Almost periodicity and its applications to Roth's theorem

Huy Tuan Pham [*]

Abstract

We give a self-contained exposition of several aspects of Croot-Sisask almost periodicity, with a special focus on its application to Roth's theorem. Using almost periodicity, we obtain a bound on the size of the largest subset of $\mathbb{Z}_n$ with no nontrivial three-term arithmetic progression of $n(\log\log n)^6/\log n$.

Contents

1   Introduction

The existence of three-term arithmetic progression in dense subsets has motivated many developments in additive combinatorics. Roth [32] first proved using Fourier analysis that any dense subset of a large interval must contain a nontrivial three-term arithmetic progression (one with distinct terms). In particular, Roth showed that the largest subset of $\mathbb{Z}_N$ with no nontrivial three-term arithmetic progression has size at most $Cn/\log\log n$. Meshulam [30] applied Fourier analysis over the finite field vector space setting and proved that the largest subset of $\mathbb{F}_3^n$ without a nontrivial three-term arithmetic progression has size at most $C3^n/n$. Many subsequent improvements to the bound were obtained by better using the spectral information [25, 40, 9, 10, 36, 5], leading to the best known bound over $\mathbb{Z}_n$ of Bloom [5], which is $n(\log\log n)^4/\log n$. Recently, Ellenberg and Gijswijt [16], building on results of Croot, Lev and Pach [12], obtained a breakthrough result over the finite field vector space setting, proving that the largest subset of $\mathbb{F}_p^n$ without a nontrivial three-term arithmetic progression has size at most $(c_p p)^n$ where c_p is a constant strictly smaller than 1. The method they use, however, is algebraic and it is orthogonal to all the developments we discuss above. Furthermore, it is not clear if any such approach can work in the setting of cyclic groups or the integers, since there is a lower bound, constructed by Behrend [3] and improved by Elkin [15], Green and Wolf [23], showing that the there exists a subset of $\mathbb{Z}_n$ without nontrivial three-term arithmetic progression which has size at least $n\exp(-c\sqrt{\log n})$.

    We discuss in this essay Croot-Sisask almost periodicity [13, 6, 38], a recent development in additive combinatorics that can get close to essentially the best known upper bound on the size of the largest subset of $\mathbb{Z}_n$ without a nontrivial three-term arithmetic progression.

Theorem 1. *There exists a constant $C > 0$ such that the following holds. Let G be an abelian group and let l be an integer. Let A, X be subsets of G such that there is a subset S with $|S + A| \le K|A|$, then there exists a subset T of S with $|T| \ge \frac{|S|}{2K^{Cpk^2/\epsilon^2}}$ such that for all $t_1, t_2, \cdots, t_{2k-1}, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_p \le \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{p/2}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

    Almost periodicity roughly says that the convolutions of two sets A and X, where A is additively structured (in the sense that A has small expansion under addition by S), is close to being periodic, in an L^p-sense. While previous approach often takes place in the spectral domain (analyzing the structure of the Fourier coefficients), almost periodicity takes place in the physical domain (i.e., working directly with the group G). To obtain this, the essential idea is to do random sampling and concentration inequalities to show that the convolution $A * X$ is actually well-approximated by $A' * X$ (appropriately scaled) where A' is a random constant-size subset of A. Using almost periodicity, we can smooth out A using the set of almost periods V so that $A * A$ and $A * A * V$ are close. Since $A * A * V$ is much smoother, we can understand $A * A * V$ and pass this information back to A. We also remark that almost periodicity applies directly to general (possible nonabelian) groups, with the same proof that we will give. However, since we focus on abelian groups in the applications we discuss, we will specialize on the abelian case for ease of notations.

    Before discussing the applications of almost periodicity, it is important to understand the limits of almost periodicity. We provide a upper bound construction on the set of almost periods, which shows that the linear dependency on p in Theorem 1 is tight. This construction has not appeared explicitly before in

the literature. We also use a previous spectral construction of Green [18] to show the tight dependency on α and ϵ for constant p. It is unknown whether the joint dependency on p and α, ϵ is tight.

Due to the tight dependency on p, the L^p almost periodicity of the convolution cannot be improved to L^∞ almost periodicity. However, under the extra assumption of bounded VC dimension, this is possible, as noticed by [39, 2]. Since the ideas around this result are fascinating and beautiful, we also include an exposition of the L^∞ almost periodicity result. The proof in [39] is a direct generalization of the proof of the L^p almost periodicity result, coupled with ideas from empirical process theory that allows for control of the supremum of empirical processes. This applies to sets with bounded VC dimension and small expansion in exactly the same way as in the L^p almost periodicity results. We introduce this result due to its direct relationship with the L^p almost periodicity results and the beautiful ideas involved in empirical process theory. The shorter proof in [2] of a very similar result under a slightly different condition and the relationship between the two results are also discussed. Again, as earlier remarked, these results can also be generalized to nonabelian groups, but we restrict our discussions to abelian groups for consistency.

Since the count of three-term arithmetic progression can be written as an inner product of $A * A$ and $2 \cdot A = \{2a, a \in A\}$, it is not surprising that such a result immediately has consequences on Roth's theorem. One of the main goals of this essay is to derive the following bound on Roth's theorem.

Theorem 2. *Let n be coprime to 2. The maximum size of a subset of $\mathbb{Z}_n$ with no nontrivial three-term arithmetic progression is at most $n(\log \log n)^6 / \log n$.*

We follow closely [6] for the exposition of this result, though we give a slightly more efficient argument that improves their constant of 7 in the exponent of the $\log \log n$ factor. It is also remarked in [6] that the bound can be improved. This is the same as the bound by Sanders [35] using almost periodicity, and it is close to the best known bound on Roth's theorem by Bloom [5], which improves the exponent on the $\log \log n$ factor to 4. Sanders' argument uses almost periodicity together with the Katz-Koester transform, which is interesting in its own right. We therefore also give an exposition of the Katz-Koester transform.

Furthermore, we can also obtain a much better bound on a generalized notion of arithmetic progression using almost periodicity. Notice that (x, y, z) forms a three-term arithmetic progression if and only if $x + z = 2y$. We can consider a similar equation in more variables, $x + y + z = 3w$. Behrend's construction generalizes and shows that there exists a set of size at least $n \exp(-c\sqrt{\log n})$ with no solution to the equation $x + y + z = 3w$ where x, y, z, w are not identical (in which case we say that the solution is nontrivial). The next theorem [38] shows that we can achieve essentially a Behrend-type bound using almost periodicity when the number of variables is at least 4.

Theorem 3. *Let n be coprime to 6. The maximum size of a subset of $\mathbb{Z}_n$ with no nontrivial solution to $x + y + z = 3w$ is at most $n \exp(-c(\log n)^{1/5})$.*

For this result, we follow [38]. While it is expectable that the bounds we get for the density of a set avoiding nontrivial solutions to $x + y + z = 3w$ to be better than the bounds in Roth's theorem, since the larger number of convolutions involved implies a smoother structure, it is still quite amazing that almost periodicity can get us all the way to a Behrend-type bound.

We briefly sketch our way to Theorem 2 and 3. The basic frame of the argument follows the density increment approach that underlies Roth's proof. Assuming that a subset A of $\mathbb{Z}_n$ is pseudorandom in a suitable sense, we can then count the number of solutions to the linear equation, which should be close to the expected number of solutions in a random set with the same density. If A deviates from the pseudorandom behavior, we show that there is a "nice subset" of the group where A has increased density. We then localize on this subset and repeat the same procedure. Roth's original proof considers arithmetic progressions as the "nice subsets", in which the argument can be easily iterated due to homogeneity of arithmetic progressions, however, this leads to certain inefficiency in the quantitative bounds. In finite field vector spaces, the situation is much nicer since the "nice subsets" are subspaces, which are plentiful. It was realized by Bourgain [9, 10] that one can consider a natural generalization of subspaces in general

abelian groups, which he calls Bohr sets. We will follow the same approach, and therefore, our general scheme is a density increment approach on Bohr sets in $\mathbb{Z}_n$.

Working with Bohr sets is, in general, much more technical than working with subgroups or subspaces, due to their non-homogeneity. However, there are several small tricks that allow us to treat Bohr sets essentially as subspaces in the arguments. We will cover these in the preliminary section on Bohr sets. Since the arguments for Bohr sets model very closely the arguments for subspaces over finite field vector spaces, we also cover the arguments over finite field vector spaces, where the main ideas are perhaps clearer. It is a common phenomenon [6, 38, 21, 43, 17] that many arguments in additive combinatorics, especially those of iterative nature, can be cleanly developed over finite field vector spaces, and then passed to general abelian groups using standard but technical machineries. We refer the the reader to the surveys [21, 43] for further discussion of this phenomenon, known as the finite field model.

Next, we briefly discuss the crux of the argument using almost periodicity. Almost periodicity is used each step to obtain density increment, replacing the spectral approach which looks at the Fourier transform for non-uniformity. However, for the iterative approach to work, we need to have a structured set, such as a subspace or a Bohr set, where we can control A, while almost periodicity produces an arbitrary set of almost periods. As such, we need to perform a bootstrapping procedure to show that we can enforce structure in the set of almost periods, and pick a large subspace or Bohr set of almost periods. This step depends crucially on the key intuition that we have a very good quantitative bound on the size of the set of almost periods, and taking sums of almost periods only grows the approximation linearly, while iterated sumsets has an exponential smoothing effect on the Fourier coefficient. To show quantitatively that we do not lose too much in the dimension when we pass down to the subspace or Bohr set, we need Chang's lemma [11], generalized to the relativised setting of sets defined locally on Bohr sets by Sanders [36]. Chang's lemma is a very useful result in additive combinatorics, which gives a tight bound on the structure of the large spectrum of a function. We dedicate a section to prove Chang's lemma and recover the machinery of Sanders. Once we have a subspace or Bohr set V of almost periods, we can easily obtain increment. Since $A * A * V$ is much smoother than $A * A$, and it approximates $A * A$, the density of three-term arithmetic progressions in A can be approximated by that of the simpler function $A * A * V$. If $A * A * V$ is very balanced, hence close to α, we can conclude that the density of three-term arithmetic progressions must be close to α^3. On the other hand, if $A * A * V$ deviates from α then we obtain density increment.

We remark that though the bounds on Roth's theorem over finite field vector spaces using the method in this essay is much weaker than what is known, the arguments over the finite field vector spaces act as model versions for those over general abelian groups. While the arguments over general abelian groups are often direct generalizations of those over finite field vector spaces, they are significantly more technical. The readers are encouraged to read the model arguments over finite field vector spaces to construct the general framework before moving to the general case where more technicalities are involved.

Structure of the essay

In Section 2, we cover the preliminaries on Fourier analysis on Bohr sets. In Section 3, we cover the almost periodicity results, following [13]. In Section 4, we describe the bootstrapping procedure to obtain a structured set of almost periods. This relies heavily on the analysis of the large spectrum (Chang's lemma) in abelian groups, and their localized and relativised version over Bohr sets (due to Sanders), which is also covered in full details. We mostly follow [36] in this exposition. In Section 5, we prove Roth's theorem and its generalization to more variables over the finite field vector space setting. In Section 6, we prove Roth's theorem and its generalization to more variables over general abelian groups, in particular, over cyclic groups and the integers. The materials for these sections are drawn from [6, 38]. In Section 7, we prove the upper bounds on the size of the set of L^p almost periods, showing the limits of the quantitative bounds in almost periodicity results. In Section 8, we introduce the recent L^∞ almost periodicity results in [39, 2], which apply to sets with bounded VC dimension. In Section 9, we state several other major

applications of almost periodicity as well as some future directions.

Acknowledgements

Notations

We write $\mathbb{E}$ for the averaging operator. If no additional information is specified, $\mathbb{E}$ averages over the uniform distribution over the appropriate domain. We use $\mathbf{E}$ for the expectation, when there is an explicit probability space involved. Correspondingly, $\mathbb{P}$ denotes the probability over the uniform distribution and $\mathbf{P}$ for the probability over the explicit probability space involved.

When we write $\mathbb{E}_x$ or $\sum_x$, we assume x ranges over a domain that should be clear from context (most often the whole group G).

Given a group G, we write

$$\|f\|_p = (\mathbb{E}_{x \in G} f(x)^p)^{1/p}.$$

In the Fourier space $\hat{G}$, we use the counting measure, so for ω defined on $\hat{G}$,

$$\|\omega\|_p = \left(\sum_{\chi \in \hat{G}} |\omega(\chi)|^p\right)^{1/p}.$$

We quite often work with general measures, for example, when we want to restrict our averaging to a Bohr set instead of the whole group. Since we only focus on discrete groups, a measure μ can be equivalently given by its density m_μ with respect to the uniform measure on G, defined so that

$$\int_G f(x) d\mu(x) = \frac{\sum_{x \in G} f(x) m_\mu(x)}{|G|},$$

where $\sum_x m_\mu(x) = |G|$. We denote

$$\|f\|_{L^p(\mu)} = \left(\int_G |f(x)|^p d\mu(x)\right)^{1/p}.$$

Note that in Fourier space, we retain our counting inner product.

For a subset A of G, we denote by A the indicator function, i.e., $A(x) = 1$ if $x \in A$ and $A(x) = 0$ otherwise.

When there is no confusion, we use Greek letters, most usually α and β, to denote either the density of a subset (A or B) of G, or the uniform measure over the subset (A or B). For example, for a subset B of G, we use β to denote $\frac{|B|}{|G|}$, or the measure with density $m_\beta(x) = \frac{B(x) \cdot |G|}{|B|}$ with respect to the uniform measure over G. We also sometimes write $m_B(x)$ to denote the density of the uniform measure over B with respect to the uniform measure over G where it is convenient, and denote $\|f\|_{L^p(B)} = \|f\|_{L^p(\beta)}$.

Given two subsets of a group G, we define the sumset

$$A + X = \{a + x, a \in A, x \in X\}.$$

We denote the scaling of a subset A of an abelian group G by

$$k \cdot A = \{ka, a \in A\}.$$

We use standard asymptotic notations. For an asymptotic parameter x, $f(x) = O(g(x))$ if there exists C such that $|f(x)| \leq C|g(x)|$. We also denote $f(x) \ll g(x)$ if $f(x) = O(g(x))$.

2 Preliminaries

In this section, we lay out the basic tools of additive combinatorics which we will use. In particular, we will discuss briefly Fourier analysis on abelian groups, which is very useful in counting solutions of linear equations in abelian groups. We then introduce Bohr sets, a crucial tool in additive combinatorics, which can be seen as approximate subgroups. Bohr sets are used to replace the role of subgroups or subspaces, and they are essential in the so-called "density increment" approach, which we will encounter in Section 6.

2.1 Convolutions

Even though the almost periodicity result that we prove applies to general groups, we will focus mostly on finite abelian groups in all applications. Therefore, we use addition as the group operation for consistency of notation. Given a group G, we define the convolution

$$f * g(x) = \mathbb{E}_{y \in G} f(y) g(-y + x) = \frac{1}{|G|} \sum_{y \in G} f(y) g(-y + x).$$

Given subsets A and X of G,

$$A * X(x) = \mathbb{E}_{y \in G} A(y) X(-y + x) = \frac{|A \cap (x - X)|}{|G|},$$

and

$$m_A * X(x) = \frac{|A \cap (x - X)|}{|A|}.$$

It is useful to note the order of magnitude of the convolutions. Note that $\mathbb{E}_x A * X(x) = \frac{|A||X|}{|G|}$, so $\mathbb{E}_x m_A * m_X(x) = 1$. Furthermore, $A * X(x) \leq \frac{\min\{|A|, |X|\}}{|G|}$. In particular,

$$\mathbb{E}_x (A * X(x))^2 \leq \frac{\min\{|A|, |X|\}}{|G|} \cdot \frac{|A||X|}{|G|} \leq \frac{|A|^2 |X|}{|G|^2}.$$

2.2 Fourier analysis on abelian groups

Given an abelian group G, let $\hat{G}$ be the group of characters $\chi : G \to \mathbb{C}$. Given a function $f : G \to \mathbb{C}$, the Fourier transform of f is defined by

$$\hat{f}(\chi) = \mathbb{E}_{x \in G}[f(x)\chi(x)].$$

We record without proof some standard facts about the Fourier transform, which are obvious upon expansion of the following identities

$$\mathbb{E}_x \chi(x) = \begin{cases} 1, & \chi = 0 \\ 0, & \chi \neq 0 \end{cases}$$

$$\sum_\chi \chi(x) = \begin{cases} |G|, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

Fourier Inversion Formula.

$$f(x) = \sum_\chi \hat{f}(\chi) \chi(-x).$$

Parseval's identity.

$$\mathbb{E}[f(x)g(x)] = \sum_\chi \hat{f}(\chi) \hat{g}(\chi).$$

Fourier transform of convolutions.

$$\widehat{f * g}(\chi) = \hat{f}(\chi)\hat{g}(\chi).$$

Translation. We define the translation operator $\tau_t f(x) = f(t + x)$.

$$\widehat{\tau_t f}(\chi) = \mathbb{E}f(t + x)\chi(x) = \chi(-t)\mathbb{E}f(x)\chi(x) = \chi(-t)\hat{f}(\chi).$$

When G is a finite abelian group, the Fourier transform can be explicitly described. Let $G = \prod_{i=1}^m \mathbb{Z}_{n_i}$. Then $\hat{G} \cong \prod_{i=1}^m \mathbb{Z}_{n_i}$, and each element $(a_1, \cdots, a_m) \in \prod_{i=1}^m \mathbb{Z}_{n_i}$ can be identified with a character via $\chi_{\underline{a}}(\underline{x}) = \exp(2\pi i \sum_{i=1}^m a_i x_i / n_i)$. We usually overload the notation to denote χ both as a character and as an element in $\prod_{i=1}^m \mathbb{Z}_{n_i}$. In particular, in a prime cyclic group $\mathbb{Z}_p$, each character can be identified with an element of $\mathbb{Z}_p$ where $\chi(x) = \exp(2\pi i \chi x / p)$. In a finite field vector space $\mathbb{F}_p^n$, each character can be identified with an element of $\mathbb{F}_p^n$ where $\chi(x) = \exp(2\pi i \chi \cdot x / p)$, where $a \cdot x = \sum_{i=1}^n a_i x_i$ is the usual dot product. These will be the most important examples of the groups that we examine in this essay.

The Fourier transform is particularly efficient at counting solutions to linear equations. This is illustrated by the following formula over $\mathbb{Z}_N$

$$\mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\prod_i f_i(x_i)] = \sum_\chi \widehat{f_n}(\chi) \prod_{i=1}^{n-1} \widehat{f_i}(a_n^{-1} a_i \chi),$$

for $a_n \in \mathbb{Z}_N^\times$. Indeed,

$$\mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\prod_i f_i(x_i)] = \mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\prod_{i=1}^n (\sum_{\chi_i} \widehat{f_i}(\chi_i)\chi_i(-x_i))]$$

$$= \mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\sum_\chi \prod_{i=1}^n \widehat{f_i}(\chi_i)\chi_i(-x_i)]$$

$$= \sum_\chi \mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\prod_{i=1}^n \widehat{f_i}(\chi_i)\chi_i(-x_i)]$$

$$= \sum_\chi \mathbb{E}_{\sum_{i=1}^n a_i x_i = 0}[\exp(-2\pi i \sum_{i=1}^n x_i \chi_i / N) \prod_{i=1}^n \widehat{f_i}(\chi_i)]$$

$$= \sum_\chi \mathbb{E}_{x_1, \cdots, x_{n-1}}[\exp(-2\pi i \sum_{i=1}^{n-1} x_i(\chi_i - a_n^{-1} a_i \chi_n)/N) \prod_{i=1}^n \widehat{f_i}(\chi_i)]$$

$$= \sum_\chi \prod_{i=1}^n \widehat{f_i}(\chi_i)\mathbb{E}_{x_i} \exp(-2\pi i x_i(\chi_i - a_n^{-1} a_i \chi_n)/N)$$

$$= \sum_\chi \widehat{f_n}(\chi) \prod_{i=1}^{n-1} \widehat{f_i}(a_n^{-1} a_i \chi).$$

The same holds over $\mathbb{F}_p^n$ where the coefficients $a_1, a_2, \cdots, a_n \in \mathbb{F}_p^\times$. This formula underlines proofs of Roth's theorem using Fourier analysis.

2.3 Bohr sets

One advantage in working over vector spaces is the existence of many subspaces that allows for an inductive approach to prove Roth's theorem. Given a general abelian group G, this is no longer true (for example, there is no nontrivial subgroup in $\mathbb{Z}$ or $\mathbb{Z}_p$), and so we need to define proper notion of well-behaving subsets

of G where the iterative approach can be carried out. The subsets should behave like a subgroup, and they should have a homogeneous structure that allows for iteration. This is the motivation for Bohr sets. Notice that a subspace of a vector space can be equivalently defined as the vanishing set of a collection of characters. This definition naturally generalizes to general abelian groups.

Definition 4. For a real number x, let $\|x\| = \inf\{|x - n|, n \in \mathbb{Z}\}$. A Bohr set $B(\Gamma, \bar{\rho})$ with frequency $\Gamma = (\gamma_1, \cdots, \gamma_d) \in \hat{G}^d$ and radius $\underline{\rho} = (\rho_1, \cdots, \rho_d) \in [0, 1]^d$ is defined by

$$B(\Gamma, \underline{\rho}) = \{x \in G : \|\arg(\gamma_i(x))/2\pi\| \leq \rho_i \forall 1 \leq i \leq d\}.$$

We call d the dimension of $B(\Gamma, \underline{\rho})$. We write $B(\Gamma, \rho)$ to denote the Bohr set where all radii are the same and equal to ρ. Furthermore, we write $B(\Gamma, \underline{\rho})_\nu = B(\Gamma, \nu\underline{\rho})$ to be the Bohr set with the scaled radii.

We remark that in most applications, we only need to consider Bohr sets with all radii being the same. An immediate consequence of the definition is that for all $x \in B(\Gamma, \underline{\rho})$,

$$|1 - \gamma_i(x)| \leq 2\pi\rho_i.$$

In fact, one can also instead define the Bohr set based on how close $\gamma_i(x)$ is to 1, which is equivalent up to constant factors. There are several nice properties of Bohr sets which we will need throughout. They are usually implicitly generalizations of corresponding properties of subgroups or subspaces. Some of the material in this section is drawn from [17, 41].

