3(x+2d)f_3(x+2d)].$$

Hence, with positive probability, the function f_3 satisfies the required properties in Theorem 3.3. ∎

To finish the proof, we prove that the parameters in Step 1 of the construction described at the beginning of the subsection can be chosen. We first prove that we can approximate any large integer with one satisfying the conditions in Theorem 3.4.

Lemma 7.3. *There exist constants $c,\alpha_0>0$ such that if $0\leq\alpha\leq\alpha_0$, $0<\epsilon\leq\alpha^7$, and r is an integer satisfying $\epsilon^{-15}\leq r\leq\mathrm{tower}(c\log(1/\epsilon))$, then we can choose $s\in[2,\lceil\log_{150}(\epsilon^{-1/4}\alpha^6/8)\rceil]$ and primes $m_1,\dots,m_s$ and q satisfying the following properties:*

- $\epsilon^{-1/3}/2\leq m_1\leq\epsilon^{-1/3}$,
- *for $i\geq2$, $n_{i-1}^6<m_i<\exp(2^{-1}\cdot64^{-2}\cdot150^{i-1}\epsilon^{1/4}n_{i-1})$ where $n_i=\prod_{j=1}^i m_j$,*
- *for $n=n_s$, we have $n\in[r(1-\epsilon^2),r]$.*

Proof. Choose a sequence of s real numbers $\tilde{m}_1,\dots,\tilde{m}_s$ satisfying the following properties: $\tilde{m}_1$ is a prime number with $\epsilon^{-1/3}/2\leq\tilde{m}_1\leq\epsilon^{-1/3}$, $\tilde{n}_{i-1}^6<\tilde{m}_i<\exp(2^{-1}\cdot64^{-2}\cdot150^{i-1}\epsilon^{1/4}\tilde{n}_{i-1})$ for $i\geq2$ with $\tilde{n}_i=\prod_{j=1}^i\tilde{m}_j$, and $\prod_{i=1}^s\tilde{m}_i=r$. The existence of $\tilde{m}_1$ is guaranteed by Bertrand's postulate. Since $\epsilon^{-15}\leq r\leq\mathrm{tower}(c\log(1/\epsilon))$ and $\epsilon\leq\alpha^7$, if we choose c,α_0 sufficiently small, it is easy to see that there exists a choice of s and $\tilde{m}_1,\dots,\tilde{m}_s$ satisfying these properties.

For each i let m_i be the largest prime such that $m_i\leq\tilde{m}_i$. So $m_1=\tilde{m}_1$, and from [1], for all $\tilde{m}_i$ large enough, $m_i\geq\tilde{m}_i-\tilde{m}_i^{0.525}$. Hence

$$n=\prod_{i=1}^s m_i\geq\tilde{m}_1\prod_{i=2}^s(\tilde{m}_i-\tilde{m}_i^{0.525})\geq r\prod_{i=2}^s(1-\tilde{m}_i^{-0.475})\geq r\exp\left(-\sum_{i=2}^s\frac{2}{\tilde{m}_i^{0.475}}\right)$$
$$\geq r\exp\left(-\frac{4}{\tilde{m}_2^{0.475}}\right)\geq r\exp(-\epsilon^{2.5})\geq r(1-\epsilon^{2.5}),$$

where we have applied the inequality $\exp(-x) \geq 1-x \geq \exp(-2x)$ for $0 \leq x \leq 1/2$, $\sum_{i=2}^{s} \frac{1}{\tilde{m}_i^{0.475}} \leq \frac{1}{\tilde{m}_2^{0.475}} \sum_{i \geq 0} \frac{1}{2^i} \leq \frac{2}{\tilde{m}_2^{0.475}}$ as $\tilde{m}_{i+1} \geq \tilde{m}_i^6 \geq 2^{1/0.475} \tilde{m}_i \geq \cdots \geq 2^{(i-1)/0.475} \tilde{m}_2$ and $\tilde{m}_i \geq \tilde{m}_2 \geq \epsilon^{-6}$ for $i \geq 2$. Thus we can choose $n = \prod_{i=1}^{s} m_i$ and $r(1-\epsilon^2) \leq n \leq r$. ∎

Using Lemma 7.3, we prove that the parameters N', q in Step 1 of the construction can be chosen.