First, we give several estimates on the size of Bohr sets and its doubling constant.

Proposition 5. *We have*

$$|B(\Gamma, \underline{\rho})| \geq |G| \prod_{i=1}^{d} \rho_i.$$

$$|B(\Gamma, 2\underline{\rho})| \leq 4^d |B(\Gamma, \underline{\rho})|.$$

Proof. Define $P : G \to \mathbb{R}^d/\mathbb{Z}^d$ by $P(g) = (\arg(\gamma_i g)/2\pi)_{i=1}^{d}$. To prove the first inequality, identiy $\mathbb{R}^d/\mathbb{Z}^d$ with $[0, 1]^d$ and note that we can partition $[0, 1]^d$ to $\prod_{i=1}^{d} \frac{1}{\rho_i}$ subboxes of diameter $\underline{\rho}$. Then there exists a subbox Q with $|P^{-1}(Q)| \geq |G| \prod_{i=1}^{d} \rho_i$. Fix any $x \in Q$, then for any $y \in Q$, $|\arg(\gamma_i x) - \arg(\gamma_i y)|/2\pi \leq \rho_i$, so $y - x \in B(\Gamma, \underline{\rho})$. Thus, $|B(\Gamma, \underline{\rho})| \geq |G| \prod_{i=1}^{d} \rho_i$.

For the second inequality, partition $\prod_{i=1}^{d}[-2\rho_i, 2\rho_i] \subseteq \mathbb{R}^d/\mathbb{Z}^d$ into 4^d subboxes with diameter $\underline{\rho}$. In each subbox, choose a point in $B(\Gamma, 2\underline{\rho})$ if it exists. Let X be the set of chosen points. Then, $X + B(\Gamma, \underline{\rho})$ covers $B(\Gamma, 2\underline{\rho})$ since for each point y in a subbox containing x, $|\arg(\gamma_i x) - \arg(\gamma_i y)|/2\pi \leq \rho_i$. Thus,

$$|B(\Gamma, 2\underline{\rho})| \leq |X||B(\Gamma, \underline{\rho})| \leq 4^d |B(\Gamma, \underline{\rho})|.$$

$\square$

Note that from the definition of Bohr sets and the triangle inequality, we have $B(\Gamma, \underline{\rho}) + B(\Gamma, \underline{\rho}') \subseteq B(\Gamma, \underline{\rho} + \underline{\rho}')$. The second estimate thus suggests that Bohr sets are fairly stable under addition. However, quantitatively, Bohr sets behave more similarly to a multidimensional arithmetic progression or a box under addition than a subspace. Indeed, it is straightforward from the definition that a Bohr set is the inverse image of a box under the group homomorphism given by the characters as we have used in the proof of the above proposition. Though Bohr sets are quite nicely behaved under addition, they are much worse than a subgroup, especially when we deal with quantitative problems where the exponential factor in the dimension becomes too large. However, observe that the addition of a very small box to a large box is much more stable (in the sense that the size only grows linearly rather than exponentially), and essentially by localizing this way, Bohr sets can be treated more like subgroups. The crucial property we need is referred to as regularity.

Definition 6. A Bohr set $B(\Gamma, \underline{\rho})$ of dimension d is regular if

$$\left| B\left(\Gamma, \left(1 + \frac{\delta}{80d}\right)\underline{\rho}\right) \setminus B\left(\Gamma, \left(1 - \frac{\delta}{80d}\right)\underline{\rho}\right) \right| \le 2\delta |B(\Gamma, \underline{\rho})|,$$

for all $0 \le \delta \le 1$.

The next proposition shows that we do not lose much in asserting that our Bohr sets are regular.

Proposition 7. *For an arbitrary Bohr set $B(\Gamma, \underline{\rho})$ of dimension d, there exists $r \in [1/2, 1]$ such that $B(\Gamma, r\underline{\rho})$ is regular.*

Proof. Let $f : [1/2, 1] \to \mathbb{R}$ be defined by $f(r) = \log_2 |B(\Gamma, r\underline{\rho})|$. Then f is increasing, and $f(1) \le 2d + f(1/2)$. We aim to prove that there exists $r_* \in [\frac{1}{2} + \frac{1}{80d}, 1 - \frac{1}{80d}]$ such that

$$f(r_2) - f(r_1) \le 40d(r_2 - r_1),$$

for all $r_1 = r_* - \frac{\delta}{80d}, r_2 = r_* + \frac{\delta}{80d}$ with $0 \le \delta \le 1$. Then, for all $0 \le \delta \le 1$,

$$\left| B\left(\Gamma, \left(r_* + \frac{\delta}{80d}\right)\underline{\rho}\right) \setminus B\left(\Gamma, \left(r_* - \frac{\delta}{80d}\right)\underline{\rho}\right) \right| = 2^{f(r_* + \delta/80d)} - 2^{f(r_* - \delta/80d)}$$

$$\le \left| B(\Gamma, r_*\underline{\rho}) \right| \cdot (2^\delta - 1)$$

$$< 2\delta |B(\Gamma, r_*\underline{\rho})|,$$

where we used that $2^x \le 1 + 2x$ for all $0 \le x \le 1$. The existence of r_* is guaranteed essentially by Hardy-Littlewood maximal inequality. Assume that there does not exist such r_*, then we have a collection of intervals $[r_1, r_2]$ covering $[\frac{1}{2} + \frac{1}{80d}, 1 - \frac{1}{80d}]$ with the property that $f(r_2) - f(r_1) > 40d(r_2 - r_1)$. By Vitali's covering lemma, we can find a finite subcollection of disjoint intervals whose total measure is at least $1/20$. Let $r_1 < r_2 < \cdots < r_{2l-1} < r_{2l}$ be an ordering of the disjoint intervals $[r_{2k-1}, r_{2k}]$, $1 \le k \le l$. Then

$$\frac{1}{20} \le \sum_{k=1}^{l} |r_{2k} - r_{2k-1}| < \frac{1}{40d} \sum_{k=1}^{l} (f(r_{2k}) - f(r_{2k-1}))$$

$$< \frac{1}{40d}\left[(f(r_{2l}) - f(r_1)) - \sum_{k=1}^{l-1} (f(2k+1) - f(2k)) \right] < \frac{f(1) - f(1/2)}{40d} \le \frac{1}{20},$$

contradiction. $\qquad\square$

We note that in the most applications, we actually only need the fact that

$$\left| B\left(\Gamma, \left(1 + \frac{\delta}{80d}\right)\underline{\rho}\right) \setminus B\left(\Gamma, \underline{\rho}\right) \right| \le 2\delta |B(\Gamma, \underline{\rho})|$$

for all special values of δ which are pre-chosen in the argument. This property can be much more easily guaranteed, simply by observing that $|B(\Gamma, r\underline{\rho})|$ is increasing, with $|B(\Gamma, \underline{\rho})|/|B(\Gamma, \underline{\rho}/2)|$ bounded, so the size of the Bohr set cannot grow too fast in all of the subintervals of length $\frac{\delta}{80d}$, and so the growth of the size of the Bohr set over one such interval must be small. We remark also that essentially localizing at the scale of $\frac{1}{d}$ is necessary, since at such scale, the exponential growth rate of the Bohr set size is upper bounded by a linear function, $(1 + \frac{c}{d})^d \le 1 + \gamma c$ for c small. This essentially explains why we need to work with Bohr sets localized at this scale in the forthcoming arguments, since instead of large error terms, we will only have small additive error terms (scaled by $|B(\Gamma, \underline{\rho})|$). Indeed, in application, we shall use the fact that $|B + B_{\delta/d}| \le |B_{1+\delta/d}| \le (1 + \delta)|B|$, so up to an error of $\delta|B|$, $B + B_{\delta/d}$ is approximately the same as B, yielding a structure closer to the invariance of subgroups under addition. The small error term allows

9

us to say, for example, that a set A of relative density α in B has relative density $\alpha(1 \pm 2\delta)$ in $B + B_{\delta/d}$. By working with several Bohr sets at the correct relative scalings, and averaging over appropriate Bohr sets, we can usually recover desirable properties and averaging effects that we have when working with groups and subgroups. The loss in the scalings, which is polynomial in the dimension of the Bohr sets, is usually negligible.

It will be useful to us later that a small change to the radius of a regular Bohr set does not significantly change regularity. We remark that the following definition is somewhat nonstandard.

Definition 8. For $\kappa \leq \frac{1}{81d}$, a Bohr set $B(\Gamma, \underline{\rho})$ of dimension d is κ-regular if

$$\left| B\left(\Gamma, \left(1 + \frac{\delta}{80d}\right)\underline{\rho}\right) \setminus B(\Gamma, \left(1 - \frac{\delta}{80d}\right)\underline{\rho}) \right| \leq 2(\delta + 81d\kappa)|B(\Gamma, \underline{\rho})|,$$

for all $0 \leq \delta \leq 1 - 81d\kappa$.

Thus, if $B(\Gamma, \underline{\rho})$ is regular then it is 0-regular. It is straightforward to verify the following proposition.

Proposition 9. *If B is a κ-regular Bohr set, then for $\eta < \frac{1}{81d} - \kappa$, $B_{1+\eta}$ is $(\kappa + \eta)$-regular.*

We remark that we only use κ-regular Bohr sets in Subsection 6.2.2 where we give a better quantitative bound for Roth's theorem compared to using only Bohr sets.

One useful property of subgroups is that we can use translates of a subgroup to tile the whole group, so that each element of the group is covered using the same number of translates. This is also essentially true for regular Bohr sets, but we need to work with Bohr sets at the appropriate scale mentioned before.

Proposition 10. *Let B be a κ-regular Bohr set of dimension d. Let $A \subseteq B$ have relative density α. Let $B' = B_\nu$ where $\nu \leq c/d$ for a constant $c < 10^{-4}$. Then*

$$\mathbb{E}_{x \in B+B'} A * m_{B'}(x) = \alpha \frac{|B|}{|B + B'|} \geq \frac{\alpha}{1 + 162d(\nu + \kappa)}.$$

In particular, for a Bohr set $B_{1-\tau} \subseteq B'' \subseteq B_{1+\tau}$ and $t \in B_\mu$ with $\tau, \mu \leq c/d$ for $c < 10^{-4}$,

$$\mathbb{E}_{x \in B''+t} A * m_{B'}(x) \geq \frac{\alpha}{1 + 162d(\tau + \kappa)} - 400d(\nu + \mu + \tau + \kappa).$$

Furthermore,
$$\mathbb{E}_{x \in B''+t} A * m_{B'}(x) \leq \alpha(1 - 162d(\tau + \kappa)).$$

Proof. We have

$$\mathbb{E}_{x \in B+B'} A * m_{B'}(x) = \frac{1}{|B + B'||B'|} \sum_{x,y}(B + B')(x)A(y)B'(x - y)$$

$$= \frac{1}{|B + B'||B'|} \sum_{y,z} A(y)B'(z)(B + B')(y + z)$$

$$= \frac{1}{|B + B'||B'|} \sum_{y,z} A(y)B'(z)$$

$$= \alpha \frac{|B|}{|B + B'|}.$$

Thus,

$$\mathbb{E}_{x\in B''+t}A*m_{B'}(x) = \frac{1}{|B''|}\sum_{x\in B+B'}(B''+t)(x)A*m_{B'}(x)$$

$$\geq \frac{1}{|B''|}\left[\sum_{x\in B+B'}A*m_{B'}(x) - |(B+B')\backslash(B''+t)|\right]$$

$$\geq \alpha\frac{|B|}{|B''|} - \frac{|(B+B')\cup(B''+t)| - |B''+t|}{|B''|}$$

$$\geq \alpha\frac{|B|}{|B_{1+\tau}|} - \frac{|B_{1+\nu+\mu+\tau}| - |B_{1+\tau}|}{|B_{1-\tau}|}$$

$$\geq \frac{\alpha}{1+162d(\tau+\kappa)} - 400d(\nu+\mu+\tau+\kappa).$$

Furthermore,

$$\mathbb{E}_{x\in B''+t}A*m_{B'}(x) = \frac{1}{|B''|}\sum_{x\in B+B'}(B''+t)(x)A*m_{B'}(x)$$

$$\leq \frac{\alpha|B|}{|B''|}$$

$$\leq \alpha(1 - 162d(\tau+\kappa)).$$

$\square$

Finally, we discuss the effect of scaling on Bohr sets.

Proposition 11. *Let k be an integer and assume that $k \nmid |G|$, so the inverse of the automorphism of G mapping $x \mapsto kx$ is well-defined. Let $B = B(\Gamma, \underline{\rho})$ be a Bohr set, and let*

$$k \cdot B = \{kb, b \in B\}.$$

Then $k \cdot B$ is a Bohr set with characters $\Gamma/k = \{\gamma/k, \gamma \in \Gamma\}$, and radius $\underline{\rho}$. Moreover, if B is κ-regular then $k \cdot B$ is κ-regular. Furthermore,

$$(k \cdot B)_\tau \subseteq B_{k\tau}.$$

Proof. Note that

$$(\gamma/k)(kx) = \gamma(x).$$

Therefore,

$$x \in B(\Gamma, \underline{\rho}) \Leftrightarrow \|\arg(\gamma_i(x))/2\pi\| \leq \rho_i \Leftrightarrow \|\arg((\gamma_i/k)(kx))/2\pi\| \leq \rho_i \Leftrightarrow kx \in B(\Gamma/k, \underline{\rho}).$$

Thus,

$$k \cdot B(\Gamma, \underline{\rho}) = B(\Gamma/k, \underline{\rho}).$$

Moreover,

$$|k \cdot B(\Gamma, \underline{\rho})| = |B(\Gamma, \underline{\rho})|,$$

so $B(\Gamma/k, \underline{\rho})$ is κ-regular if $B(\Gamma, \underline{\rho})$ is κ-regular.

Finally, if

$$\|\arg((\gamma_i/k)(kx))/2\pi\| \leq \tau\rho_i$$

then

$$\|\arg(\gamma_i(kx))/2\pi\| \leq k\|\arg(\gamma_i(x))/2\pi\| \leq k\tau\rho_i,$$

so

$$(k \cdot B)_\tau \subseteq B_{k\tau}.$$

$\square$

3 Almost Periodicity

In this section, we discuss L^p almost periodicity results for finite p. The results apply to general groups with no extra cost. However, since we focus on abelian groups in the subsequent applications, and for consistency of notations, we will only write the proof in the abelian case. We will roughly follow [13, 6], however, instead of combinatorial moment estimates for the binomial distribution, we will use exponential concentration, an idea which also appears in [4].

The crux of the almost periodicity result is that the convolution of a function with a well-structured set, particularly a set with small expansion under set addition, is almost periodic in a large set of directions. Note that in particular, this applies to dense subsets of G, which satisfies $|A + G| \leq \frac{|G|}{|A|}|A|$.

Theorem 12 (L^p almost periodicity). *There exists a constant $C > 0$ such that the following holds. Let G be a group and let l be an integer. Let A, X be subsets of G such that there is a subset S with $|S+A| \leq K|A|$, then there exists a subset T of S with $|T| \geq \frac{|S|}{2K^{Cpk^2/\epsilon^2}}$ such that for all $t_1, t_2, \cdots, t_{2k-1}, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_p \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{p/2}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

For some examples of sets satisfying the small expansion constraint $|S + A| \leq K|A|$, one can take X to be a set with small doubling, such as a generalized arithmetic progression in $\mathbb{Z}_n$, and take A to be an arbitrary dense subset of X and take S to be X. One can think about almost periodicity as a way to convert a density/packing condition to an additive condition of L^p almost periodicity on two-fold convolutions. Indeed, the proof technique we discuss next will be reminiscent of this intuition.

To illustrate the proof technique, we first cover the case $p = 2$, where things can be done fairly cleanly. The main idea of the proof is to show that for most $C \subseteq A$ of constant size, $m_C * X(x)$ is close to $m_A * X(x)$, so in fact, most C would give us a sufficiently good approximation of $m_A * X(x)$. To establish this, we simply take a random C, and use Chebyshev's inequality, computing the variance of $m_C * X(x)$. Next, we use the fact that A is structured (having small expansion under addition by S), to say that many of the good C are translates of a single set. But if $m_{s+C} * X$ and $m_C * X$ are both good approximations for $m_A * X$, then since $m_{s+C} * X$ is a translation of $m_C * X$, $m_A * X$ must be approximately invariant under this translation. This establish one almost period. To get the sumset structure of the almost periods, we simply use triangle inequality, which tells us that the sum of many good almost periods is only slightly worse as an almost period.

Theorem 13 (L^2 almost periodicity). *Let G be a group and let k be an integer. Let A, X be subsets of G such that there is a subset S with $|S + A| \leq K|A|$. Then there exists a subset T of S with $|T| \geq \frac{|S|}{2K^{8k^2/\epsilon^2}}$ such that for all $t_1, t_2, \cdots, t_{2k-1}, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_2 \leq \frac{\epsilon|A|\sqrt{|X|}}{|G|}.$$

Proof. Recall that

$$m_A * X(x) = \frac{|G|}{|A|} \frac{\sum_{y \in G} A(y)X(-y + x)}{|G|}.$$

Let $n = \lceil 8k^2/\epsilon^2 \rceil$. Let $\bar{a} = (a_1, a_2, \cdots, a_n)$ be a tuple of elements of A of size n chosen independently and uniformly at random. Consider $m_{\bar{a}}(x) = |G| \frac{\sum_{i \leq n} I(a_i = x)}{n}$. Note that $\mathbf{E}m_{\bar{a}}(x) = |G| \frac{n \cdot \frac{1}{|A|} A(x)}{n} = |G| \frac{A(x)}{|A|} =$

$m_A(x)$. Thus

$$m_{\bar{a}} * X(x) = \frac{\sum_{y \in G} m_{\bar{a}}(y) X(-y + x)}{|G|}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{y} I(a_i = y) X(-y + x)$$

$$= \frac{1}{n} \sum_{i=1}^{n} X(-a_i + x).$$

Note that $X(-a_i + x)$ is 1 if and only if $A(a_i) X(-a_i + x) = 1$, so the probability that $X(-a_i + x) = 1$ is exactly $m_A * X(x)$. Moreover, $\{X(-a_i + x), 1 \le i \le n\}$ are independent random variables. Thus, we immediately obtain

$$\mathbf{E}[m_{\bar{a}} * X(x)] = m_A * X(x),$$

and

$$\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^2]$$

$$= \mathbf{E}\left[\left(\frac{1}{n} \sum_{i=1}^{n} (X(-a_i + x) - m_A * X(x))\right)^2\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \mathbf{E}\left[(X(-a_i + x) - m_A * X(x))^2\right]$$

$$= \frac{1}{n}\left[(1 - m_A * X(x))^2 \cdot m_A * X(x) + (0 - m_A * X(x))^2 \cdot (1 - m_A * X(x))\right]$$

$$= \frac{1}{n}\left[m_A * X(x) - (m_A * X(x))^2\right].$$

Hence,

$$\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^2] \le \frac{m_A * X(x)}{n}.$$

Thus,

$$\mathbf{E}[\sum_{x} (m_{\bar{a}} * X(x) - m_A * X(x))^2] \le \frac{1}{n} \sum_{x} m_A * X(x) = \frac{1}{n} \frac{|G|}{|A|} \frac{|A||X|}{|G|} = \frac{|X|}{n}.$$

By Markov inequality, the probability that $\sum_{x}(m_{\bar{a}} * X(x) - m_A * X(x))^2 \ge \frac{2|X|}{n}$ is at most $\frac{1}{2}$. Thus, there are at least $\frac{|A|^n}{2}$ tuples $\bar{a}$ such that $\sum_{x}(m_{\bar{a}} * X(x) - m_A * X(x))^2 < \frac{2|X|}{n}$. Let G be the set of such $\bar{a}$. For each $\bar{a} \in G$, let $S(\bar{a}) = \{(s + a_1, s + a_2, \cdots, s + a_n), s \in S\}$. The tuples $(s + a_1, s + a_2, \cdots, s + a_n)$ are distinct for distinct s, and each of them consist of elements in $|S + A| \le K|A|$. Thus, there is a tuple in $\binom{S+A}{n}$ which occurs in at least $\frac{|S||A|^n/2}{|S+A|^n} \ge \frac{|S|}{2K^n}$ sets $S(\bar{a})$. Let $\bar{s} = (s_1, s_2, \cdots, s_n)$ be such a tuple. Then there are at least $\frac{|S|}{2K^n}$ distinct tuples $\bar{a} \in G$ such that $\bar{s} = \bar{a} + t$ for some $t \in S$. Then for a subset T of S of size at least $\frac{|S|}{2K^n}$, for all $t \in T$, we have $\sum_{x}(m_{\bar{s}-t} * X(x) - m_A * X(x))^2 \le \frac{2|X|}{n}$.

Moreover, for $t_1, t_2 \in T$,

$$m_{-t_2+\bar{s}} * X(t_1 - t_2 + x) = \frac{1}{n} \sum_{i=1}^{n} X(-s_i + t_1 + x) = m_{-t_1+\bar{s}} * X(x),$$

so by the triangle inequality,

$$\|m_A * X(t_1 - t_2 + \cdot) - m_A * X(\cdot)\|_2$$
$$\leq \|m_A * X(t_1 - t_2 + \cdot) - m_{-t_2+\bar{s}} * X(t_1 - t_2 + \cdot)\|_2 + \|m_{-t_1+\bar{s}} * X(\cdot) - m_A * X(\cdot)\|_2$$
$$= \|m_A * X(\cdot) - m_{-t_2+\bar{s}} * X(\cdot)\|_2 + \|m_{-t_1+\bar{s}} * X(\cdot) - m_A * X(\cdot)\|_2$$
$$\leq 2\sqrt{\frac{2|X|}{n}}.$$

By induction, we now have

$$\|m_A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_2$$
$$\leq \|m_A * X(t_1 - t_2 + \cdots + t_{2k-3} - t_{2k-2} + \cdot) - m_A * X(\cdot)\|_2 + \|m_A * X(t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_2$$
$$\leq 2k\sqrt{\frac{2|X|}{n}} \leq \epsilon\sqrt{|X|},$$

by our choice of n. $\qquad\square$

We now give the proof of the almost periodicity result for general p. As in the proof in the case $p = 2$, we only need to establish a concentration result for $\|m_{\bar{a}} * X - m_A * X\|_p$. This can be obtained quite accurately since we are simply computing moments of binomial random variables, and these exhibit strong concentration from Hoeffding's bound. We give a tight estimate for the exponential concentration of such random variables. Then we proceed through a standard but somewhat tedious path to get bounds for moments from this. This is essentially how one can prove Marcinkiewicz inequality, and it can be applied to a much wider range of distributions. Thus, we think that there is certain benefit in getting estimates this way rather than more combinatorial methods of obtaining moments for binomial random variables. The bounds we get are comparable to the best available from direct moment bounds for binomial random variables. In certain situations, an exponential tail form of almost periodicity instead of L^p almost periodicity, which is straightforward from our proof, can also be useful. We refer the reader to [4] for further details.

Proof of Theorem 12. As in the $p = 2$ case, let $\bar{a} = (a_1, a_2, \cdots, a_n)$ be a tuple of elements of A chosen independently and uniformly at random, and let $m_{\bar{a}}(x) = |G|^{\frac{\sum_{i \leq n} 1(a_i = x)}{n}}$. Then

$$m_{\bar{a}} * X(x) = \frac{1}{n} \sum_{i=1}^{n} X(-a_i + x),$$

where $\{X(-a_i + x), 1 \leq i \leq n\}$ are independent Bernoulli random variables which receive the value 1 with probability $m_A * X(x)$. Thus $\mu_{\bar{a}} * X(x)$ is a Binomial random variable, and $\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^p]$ is the centered pth moment of such a random variable. Instead of computing this explicitly, we upper bound the pth moment by the tail bound using exponential concentration.

Assume that $m_A * X(x) > 0$, in the case $m_A * X(x) = 0$, we trivially have $\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^p] = 0$.

Notice that

$$\mathbf{E}[\exp(\lambda m_{\bar{a}} * X(x) - \lambda m_A * X(x))] = \mathbf{E}\left[\prod_{i=1}^{n} \exp\left(\frac{\lambda}{n} X(-a_i + x) - \frac{\lambda}{n} m_A * X(x)\right)\right]$$
$$= \prod_{i=1}^{n} \mathbf{E}\left[\exp\left(\frac{\lambda}{n} X(-a_i + x) - \frac{\lambda}{n} m_A * X(x)\right)\right],$$

14

by independence. We can compute

$$\mathbf{E}\left[\exp\left(\frac{\lambda}{n}X(-a_i + x) - \frac{\lambda}{n}m_A * X(x)\right)\right]$$

$$= m_A * X(x)\exp\left(\frac{\lambda}{n} - \frac{\lambda}{n}m_A * X(x)\right) + (1 - m_A * X(x))\exp\left(-\frac{\lambda}{n}m_A * X(x)\right)$$

$$= \exp(-\lambda m_A * X(x)/n)\left(1 + m_A * X(x)(\exp(\lambda/n) - 1)\right)$$

$$\leq \exp(-\lambda m_A * X(x)/n)\exp\left(m_A * X(x)(\exp(\lambda/n) - 1)\right).$$

Thus, we get

$$\exp(-\lambda t)\mathbf{E}[\exp(\lambda m_{\bar{a}} * X(x) - \lambda m_A * X(x))]$$

$$\leq \exp\left(-\lambda t - \lambda m_A * X(x) + nm_A * X(x)\left(\exp\left(\frac{\lambda}{n}\right) - 1\right)\right)$$

Taking the derivative with respect to λ, the minimum value of $f_t(\lambda) = -\lambda(t + m_A * X(x)) + nm_A * X(x)\left(\exp\left(\frac{\lambda}{n}\right) - 1\right)$ is obtained at λ_t such that $\exp(\lambda_t/n)m_A * X(x) = t + m_A * X(x)$, so

$$\exp(-\lambda_t t)\mathbf{E}[\exp(\lambda_t m_{\bar{a}} * X(x) - \lambda_t m_A * X(x))]$$

$$\leq \exp(-\lambda_t(t + m_A * X(x)) + nt)$$

$$= \exp(-n(t + m_A * X(x))\log(1 + t/m_A * X(x)) + nt).$$

Noting that $\lambda_t > 0$ for $t > 0$, by Markov's inequality, we have

$$\mathbf{P}(m_{\bar{a}} * X(x) - m_A * X(x) \geq t) \leq \exp(-\lambda_t t)\mathbf{E}[\exp(\lambda_t m_{\bar{a}} * X(x) - \lambda_t m_A * X(x))].$$

For the lower tail, changing the sign, we get

$$\exp(\lambda t)\mathbf{E}[\exp(\lambda m_{\bar{a}} * X(x) - \lambda m_A * X(x))]$$

$$\leq \exp\left(\lambda t - \lambda m_A * X(x) + nm_A * X(x)\left(\exp\left(\frac{\lambda}{n}\right) - 1\right)\right)$$

If $0 \leq t < m_A * X(x)$, taking the derivative with respect to λ, the minimum value of $f_t(\lambda) = \lambda(t - m_A * X(x)) + nm_A * X(x)\left(\exp\left(\frac{\lambda}{n}\right) - 1\right)$ is obtained at λ_t such that $\exp(\lambda_t/n)m_A * X(x) = m_A * X(x) - t$. Thus

$$\exp(\lambda_t t)\mathbf{E}[\exp(\lambda_t m_{\bar{a}} * X(x) - \lambda_t m_A * X(x))]$$

$$\leq \exp(\lambda_t(t - m_A * X(x)) + nt)$$

$$= \exp(-n(m_A * X(x) - t)\log(1 - t/m_A * X(x)) + nt).$$

Noting that $\lambda_t < 0$ for $t > 0$, by Markov's inequality, we have

$$\mathbf{P}(m_{\bar{a}} * X(x) - m_A * X(x) \leq -t) \leq \exp(-\lambda_t t)\mathbf{E}[\exp(\lambda_t m_{\bar{a}} * X(x) - \lambda_t m_A * X(x))].$$

If $t > m_A * X(x)$ then $\mathbf{P}(m_{\bar{a}} * X(x) - m_A * X(x) \leq -t) = 0$ since $m_{\bar{a}} * X(x) \geq 0$.

Thus, letting $u = t/m_A * X(x)$, we get

$$\mathbf{P}(|m_{\bar{a}} * X(x) - m_A * X(x)| \geq t)$$

$$\leq \exp(-nm_A * X(x)[(u + 1)\log(u + 1) - u]) + I(u \leq 1)\exp(-nm_A * X(x)[(1 - u)\log(1 - u) + u])$$

Note that $(u + 1)\log(u + 1) - u$ is nonnegative and increasing for $u \geq 0$,

$$(u + 1)\log(u + 1) - u \geq \frac{u^2}{4}$$

when $0 \leq u \leq 4$, and
$$(u+1)\log(u+1) - u \geq u$$
for $u \geq 4$. Furthermore, $(1-u)\log(1-u) + u$ is nonnegative and increasing for $u \in [0,1]$, and
$$(1-u)\log(1-u) + u \geq \frac{u^2}{4}$$
for $0 \leq u \leq 1$. Thus,

$$\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^p]$$
$$= p\int_0^\infty t^{p-1}\mathbf{P}(|m_{\bar{a}} * X(x) - m_A * X(x)| \geq t)dt$$
$$\leq p(m_A * X(x))^p \int_0^\infty u^{p-1}\exp(-nm_A * X(x)[(u+1)\log(u+1) - u])du$$
$$\quad + p(m_A * X(x))^p \int_0^1 u^{p-1}\exp(-nm_A * X(x)[(1-u)\log(1-u) + u])du$$
$$\leq 2p(m_A * X(x))^p \int_0^4 u^{p-1}\exp(-nm_A * X(x)u^2/4)du + p(m_A * X(x))^p \int_4^\infty u^{p-1}\exp(-nm_A * X(x)u)du.$$

We first observe

$$p\int_4^\infty u^{p-1}\exp(-nm_A * X(x)u)du$$
$$\leq p\int_0^\infty u^{p-1}\exp(-nm_A * X(x)u)du$$
$$= p(p-1)!(nm_A * X(x))^{-p}$$
$$= p!n^{-p}(m_A * X(x))^{-p}$$
$$\leq \left(\frac{p}{en}\right)^p (m_A * X(x))^{-p}.$$

In the range $u \leq 4$,

$$\int_0^4 u^{p-1}\exp(-nm_A * X(x)u^2/4)du$$
$$\leq p\int_0^\infty u^{p-1}\exp(-nm_A * X(x)u^2/4)du$$
$$< \frac{p}{2}(nm_A * X(x)/2)^{-p/2} \cdot (p-1)!!$$
$$< \frac{1}{2}\left(\frac{2p}{en}\right)^{p/2}(m_A * X(x))^{-p/2}.$$

where we used the moment estimate for a Gaussian random variable in the second inequality. Thus,

$$\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^p]$$
$$< \left(\frac{p}{en}\right)^p + \left(\frac{2p}{en}\right)^{p/2}(m_A * X(x))^{p/2}.$$

Hence,

$$\mathbf{E}[\|m_{\bar{a}} * X(\cdot) - m_A * X(\cdot)\|_p^p]$$
$$< \left(\frac{p}{en}\right)^p + \left(\frac{2p}{en}\right)^{p/2}(m_A * X(x))^{p/2}.$$

The rest of the proof follows exactly as in the $p = 2$ case. $\qquad\square$

When working with general abelian groups, in particular, cyclic groups, we have to deal more generally with Bohr sets instead of subgroups. In this case, we usually localize, and therefore need to develop a version of almost periodicity that localizes over a measure. The proof of this result is essentially identical to the above proof, though there is extra complication arising from shifting the local measure, particularly when we use the triangle inequality. Due to the translations involved in applying the triangle inequality to the local measure, we need the following definition.

Definition 14. For a measure τ, let $\tau + x$ be the measure such that $(\tau + x)(X) = \tau(X - x)$. A measure ν is an S-upper envelope for a measure τ if and only if

$$m_\nu(x) \geq m_{\tau+s}(x)$$

for all $x \in G$ and $s \in S$.

Theorem 15. *There exists a constant $C > 0$ such that the following holds. Let G be a group and let k be an integer. Let A, X be subsets of G such that there is a subset S with $|S + A| \leq K|A|$. Let τ be a positive measure and ν a $k(S - S)$-upper envelope for τ. Then there exists a subset T of S with $|T| \geq \frac{|S|}{2K^{Cpk^2/\epsilon^2}}$ such that for all $t_1, t_2, \cdots, t_{2k-1}, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_{L^p(\tau)} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{L^{p/2}(\nu)}^{1/2} + \epsilon^2 \frac{|A|}{|G|} \|1\|_{L^1(\nu)}.$$

Proof. The only change compared to the above proofs is where we apply the triangle inequality, since we only do a local averaging here and need to be more careful with the change of variables. The above proof gives

$$\mathbf{E}[(m_{\bar{a}} * X(x) - m_A * X(x))^p] < \left(\frac{p}{en}\right)^p + \left(\frac{2p}{en}\right)^{p/2} (m_A * X(x))^{p/2}.$$

Let ν be an arbitrary positive measure. By averaging x with respect to $L^p(\nu)$, we get

$$\mathbf{E}[\|m_{\bar{a}} * X - m_A * X\|_{L^p(\nu)}^p] < \left(\frac{p}{en}\right)^p \|1\|_{L^1(\nu)} + \left(\frac{2p}{en}\right)^{p/2} \|m_A * X\|_{L^{p/2}(\nu)}^{p/2}.$$

Similar to the proof in the case $p = 2$, we get a tuple $\bar{s}$ and a subset T of S of size at least $\frac{|S|}{2K^n}$ such that for all $t \in T$,

$$\|m_{\bar{s}-t} * X(\cdot) - m_A * X(\cdot)\|_{L^p(\nu)} < \frac{2p}{en} \|1\|_{L^1(\nu)} + \sqrt{\frac{4p}{en}} \|m_A * X\|_{L^{p/2}(\nu)}^{1/2}.$$

Let τ be another measure. Then

$$\|m_A * X(t_1 - t_2 + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)}$$
$$\leq \|m_A * X(t_1 - t_2 + \cdot) - m_{-t_2+\bar{s}} * X(t_1 - t_2 + \cdot)\|_{L^p(\tau)} + \|m_{-t_1+\bar{s}} * X(\cdot) - m_A * X(\cdot)\|_{L^p(\tau)}$$
$$= \|m_A * X(\cdot) - m_{-t_2+\bar{s}} * X(\cdot)\|_{L^p(\tau+t_1-t_2)} + \|m_{-t_1+\bar{s}} * X(\cdot) - m_A * X(\cdot)\|_{L^p(\tau)}.$$

Similarly,

$$\|m_A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)}$$
$$\leq \|m_A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - m_A * X(t_{2k-1} - t_{2k} + \cdot)\|_{L^p(\tau)}$$
$$\quad + \|m_A * X(t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)}$$
$$\leq \|m_A * X(t_1 - t_2 + \cdots + t_{2k-3} - t_{2k-2} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau+t_{2k-1}-t_{2k})}$$
$$\quad + \|m_A * X(t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)}$$
$$\leq \cdots$$
$$\leq \|m_A * X(t_1 - t_2 + \cdot) - m_A * X(\cdot)\|_{L^p(\tau+t_3-t_4+\cdots+t_{2k-1}-t_{2k})}$$
$$\quad + \|m_A * X(t_3 - t_4 + \cdot) - m_A * X(\cdot)\|_{L^p(\tau+t_5-t_6+\cdots+t_{2k-1}-t_{2k})}$$
$$\quad + \cdots + \|m_A * X(t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)}.$$

17

Thus, if we let ν be a measure such that $m_\nu(x) \geq m_{\tau+k(T-T)}(x)$ for all x, then

$$\|m_A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - m_A * X(\cdot)\|_{L^p(\tau)} \leq k \cdot \left(\frac{2p}{en} \|1\|_{L^1(\nu)} + \sqrt{\frac{4p}{en}} \|m_A * X\|_{L^{p/2}(\nu)}^{1/2} \right).$$

$\square$

We give some remarks on generalizing the above results from subsets of G to general bounded functions. In the proof of Theorem 12, we use quite exact information about the exponential moment of binomial random variables. If we wish to generalize the above results from indicator functions to general bounded functions, we need to slightly generalize this step. However, we make a simple observation that the binomial case is essentially the worst case for the exponential moment, thus the same argument would go through in the general case.

Lemma 16. *Let X be a random variable taking values in $[0, 1]$. Then*

$$\mathbf{E}(\exp(\lambda X)) \leq \exp(\lambda)\mathbf{E}X + (1 - \mathbf{E}X).$$

Proof. Since $\exp(\lambda \cdot)$ is convex,

$$\exp(\lambda x) \leq x \exp(\lambda) + (1 - x)\exp(0).$$

Hence,

$$\mathbf{E}(\exp(\lambda X)) \leq \exp(\lambda)\mathbf{E}X + (1 - \mathbf{E}X).$$

$\square$

Using this observation, we can run through the entire argument to replace X by a function $f : G \to [0, 1]$.

The role of A can also be generalized to weighted sets, i.e., we can replace a set A by a function $f : G \to [0, 1]$. The assumption $|S + A| \leq K|A|$ needs to be changed to the condition

$$\mathbb{E}[\sup_{s \in S} f(x + s)] \leq K\mathbb{E}[f(x)].$$

In particular, this holds with $K = 1/\alpha$ for all $f : G \to [0, 1]$ with $\mathbb{E}f = \alpha$.

The only change in the proof of the almost periodicity results is that we need to sample our n-tuples of independent elements with probability given by f.

4 Bootstrapping

In this section, we bootstrap the almost periodicity result to obtain a well-structured set of almost periods. We have so far deduced that convolutions involving a nice set has a large set of almost periods. However, in the applications that follow, we often have iterative procedures, and we would like to have a set of almost periods that is more structured so that it can be easily iterated on. To be precise, we will construct a set of almost periods that is a subspace (in the finite field vector space case), or a Bohr set in the general case. This sounds daunting given the method we used in the last section, however, our a priori set of almost periods actually exhibits a fair amount of structure. We obtained a large set of almost periods that contain a highly iterated sumset of a large set, which is a fairly smooth object (for example, Bogolyubov-Ruzsa tells us that $2A - 2A$ contains a large subspace for dense A).

We will exactly use the smoothing effect of taking iterated sumset to deduce a large subspace or Bohr set of almost periods. In particular, taking sumset corresponds to taking product in the spectral domain, therefore taking the k-fold sumset of almost periods (which leads to only a polynomial loss in the bound

on the density of almost periods following the results in the previous section) has an exponential effect on the Fourier coefficients of the set of almost periods. To get a subspace or Bohr set, we essentially pick out the large Fourier coefficients of the set of almost periods and construct a subspace or Bohr set which annihilate the large spectrum. This leaves us with the task of understanding the large Fourier coefficients of a set, and how to efficiently annihilate the large spectrum. In the finite vector space case (and in fact, over general abelian groups), this follows from a result known as Chang's lemma, which gives a bound on the dimension of the large spectrum that significantly improves over Parseval's bound in our regime of interest. The general case is harder, since we have to iterate our argument on Bohr sets, which are not homogeneous objects like subgroups. Chang's lemma is generalized to work with sets defined locally on Bohr sets by an argument of Sanders. This looks more technical, but essentially generalizes the ingredients that we need in the finite field vector space case. We remark that the generalization of Chang's lemma to general abelian groups is in fact straightforward, as we shall see in Subsection 4.1. The main difficulty in the work of Sanders is to generalized Chang's lemma from the global setting to the local and relativised setting of sets defined on Bohr sets. As such, we sometimes refer to Subsection 4.1 as the global setting, whereas Subsection 4.2 is the local setting. In Subsection 4.1, we exposit Chang's lemma in the global setting, and particularly in finite field vector spaces, and use this to obtain our bootstrapped subspace of almost periods. In Subsection 4.2, we cover Sanders' generalization of Chang's lemma [36] and obtain our bootstrapped Bohr set of almost periods. We draw upon materials from [36, 41, 6].

4.1 Finite field vector spaces

4.1.1 A global version of Chang's lemma

The ϵ-spectrum of a function $f : G \to \mathbb{R}$ is defined by $\Delta_\epsilon(f) = \{\chi \in \hat{G} : \hat{f}(\chi) \geq \epsilon \|f\|_1\}$. Parseval's identity easily yields an upper bound on the size of the ϵ-spectrum

$$|\Delta_\epsilon(f)| \leq \left(\frac{\|f\|_2}{\epsilon \|f\|_1} \right)^2.$$

This bound is optimal, for example, if f is the indicator function of a subspace V, then $\hat{f}(\chi) = 0$ for all $\chi \notin V^\perp$, and $\hat{f}(\chi) = \frac{|V|}{|G|}$ for $\chi \in V^\perp$. Also, $\|f\|_1 = \frac{|V|}{|G|}$ and $\|f\|_2 = \sqrt{\frac{|V|}{|G|}}$, so $|\Delta_\epsilon(f)| = |V^\perp| = \frac{|G|}{|V|}$ for $\epsilon = 1$, which is equal to $\left(\frac{\|f\|_2}{\epsilon\|f\|_1} \right)^2$. However, observe that while the size of $\Delta_\epsilon(f)$ is potentially large, its dimension in the above case is only $\log_p |\Delta_\epsilon(f)|$. Thus, the large spectrum is highly structured, and indeed this intuition is quantified by Chang's lemma. Note that when f is an indicator function of a set A of density α, $\frac{\|f\|_2}{\|f\|_1} = \frac{1}{\sqrt{\alpha}}$ is large. Chang's lemma shows that in such a case, in fact, the dimension of $\Delta_\epsilon(f)$ can be more efficiently bounded. This is extremely useful in our forthcoming applications on Roth's theorem, as we will work with relatively sparse sets, and we do not want the codimension to grow too fast when passing to a subspace which is orthogonal to the large spectrum.

Theorem 17 (Chang's lemma). *Let $f : \mathbb{F}_p^n \to [0, 1]$. Then $\dim \Delta_\epsilon(f) \ll \epsilon^{-2} \log \left(2\frac{\|f\|_2}{\|f\|_1} \right)$.*

While Theorem 17 is stated over finite field vector spaces, this version of Chang's lemma generalizes directly to general abelian groups, which we state at the end of the subsection. Proofs of Chang's lemma can be found in [11, 19, 36, 41]. Chang [11] first proved this to prove a better bound on Freiman's theorem, though it is remarked that similar ideas can already be found in [8, 33]. Chang's lemma has proved to be useful in many other important problems in additive combinatorics, for example, Roth's theorem [5, 36], arithmetic progression in sumsets [19], and the structure of boolean functions with small l^1 norm [22]. A short proof of Chang's lemma over $\mathbb{F}_2^n$ can be found in [27] using entropy. In fact, one can also show Chang's lemma and more refined variants, including Bloom's version of Chang's lemma leading to the best known bound on Roth's theorem [5], using entropy [28, 44]. We refer the reader to the survey [44] for more details.

We give here the standard proof using Rudin's inequality and duality. We prove that the largest set of characters that are "independence" in a suitable sense in the large spectrum is small. This tells us that the dimension of $\Delta_\epsilon(f)$ must be small. To get a very good upper bound on the number of independent characters, we show that a function consisting of independent characters must exhibit a fair amount of cancellation, therefore concentrate around 0 in a suitable sense. In particular, we control the pth moment of such a function via the 2nd moment of the Fourier coefficients (at the independent characters), and use duality to control the 2nd moment of the Fourier coefficients using $p/(p-1)$th moment of the function, which is much smaller than Parseval's bound, yielding the desired upper bound on the dimension of the large spectrum.

The notion of independence that is sufficient for the proof to work is *dissociativity*, which is essentially a weak form of linear independence. While the role of this weaker notion is not clear here, it is essential to generalize the results to general abelian groups where linear independence is not defined.