Lemma 7.4. *Let $N \geq \epsilon^{-15}$. There exist q, N' such that N'/q is prime, q satisfies the conditions in Theorem 3.4, and $(1-\epsilon^2)N \leq N' \leq N$.*

Proof. We choose p to be a prime number in $(N^{1/5}, \sqrt{\epsilon^4 \alpha^3 (1-\epsilon) N})$. Let $r = \lfloor N(1-\epsilon^2/4)/p \rfloor \geq N(1-\epsilon^2/2)/p \geq \epsilon^{-7}$. Applying Lemma 7.3 with the above choice of r and with $\epsilon_{7.3} = 4\epsilon$, we find n such that $(1-\epsilon^{2.5})r \leq n \leq r$ and n further satisfies the conditions in Theorem 3.4, applied with $\alpha_{3.4} \in [\alpha, \alpha(1+\epsilon)]$ and $\epsilon_{3.4} \leq 4\epsilon$. Then we let $N' = pn$ and $q = n$.

We have

$$N' = pn \geq (1-\epsilon^{2.5}) rp \geq (1-\epsilon^2/2)(1-\epsilon^{2.5}) N \geq (1-\epsilon^2) N,$$

thus $(1-\epsilon^2)N \leq N' \leq N$, finishing the proof. ∎

Appendix A. Proof of Theorem 1.5 from Theorem 3.3

In this appendix, we show how to deduce Theorem 1.5 from its functional version, Theorem 3.3.

We first show that Theorem 1.5 holds if it holds in the case $N \geq \epsilon^{-15}$. Assume that $N < \epsilon^{-15}$. Recall that $N(\alpha)$ denotes the least positive integer such that if $N \geq N(\alpha)$ then any $A \subset [N]$ with $|A| \geq \alpha N$ contains a nontrivial 3-AP. If $N < N(\alpha)$, there exists a subset A of $[N]$ with $|A| \geq \alpha N$ and A does not contain a nontrivial 3-AP. In this case, for all $d \neq 0$,

$$\mathsf{E}_{x \in [N-2d]} [A(x) A(x+d) A(x+2d)] = 0,$$

so the conclusion of Theorem 1.5 holds. By Behrend's bound [2,10], we find that $N(\alpha) \geq \exp((\log(1/\alpha))^2/6)$ for $\alpha > 0$ sufficiently small. For $N \geq N(\alpha) \geq \exp((\log 1/\alpha)^2/6)$, let $\epsilon_0 = N^{-1/15}$ so $N = \epsilon_0^{-15}$. We have $\epsilon_0 > \epsilon$ and $\epsilon_0 \leq \exp(-(\log 1/\alpha)^2/90) \leq \alpha^{12}$ (assuming α_0 is small enough). By choosing α_0 in Theorem 1.5 small enough, we may assume that for all $x \leq \alpha_0$, $x^{-15} < \text{tower}(c \log(1/x))$. Then $N = \epsilon_0^{-15} < \text{tower}(c \log(1/\epsilon_0))$ as $\epsilon_0 < \alpha \leq \alpha_0$. Thus, $\epsilon_0^{-15} \leq N \leq \text{tower}(c \log(1/\epsilon_0))$, so we can apply Theorem 1.5 with ϵ_0 in place of ϵ to obtain the desired set A. The same argument also shows that we only need to prove Theorem 3.3 when $N \geq \epsilon^{-15}$.

We next discuss how to obtain a set A with the properties in Theorem 1.5 from Theorem 3.3 when $N \geq \epsilon^{-15}$. This follows via a standard sampling argument which is essentially similar to that in [13, Lemma 9]. However, there are some small differences

to the argument which we now highlight. Given a function $f : [N] \to [0,1]$ such that the density of 3-APs with common difference d of f is small for all $0 < d < N/2$, we sample a set A where each element $x \in [N]$ is in A with probability $f(x)$ independent of each other. If the density of 3-APs with common difference d in A is concentrated around its expectation, which is the density of 3-APs with common difference d of f, then it is small with high probability. However, for d near $N/2$, there are very few 3-APs with common difference d, and we do not have sufficiently strong concentration to be able to take a union bound over all such d. To get around this, we define a function f' which is 0 for all x close to N, and which is equal to f elsewhere, and sample the set A from f'. This ensures that for common differences d which are close to $N/2$, the set A contains very few 3-APs with common difference d.