Definition 18. A set of vectors Λ is called *dissociated* if any linear combination of the vectors with coefficients $\{-1, 0, 1\}$ where the coefficients are not identically zero is nonzero.

Note that if Λ is dissociated, then for any $\omega, a : \Lambda \to \mathbb{C}$,

$$\mathbb{E}_x \prod_{\lambda \in \Lambda} (1 + a(\lambda)\Re(\omega(\lambda)\lambda(x))) = \sum_{\epsilon_\lambda \in \{-1,0,1\}} \mathbb{E}_x c(a(\lambda), |\epsilon_\lambda|) \prod_\lambda c(\omega(\lambda)\lambda(x), \epsilon_\lambda) = 1,$$

where $c(x, i) = \begin{cases} x, & i = 1 \\ \bar{x}, & i = -1 \\ 1, & i = 0 \end{cases}$.

We use the above observation to prove Rudin's inequality, which says that functions formed by a combination of dissociated characters must have small exponential moment.

Theorem 19. *Let Λ be a dissociated set of characters, then*

$$\mathbb{E}_x \exp\left(\sigma \Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda(x)) \right) \leq e^{\sigma^2 \|\omega\|_2^2 / 2}.$$

Proof. We linearize the exponential function via the basic inequality $\exp(tx) \leq \cosh x + t \sinh x$,

$$\mathbb{E}_x \exp\left(\sigma \Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda(x)) \right) \leq \mathbb{E}_x \prod_{\lambda \in \Lambda} \left[\cosh(\sigma|\omega(\lambda)|) + \frac{\Re(\omega(\lambda)\lambda(x))}{|\omega(\lambda)|} \sinh(\sigma|\omega(\lambda)|) \right]$$

$$= \prod_{\lambda \in \Lambda} \cosh(\sigma|\omega(\lambda)|) \mathbb{E}_x \prod_{\lambda \in \Lambda} \left(1 + \frac{\sinh(\sigma|\omega(\lambda)|)}{|\omega(\lambda)| \cosh(\sigma|\omega(\lambda)|)} \Re(\omega(\lambda)\lambda(x)) \right)$$

$$= \prod_{\lambda \in \Lambda} \cosh(\sigma|\omega(\lambda)|).$$

Using $\cosh x \leq e^{x^2/2}$, we get the claimed bound. $\qquad \square$

Rudin's inequality immediately shows that a combination of dissociated characters cannot receive large value frequently. To deduce Chang's lemma from Rudin's inequality, we use duality, passing from exponential bound to bounds on L^p norms.

Proposition 20. *We have*

$$\| \sum_{\lambda \in \Lambda} \omega(\lambda)\lambda \|_p \leq 2p^{1/2} \|\omega\|_2.$$

Proof. Fix $t \geq 0$. From Rudin's inequality, we have

$$\mathbb{P}_x(\Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda) \geq t) \leq e^{\sigma^2 \|\omega\|_{L^2}^2 / 2 - \sigma t},$$

which is bounded by $\exp\left(-\frac{t^2}{2\|\omega\|_2^2}\right)$ when $\sigma = \frac{t}{\|\omega\|_2^2}$. Replacing $\omega(\lambda)$ with $-\omega(\lambda)$, we get

$$\mathbb{P}_x(\Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda) \leq -t) \leq e^{\sigma^2 \|\omega\|_2^2 / 2 - \sigma t}.$$

Hence,

$$\mathbb{P}_x(|\Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda)| \geq t) \leq 2\exp\left(-\frac{t^2}{2\|\omega\|_2^2}\right).$$

Replacing $\omega(\lambda)$ with $i\omega(\lambda)$, we get

$$\mathbb{P}_x(|\Im(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda)| \geq t) \leq 2\exp\left(-\frac{t^2}{2\|\omega\|_2^2}\right).$$

Hence,

$$\mathbb{P}_x\left(\left|\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda\right| \geq t\right) \leq 4\exp\left(-\frac{t^2}{4\|\omega\|_2^2}\right).$$

Therefore,

$$\|\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda\|_p^p = p\int_0^\infty t^{p-1}\mathbb{P}_x(|\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda| \geq t)dt$$

$$\leq p\int_0^\infty t^{p-1} \cdot 4\exp\left(-\frac{t^2}{4\|\omega\|_2^2}\right)dt$$

$$\leq 4\frac{p}{2}(p-1)!!(2\|\omega\|_2)^p$$

$$\leq \left(\frac{4p\|\omega\|_2^2}{e}\right)^{p/2}$$

Hence,

$$\|\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda\|_p \leq 2p^{1/2}\|\omega\|_2.$$

$\square$

The final step to deduce Chang's lemma is to use duality. The estimates above say that the map $L^2(\Lambda) \to L^p(G)$ defined by $\omega \mapsto \sum_{\lambda \in \Lambda} \omega(\lambda)\lambda$ has norm at most $2p^{1/2}$. Let $q = \frac{p}{p-1}$. Note that the dual map on the dual spaces $L^q(G) \to L^2(\Lambda)$ is $g \mapsto (\hat{g}(\lambda))_{\lambda \in \Lambda}$, since $\int_\Lambda \hat{g}(\lambda)\omega(\lambda) = \int_G g(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda)$ by Parseval's identity. The following result allows us to pass from the bound on the norm of a linear map to a bound on the norm of the dual map.

Lemma 21. *Let $A : X \to Y$ be a bounded linear map, where X, Y are reflexive Banach spaces, then $A^* : Y^* \to X^*$ defined by $A^*l(\cdot) = l(A\cdot)$ has $\|A^*\| \leq \|A\|$.*

Proof. Clearly

$$\|A^*l\|_{X^*} = \sup_{\|x\|_X \leq 1} |l(Ax)| \leq \sup_{\|x\|_X \leq 1} (\|l\|_{Y^*}\|Ax\|_{Y^*}) \leq \|A\|\|l\|_{X^*}.$$

Hence, $\|A^*\| \leq \|A\|$. $\square$

This duality allows us to control the L^2 norm of $(\hat{g}(\lambda))_{\lambda \in \Lambda}$ using the L^q norm of g, immediately yielding Chang's lemma.

Proof of Theorem 17. By duality,
$$\left(\sum_{\lambda \in \Lambda} |\hat{f}(\lambda)|^2\right)^{1/2} \leq 2p^{1/2}\|f\|_q.$$

Assume that Λ is a dissociated subset of $\Delta_\epsilon(f)$, we get
$$\sum_{\lambda \in \Lambda} |\hat{f}(\lambda)|^2 \geq |\Lambda|\epsilon^2(\mathbb{E}_x f)^2,$$

so
$$|\Lambda| \leq \frac{4p\|f\|_q^2/\|f\|_1^2}{\epsilon^2}.$$

Note that by convexity of norms,
$$\|f\|_q \leq \|f\|_2^{q-1}\|f\|_1^{2-q},$$

so
$$|\Lambda| \leq \frac{4p(\|f\|_2/\|f\|_1)^{2/(p-1)}}{\epsilon^2}.$$

Choosing $p = \log(2\|f\|_2/\|f\|_1) + 1$, we get
$$|\Lambda| \leq \frac{50\log(2\|f\|_2/\|f\|_1)}{\epsilon^2}.$$

In particular,
$$\dim \Delta_\epsilon(f) \leq \frac{50\log(2\|f\|_2/\|f\|_1)}{\epsilon^2}.$$

$\square$

The identical proof gives Chang's lemma over general abelian groups. For this, we need to adapt the notion of dimension with a spanning notion corresponding to dissociativity.

Theorem 22. *Let $f : G \to [0,1]$ for a finite abelian group G. Then there exists a set S of size at most $O\left(\epsilon^{-2}\log\left(2\frac{\|f\|_2}{\|f\|_1}\right)\right)$ such that*
$$\Delta_\epsilon(f) \subseteq \{\sum_{s \in S} \epsilon_s s, \epsilon_s \in \{-1,0,1\}\}.$$

Proof. Following the proof of Theorem 17, we find that the largest dissociated subset of $\Delta_\epsilon(f)$ has size at most $O\left(\epsilon^{-2}\log\left(2\frac{\|f\|_2}{\|f\|_1}\right)\right)$. Let S be such a subset. By maximality, for any $\lambda \in \Delta_\epsilon(f)$, $S \cup \{\lambda\}$ is not dissociated, hence,
$$\epsilon_\lambda \lambda + \sum_{s \in S} \epsilon_s s = 0$$

for some $\epsilon_s, \epsilon_\lambda \in \{-1,0,1\}$. Since S is dissociated, $\epsilon_\lambda \neq 0$. Hence,
$$\lambda = \pm\sum_{s \in S} \epsilon_s s \in \{\sum_{s \in S} \epsilon_s s, \epsilon_s \in \{-1,0,1\}\}.$$

Thus,
$$\Delta_\epsilon(f) \subseteq \{\sum_{s \in S} \epsilon_s s, \epsilon_s \in \{-1,0,1\}\}.$$

$\square$

We remark that Chang's lemma cannot generally be further improved, as shown by Green's construction [18], Theorem 50.

Theorem 17 and 22 imply the following results on annihilation of the large spectrum, showing that we can find a large subspace or Bohr set over which the characters in the large spectrum have value close to 1.

Theorem 23. *Let G be an abelian group and $f : G \to [0,1]$. If $G = \mathbb{F}_p^n$, there is a subspace V of codimension*

$$d \ll \epsilon^{-2} \log \left(2 \frac{\|f\|_2}{\|f\|_1} \right)$$

such that for all $\chi \in \Delta_\epsilon(f)$ and $v \in V$,

$$|1 - \chi(v)| = 0.$$

Otherwise, there is a regular Bohr set B of dimension

$$d \ll \epsilon^{-2} \log \left(2 \frac{\|f\|_2}{\|f\|_1} \right),$$

and radius

$$\rho \gg \frac{\delta \epsilon^2}{\log \left(2 \frac{\|f\|_2}{\|f\|_1} \right)},$$

such that for all $\chi \in \Delta_\epsilon(f)$ and $b \in B$,

$$|1 - \chi(b)| \leq \delta.$$

Proof. In the case $G = \mathbb{F}_p^n$, we simply take $V = (\Delta_\epsilon(f))^{\perp}$. In the general case, let S be the set from Theorem 22. Let $B = B\left(S, c\delta\epsilon^2 / \log\left(2\|f\|_2/\|f\|_1\right)\right)$ for some small constant c. For any $\chi \in \Delta_\epsilon(f)$, we can write

$$\chi = \sum_{s \in S} \epsilon_s s.$$

Thus, for $b \in B$, enumerating $S = \{s_1, \cdots, s_l\}$,

$$|1 - \chi(b)| = |1 - (\epsilon_{s_1} s_1)(b) + (\epsilon_{s_1} s_1)(b)(1 - (\epsilon_{s_2} s_2)(b)) + \cdots + (\sum_{k \leq l-1} \epsilon_{s_k} s_k)(b)(1 - (\epsilon_{s_l} s_l)(b))|$$

$$\leq \sum_{s \in S} |1 - (\epsilon_s s)(b)|$$

$$\leq \frac{c\delta\epsilon^2}{\log\left(2\|f\|_2/\|f\|_1\right)} \cdot |S|$$

$$\leq \delta,$$

where the last inequality follows is c is chosen sufficiently small. $\qquad \square$

4.1.2 Bootstrapping over finite field vector spaces

We now describe how to get the bootstrapped set of almost periods over $\mathbb{F}_p^n$. By Theorem 12, we can find a large set T with the property that

$$\|A * X(x + z) - A * X(x)\|_p \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{p/2}^{1/2} + \epsilon^2 \frac{|A|}{|G|}$$

for all $z \in kT - kT$. Notice that when we increase k, the dependency of the almost periodicity result on k is polynomial. However, repeated sumset has an exponential smoothing effect on the Fourier coefficients. We pick out the subspace V to be the orthogonal complement of the large spectrum of the set of almost periods, as in Theorem 23.

Theorem 24. *There exists a constant $C > 0$ such that the following holds. Let $G = \mathbb{F}_p^n$. Let A, X be subsets of G such that $\|A\|_1 \geq \alpha$, then there exists a subspace V of codimension*

$$d \leq Cp\epsilon^{-2}(\log \alpha^{-1})(\log \epsilon^{-1} + \log \alpha^{-1})^2$$

such that for all $v \in V$,

$$\|A * X(v + \cdot) - A * X(\cdot)\|_p \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{p/2}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

Proof. The core idea of bootstrapping is that the set of almost periods is fairly stable under taking sumsets. Let $k = \log(\epsilon^{-1}\alpha^{-1/2})$. Let T be the set of almost periods, $|T| \geq \frac{|G|}{\alpha^{-Cpk^2\epsilon^{-2}}}$. Let $V = \text{span}\{\Delta_{1/2}(T-T)\}^T$, which by Chang's lemma, $\text{codim}(V) \ll k^2\epsilon^{-2}\log(1/\alpha)$. Let

$$\tau_\pm^{*(k)}(x) = m_T * m_{-T} * \cdots * m_T * m_{-T}(x),$$

where there are k convolutions of $m_T * m_{-T}$. Note that $\tau_\pm^{*(k)}$ is the density with respect to the uniform measure on G of a probability distribution supported on $kT - kT$. We then have

$$\|A * X(v + \cdot) - A * X(\cdot)\|_p$$
$$\leq \|A * X * \tau_\pm^{*(k)}(\cdot) - A * X(\cdot)\|_p + \|A * X * \tau_\pm^{*(k)}(v + \cdot) - A * X(v + \cdot)\|_p$$
$$\quad + \|A * X * \tau_\pm^{*(k)}(v + \cdot) - A * X * \tau_\pm^{*(k)}(\cdot)\|_p$$
$$\leq 2\left(\epsilon\frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{p/2}^{1/2} + \epsilon^2\frac{|A|}{|G|}\right) + \|A * X * \tau_\pm^{*(k)}(v + \cdot) - A * X * \tau_\pm^{*(k)}(\cdot)\|_p,$$

where the second inequality follows since $\tau_\pm^{*(k)}$ is a probability distribution supported on $kT - kT$.

Note that for $v \in V$, $1 - \chi(-v) = 0$ for all $\chi \in \Delta_{1/2}(T - T)$, thus

$$|A * X * \tau_\pm^{*(k)}(v + x) - A * X * \tau_\pm^{*(k)}(x)| = |\sum_\chi \widehat{A}(\chi)\widehat{X}(\chi)(\widehat{m_T}(\chi)\widehat{m_{-T}}(\chi))^k\chi(-x)(\chi(-v) - 1)|$$
$$\leq \sum_\chi |\widehat{A}(\chi)\widehat{X}(\chi)(\widehat{m_T}(\chi)\widehat{m_{-T}}(\chi))^k||1 - \chi(-v)|$$
$$\leq 2\sum_{\chi \notin \Delta} |\widehat{A}(\chi)\widehat{X}(\chi)||\widehat{m_T}(\chi)|^{2k}$$
$$\leq 2(1/2)^{2k}(\sum_\chi \widehat{A}(\chi)^2)^{1/2}(\sum_\chi \widehat{X}(\chi)^2)^{1/2}$$
$$\leq 2^{1-2k}\alpha^{1/2}.$$

By our choice of k, we get

$$|A * X * \tau_\pm^{*(k)}(v + x) - A * X * \tau_\pm^{*(k)}(x)| \leq \epsilon^2\frac{|A|}{|G|}.$$

Thus,

$$\|A * X(v + \cdot) - A * X(\cdot)\|_p \leq 3\left(\epsilon\frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{p/2}^{1/2} + \epsilon^2\frac{|A|}{|G|}\right).$$

Finally, we replace ϵ by $\epsilon/3$ to arrive at the claimed bound. $\qquad\square$

4.2 General abelian groups

4.2.1 Annihilating the large spectrum with Bohr sets

In this section, we will cover Sanders' analog of Chang's lemma that applies to Bohr sets [36]. Following [36], we will introduce the tools to prove a local and relativised version of Chang's lemma where the function is defined only on a Bohr set. Sanders' result says that the large spectrum of a function defined on a Bohr set is annihilated by a large sub-Bohr set of small relative codimension. This is an analog of the condition of having small codimension in finite field vector spaces, as in Theorem 23.

In the proof of Chang's lemma over groups, dissociativity is used crucially in Rudin's inequality, where we use an averaging over the whole group together with the dissociativity condition. In the upcoming version of Chang's lemma with Bohr sets, we cannot average over the whole group, so we instead use the averaging condition as the definition of dissociativity. In fact, the more complicated averaging operator is the only change in this setting compared to the previous setting.

Definition 25. Let μ be a probability distribution over G. A set Λ is (K, μ)-dissociated if for all $\omega : \Lambda \to \mathbb{C}^\times$,

$$\int_G \prod_{\lambda \in \Lambda} (1 + \Re(\omega(\lambda)\lambda(x))) d\mu(x) \leq \exp(K).$$

By exactly the same proof as in the finite field case, we get the following version of Rudin's inequality.

Lemma 26. *Suppose Λ is (K, μ)-dissociated,*

$$\int_G \exp(\sigma\Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda(x))) d\mu(x) \leq \exp(K + \sigma^2\|\omega\|_2^2/2).$$

Proof. Using $\exp(tx) \leq \cosh x + t \sinh x$,

$$\int_G \exp(\sigma\Re(\sum_{\lambda \in \Lambda} \omega(\lambda)\lambda(x))) d\mu(x) \leq \int_G \prod_{\lambda \in \Lambda} \left[\cosh(\sigma|\omega(\lambda)|) + \frac{\Re(\omega(\lambda)\lambda(x))}{|\omega(\lambda)|} \sinh(\sigma|\omega(\lambda)|) \right] d\mu(x)$$

$$= \prod_{\lambda \in \Lambda} \cosh(\sigma|\omega(\lambda)|) \int_G \prod_{\lambda \in \Lambda} \left(1 + \Re\left(\frac{\omega(\lambda)\sinh(\sigma|\omega(\lambda)|)}{|\omega(\lambda)|\cosh(\sigma|\omega(\lambda)|)} \lambda(x) \right) \right) d\mu(x)$$

$$\leq \exp(\sigma^2\|\omega\|_2^2/2 + K).$$

$\square$

Deducing the p-moment from exponential concentration as in the global case, we get

$$\| \sum_{\lambda \in \Lambda} \omega(\lambda)\lambda \|_{L^p(\mu)} \leq 2p^{1/2} \exp(K/p)\|\omega\|_2.$$

We would like to next deduce using duality an upper bound on the dimension of the large spectrum of Chang's type using duality. However, we need to slightly change our definition of the large spectrum since we are working with a general measure. We define our large spectrum according to what is given by duality. Note that the dual of the map $L^2(\Lambda) \to L^p(\mu)$ mapping $\omega \mapsto \sum_{\lambda \in \Lambda} \omega(\lambda)\lambda$ is the map $L^q(\mu) \to L^2(\Lambda)$ mapping $g \mapsto (\widehat{gm_\mu}(\lambda))_{\lambda \in \Lambda}$. Define the large spectrum

$$\Delta_\epsilon(f) = \{\chi : |\widehat{fm_\mu}(\chi)| \geq \epsilon\|f\|_{L^1(\mu)}\}.$$

Via duality, as in the finite field case, we obtain the following bound,

$$\|(\widehat{gm_\mu}(\lambda))_{\lambda \in \Lambda}\|_2 \leq 2p^{1/2} \exp(K/p)\|g\|_{L^q(\mu)},$$

where q is the conjugate index $q = \frac{p}{p-1}$. We use this to bound the largest dissociated subset of the large spectrum.

25

Lemma 27. *Let $f : G \to [0,1]$ be a function supported on B. Recall that β denotes the normalized measure of B, having density $m_\beta(x) = |G|\frac{B(x)}{|B|}$ with respect to the uniform measure on G. Then any $(1,\beta)$-dissociated subset of $\Delta_\epsilon(f)$ has size at most $O\left(\frac{\log(2\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)})}{\epsilon^2}\right)$.*

Proof. Let Λ be a $(1,\beta)$-dissociated subset of $\Delta_\epsilon(f)$. By the duality bound,

$$|\Lambda|\epsilon^2\|f\|^2_{L^1(\beta)} \leq 2p^{1/2}\exp(1/p)\|f\|_{L^q(\mu)} \leq 2p^{1/2}\exp(1/p)\|f\|^{q-1}_{L^2(\mu)}\|f\|^{2-q}_{L^1(\mu)},$$

so

$$|\Lambda| \leq \frac{4p\exp(2/p)(\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)})^{-2/(p-1)}}{\epsilon^2}.$$

Choosing $p = \log(2\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)}) + 1$, we obtain the claimed bounds. $\qquad\square$

We have so far created our definitions to match exactly what goes on in the global setting. In the last step, we need to pass this information back to the Bohr set. It is not a priori clear that our analytic definition of dissociativity has any relationship to more combinatorial notions of dimension, which allows for annihilation, as seen in the previous subsection. We will derive directly from analytic dissociativity an annihilation result, in a way that is quite reminiscent of the covering argument in the previous subsection.

Lemma 28. *Assume that B is a κ-regular Bohr set with dimension d with $\kappa \leq \frac{10^{-4}}{d}$, and $S \subseteq \hat{G}$ is a subset of character so that the largest $(1,\beta)$-dissociated subset of S has size at most k. Then there is a $(1,\beta)$-dissociated subset Λ of S such that for all $x \in B_\rho \cap B(\Lambda, \nu)$ and $\gamma \in S$,*

$$|1 - \gamma(x)| \leq C(k\nu + \rho d^2(k+1)).$$

For a reason that will be clear in the proof of the lemma, we will work with a highly smoothed version of β instead of working directly with β. Let L be a large constant to be chosen later, and let

$$m_{\beta^+} = m_{B_{1+L\rho'}} * m_{B_{\rho'}} * \cdots * m_{B_{\rho'}},$$

where there are L terms $m_{B_{\rho'}}$. Since B is a regular Bohr set, m_{β^+} is close to m_β. In fact, for $x \in B$, $y_1, y_2, \cdots, y_L \in B_{\rho'}$, $x - \sum_{i=1}^L y_i \in B_{1+L\rho'}$, so

$$m_{\beta^+}(x) \geq \frac{|G|}{|B_{1+L\rho'}|} \geq \frac{|B|}{|B_{1+L\rho'}|}m_\beta(x).$$

By regularity of B, $|B_{1+L\rho'}| \leq (1 + 81dL\rho' + 81d\kappa)|B|$, as long as $L\rho' + \kappa \leq \frac{1}{81d}$. Choose ρ' so that $\frac{1}{4} \leq 81d(L\rho' + \kappa) \leq \frac{1}{2}$ and $B_{\rho'}$ is regular. Since β^+ is essentially an upper envelope for β, it is not surprising that dissociated subsets with respect to β^+ are essentially dissociated subsets with respect to β. In particular,

$$\int_G \prod_{\lambda \in \Lambda}(1 + \Re(\omega(\lambda)\lambda(x)))d\beta(x) \leq \frac{3}{2}\int_G \prod_{\lambda \in \Lambda}(1 + \Re(\omega(\lambda)\lambda(x)))d\beta^+(x)$$

$$\leq \exp(1/2)\int_G \prod_{\lambda \in \Lambda}(1 + \Re(\omega(\lambda)\lambda(x)))d\beta^+(x).$$

Hence, a set T which is $(1/2, \beta^+)$-dissociated is $(1,\beta)$-dissociated.

We construct Λ via an iterative process. This is essentially an analog of the fact that one can get a covering result out of a maximal dissociated set used in the previous subsection. We define $\eta_i = \frac{i}{2(k+1)}$, and construct sets Λ_i iteratively so that Λ_i is (η_i, β^+)-dissociated. In particular, we initialize $\Lambda_0 = \emptyset$. If there is $\gamma \in S\backslash\Lambda_i$ such that $\Lambda_i \cup \{\gamma\}$ is (η_{i+1}, β^+)-dissociated then we set $\Lambda_{i+1} = \Lambda_i \cup \{\gamma\}$. If no such γ

26

exists, we stop the process. Observe that if the process continues for at least $k + 1$ steps, then we find a $(\frac{1}{2}, \beta^+)$-dissociated subset of S of size at least $k + 1$, hence also $(1, \beta)$-dissociated, which is a contradiction. Thus we must have stopped at some step $i \leq k$. Then there is a set Λ_i such that Λ_i is (η_i, β^+)-dissociated, but $\Lambda_i \cup \{\gamma\}$ is not (η_{i+1}, β^+)-dissociated for all $\gamma \in S \backslash \Lambda_i$. We let $\Lambda = \Lambda_i$ and prove that it has the desired property.

Proof of Lemma 28. By our construction, for all $\gamma \in S \backslash \Lambda$, we can find $\omega : \Lambda \cup \{\gamma\} \to \mathbb{C}^\times$ such that

$$\fint_G (1 + \Re(\omega(\gamma)\gamma(x))) \prod_{\lambda \in \Lambda} (1 + \Re(\omega(\lambda)\lambda(x))) d\beta^+(x) > \exp(\eta_{i+1}).$$

However,

$$\fint_G \prod_{\lambda \in \Lambda} (1 + \Re(\omega(\lambda)\lambda(x))) d\beta^+(x) \leq \exp(\eta_i).$$

Thus,

$$\fint_G \Re(\omega(\gamma)\gamma(x)) \prod_{\lambda \in \Lambda} (1 + \Re(\omega(\lambda)\lambda(x))) d\beta^+(x) > \exp(\eta_{i+1}) - \exp(\eta_i).$$

We have

$$\fint_G \Re(\omega(\gamma)\gamma(x)) \prod_{\lambda \in \Lambda} (1 + \Re(\omega(\lambda)\lambda(x))) d\beta^+(x)$$

$$= \sum_{\epsilon_\gamma \in \{\pm 1\}, \epsilon_\lambda \in \{0, \pm 1\}} c(\omega(\gamma), \epsilon_\gamma) \prod_{\lambda \in \Lambda} c(\omega(\lambda), \epsilon_\lambda) \int c(\gamma(x), \epsilon_\gamma) \prod_{\lambda \in \Lambda} c(\lambda(x), \epsilon_\lambda) d\beta^+(x)$$

$$= \sum_{\epsilon_\gamma \in \{\pm 1\}, \epsilon_\lambda \in \{0, \pm 1\}} c(\omega(\gamma), \epsilon_\gamma) \prod_{\lambda \in \Lambda} c(\omega(\lambda), \epsilon_\lambda) \widehat{m_{\beta^+}}\left(\epsilon_\gamma \gamma + \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda\right).$$

Thus,

$$\exp(\eta_{i+1}) - \exp(\eta_i) \leq \sum_{\epsilon_\gamma \in \{\pm 1\}, \epsilon_\lambda \in \{0, \pm 1\}} \left| \widehat{m_{\beta^+}}\left(\epsilon_\gamma \gamma + \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda\right) \right|$$

$$\leq \sum_{\epsilon_\gamma \in \{\pm 1\}, \epsilon_\lambda \in \{0, \pm 1\}} \left| \widehat{m_{B_{\rho'}}}\left(\epsilon_\gamma \gamma + \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda\right) \right|^L$$

$$= 2 \sum_{\epsilon_\lambda \in \{0, \pm 1\}} \left| \widehat{m_{B_{\rho'}}}\left(\gamma - \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda\right) \right|^L.$$

Choosing $L = \lceil \log_2 3^k 2(k + 1) \rceil$, and note that

$$\exp(\eta_{i+1}) - \exp(\eta_i) = \exp\left(\frac{i+1}{2(k+1)}\right) - \exp\left(\frac{i}{2(k+1)}\right) \geq \frac{1}{2(k+1)},$$

we get $\sup_{\epsilon_\lambda \in \{0, \pm 1\}} |\widehat{m_{B_{\rho'}}}(\gamma - \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda)| \geq \frac{1}{2}$. Let $\chi = \gamma - \sum_{\lambda \in \Lambda} \epsilon_\lambda \lambda$ where $|\widehat{m_{B_{\rho'}}}(\chi)| \geq \frac{1}{2}$. Note that if

$x \in B_\rho$ for $\rho \leq \frac{\rho'}{80d}$, then

$$|\widehat{m_{B_{\rho'}}}(\chi)(1 - \chi(x))| = \frac{1}{|B_{\rho'}|} \left| \sum_y (\chi(y) - \chi(x + y)) B_{\rho'}(y) \right|$$

$$= \frac{1}{|B_{\rho'}|} \left| \sum_{y \in B_{\rho'}} \chi(y) - \sum_{y \in B_{\rho'} + x} \chi(y) \right|$$

$$\leq \frac{|B_{\rho' + \rho} \setminus B_{\rho' - \rho}|}{|B_{\rho'}|}$$

$$\leq 160d \frac{\rho}{\rho'},$$

by regularity of $B_{\rho'}$. Thus,

$$|1 - \chi(x)| \leq 320d\rho/\rho'.$$

Enumerating $\Lambda = \{\lambda_1, \cdots, \lambda_i\}$, and using the triangle inequality, we get

$$|1 - \gamma(x)| = |1 - \chi(x)(\sum_{j=1}^{i} \epsilon_{\lambda_j} \lambda_j)(x)|$$

$$\leq |1 - \chi(x)| + \sum_{l=1}^{i} |\chi(x)| \prod_{j=1}^{l-1} |\lambda_l(x)| \cdot |1 - \lambda_l(x)|$$

$$\leq |1 - \chi(x)| + \sum_{j=1}^{i} |1 - \lambda_j(x)|.$$

Thus, if $x \in B_\rho \cap B(\Lambda, \nu)$,

$$|1 - \gamma(x)| \leq 320d\rho/\rho' + k2\pi\nu.$$

Plugging in $\rho' \geq \frac{\frac{1}{640d} - \kappa}{L} \geq \frac{10^{-3}}{dL}$ and $L = \lceil \log_2 3^k 2(k+1) \rceil$, we get the desired conclusion for all $\gamma \in S \setminus \Lambda$. For $\gamma \in \Lambda$, the conclusion is obvious. $\square$

Combining Lemma 27 and 28 readily gives the proof of the following relativised version of Chang's lemma.

Theorem 29. *There exists a constant C such that the following holds. Let B be a κ-regular Bohr set with dimension d, radius ρ, and $\kappa \leq 10^{-4}/d$. Let $f : B \to [0, 1]$ be a function defined on B. We can find a Bohr set B' of dimension*

$$d' \leq d + C \frac{\log(\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)})}{\epsilon^2},$$

and radius

$$\rho' \geq c \frac{\rho\delta\epsilon^2}{d^2 \log(\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)})},$$

such that for all $\chi \in \Delta_\epsilon(f)$ and $x \in B'$,

$$|1 - \chi(x)| \leq \delta.$$

Furthermore, when $f = T$ is the indicator function of a set T, then

$$\|f\|_{L^2(\beta)}/\|f\|_{L^1(\beta)} = |B|/|T|.$$

4.2.2 Bootstrapping over abelian groups

In the following, we assume that G is an abelian group of odd order, so the map $\gamma \in \hat{G} \mapsto \gamma/2 \in \hat{G}$ is well-defined. The next theorem is the general relativised bootstrapping analog of Theorem 24.

Theorem 30. *There exists a constant $C > 0$ such that the following holds. Let B be a κ-regular Bohr set of dimension d and radius ρ where $\kappa \leq \frac{10^{-4}}{d}$. Let μ be so that $\mu \leq \frac{10^{-3}}{d\log(\alpha\epsilon)^{-1}}$ and B_μ is regular. Let $\bar{\beta}$ be the measure supported on $B_{1+2\mu\log(\alpha\epsilon)^{-1}}$ with density $m_{\bar{\beta}}(x) = |G|\frac{B_{1+2\mu\log(\alpha\epsilon)^{-1}}(x)}{|B|}$ with respect to the uniform measure on G. Let $\tilde{B}$ be a regular Bohr set such that $B_\eta \subseteq \tilde{B}_{\theta\eta}$ for all $1/\theta \geq \eta \geq 0$. Let A be a subset of $\tilde{B}$ such that $\|A\|_{L^1(\tilde{B})} \geq \alpha$. Assume that $\log(\epsilon^{-1}\alpha^{-1})\mu d\theta \leq 10^{-4}$.*

Then there exists a Bohr set B' with dimension

$$d' \leq d + Cp\epsilon^{-2}\log(\epsilon^{-1}\alpha^{-1})^2\log\alpha^{-1},$$

and radius

$$\rho' \geq c\mu\rho\frac{\epsilon^6\alpha^{1/2}}{pd^2\log(\epsilon^{-1}\alpha^{-1})^2\log\alpha^{-1}},$$

such that for all $b \in B'$,

$$\|A * X(b + \cdot) - A * X(\cdot)\|_{L^p(B)} \leq \epsilon\frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{L^{p/2}(\bar{\beta})}^{1/2} + \epsilon^2\frac{|A|}{|G|}.$$

Proof. The proof follows the same ideas as in the finite field case. Let $k = \log(\epsilon^{-1}\alpha^{-1})$, $S = B_\mu$. Then $B + k(S - S) \subseteq B_{1+2k\mu}$, so

$$|S + A| \leq |\tilde{B} + B_{2k\mu}| \leq \frac{|\tilde{B}_{1+2k\mu\theta}|}{|\tilde{B}|\alpha}|A| \leq \frac{1 + 320k\mu d\theta}{\alpha}|A| \leq \frac{2}{\alpha}|A|.$$

Let T be the set of almost periods in Theorem 15, $|T| \geq \frac{|S|}{\alpha^{-Cpk^2\epsilon^{-2}}}$. Note that $\bar{\beta}$ is a $k(S - S)$-upper envelope for β, and

$$\|1\|_{L^p(\bar{\beta})} = \frac{|B_{1+2k\mu}|}{|B|} \leq 2.$$

Let

$$\tau_\pm^{*(k)}(x) = m_T * m_{-T} * \cdots * m_T * m_{-T}(x),$$

where there are k convolutions of $m_T * m_{-T}$.

Then

$$\|A * X(b + \cdot) - A * X(\cdot)\|_{L^p(B)}$$
$$\leq \|A * X * \tau_\pm^{*(k)}(\cdot) - A * X(\cdot)\|_{L^p(B)} + \|A * X * \tau_\pm^{*(k)}(b + \cdot) - A * X(v + \cdot)\|_{L^p(B)}$$
$$\quad + \|A * X * \tau_\pm^{*(k)}(b + \cdot) - A * X * \tau_\pm^{*(k)}(\cdot)\|_{L^p(B)}$$
$$\leq 2\left(\epsilon\frac{|A|^{1/2}}{|G|^{1/2}}\|A * X\|_{L^{p/2}(\bar{\beta})}^{1/2} + 2\epsilon^2\frac{|A|}{|G|}\right) + \|A * X * \tau_\pm^{*(k)}(b + \cdot) - A * X * \tau_\pm^{*(k)}(\cdot)\|_{L^p(B)}.$$

Note that

$$|A * X * \tau_\pm^{*(k)}(b + x) - A * X * \tau_\pm^{*(k)}(x)|$$
$$= |\sum_\chi \widehat{A}(\chi)\widehat{X}(\chi)(\widehat{m_T}(\chi)\widehat{m_{-T}}(\chi))^k\chi(-x)(\chi(-b) - 1)|$$
$$\leq \sum_\chi |\widehat{A}(\chi)\widehat{X}(\chi)(\widehat{m_T}(\chi)\widehat{m_{-T}}(\chi))^k||1 - \chi(-b)|.$$

By Theorem 29, we can find a Bohr set B' with dimension

$$d' \leq d + C \log \frac{|B_\mu|}{|S|} \leq d + Cpk^2\epsilon^{-2} \log \alpha^{-1}$$

and radius

$$\rho' \geq c\mu\rho \frac{\epsilon^2 \delta}{d^2 pk^2 \epsilon^{-2} \log \alpha^{-1}}$$

such that for all $\chi \in \Delta_{1/2}(T)$, $b \in B'$,

$$|1 - \chi(-b)| \leq \delta.$$

Then for $b \in B'$,

$$|A * X * \tau_\pm^{*(k)}(b+x) - A * X * \tau_\pm^{*(k)}(x)|$$
$$\leq \delta \sum_{\chi \in \Delta_{1/2}(T)} |\widehat{A}(\chi)\widehat{X}(\chi)| + 2(1/2)^{2k} \sum_{\chi \notin \Delta_{1/2}(T)} |\widehat{A}(\chi)\widehat{X}(\chi)|$$
$$\leq \max(\delta, 2^{1-2k})\sqrt{\frac{|A|\,|X|}{|G|\,|G|}}.$$

Choose $\delta = \epsilon^2 \alpha^{1/2}$, and by our choice of k, we get

$$|A * X * \tau_\pm^{*(k)}(b+x) - A * X * \tau_\pm^{*(k)}(x)| \leq \epsilon^2 \frac{|A|}{|G|}.$$

Hence,

$$\|A * X * \tau_\pm^{*(k)}(b+\cdot) - A * X * \tau_\pm^{*(k)}(\cdot)\|_{L^p(B)} \leq \epsilon^2 \frac{|A|}{|G|}.$$

Replacing ϵ by $\epsilon/8$, we arrive at the required conclusion. $\qquad\square$

The above theorem is not convenient to use due to the measure $\bar{\beta}$. Using regularity, we can replace $\bar{\beta}$ by β, at the cost of a worse bound on the radius. We remark that the following theorem, as stated, only applies to regular Bohr sets and not κ-regular Bohr sets.

Theorem 31. *There exists a constant $C > 0$ such that the following holds. Let B be a regular Bohr set of dimension d and radius ρ. Let $\tilde{B}$ be a regular Bohr set such that $B_\eta \subseteq \tilde{B}_{\theta\eta}$ for all $1/\theta \geq \eta \geq 0$. Let A be a subset of $\tilde{B}$ such that $\|A\|_{L^1(\tilde{B})} \geq \alpha$. Assume that $\theta \leq \alpha^{-3}\epsilon^{-3}$.*
Then there exists a Bohr set B' with dimension

$$d' \leq d + Cp\epsilon^{-2} \log(\epsilon^{-1}\alpha^{-1})^2 \log \alpha^{-1},$$

and radius

$$\rho' \geq c\rho \left(\frac{\epsilon^{p+1}\alpha}{pd \log(\epsilon^{-1}\alpha^{-1}) \log \alpha^{-1}} \right)^6,$$

such that for all $b \in B'$,

$$\|A * X(b+\cdot) - A * X(\cdot)\|_{L^p(B)} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{L^{p/2}(B)}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

Proof. We need to control $\|A * X\|_{L^{p/2}(\bar{\beta})}^{1/2}$ in terms of $\|A * X\|_{L^{p/2}(B)}^{1/2}$ by choosing μ in Theorem 30 appropriately. Note that $A * X(x) \leq \frac{|A|}{|G|}$, and

$$\|A * X\|_{L^{p/2}(\bar{\beta})}^{p/2} = \|A * X\|_{L^{p/2}(B)}^{p/2} + \frac{1}{|B|} \sum_{x \in B_{1+2k\mu} \backslash B} |A * X(x)|^{p/2}$$

$$\leq \|A * X\|_{L^{p/2}(B)}^{p/2} + 160 dk\mu\alpha^{p/2}.$$

Thus,

$$\|A * X(b + \cdot) - A * X(\cdot)\|_{L^p(B)} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{L^{p/2}(B)}^{1/2} + \epsilon^2 \frac{|A|}{|G|} + (160 dk\mu)^{1/p} \epsilon \frac{|A|}{|G|}.$$

Choosing $\mu = \frac{\epsilon^p}{160 kd}$, we get

$$\|A * X(b + \cdot) - A * X(\cdot)\|_{L^p(B)} \leq 2 \left(\epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * X\|_{L^{p/2}(B)}^{1/2} + \epsilon^2 \frac{|A|}{|G|} \right).$$

Replacing ϵ by $\epsilon/2$, we arrive at the conclusion. $\qquad\square$

5 Roth's theorem over finite field vector spaces

5.1 The density increment strategy

The density increment strategy dates back to Roth's original proof [32] of the existence of three-term arithmetic progressions in dense subsets of the integers, using Fourier analysis. The simpler proof in the finite field setting was obtained by Meshulam [30]. The rough idea is that if a set A behaves like a random set, then A necessarily contains many three-term arithmetic progressions. If A contains significantly fewer three-term arithmetic progressions, then A must have increased density on a structured set. The notion of pseudorandomness in the original proofs of Roth's theorem that suffices to control three-term arithmetic progression density is Fourier uniformity. The increment accordingly proceeds via analysis of the Fourier coefficients in the spectral domain. In later subsections, we will replace this crucial step by almost periodicity and obtain density increment directly from the physical space without going into the spectral domain. Of course, one may say that spectral analysis is still crucial in bootstrapping. However, in this form, it has much less effect on the density increment step. In this subsection, we first lay out the basic framework behind the density increment, and record the dependency of the quantitative bound in the final density on the parameters in the density increment step.

Theorem 32. *Let $\mathcal{P}$ be a property of subsets of $\mathbb{F}_p^n$ which is translation invariant and upward closed, i.e., for each $t \in \mathbb{F}_p^n$, A has $\mathcal{P}$ iff $A + t$ has $\mathcal{P}$, and if A has $\mathcal{P}$ then all $A' \supseteq A$ has $\mathcal{P}$. Assume that the following holds. Given any subset A of $\mathbb{F}_p^n$ with density α and $\mathcal{P}$ does not hold, then either $n < C \log(C\alpha^{-1})$ or there exists a subspace V of codimension bounded by $C\alpha^{-\mu}(\log C\alpha^{-1})^m$ and x such that $A * m_V(x) \geq (1 + c)\alpha$.*
Then any set A of $\mathbb{F}_p^n$ which does not have $\mathcal{P}$ must have density at most $C_{32}(C, c, \mu, m)(\log n)^{m/\mu} n^{-1/\mu}$.

Proof. The proof follows from a simple iteration. Assume that the set A we start with does not have $\mathcal{P}$. Assume that $n > C \log(C\alpha^{-1})$. Let α be the density of A, then we can find a subspace V_1 of codimension at most $C\alpha^{-\mu}(\log C\alpha^{-1})^m$ and a translate of the subspace where the density of A is at least $(1 + c)\alpha$. By translation invariance, we can restrict our attention to the subspace $V_1 \cong \mathbb{F}_p^{n_1}$ with $n_1 \geq n - \alpha^{-\mu}(\log \alpha^{-1})^m$, and $A_1 = A \cap (V_1 + x_1)$ having density $\alpha_1 \geq (1 + c)\alpha$ in $V_1 + x_1$. By upward closeness of $\mathcal{P}$, A_1 does not have $\mathcal{P}$. As long as $n_1 > C \log(C\alpha_1^{-1})$, we can find a subspace V_2 of V_1 of codimension at most $C\alpha_1^{-\mu}(\log C\alpha_1^{-1})^m < C\alpha^{-\mu}(\log C\alpha^{-1})^m$ and a translate of V_2 where $A_1 \cap (V_2 + x_2)$

has density $\alpha_2 \geq (1 + c)\alpha_1$ in V_1. This can be iterated as long as $\dim V_i > C \log(C\alpha_i^{-1})$. Note that the number of iterations is bounded above by $\log_{1+c}(1/\alpha)$ since clearly the density of a set A_i in V_i is at most 1. Moreover $\dim V_i \geq n - C \sum_{j=1}^{i} \alpha_j^{-\mu}(\log C\alpha_j^{-1})^m$. Thus, at the last step t before the iteration is forced to terminate, we must have

$$n - C\sum_{i=1}^{t} \alpha_i^{-\mu}(\log C\alpha_i^{-1})^m \leq C \log C\alpha^{-1},$$

where $\alpha_{i+1} \geq (1+c)\alpha_i$ for all $i < t$, so

$$n \leq C \log(C\alpha^{-1}) + C'\alpha^{-\mu}(\log C\alpha^{-1})^m.$$

Hence, we can easily check that

$$\alpha \leq \frac{C_{32}(\log n)^{m/\mu}}{n^{1/\mu}},$$

for some constant C_{32} depending on C, c, μ, m. $\qquad\square$

In the next subsections, we will turn to establish the density increment condition in Theorem 32, thereby establishing Roth's theorem in this setting. We note that over the finite field setting, much better bounds for Roth's theorem are known using the polynomial method [12, 16], which is a completely different framework from what we discuss here. There, it is shown that the largest subset of $\mathbb{F}_p^n$ with no nontrivial three-term arithmetic progression has size at most $(c_p p)^n$ for a constant $c_p < 1$. However, the polynomial method cannot be applied to the general abelian groups while what we describe here generalizes to general abelian groups via standard techniques. In fact, such a strong bound is not true in general abelian groups, where the Behrend construction gives a lower bound which is strictly bigger than any power smaller than 1 of the group size. The methods we discuss will lead to a bound of the form $Cp^n(\log n)^4/n$, which is much weaker. In fact, the bound we obtain is weaker than Meshulam's bound, which has the form Cp^n/n. However, this method directly generalizes to essentially the best bound for the problem over general abelian groups. Thus, one should think about the results in this section as being the model for the argument over general abelian groups instead of focusing on the exact quantitative bounds.