We now carry out the details. By Theorem 3.3 applied with α replaced by $\alpha + 2\epsilon$ and ϵ replaced by $12\epsilon/\alpha^3 \le \alpha^7$, we can find a function $f : [N] \to [0,1]$ such that for any $0 < d < N/2$,

$$\mathsf{E}_{x \in [N-2d]}[f(x)f(x+d)f(x+2d)] \le (\alpha + 2\epsilon)^3(1 - 12\epsilon/\alpha^3).$$

Define $f' : [N] \to [0,1]$ by $f'(x) = 0$ if $x \ge N(1-\epsilon)$ and $f'(x) = f(x)$ otherwise. We define A to be a random subset of $[N]$ where each $x \in [N]$ is in A with probability $f'(x)$, independently of the other elements. In particular, A does not contain x if $x \ge N(1-\epsilon)$. Hence, if the common difference d is larger than $N(1-\epsilon)/2$ then

$$\mathsf{E}_{x \in [N-2d]}[A(x)A(x+d)A(x+2d)] = 0.$$

If d is at most $N(1-\epsilon)/2$, then $N - 2d \ge \epsilon N$, so by Hoeffding's inequality, with probability at least $1 - \exp(-\epsilon^2(N-2d)) \ge 1 - \exp(-\epsilon^3 N)$,

$$\left|\mathsf{E}_{x \in [N-2d]}[A(x)A(x+d)A(x+2d)] - \mathsf{E}_{x \in [N-2d]}[f''(x)f''(x+d)f''(x+2d)]\right| \le \epsilon.$$

Furthermore, note that

$$\begin{aligned}\mathsf{E}_{x \in [N-2d]}[f''(x)f''(x+d)f''(x+2d)] &\le \mathsf{E}_{x \in [N-2d]}[f(x)f(x+d)f(x+2d)] \\ &\le (\alpha + 2\epsilon)^3(1 - 12\epsilon/\alpha^3) \le \alpha^3 - 2\epsilon.\end{aligned}$$

By Hoeffding's inequality, with probability at least $1 - \exp(-\epsilon^2 N)$, the density of A is at least

$$\mathsf{E}_x[f'(x)] - \epsilon \ge \mathsf{E}_x[f(x)] - \epsilon N/N - \epsilon = \alpha.$$

Thus, by the union bound, with probability at least $1 - N\exp(-\epsilon^3 N) - \exp(-\epsilon^2 N)$, A is a set with density at least α such that for all $0 < d < N/2$,

$$\mathsf{E}_{x \in [N-2d]}[A(x)A(x+d)A(x+2d)] \le \alpha^3 - \epsilon.$$

Since $N > \epsilon^{-15}$, for ϵ sufficiently small we have $1 - N\exp(-\epsilon^3 N) - \exp(-\epsilon^2 N) > 0$. This gives Theorem 1.5, assuming Theorem 3.3.

Funding. Fox was supported by a Packard Fellowship and by NSF award DMS-1855635.

Pham was supported by a Two Sigma Fellowship.

Zhao was supported by NSF awards DMS-1764176 and CAREER award DMS-2044606, a Sloan Research Fellowship, and the MIT Solomon Buchsbaum Fund.

References

[1] Baker, R. C., Harman, G., Pintz, J.: The difference between consecutive primes. II. Proc. London Math. Soc. (3) **83**, 532–562 (2001) Zbl 1016.11037 MR 1851081

[2] Behrend, F. A.: On sets of integers which contain no three terms in arithmetical progression. Proc. Nat. Acad. Sci. U.S.A. **32**, 331–332 (1946) Zbl 0060.10302 MR 18694

[3] Bergelson, V., Host, B., Kra, B.: Multiple recurrence and nilsequences. Invent. Math. **160**, 261–303 (2005) Zbl 1087.28007 MR 2138068

[4] Berger, A.: Popular differences for corners in abelian groups. Math. Proc. Cambridge Philos. Soc. **171**, 207–225 (2021) Zbl 07395376 MR 4268809

[5] Bloom, T. F.: A quantitative improvement for Roth's theorem on arithmetic progressions. J. London Math. Soc. (2) **93**, 643–663 (2016) Zbl 1364.11024 MR 3509957

[6] Bloom, T. F., Sisask, O.: Logarithmic bounds for Roth's theorem via almost-periodicity. Discrete Anal. **2019**, art. 4, 20 pp. Zbl 1472.11055 MR 3953879

[7] Bourgain, J.: On triples in arithmetic progression. Geom. Funct. Anal. **9**, 968–984 (1999) Zbl 0959.11004 MR 1726234

[8] Bourgain, J.: Roth's theorem on progressions revisited. J. Anal. Math. **104**, 155–192 (2008) Zbl 1155.11011 MR 2403433

[9] Conlon, D., Fox, J.: Bounds for graph regularity and removal lemmas. Geom. Funct. Anal. **22**, 1191–1256 (2012) Zbl 1256.05114 MR 2989432

[10] Elkin, M.: An improved construction of progression-free sets. Israel J. Math. **184**, 93–128 (2011) Zbl 1280.11008 MR 2823971