This section is based on materials from [6, 38].

5.2 Roth's theorem in finite field vector spaces

In this section, we will prove the density increment condition in Theorem 32, using the bootstrapped version of almost periodicity to replace spectral analysis.

Theorem 33. *Let p be an odd prime. Let A be a subset of $\mathbb{F}_p^n$ with density α. If $\mathbb{E}_{x,d}A(x)A(x+d)A(x+2d) < \frac{\alpha^3}{2}$, we can find a subspace V of codimension at most $C\alpha^{-1}(\log C\alpha^{-1})^4$ such that there exists a translate $V + x$ of V where $\frac{|A \cap (V+x)|}{|V|} \geq \frac{3}{2}\alpha$.*

Proof. By the hypothesis,

$$\mathbb{E}_{x,d}A(x)A(x+d)A(x+2d) = \mathbb{E}_x A * A(x)(2 \cdot A)(x) < \frac{\alpha^3}{2}.$$

Using the almost periodicity result, we get a good approximation of $A * A$ which is a constant function on translates of a large subspace. In particular, apply Theorem 24 with $p = 20\log(\alpha^{-1})$, we get a subspace V of codimension at most $Cp\epsilon^{-2}(\log\alpha^{-1})(\log\epsilon^{-1} + \log\alpha^{-1})^2$ such that

$$\|A * A * m_V - A * A\|_p \leq \epsilon\sqrt{\alpha}\|A * A\|_{p/2}^{1/2} + \epsilon^2\alpha.$$

Our argument will proceed as follows. Assuming that $\|A * A\|_{p/2}^{1/2}$ is sufficiently small, then $A * A$ is well-approximated by $A * A * m_V$, which is much smoother. We can show that if there is no density increment in a translate of V, then $A * A * m_V$ is essentially close to α, meaning that the value of $\mathbb{E}_x A * A * m_V(x)(2 \cdot A)(x)$ is close to α^3, so $\mathbb{E}_x A * A(x)(2 \cdot A)(x)$ is also large. On the other hand, if $\|A * A\|_{p/2}^{1/2}$ is large, we directly get density increment, since this implies that $\|A * A * m_V\|_p$ is large.

Let $c = 1/2$. Our aim is to show that there exists x such that $A * m_V(x) \geq (1+c)\alpha$. Assume otherwise that $\|A * m_V\|_\infty < (1+c)\alpha$. By a basic averaging argument, we will show that $A * A * m_V(x) \geq (1-c^2)\alpha^2$ for all x. Indeed, notice that $m_V * m_V = m_V$, and $\mathbb{E}_x A * m_V(x) = \alpha$. Moreover, by our assumption, for all y, z,

$$[(A * m_V)(y) - (1+c)\alpha][(A * m_V)(z) - (1+c)\alpha] \geq 0,$$

so

$$(A * m_V)(y)(A * m_V)(z) \geq (1+c)\alpha[(A * m_V)(y) + (A * m_V)(z)] - (1+c)^2\alpha^2.$$

By averaging this inequality,

$$
\begin{aligned}
A * A * m_V(x) &= (A * m_V) * (A * m_V)(x) \\
&= \mathbb{E}_y (A * m_V)(y)(A * m_V)(x - y) \\
&\geq (1+c)\alpha \mathbb{E}_y[(A * m_V)(y) + (A * m_V)(x - y)] - (1+c)^2\alpha^2 \\
&= 2(1+c)\alpha^2 - (1+c)^2\alpha^2 \\
&\geq (1-c^2)\alpha^2.
\end{aligned}
$$

Thus,

$$\mathbb{E}_x A * A * m_V(x)(2 \cdot A)(x) \geq (1-c^2)\alpha^2 \mathbb{E}_x(2 \cdot A)(x) = (1-c^2)\alpha^3.$$

The fact that $A * A * m_V$ approximates $A * A$ in L^p allows us to pass this approximate information to a lower bound on $\mathbb{E}_x A * A(x)(2 \cdot A)(x)$. By Holder's inequality,

$$
\begin{aligned}
|\mathbb{E}_x(A * A - A * A * m_V(x))(2 \cdot A)(x)| &\leq \|A * A - A * A * m_V\|_p \|(2 \cdot A)\|_{p/(p-1)} \\
&\leq \epsilon\sqrt{\alpha}(\|A * A\|_{p/2}^{1/2} + \epsilon\sqrt{\alpha})\alpha^{1-1/p}
\end{aligned}
$$

If $\|A * A\|_{p/2} \leq 10\alpha^2$, then choosing $\epsilon = \sqrt{\alpha}/100$, we get

$$|\mathbb{E}_x(A * A - A * A * m_V(x))(2 \cdot A)(x)| \leq \alpha^3/4,$$

so

$$\mathbb{E}_x A * A(x)(2 \cdot A)(x) \geq (1 - c^2 - 1/4)\alpha^3 = \alpha^3/2,$$

which contradicts the assumption.

Consider the case $\|A * A\|_{p/2} > 10\alpha^2$. Choose $\epsilon = \sqrt{\alpha}/100$. Then, $\|A * A\|_p \geq \|A * A\|_{p/2} > 10\alpha^2$, so we get

$$
\begin{aligned}
\|A * A * m_V\|_p &\geq \|A * A\|_p - \|A * A * m_V - A * A\|_p \\
&\geq \|A * A\|_p - \epsilon\sqrt{\alpha}\|A * A\|_{p/2}^{1/2} - \epsilon^2\alpha \\
&\geq \|A * A\|_p^{1/2}(\|A * A\|_p^{1/2} - \alpha/100) - \alpha^2/100^2 \\
&> 5\alpha^2.
\end{aligned}
$$

However, if $\|A * m_V\|_\infty \leq (1+c)\alpha$ then $\|A * A * m_V\|_\infty = \|A * m_V * A * m_V\|_\infty \leq (1+c)^2\alpha^2$, contradicting the above inequality. $\qquad\square$

Corollary 34. *The largest subset of $\mathbb{F}_p^n$ with no nontrivial three-term arithmetic progression has size at most $p^n(\log n)^4/n$.*

Proof. This follows from Theorem 32 and 33. In particular, let $\mathcal{P}$ be the property of containing a nontrivial three-term arithmetic progression, and note that if A does not have $\mathcal{P}$, then

$$\mathbb{E}_{x,d}A(x)A(x+d)A(x+2d) \leq \frac{|A|}{|G|^2},$$

so

$$\mathbb{E}_{x,d}A(x)A(x+d)A(x+2d) < \frac{\alpha^3}{2}$$

as long as $n > 2\log(2\alpha^{-1})$. $\qquad\square$

5.3 Behrend-type bound for Roth's theorem in four variables over finite field vector spaces

In this section, we show that for Roth's theorem in four variables $x + y + z = 3w$, we have a much better quantitative bound from almost periodicity, matching the form of Behrend's lower bound. Notice that the number of solutions to $x + y = 2z$ can be rewritten as $|G|^2\mathbb{E}_{x,y}A(x)A(y)(2 \cdot A)(x+y)$. However, the number of solutions to $x+y+z = 3w$ is at least $|G|^2\mathbb{E}_{x,y}A(x)(A+A)(y)(3\cdot A)(x+y)$. The main difference between Roth's theorem in four (or more) variables and Roth's theorem in three variables is that in four variables, we get to replace one copy of A with $A + A$. For an unstructured set A with reasonably large density, $A + A$ is expected to have constant density in G, which makes the convolution $(A + A) * A$ much simpler to understand. In particular, since $A + A$ is dense, $(A + A) * A$ has order α. To obtain an almost periodicity result of the right order on $(A + A) * A$, we can take ϵ to be small constant, noting that upper bounds on $(A + A) * A$ can be obtained trivially by majorizing $A + A$ by G, losing only a constant factor. This allows us to obtain the much better quantitative bound with a simpler proof.

Theorem 35. *Let p be a prime coprime to 6. Let A be a subset of $\mathbb{F}_p^n$ with density α. If $\mathbb{E}_{x,y}A(x)(A + A)(y)(3 \cdot A)(x + y) < \frac{\alpha^2}{16}$, we can find a subspace V of codimension at most $C(\log C\alpha^{-1})^4$ such that there exists a translate $V + x$ of V where $\frac{|A\cap(V+x)|}{|V|} \geq \frac{3}{2}\alpha$.*

Proof. The proof follows along the lines of the proof of Theorem 33. However, we get extra saving here due to the fact that we can expect $(A + A)$ to have constant density. By the hypothesis,

$$\mathbb{E}_{x,y}A(x)(A + A)(y)(3 \cdot A)(x + y) = \mathbb{E}_x A * (A + A)(x)(3 \cdot A)(x) < \frac{\alpha^2}{16}.$$

Apply Theorem 24 with $p = \log_2(\alpha^{-1})$, we get a subspace V of codimension at most $Cp\epsilon^{-2}(\log \alpha^{-1})(\log \epsilon^{-1} + \log \alpha^{-1})^2$ such that

$$\|A * (A + A) * m_V - A * (A + A)\|_p \leq \epsilon\sqrt{\alpha}\|A * (A + A)\|_{p/2}^{1/2} + \epsilon^2\alpha.$$

Since $\|(A + A) * A\|_{p/2} \leq \|G * A\|_{p/2} = \alpha$,

$$\|A * (A + A) * m_V - A * (A + A)\|_p \leq 2\epsilon\alpha.$$

Let $c = 1/12$. If there exists x such that $A * m_V(x) \geq (1 + c)\alpha$, we are done. Assume otherwise that $\|A * m_V\|_\infty < (1+c)\alpha$. Since $\mathbb{E}_x A * m_V(x) = \alpha$, this means that $A * m_V(x) \geq \frac{3\alpha}{4}$ for at least a $\frac{3}{4}$-fraction of $x \in G$.

If $|A + A| \geq \frac{|G|}{2}$, then for all x,

$$(A + A) * A * m_V(x) \geq \frac{1}{4}\frac{3\alpha}{4} > \frac{\alpha}{8},$$

so

$$\mathbb{E}_x(A+A) * A * m_V(x)(3 \cdot A)(x) > \frac{\alpha^2}{8}.$$

By Holder's inequality,

$$
\begin{aligned}
&|\mathbb{E}_x((A+A) * A - (A+A) * A * m_V(x))(3 \cdot A)(x)| \\
&\leq \|(A+A) * A - (A+A) * A * m_V\|_p \|3 \cdot A\|_{p/(p-1)} \\
&\leq 2\epsilon\alpha\alpha^{1-1/p}
\end{aligned}
$$

By choosing $\epsilon = \frac{1}{64}$, we get

$$|\mathbb{E}_x((A+A) * A - (A+A) * A * m_V(x))(3 \cdot A)(x)| \leq \frac{\alpha^2}{16},$$

so

$$\mathbb{E}_x(A+A) * A(x)(3 \cdot A)(x) \geq \frac{\alpha^2}{8} - \frac{\alpha^2}{16} = \frac{\alpha^2}{16},$$

which contradicts the assumption.

Consider the case $|A+A| < \frac{|G|}{2}$. Observe the crucial property that for $x \in -A$,

$$(-A - A) * A(x) = \mathbb{E}_y(-A - A)(x - y)A(y) = \mathbb{E}_y A(y) = \alpha.$$

The intuition is that if $|A+A|$ is small, then we expect $(-A - A) * A$ to be much smaller than α. Indeed, smoothing out by the subspace of almost periods V, we get $(-A - A) * A * m_V$ is much smaller than α, assuming no density increment on a translate of V. Since $(-A - A) * A * m_V$ is close to $(-A - A) * A$, we get a contradiction by taking an inner product against A, using Holder's inequality as usual. We now go back to the rigorous argument.

We apply the almost periodicity result, Theorem 24, with $\epsilon = 1/16$ to get

$$\|(-A - A) * A * m_V - (-A - A) * A\|_p \leq \frac{\alpha}{8}.$$

By Holder's inequality

$$|\mathbb{E}_x((-A - A) * A * m_V(x) - (-A - A) * A(x))(-A)(x)| \leq \frac{\alpha}{8}\alpha^{1-1/p} \leq \frac{\alpha^2}{4}.$$

Hence,

$$
\begin{aligned}
\mathbb{E}_x(-A - A) * A * m_V(x)(-A)(x) &\geq \mathbb{E}_x(-A - A) * A(x)(-A)(x) - \frac{\alpha^2}{4} \\
&= \alpha^2 - \frac{\alpha^2}{4} \\
&= \frac{3\alpha^2}{4},
\end{aligned}
$$

where the first equality follows from our previous observation that $(-A - A) * A(x) = \alpha$ for $x \in -A$. However, if $\|A * m_V\|_\infty \leq (1+c)\alpha$ then

$$\|(-A - A) * A * m_V\|_\infty \leq \|A * m_V\|_\infty \|(-A - A)\|_1 \leq \frac{1+c}{2}\alpha = \frac{13\alpha}{24},$$

so

$$\mathbb{E}_x(-A - A) * A * m_V(x)(-A)(x) \leq \frac{13\alpha^2}{24},$$

contradicting the above inequality. $\qquad\square$

Corollary 36. *The largest subset of $\mathbb{F}_p^n$ with no nontrivial solution to $x + y + z = 3w$ has size at most $n \exp(-c(\log n)^{1/4})$.*

Proof. This follows from Theorem 32 and 35. In particular, let $\mathcal{P}$ be the property of containing a nontrivial solution to $x + y + z = 3w$, and note that if A does not have $\mathcal{P}$, then

$$\mathbb{E}_{x,y} A(x)(A+A)(y)(3 \cdot A)(x+y) \leq \frac{|A|}{|G|^2},$$

so

$$\mathbb{E}_{x,y} A(x)(A+A)(y)(3 \cdot A)(x+y) < \frac{\alpha^2}{16}$$

as long as $n > \log(16\alpha^{-1})$. $\qquad\square$

6 Roth's theorem in general abelian groups

6.1 The density increment strategy

We use a similar density increment strategy to prove Roth's theorem over general abelian groups. The technical details in the general setting are much more complicated than those in the finite field vector space model. In general abelian groups, subspaces are replaced by Bohr sets, which are more complicated due to their non-homogeneity. When dealing with Bohr sets, it is crucial to localize on different Bohr sets at appropriate scales to make use of the regularity condition of Bohr sets, which allows us to treat Bohr sets as subgroups in a certain way. For this reason, the density increment step needs to take into account the different scales that we need in order to run the argument. Again, we follow [6].

Theorem 37. *Let G be an abelian group of size n. Let $\mathcal{P}$ be a property which is translation invariant and upward closed. Let w, m, μ, k, C, C', c be constants. Assume that the following holds. Let A be a subset of a regular Bohr set $B = B(\Gamma, \rho)$ with density α in B. Let A' be a subset of A such that $A' \subseteq B_\nu$, where B_ν is regular and $\nu d \leq 10^{-4}$, and the density of A' in B_ν is at least α. Assume that either $\mathcal{P}$ holds, or $|B(\Gamma, \rho)| < C\alpha^{-w}$, or there exists a Bohr set $B''(\Gamma'', \rho'')$ of codimension bounded by*

$$d + C\alpha^\mu (\log C\alpha^{-1})^m$$

and radius

$$\rho'' \geq c\rho\nu \exp(-C'(\log C\alpha^{-1})^k)$$

and x such that $A * m_{B''}(x) \geq (1+c)\alpha$.
Then any set A without $\mathcal{P}$ must have size at most $C_{37} n (\log \log n)^{(m+1)/\mu} (\log n)^{-1/\mu}$.

Proof. Before starting the proof, we remark that there are many constants involved which are not quantitatively important. The reader only needs to keep in mind that all of them are absolute constants that depend only on the constants given in the theorem statement in an appropriate way. The proof follows from a simple iteration of the assumption. However, to match the situation in the assumption, from a dense subset of a Bohr set, we need to find a pair of Bohr sets with one being smaller, where the set is dense in both sets in the pair. Let $c'', \tilde{c}$ be small positive constants depending on c and α to be chosen later. Let c' be a positive constant chosen so that $(1+c)(1-c') > 1$. Let $c'', \tilde{c}$ be positive constants depending on c' and α. We first guarantee the existence of this pair of Bohr sets.

Claim 38. There exists a choice of $c'', \tilde{c}, c'$ such that given A of density α in B, either we can find two Bohr sets $B_1 = B_{\delta_1}, B_2 = B_{\delta_2}$ such that B_1, B_2 are regular, $\tilde{c}/2 \leq \delta_1 d \leq \tilde{c}$ and $\tilde{c}\delta_1/2 \leq \delta_2 d \leq \tilde{c}\delta_1$, and $A * m_{B_1}(x) \geq (1-c')\alpha$ and $A * m_{B_2}(x) \geq (1-c')\alpha$; or we can find increment $\max\{\|A * m_{B_1}\|_\infty, \|A * m_{B_2}\|_\infty\} \geq (1+c'')\alpha$.

Proof of Claim 38. Choose δ_1, δ_2 so that $\tilde{c}/2 \le \delta_1 d \le \tilde{c}$ and $\tilde{c}\delta_1/2 \le \delta_2 d \le \tilde{c}\delta_1$ and $B_{\delta_1}, B_{\delta_2}$ are regular. If there exists x such that $A * m_{B_1}(x) > (1 + c'')\alpha$ or $A * m_{B_2}(x) > (1 + c'')\alpha$ then we are done. Assume otherwise that $A * m_{B_1}(x) \le (1 + c'')\alpha$ and $A * m_{B_2}(x) \le (1 + c'')\alpha$ for all x. By Proposition 10,

$$\mathbb{E}_{x \in B} A * m_{B_1}(x) \ge \alpha - 80\tilde{c} \ge (1 - \frac{80\tilde{c}}{\alpha})\alpha.$$

Let $c' = \frac{c''}{3} + \frac{320\tilde{c}}{3\alpha}$ and choose $c'', \tilde{c}$ sufficiently small so that $(1 - c')(1 + c) > 1$. It is easy to check that the choice of c' guarantees that at least a $\frac{3}{4}$-fraction of $x \in B$ satisfies $A * m_{B_1}(x) \ge (1 - c')\alpha$. Similarly, at least a $\frac{3}{4}$-fraction of $x \in B$ satisfies $A * m_{B_2}(x) \ge (1 - c')\alpha$. Thus, there must exists $x \in B$ such that $A * m_{B_1}(x) \ge (1 - c')\alpha$ and $A * m_{B_2}(x) \ge (1 - c')\alpha$. This finishes the proof of the claim. $\square$

Assume that the set A we start with does not have $\mathcal{P}$, and assume that $|G| > C\alpha^{-w}$. By Claim 38, we either find a smaller Bohr set where we have density increment, or we can find a pair of Bohr sets where A has relative density at least $(1 - c')\alpha$ in both sets in the pair, and we can apply the assumption in the theorem statement. This gives a Bohr set B'' and a translate of B'' where the density of A is at least $(1 + c)(1 - c')\alpha$. Then, in both cases, we can find a translate of a Bohr set B'' of codimension at most

$$C\alpha^\mu(\log C\alpha^{-1})^m$$

and radius at least

$$c\rho \exp(-C'(\log C'\alpha^{-1})^k)$$

such that the density of A in an affine translate of B'' is at least $\min\{(1+c)(1-c')\alpha, (1+c'')\alpha\} = (1+\tau)\alpha$ for some absolute constant $\tau > 0$. Let $B_1 = B''$ and $A_1 = A \cap (B_1 + x)$ where the density of A_1 in $B_1 + x$ is at least $(1 + \tau)\alpha$. We then iterate this process.

The number of iterations is bounded above by $\log \alpha^{-1}$ since clearly the density of a set in an affine translate of a Bohr set is at most 1. Moreover the dimension of B_i satisfies the bound

$$d_i \le d_{i-1} + C\alpha_i^{-\mu}(\log C\alpha_i^{-1})^m,$$

and the radius of B_i satisfies the bound

$$\rho_i \ge \bar{c}\rho_{i-1} \exp(-C'(\log C'\alpha_i^{-1})^k)/d_i.$$

Here $1 \ge \alpha_i \ge (1 + \tau)\alpha_{i-1}$. Thus, by summing the geometric series,

$$d_i \le \sum_{j=0}^{i-1} C\alpha_j^{-\mu}(\log C\alpha_j^{-1})^m \le \tilde{C}\alpha^{-\mu}(\log \tilde{C}\alpha^{-1})^m.$$

Let t be the last step before the iteration is forced to terminate. Then $t \le \bar{C}\log(\bar{C}\alpha^{-1})$. The final radius satisfies the bound

$$\rho_t \ge \exp(t \log \bar{c} - \sum_{i=1}^t C'(\log C'\alpha_i^{-1})^k - \sum_{i=1}^t \log d_i) \ge \exp(-\tilde{C}'(\log \tilde{C}'\alpha^{-1})^{k+1}).$$

Hence,

$$|B_t| \ge |G| \exp(-\tilde{C}''\alpha^{-\mu}(\log \tilde{C}''\alpha^{-1})^{m+k+1}).$$

Moreover, we must have

$$|B_t| \le C\alpha^{-w},$$

so

$$n \le C\alpha^{-w} \exp(\tilde{C}''\alpha^{-\mu}(\log \tilde{C}''\alpha^{-1})^{m+k+1}).$$

From this, we can easily check that for n sufficiently large,

$$\alpha \le C_{37} \frac{(\log \log n)^{(m+k+1)/\mu}}{(\log n)^{1/\mu}}.$$

$\square$

6.2 Roth's theorem in general abelian groups

In this subsection, we establish the density increment condition in Theorem 37. The ideas model the proof in the finite field vector space setting, however replacing subspaces with Bohr sets at the appropriate scales. We first give in a simple version following [6], which is simple and clean but involving a wasteful step. This leads to a slightly worse quantitative bound for Roth's theorem. In Subsection 6.2.2, we discuss our attempt to avoid the wasteful step, leading to a better quantitative bound. Exactly the same bound is achieved by an argument of Sanders [35] using almost periodicity coupled with the Katz-Koester transform. Though this proof is longer, the Katz-Koester transform is itself very interesting, and thus we give this proof in Subsection 6.2.3.

6.2.1 The first attempt

In this section, we discuss the density increment lemma for Roth's theorem in general abelian groups following [6], which yields a slightly worse quantitative bound, but provides a good illustration of the ideas and techniques involved.

Theorem 39. *Let $B = B(\Gamma, \rho)$ be a regular Bohr set of dimension d, and let $B' = B_\nu$ be a regular Bohr set where $\nu d \leq 10^{-6}$. Let $A \subseteq B$ and $A' \subseteq B'$ such that $\beta(A) \geq \alpha$ and $\beta'(A') \geq \alpha$. If $\mathbb{E}A(x)A'(x + d)A(x + 2d) < \frac{\alpha^3}{2}\beta\beta'$, we can find a regular Bohr set B'' of codimension at most*

$$d + O(\alpha^{-1}(\log \alpha^{-1})^4)$$

and radius at least

$$c\rho\nu \frac{\alpha^{O(\log \alpha^{-1})}}{d^2},$$

such that there exists a translate $B'' + x$ of B'' where $\frac{|A \cap (B'' + x)|}{|B''|} \geq \frac{9}{8}\alpha$.

Proof. By assumption,

$$\mathbb{E}_{x,d}A(x)A'(x + d)A(x + 2d) = \mathbb{E}_x A * A(x)(2 \cdot A')(x) < \frac{\alpha^3}{2}\beta\beta'.$$

Let $p = \log_2 \alpha^{-1}$. By Proposition 11, $2 \cdot B'$ is regular Bohr set and $(2 \cdot B')_\eta \subseteq B'_{2\eta} \subseteq B_{2\eta\nu}$. Apply Theorem 31 with the sets $A_{30} = A, X_{30} = A, \tilde{B}_{30} = B$, and $B_{30} = 2 \cdot B'$, we get a regular Bohr set B''_0 such that for all $b'' \in B''_0$,

$$\|A * A(b'' + \cdot) - A * A(\cdot)\|_{L^p(2 \cdot B')} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}}\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

For a reason which will be clear in the course of the proof, we let B'' be a regular Bohr set such that $B'' \subseteq (B''_0)_{1/2}$, so $B'' + B'' \subseteq B''_0$. We can choose B'' to satisfy the required bounds on the dimension and radius, and furthermore,

$$\|A * A * m_{B''} * m_{B''} - A * A\|_{L^p(2 \cdot B')} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}}\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

As in the finite field case, the general scheme for the proof is as follows. We first establish a lower bound on $A * A * m_{B''} * m_{B''}$ assuming no density increment. Then, assuming $\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2}$ is small, we obtain that $\mathbb{E}_x A * A(x)(2 \cdot A')(x) \approx \mathbb{E}_x A * A * m_{B''} * m_{B''}(x)(2 \cdot A')(x) \approx \alpha^3\beta\beta'$. If otherwise $\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2}$ is large, we directly obtain density increment by proving $\|A * A * m_{B''} * m_{B''}\|_{L^p(B')}$ is large. The only reason we use two copies of $m_{B''}$ in the convolution is that in the finite field vector space case, $m_V * m_V = m_V$,

however, this is not true for Bohr sets, but can easily be achieved by reducing the radius of the Bohr set by a constant factor.

Let $c = 1/8$. Assume that $\|A * m_{B''}\|_\infty \leq (1 + c)\alpha$. Then for all x, y,

$$((1 + c)\alpha - A * m_{B''}(x))(((1 + c)\alpha - A * m_{B''}(y)) \geq 0.$$

Furthermore, by Proposition 10, we have

$$\mathbb{E}_{t \in B + B''} A * m_{B''}(t) \geq \frac{\alpha}{1 + 80\nu d},$$

and for $x \in 2 \cdot B' \subseteq B_{2\nu}$,

$$\mathbb{E}_{t \in B + B''} A * m_{B''}(x - t)$$
$$\geq \frac{\alpha}{1 + 162d \cdot 4\nu} - 400d \cdot 6\nu$$
$$\geq \frac{15\alpha}{16}.$$

Thus,

$$A * A * m_{B''} * m_{B''}(x) = \mathbb{E}_t A * m_{B''}(t) A * m_{B''}(x - t)$$
$$\geq \frac{|B + B''|}{|G|} \mathbb{E}_{t \in B + B''} A * m_{B''}(t) A * m_{B''}(x - t)$$
$$\geq (1 + c)\alpha\beta \mathbb{E}_{t \in B + B''} [A * m_{B''}(t) + A * m_{B''}(x - t) - (1 + c)\alpha]$$
$$\geq (1 + c)\alpha\beta \cdot \left(\frac{15}{8}\alpha - \frac{9}{8}\alpha\right)$$
$$\geq \frac{3}{4}\alpha^2\beta.$$

By Holder's inequality,

$$\frac{|G|}{|B'|} \mathbb{E}_x (A * A * m_{B''} * m_{B''} - A * A)(x)(2 \cdot A')(x) \leq \|A * A * m_{B''} * m_{B''} - A * A\|_{L^p(B')} \|2 \cdot A'\|_{L^q(B')}.$$

Hence,

$$\mathbb{E}(A * A * m_{B''} * m_{B''} - A * A)(2 \cdot A') \leq \beta' \left(\epsilon\sqrt{\alpha\beta}\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2} + \epsilon^2\alpha\beta\right)\alpha^{1/q}$$
$$\leq 2\beta'\alpha\epsilon(\sqrt{\alpha\beta}\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2} + \epsilon\alpha\beta).$$

If $\|A * A\|_{L^{p/2}(2 \cdot B')} \leq 10\alpha^2\beta$, picking $\epsilon = \alpha^{1/2}/4$, we get

$$\mathbb{E}(A * A * m_{B''} * m_{B''} - A * A)(2 \cdot A') \leq \alpha^3\beta\beta'/4.$$

Thus,

$$\mathbb{E}_x A * A(x)(2 \cdot A')(x) \geq \mathbb{E}_x A * A * m_{B''} * m_{B''}(x)(2 \cdot A')(x) - \frac{\alpha^3\beta\beta'}{4}$$
$$\geq \frac{3}{4}\alpha^2\beta\mathbb{E}_x(2 \cdot A')(x) - \frac{\alpha^3\beta\beta'}{4}$$
$$= \frac{\alpha^3\beta\beta'}{2}.$$

Otherwise, $\|A * A\|_{L^{p/2}(2 \cdot B')} > 10\alpha^2\beta$, so $\|A * A\|_{L^p(2 \cdot B')} \geq \|A * A\|_{L^{p/2}(2 \cdot B')} > 10\alpha^2\beta$. Choosing $\epsilon = \alpha^{1/2}/4$,

$$\|A * A * m_{B''} * m_{B''}\|_{L^p(2 \cdot B')} \geq \|A * A\|_{L^p(2 \cdot B')} - \epsilon\sqrt{\alpha\beta}\|A * A\|_{L^{p/2}(2 \cdot B')}^{1/2} - \epsilon^2\alpha\beta$$
$$\geq \|A * A\|_{L^p(2 \cdot B')}/2$$
$$> 2\alpha^2\beta.$$

If $\|A * m_{B''}\|_\infty \leq (1+c)\alpha$ then noticing that $A * m_{B''}$ is supported on $B + B''$, we have

$$\|A * A * m_{B''} * m_{B''}\|_{L^p(2 \cdot B')}^p = \frac{1}{|B'|}\sum_{x \in B'}(\mathbb{E}_t A * m_{B''}(t)A * m_{B''}(x-t))^p$$
$$\leq \frac{1}{|B'|}|B'|\left(\frac{|B+B''|}{|G|}(1+c)^2\alpha^2\right)^p$$
$$\leq (2\beta\alpha^2)^p,$$

contradiction. $\qquad\square$

Corollary 40. *The largest subset of $\mathbb{Z}_n$ with no nontrivial three-term arithmetic progression has size at most $Cn(\log\log n)^7/\log n$.*

Proof. Let $\mathcal{P}$ be the property of not containing a nontrivial three-term arithmetic progression. If $\frac{|A'|}{|B'|} > \frac{9}{8}\alpha$ then we immediately obtain density increment. Otherwise, assume that A contains no nontrivial three-term arithmetic progression. Then

$$\mathbb{E}_{x,d}A(x)A'(x+d)A(x+2d) \leq \frac{|A'|}{|G|^2} \leq \frac{9}{8}\frac{\alpha\beta'}{|G|} < \frac{\alpha^3\beta\beta'}{2},$$

as long as $|B| > 3\alpha^{-2}$. We get the desired bound from Theorem 37. $\qquad\square$

6.2.2 Improving the number of $\log\log$ factors

We would like to improve the quantitative bound in the previous subsection and remove one $\log\log$ factor. This coincides with the best known bound using almost periodicity for Roth's theorem by Sanders [35].

To get the improvement, we will refine our bound on the radius of the Bohr set in each step of the increment. We used a clean form of the bootstrapping result (Theorem 31) in the previous subsection, but this leads to a large loss in the radius of the Bohr set. By using instead Theorem 30 and κ-regular Bohr sets, we can get a better quantitative bound. This is the only place that we need our nonstandard notion of κ-regular Bohr set.

Theorem 41. *Let $B = B(\Gamma, \rho)$ be a regular Bohr set of dimension d, and let $B' = B_\nu$ be a regular Bohr set where $\nu d \leq 10^{-6}$. Let $A \subseteq B$ and $A' \subseteq B'$ such that $\beta(A) \geq \alpha$ and $\beta'(A') \geq \alpha$. If $\mathbb{E}_{x,d}A(x)A'(x+d)A(x+2d) < \frac{\alpha^3}{2}\beta\beta'$, we can find a regular Bohr set B'' of codimension at most*

$$d + O(\alpha^{-1}(\log\alpha^{-1})^4)$$

and radius at least

$$c\rho\nu\left(\frac{\alpha}{d}\right)^{O(1)},$$

such that there exists a translate $B'' + x$ of B'' where $\frac{|A \cap (B''+x)|}{|B''|} \geq \frac{17}{16}\alpha$.

Proof. The proof follows the lines of the proof given in the previous subsubsection. We will highlight the main differences.

Let $p = \log_2 \alpha^{-1}$. Apply Theorem 30 with the sets $A_{30} = A, X_{30} = A, \tilde{B}_{30} = B, B_{30} = 2 \cdot B'$, and $\tau_{30} = \tau$ to be chosen later, we get a regular Bohr set B_0'' such that for all $b'' \in B_0''$,

$$\|A * A(b'' + \cdot) - A * A(\cdot)\|_{L^p(2 \cdot B')} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * A\|_{L^{p/2}(\bar{\beta}')}^{1/2} + \epsilon^2 \frac{|A|}{|G|},$$

where $\bar{\beta}'$ is a measure with mass $m_{\bar{\beta}'}(x) = |G| \frac{(2 \cdot B')_{1+2\tau \log(\alpha\epsilon)^{-1}}(x)}{|2 \cdot B'|}$. Let B'' be a regular Bohr set such that $B'' \subseteq (B_0'')_{1/2}$, so $B'' + B'' \subseteq B_0''$. Then

$$\|A * A * m_{B''} * m_{B''} - A * A\|_{L^p(2 \cdot B')} \leq \epsilon \frac{|A|^{1/2}}{|G|^{1/2}} \|A * A\|_{L^{p/2}(\bar{\beta}')}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

In the case $\|A * A\|_{L^{p/2}(\bar{\beta}')} \leq 10\alpha^2\beta$, we proceed exactly as in the proof of Theorem 39 to get a translate of B'' where $\frac{|A \cap (B'' + x)|}{|B''|} \geq \frac{9}{8}\alpha$.

Consider the case $\|A * A\|_{L^{p/2}(\bar{\beta}')} > 10\alpha^2\beta$, and assume that $\|A * m_{B''}\|_\infty \leq \frac{9}{8}\alpha$. First, consider the case $\|A * A\|_{L^p(2 \cdot B')} \geq \|A * A\|_{L^{p/2}(2 \cdot B')} > \|A * A\|_{L^{p/2}(\bar{\beta}')}/3$. Choose $\epsilon = \alpha^{1/2}/16$, we get

$$\|A * A * m_{B''} * m_{B''}\|_{L^p(2 \cdot B')} \geq \|A * A\|_{L^p(2 \cdot B')} - \epsilon\sqrt{\alpha\beta}\|A * A\|_{L^{p/2}(\bar{\beta}')}^{1/2} - \epsilon^2\alpha\beta$$
$$\geq 2\|A * A\|_{L^p(2 \cdot B')}/3$$
$$> 2\alpha^2\beta.$$

We arrive at a contradiction as in the proof of Theorem 39.

Otherwise, we have $\|A * A\|_{L^{p/2}(2 \cdot B')} \leq \|A * A\|_{L^{p/2}(\bar{\beta}')}/3$. This is the only case where we do not arrive directly at density increment, but it gives a significant increment on the norm of $A * A$ by passing to only a slightly larger Bohr set. Let $B_1' = B'_{1+2\tau \log(\alpha\epsilon)^{-1}}$. Let $\kappa_1 = 2\tau \log(\alpha\epsilon)^{-1}$. Then B_1' is κ_1-regular and

$$\|A * A\|_{L^{p/2}(2 \cdot B_1')} = \left(\frac{|B'|}{|B_1'|}\right)^{2/p} \|A * A\|_{L^{p/2}(\bar{\beta}')} \geq \max\{2\|A * A\|_{L^{p/2}(2 \cdot B')}, 5\alpha^2\beta\}.$$

We want to replace B' with B_1' and run through the same argument. Note that

$$\beta_1'(A') \geq \alpha \frac{|B'|}{|B_1'|} \geq \frac{\alpha}{1 + 320\tau \log(\alpha\epsilon)^{-1}}.$$

Since Theorem 30 applies to κ-regular Bohr sets for $\kappa \leq 10^{-4}/d$, as long as $\kappa_1 \leq 10^{-4}/d$ and $(1 + \kappa_1)\nu d \leq 10^{-4}$, we can repeat the same argument as above with B' replaced by B_1' and α replaced by $\alpha_1 = \frac{\alpha}{1 + 320\tau \log(\alpha\epsilon)^{-1}}$.

Iterate the argument above, we obtain Bohr sets $B_i' = B'_{\prod_{j=0}^i (1+2\tau \log(\alpha_i\epsilon)^{-1})}$ which are κ_i-regular for $\kappa_i = \prod_{j=0}^i (1 + 2\tau \log(\alpha_i\epsilon)^{-1}) - 1$, and

$$\|A * A\|_{L^{p/2}(2 \cdot B_i')} \geq 2^{i-1}\|A * A\|_{L^{p/2}(2 \cdot B_1')} \geq 2^{i-1} \cdot 5\alpha^2\beta,$$

and furthermore

$$\beta_i'(A' \cap B_i') \geq \alpha_i = \frac{\alpha_{i-1}}{1 + 320\tau \log(\alpha_i\epsilon)^{-1}}.$$

We terminate the process when $\kappa_i d > 10^{-4}$, or $(1+\kappa_i)\nu d > 10^{-4}$, or $\|A*A\|_{L^{p/2}(2 \cdot B_i')} \geq \|A*A\|_{L^{p/2}(2 \cdot B_{i+1}')}/2$.

41

Note that $\|A * A\|_\infty \leq \alpha\beta$, we get that we must terminate after at most $C' \log(1/\alpha)$ iterations. Choose $\tau = c'(\log(\alpha\epsilon)^{-3}d)^{-1}$. For all $i \leq C' \log(1/\alpha)$, we have by induction that

$$\alpha_i \geq \alpha_{i-1}(1 - 320\tau \log(\alpha_i\epsilon)^{-1}) \geq \alpha_{i-1}(1 - 640\tau \log(\alpha\epsilon)^{-1}) \geq (1 - 640\tau \log(\alpha\epsilon)^{-1})^i\alpha,$$

since

$$(1 - 640\tau \log(\alpha\epsilon)^{-1})^i\alpha \geq \alpha^2.$$

Thus $\log(\alpha_i\epsilon)^{-1} \leq 2\log(\alpha\epsilon)^{-1}$ for all $i \leq C' \log(1/\alpha)$. Furthermore

$$\prod_{j=0}^{i}(1 + 2\tau \log(\alpha_i\epsilon)^{-1}) \leq (1 + 4\tau \log(\alpha\epsilon)^{-1})^{C \log(1/\alpha)} < 1 + 10^{-4}/d < 2,$$

so $\kappa_i d \leq 10^{-4}$ and $(1 + \kappa_i)\nu d \leq 10^{-4}$ for all $i \leq C' \log(1/\alpha)$. Thus, upon termination, we must have $\|A * A\|_{L^{p/2}(2 \cdot B_i')} \geq \|A * A\|_{L^{p/2}(2 \cdot B_{i-1}')}/2$.

In this case, by the argument above, we immediately get a density increment

$$\frac{|A \cap B_i''|}{|B_i''|} \geq \frac{9}{8}\alpha_i > \frac{9}{8}\alpha(1 - 640\tau \log(\alpha\epsilon)^{-1})^{C \log(1/\alpha)} \geq \frac{17}{16}\alpha$$

on a translate of B_i'', where the codimension of B_i'' is at least

$$d + O(\alpha_i^{-1}(\log \alpha_i^{-1})^4) = d + O(\alpha^{-1}(\log \alpha^{-1})^4)$$

and the radius of B_i'' is at least

$$c\rho\tau \frac{\alpha^{O(1)}}{d^2} = c\rho \left(\frac{\alpha}{\log(\alpha^{-1})d}\right)^{O(1)}.$$

This completes the argument. $\qquad\square$

This improved density increment step immediately gives an improved bound on Roth's theorem.

Corollary 42. *The largest subset of $\mathbb{Z}_n$ with no nontrivial three-term arithmetic progression has size at most $Cn(\log \log n)^6/\log n$.*

6.2.3 Improving the number of $\log \log$ factors via the Katz-Koester transform

In this subsection, we exposit a different way to arrive at the same bound as in the previous subsection from the work of Sanders [35]. In fact, this result is key in Sanders' proof, since he used a much weaker version of almost periodicity. While it does not give an improvement on the quantitative bound, it uses an interesting result, the Katz-Koester transform. The main purpose of this subsection is to introduce the Katz-Koester transform, following [35]. The main idea of the Katz-Koester transform is to replace sets A, A' by sets L and S such that $L + S \subseteq A + A'$, but L has constant density while S has density not too small. As observed in Subsection 5.3, constant density sets allow us to run almost periodicity arguments much more easily. The core of the argument is the following simple observation. Given L and S such that $L + S \subseteq A + A'$, for any $x \in G$, $(L \cup (A - x)) + (S \cap (x - A')) \subseteq A + A'$. This increases the density of L, and as long as the density of S does not decrease too much, we can iterate this argument to obtain the desired constant density set L. Thus, we aim to find x such that $|L \cup (A - x)| = |L| + |(A - x)\backslash L|$ is large while $|S \cap (x - A')|$ is not too small. If A is not very structured, we expect many x to satisfy $|(A - x)\backslash L|$ is large. Moreover, by averaging, few x satisfy that $S \cap (x - A')$ has density decreasing by more than a factor of $2/\alpha'$. Thus, we must be able to find x for which $|(A - x)\backslash L|$ is large and $|S \cap (x - A')|$ is not too small.

Before going into the details of the Katz-Koester transform, we briefly comment on how it leads directly to more efficient almost periodicity results. Almost periodicity gives $\|A' * A * m_{B''} - A' * A\|_{L^p(B)} \leq$

$\epsilon^2 \alpha \beta' + \epsilon \sqrt{\alpha \beta'} \|A' * A\|_{L^{p/2}(\bar\beta)}^{1/2}$. In using almost periodicity to prove Roth's theorem, we needed to separate the cases where $\|A' * A\|_{L^{p/2}(\bar\beta)}$ is large or small, with the case where it is small being easier. If A', A both have density α, it is crucial that $\|A' * A\|_{L^{p/2}(\bar\beta)}$ is much smaller than the trivial bound (on the supremum of $A' * A$) and closer to the bound when A' and A behave like random sets. However, if we replace A' by S and A by L where L has constant density (using $S * L$ as a lower bound for $A' * A$), we get no saving in using the trivial bound or the bound when S and L behave like random sets. Thus, we remove the need to estimate $\|S * L\|_{L^{p/2}(\bar\beta)}$ and use the trivial bound instead, allowing for an easier argument.

We now state the iterative step to obtain the Katz-Koester transform.

Lemma 43. *Let B, B', B'' be regular Bohr sets. Assume that B, B', B'' has dimension bounded by d, B has radius ρ, $B' \subseteq B_{\nu'}$ and $B'' \subseteq B'_{\nu''}$ where $10^{-8} \leq \nu'd, \nu''d \leq 10^{-4}$. Let $A \subseteq B$, $A' \subseteq B'$ be such that $\beta(A) \geq \alpha$ and $\beta'(A') \geq \alpha$. Let $L \subseteq B$, $S \subseteq B''$ be such that $\beta(L) = \lambda$ and $\beta''(S) = \sigma$. Assume that $\lambda \leq c$ for an absolute constant c. Then one of the following must hold.*

1. *There is a regular Bohr set B''' of codimension at most*

$$d + O(\alpha^{-1} \log \alpha^{-1})$$

   and radius at least

$$\rho \left(\frac{\alpha}{d}\right)^{O(1)}$$

   such that $\|A * m_{B'''}\|_\infty \geq \alpha(1 + c)$.