[11] Fox, J., Lovász, L. M.: A tight lower bound for Szemerédi's regularity lemma. Combinatorica **37**, 911–951 (2017) Zbl 1413.05316 MR 3737374

[12] Fox, J., Pham, H. T.: Popular progression differences in vector spaces II. Discrete Anal. **2019**, art. 16, 39 pp. Zbl 1473.11058 MR 4042159

[13] Fox, J., Pham, H. T.: Popular progression differences in vector spaces. Int. Math. Res. Notices **2021**, 5261–5289 Zbl 1481.11013 MR 4241128

[14] Fox, J., Sah, A., Sawhney, M., Stoner, D., Zhao, Y.: Triforce and corners. Math. Proc. Cambridge Philos. Soc. **169**, 209–223 (2020) Zbl 07395402 MR 4120790

[15] Fox, J., Sudakov, B.: Dependent random choice. Random Structures Algorithms **38**, 68–99 (2011) Zbl 1215.05159 MR 2768884

[16] Gowers, W. T.: Lower bounds of tower type for Szemerédi's uniformity lemma. Geom. Funct. Anal. **7**, 322–337 (1997) Zbl 0876.05053 MR 1445389

[17] Gowers, W. T.: A new proof of Szemerédi's theorem. Geom. Funct. Anal. **11**, 465–588 (2001) Zbl 1028.11005 MR 1844079

[18] Green, B.: A Szemerédi-type regularity lemma in abelian groups, with applications. Geom. Funct. Anal. **15**, 340–376 (2005) Zbl 1160.11314 MR 2153903

[19] Green, B., Tao, T.: An arithmetic regularity lemma, an associated counting lemma, and applications. In: An Irregular Mind, Bolyai Soc. Math. Stud. 21, János Bolyai Math. Soc., Budapest, 261–334 (2010) Zbl 1222.11015 MR 2815606

[20] Green, B., Wolf, J.: A note on Elkin's improvement of Behrend's construction. In: Additive Number Theory, Springer, New York, 141–144 (2010) Zbl 1261.11013 MR 2744752

[21] Heath-Brown, D. R.: Integer sets containing no arithmetic progressions. J. London Math. Soc. (2) **35**, 385–394 (1987) Zbl 0589.10062 MR 889362

[22] Hosseini, K., Lovett, S., Moshkovitz, G., Shapira, A.: An improved lower bound for arithmetic regularity. Math. Proc. Cambridge Philos. Soc. **161**, 193–197 (2016) Zbl 1371.11026 MR 3530502

[23] Mandache, M.: A variant of the Corners theorem. Math. Proc. Cambridge Philos. Soc. **171**, 607–621 (2021) Zbl 1486.11015 MR 4324960

[24] Moshkovitz, G., Shapira, A.: A short proof of Gowers' lower bound for the regularity lemma. Combinatorica **36**, 187–194 (2016) Zbl 1399.05128 MR 3516883

[25] Roth, K. F.: On certain sets of integers. J. London Math. Soc. **28**, 104–109 (1953) Zbl 0050.04002 MR 51853

[26] Sah, A., Sawhney, M., Zhao, Y.: Patterns without a popular difference. Discrete Anal. **2021**, art. 8, 30 pp. Zbl 07397977 MR 4293329

[27] Sanders, T.: On Roth's theorem on progressions. Ann. of Math. (2) **174**, 619–636 (2011) Zbl 1264.11004 MR 2811612

[28] Sanders, T.: On certain other sets of integers. J. Anal. Math. **116**, 53–82 (2012) Zbl 1280.11009 MR 2892617

[29] Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. Acta Arith. **27**, 199–245 (1975) Zbl 0303.10056 MR 369312

[30] Szemerédi, E.: Regular partitions of graphs. In: Problèmes combinatoires et théorie des graphes (Orsay, 1976), Colloq. Internat. CNRS 260, CNRS, Paris, 399–401 (1978) Zbl 0413.05055 MR 540024

[31] Szemerédi, E.: Integer sets containing no arithmetic progressions. Acta Math. Hungar. **56**, 155–158 (1990) Zbl 0721.11007 MR 1100788

[32] Tao, T., Vu, V.: Additive Combinatorics. Cambridge Stud. Adv. Math. 105, Cambridge Univ. Press, Cambridge (2006) Zbl 1127.11002 MR 2289012

[33] Varnavides, P.: On certain sets of positive density. J. London Math. Soc. **34**, 358–360 (1959) Zbl 0088.25702 MR 106865