2. *There is a set $L' \subseteq B$ and $S' \subseteq B''$ with $\beta'(L') \geq \lambda + \alpha/4$ and $\beta''(S') \geq \alpha\sigma/2$ such that for all $x \in G$,*

$$L' * S'(x) \leq L * S(x) + A * A'(x).$$

The proof of the lemma follows from the heuristics we discussed above.

Lemma 44. *Using the same set-up as Lemma 43, if more than $\alpha|B'|/8$ elements $x \in B'$ has*

$$L * (-A)(x) \geq \alpha\beta/2,$$

then there is a regular Bohr set $B''' \subseteq B_{\nu'''}$ with $\nu'''d \leq 10^{-4}$, having codimension at most

$$d + O(\alpha^{-1} \log \alpha^{-1})$$

and radius at least

$$\rho \left(\frac{\alpha}{d}\right)^{O(1)}$$

such that $\|A * m_{B'''}\|_\infty \geq \alpha(1 + c)$.

Proof. Let $T = \{x \in B' : L * (-A)(x) \geq \alpha\beta/2\}$ and assume that $|T| \geq \alpha|B'|/8$. We have

$$\mathbb{E}_x L * (-A)(x)T(x) \geq \frac{|T|}{|G|}\alpha\beta/2.$$

Since we work with local information, we need to normalize $-A$. Observe that

$$\mathbb{E}_x L * B(x)T(x) = \mathbb{E}_{x,y}L(y)T(x)[B(x-y) - (B+B')(x-y)] + \mathbb{E}_{x,y}L(y)T(x)(B+B')(x-y)$$
$$= \mathbb{E}_{x,y}L(y)T(x)[B(x-y) - (B+B')(x-y)] + \lambda\beta|T|/|G|,$$

since $(B + B')(x - y) = 1$ for all $y \in L \subseteq B, x \in T \subseteq B'$. By regularity,

$$
\begin{aligned}
|\mathbb{E}_{x,y} L(y) T(x) [B(x - y) - (B + B')(x - y)]| &\leq \mathbb{E}_{x,z} T(x) |B(z) - (B + B')(z)| \\
&= \frac{|T|}{|G|} \frac{|(B + B') \backslash B|}{|G|} \\
&\leq 80 d\nu\beta |T|/|G| \\
&\leq \frac{1}{8} \beta |T|/|G|.
\end{aligned}
$$

Thus,

$$
\mathbb{E}_x L * [(-A) - \alpha B](x) T(x) \geq \frac{|T|}{|G|} \frac{\alpha\beta}{4},
$$

assuming λ is small enough. Then

$$
\sum_\chi \hat{L}(\chi) (\widehat{-A} - \alpha\hat{B})(\chi) \hat{T}(\chi) \geq \frac{|T|}{|G|} \frac{\alpha\beta}{4}.
$$

By Cauchy-Schwartz,

$$
\sum_\chi |(\widehat{-A} - \alpha\hat{B})(\chi) \hat{T}(\chi)|^2 \geq \frac{(\alpha\beta |T|/4|G|)^2}{\lambda\beta} = \frac{\alpha^2\beta |T|^2}{16\lambda |G|^2}.
$$

Furthermore,

$$
\begin{aligned}
\sum_{\chi \notin \Delta_{\alpha^{1/2}}(T)} |(\widehat{-A} - \alpha\hat{B})(\chi)|^2 |\hat{T}(\chi)|^2 &\leq \alpha \frac{|T|^2}{|G|^2} \sum_\chi |\widehat{-A}(\chi) - \alpha\hat{B}(\chi)|^2 \\
&= \frac{\alpha |T|^2}{|G|^2} (\alpha\beta + \alpha^2\beta - 2\alpha(\alpha\beta)\beta) \\
&\leq \frac{2\alpha^2\beta |T|^2}{|G|^2}.
\end{aligned}
$$

Assuming λ is sufficiently small, we get

$$
\sum_{\chi \in \Delta_{\alpha^{1/2}}(T)} |(\widehat{-A} - \alpha\hat{B})(\chi)|^2 |\hat{T}(\chi)|^2 > \frac{2\alpha^2\beta |T|^2}{|G|^2}.
$$

Thus,

$$
\sum_{\chi \in \Delta_{\alpha^{1/2}}(T)} |(\widehat{-A} - \alpha\hat{B})(\chi)|^2 > 2\alpha^2\beta.
$$

This non-uniformity in the Fourier coefficients of $\hat{A}$ allows us to get density increment. By Theorem 29, we can find a Bohr set B''' such that $|1 - \chi(x)| \leq \frac{1}{2}$ for all $x \in B'''$ and $\chi \in \Delta_{\alpha^{1/2}}(T)$. Then

$$
\sum_{\chi \in \Delta_{\alpha^{1/2}}(T)} |(\widehat{-A} - \alpha\hat{B})(\chi)|^2 |\widehat{m_{B'''}}(\chi)|^2 \geq \frac{\alpha^2\beta}{2}.
$$

But this can be rewritten as

$$
\mathbb{E}_x |((-A) - \alpha B) * m_{B'''}(x)|^2 \geq \frac{\alpha^2\beta}{2}.
$$

We have

$$\mathbb{E}_x|((-A) - \alpha B) * m_{B'''}(x)|^2$$
$$= \mathbb{E}_x|(-A) * m_{B'''}(x)|^2 + \alpha^2\mathbb{E}_x|B * m_{B'''}(x)|^2 - 2\alpha\mathbb{E}_x(-A) * m_{B'''}(x)B * m_{B'''}(x)$$
$$= \mathbb{E}_x|(-A) * m_{B'''}(x)|^2 - \alpha^2\mathbb{E}_x|B * m_{B'''}(x)|^2 + 2\alpha[\mathbb{E}_xB * m_{B'''}(x)\alpha - \mathbb{E}_xB * m_{B'''}(x)(-A) * m_{B'''}(x)].$$

We have

$$\mathbb{E}_xB * m_{B'''}(x) = \mathbb{E}_x\mathbb{E}_{y\in B'''}B(x - y) = \beta\mathbb{E}_{z\in B, y\in B'''}1 = \beta,$$

and furthermore, $B * m_{B'''}$ is supported on $B + B'''$ and for $x \in B_{1-\nu'''}$,

$$B * m_{B'''}(x) = \mathbb{E}_{y\in B'''}B(x - y) = 1.$$

Since

$$\mathbb{E}_xB * m_{B'''}(x)(-A) * m_{B'''}(x)$$

$$= \frac{1}{|G|}\left[\sum_{x\in B_{1+\nu'''}}(-A) * m_{B'''}(x) - \sum_{x\in B_{1+\nu'''}\backslash B_{1-\nu'''}}(1 - B * m_{B'''}(x))(-A) * m_{B'''}(x)\right],$$

and by Proposition 10,

$$\sum_{x\in B_{1+\nu'''}}(-A) * m_{B'''}(x) = |A|,$$

so

$$|\mathbb{E}_xB * m_{B'''}(x)(-A) * m_{B'''}(x) - \alpha\beta| \leq \frac{1}{|G|}\sum_{x\in B_{1+\nu'''}\backslash B_{1-\nu'''}}(-A) * m_{B'''}(x)$$

$$\leq \frac{|B_{1+\nu'''}\backslash B_{1-\nu'''}|}{|G|}$$

$$\leq 160\beta d\nu'''.$$

Thus

$$|\mathbb{E}_xB * m_{B'''}(x)\alpha - \mathbb{E}_xB * m_{B'''}(x)(-A) * m_{B'''}(x)| \leq 160\beta d\nu'''.$$

Similarly,

$$\alpha^2\mathbb{E}_x|B * m_{B'''}(x)|^2 \geq (1 - 10^{-3})\alpha^2\beta.$$

Assuming $\nu''' < c\alpha/d$, we get

$$\mathbb{E}_x|(-A) * m_{B'''}(x)|^2 \geq \frac{5}{4}\alpha^2\beta.$$

Hence,

$$\|(-A) * m_{B'''}\|_\infty \geq \frac{5}{4}\alpha,$$

since $\mathbb{E}_x(-A) * m_{B'''}(x) = \alpha\beta$.

It is easy to check that B''' can be chosen to satisfy the required bounds on the codimension and radius. $\qquad\square$

Next, we turn to show that a large fraction of elements x satisfy $S * A(x)$ is not too small by a simple averaging argument.

Lemma 45. *Using the same set-up as Lemma 43, at least $\alpha|B'|/4$ elements $x \in B'$ satisfies $S*(-A')(-x) \geq \sigma\alpha/2$.*

Proof. As in Proposition 10

$$\mathbb{E}_x(B' + B'')(-x)S * (-A')(-x) = \frac{|S||A'|}{|G|^2} \geq \sigma\alpha\beta''\beta'.$$

By regularity,

$$\mathbb{E}_x B'(-x)S * (-A')(-x) \geq \sigma\alpha\beta''\beta' - \sigma\beta'' \frac{|(B' + B'')\backslash B'|}{|G|} \geq \frac{\sigma\alpha\beta''\beta'}{1 + 10^{-3}},$$

assuming $\rho''/\rho' \leq cd/\alpha$.

Thus the number of $x \in B'$ with $S * (-A')(-x) \geq \sigma\alpha\beta''/2$ is at least $\alpha|B'|/4$, since $\|S * (-A')\|_\infty \leq \sigma\beta''$. $\qquad\square$

Proof of Lemma 43. Combining Lemma 44 and Lemma 45, we either get increment, or we get at least one element $x \in B'$ such that

$$L * (-A)(-x) \leq \alpha\beta/2,$$

and

$$S * (-A')(x) \geq \sigma\alpha\beta''/2.$$

Define $L' = L \cup (A - x)$ and $S' = S \cap (A' + x)$. Then

$$\beta(L') \geq \lambda + \beta((A - x) \cap B) - \alpha/2 \geq \lambda + \alpha/2 - \frac{|(B + B')\backslash B|}{|B|} \geq \lambda + \alpha/4,$$

and

$$\beta''(S') \geq \alpha\sigma/2.$$

Finally,

$$\begin{aligned}
L' * S'(y) &\leq L * S'(y) + (A - x) * S'(y) \\
&\leq L * S(y) + (A - x) * (A' + x)(y) \\
&\leq L * S(y) + A * A'(y).
\end{aligned}$$

$\qquad\square$

We initiate $L = \emptyset$ and $S = B''$. By iterating Lemma 43 at most $O(\alpha^{-1})$ times, we obtain the Katz-Koester transform.

Theorem 46 (Katz-Koester transform)**.** *Let B, B', B'' be regular Bohr sets. Assume that B, B', B'' has dimension bounded by d, B has radius ρ, $B' \subseteq B_{\nu'}$ and $B'' \subseteq B'_{\nu''}$ where $10^{-8} \leq \nu'd, \nu''d \leq 10^{-4}$. Let $A \subseteq B$, $A' \subseteq B'$ be such that $\beta(A) \geq \alpha$ and $\beta'(A') \geq \alpha$. Then one of the following must hold.*

1. *There is a regular Bohr set B''' of codimension at most*

$$d + O(\alpha^{-1} \log \alpha^{-1})$$

   and radius at least

$$\rho\left(\frac{\alpha}{d}\right)^{O(1)}$$

   such that $\|A * m_{B'''}\|_\infty \geq \alpha(1 + c)$.

2. *There exists $L \subseteq B$ and $S \subseteq B''$ such that $\beta(L) \geq \Omega(1)$, $\beta''(S) \geq \alpha^{O(\alpha^{-1})}$ such that for all y*

$$L * S(y) \leq O(\alpha^{-1})A * A'(y).$$

This is a relativised version of the heuristics we described at the beginning of the subsection.

With this result at hand, we can use almost periodicity to give an alternative proof of Theorem 41. With the same setting as in Theorem 41, we can equivalently write

$$\mathbb{E}_x A * A(x)(2 \cdot A')(x) = \mathbb{E}_x A * (2 \cdot A')(x)(-A)(x).$$

We can find sets L, S as in Theorem 46, or density increment. Assume that we can find L and S. Let $\beta(L) = \lambda, \beta''(S) = \sigma$. Apply Theorem 31 with $p_{31} = \log(1/\alpha)$, $\epsilon_{31} = c(\lambda)$ an absolute constant, $A_{31} = S$, $X_{31} = L$, $\tilde{B}_{31} = B''$ and $B_{31} = B$ (note that $B_\eta \subseteq B''_{\theta\eta}$ for $\theta \leq c/d^3$ which can be assumed to be smaller than α^{-3}), we get a Bohr set B''' such that

$$\|S * L - S * L * m_{B'''} * m_{B'''}\|_{L^p(B)} \leq 2\epsilon^2 \frac{|S|}{|G|} \leq c\lambda\sigma\beta''.$$

If

$$\mathbb{E}_{x \in B} S * L(x)(-A)(x) \geq \lambda\sigma\beta''\alpha/2,$$

we also get a lower bound on $\mathbb{E}A * (2 \cdot A')(x)(-A)(x)$. Assume otherwise, then

$$\mathbb{E}_{x \in B} S * L * m_{B'''} * m_{B'''}(x)(-A)(x) \leq \frac{3\lambda\sigma\alpha\beta''}{4}.$$

But

$$\mathbb{E}_{x \in B} S * L * m_{B'''} * m_{B'''}(x)(-A)(x) = \frac{1}{\beta}\mathbb{E}_x S * L * m_{B'''}(x)(-A) * m_{B'''}(x).$$

Furthermore, by regularity,

$$\mathbb{E}_{x \in B} S * L * m_{B'''}(x) \geq (1 - c)\lambda\sigma\beta'',$$

and

$$\mathbb{E}_{x \in B} m_{B'''} * (-A)(x) \geq (1 - c)\alpha.$$

Hence,

$$\mathbb{E}_{x \in B}(\alpha - m_{B'''} * (-A)(x)) S * L * m_{B'''}(x) \geq \frac{\lambda\sigma\alpha\beta''}{4}.$$

Assuming that $\|m_{B'''} * (-A)\|_\infty \leq (1 + c)\alpha$, we get

$$\mathbb{E}_{x \in B} |\alpha - m_{B'''} * (-A)(x)| \leq 2c\alpha.$$

Thus

$$|\mathbb{E}_{x \in B}(\alpha - m_{B'''} * (-A)(x)) S * L(x)| \leq \|S * L\|_\infty \mathbb{E}_{x \in B} |\alpha - m_{B'''} * (-A)(x)| \leq \sigma\beta'' \cdot 2c\alpha.$$

Choosing c appropriately (depending on the absolute constant lower bound for λ), we arrive at the desired contradiction.

Let us now consider the quantitative bound coming from this argument. If we can find sets L and S, the Bohr set where we get increment as codimension at most $d + O(\alpha^{-1}(\log \alpha^{-1})^4)$ (since $\sigma \geq \alpha^{O(\alpha^{-1})}$), and radius at least $c\rho \left(\frac{1}{d \log \alpha^{-1}}\right)^{O(1)}$. In our application, d is certainly on the order of $\alpha^{-\Theta(1)}(\log \alpha^{-1})^{\Theta(1)}$. If we instead get increment within the Katz-Koester transform, then we get increment on a Bohr set of codimension at most $d + O(\alpha^{-1} \log \alpha^{-1})$ and radius at least $c\rho \left(\frac{\alpha}{d}\right)^{O(1)}$. Combining these gives exactly the same quantitative bound for Roth's theorem as in the previous subsection. We note that the claimed bound in [35] is $n(\log \log n)^5 / \log n$. However, we think that this is not possible. At the first step of the iteration, the Bohr set has dimension $\Omega(\alpha^{-1}(\log \alpha^{-1})^4)$. In each subsequent step, the radius of this Bohr set can shrink by a factor of $\alpha^{-\Omega(1)}$. Thus, after $\log \alpha^{-1}$ iterations, the density of the Bohr set can be as small as $O(\exp(-\alpha^{-1}(\log \alpha^{-1})^6))$.

6.3 Behrend-type bound for Roth's theorem in four variables over general abelian groups

In this subsection, we show that for Roth's theorem in four variables, we again have a Behrend-type upper bound on the density of sets avoiding nontrivial solutions to the equation $x + y + z = 3w$. The rationale behind the improved quantitative bound is due to the appearance of $A + A$ in the convolutions, which is expected to be dense in the Bohr set, exactly as in the finite field vector space case of Subsection 5.3.

Theorem 47. *Let $B = B(\Gamma, \rho)$ be a regular Bohr set of dimension d, and let $B' = B_\nu$ where $\nu \leq 10^{-6}/d$. Let $A \subseteq B$ and $A' \subseteq B'$ such that $\beta(A) \geq \alpha$ and $\beta'(A') \geq \alpha$. Assume that $A' \subseteq A$. If $\mathbb{E}_x A * (A + A)(x)(3 \cdot A')(x) < \frac{\alpha^2 \beta \beta'}{8}$, we can find a regular Bohr set B'' of codimension at most*

$$d + O((\log \alpha^{-1})^4)$$

and radius at least

$$c \rho \nu \alpha^{O(1)}$$

such that there exists a translate $B'' + x$ of B'' where $\frac{|A \cap (B'' + x)|}{|B''|} \geq \frac{17}{16}\alpha$.

Proof. Let $p = \log \alpha^{-1}$. Apply Theorem 31 with the sets $A_{30} = A, X_{30} = A + A, \tilde{B}_{30} = B$, and $B_{30} = 3 \cdot B'$, we get a regular Bohr set B_0'' such that for all $b'' \in B_0''$,

$$\|A * (A + A)(b'' + \cdot) - A * (A + A)(\cdot)\|_{L^p(3 \cdot B')} \leq \epsilon \sqrt{\frac{|A|}{|G|}} \|A * (A + A)\|_{L^{p/2}(3 \cdot B')}^{1/2} + \epsilon^2 \frac{|A|}{|G|}.$$

Let B'' be a regular Bohr set such that $B'' \subseteq (B_0'')_{1/2}$, so $B'' + B'' \subseteq B_0''$.

Since $\|(A + A) * A\|_{L^{p/2}(3 \cdot B')} \leq \|G * A\|_{L^{p/2}(3 \cdot B')} = \alpha\beta$,

$$\|A * (A + A) * m_{B''} - A * (A + A)\|_{L^p(3 \cdot B')} \leq 2\epsilon\alpha\beta.$$

As in the finite field vector space case, the proof follows along the lines of the proof of Theorem 37. However, we get extra saving here due to the fact that we can expect $(A + A)$ to have much larger density.

Let $c = \frac{1}{16}$. If there exists x such that $A * m_{B''}(x) \geq (1 + c)\alpha$, we are done. Assume otherwise that $\|A * m_{B''}\|_\infty < (1 + c)\alpha$. By Proposition 10,

$$\mathbb{E}_{t \in B + B''} A * m_{B''}(t) \geq \frac{15\alpha}{16}.$$

Thus $A * m_{B''}(x) \geq \frac{7\alpha}{8}$ for at least a $\frac{7}{8}$-fraction of $x \in B + B''$.

If $|(A + A) \cap (B + B'')| \geq \frac{3|B+B''|}{4}$, then for all $x \in 3 \cdot B'$,

$$
\begin{aligned}
&|(x - (A + A) \cap (B + B'')) \cap (B + B'')| \\
&\geq |(A + A) \cap (B + B'')| + |B + B''| - |B + B'' + 3 \cdot B'| \\
&\geq \frac{3|B + B''|}{4} - 4\nu d|B| \\
&\geq \frac{5|B + B''|}{8}.
\end{aligned}
$$

Hence, for all $x \in 3 \cdot B'$,

$$(A + A) * A * m_{B''}(x) \geq \left(\frac{5}{8} + \frac{7}{8} - 1\right)\frac{7\alpha\beta}{8} > \frac{\alpha\beta}{4},$$

48

so

$$\mathbb{E}_x(A+A) * A * m_{B''}(x)(3 \cdot A')(x) > \frac{\alpha^2 \beta \beta'}{4}.$$

By Holder's inequality,

$$\frac{|G|}{|3 \cdot B'|}|\mathbb{E}_x((A+A) * A - (A+A) * A * m_{B''}(x))(3 \cdot A')(x)|$$

$$\leq \|(A+A) * A - (A+A) * A * m_{B''}\|_{L^p(3 \cdot B')}\|(3 \cdot A')\|_{L^{p/(p-1)}(3 \cdot B')}$$

$$\leq 2\epsilon\alpha\beta(1+c)\alpha$$

By choosing $\epsilon = \frac{1}{64}$, we get

$$|\mathbb{E}_x((A+A) * A - (A+A) * A * m_{B''}(x))(3 \cdot A')(x)| \leq \frac{\alpha^2 \beta \beta'}{8},$$

so

$$\mathbb{E}_x(A+A) * A(x)(3 \cdot A')(x) \geq \frac{\alpha^2 \beta \beta'}{4} - \frac{\alpha^2 \beta \beta'}{8} = \frac{\alpha^2 \beta \beta'}{8},$$

which contradicts the assumption.

Consider the case $|(A+A) \cap (B+B'')| < \frac{3|B+B''|}{4}$. We proceed as in the proof of Theorem 5.3, noticing that for $x \in -A'$,

$$(-A - A) * A(x) = \mathbb{E}_y(-A - A)(x - y)A(y) = \mathbb{E}_y A(y) = \alpha\beta.$$

Note that our assumption that $A + A$ is small implies

$$|(-A - A) \cap (B + B' + B')| = |(A+A) \cap (B + B' + B')| < \frac{3|B + B''|}{4} + |B + B' + B'\backslash B| < \frac{7|B|}{8}.$$

Apply Theorem 31 with $\epsilon = 10^{-4}$ and the sets $A_{30} = A, X_{30} = -A - A, \tilde{B}_{30} = B$, and $B_{30} = B'$ to get a regular sub-Bohr set B''' of B' such that

$$\|(-A - A) * A * m_{B'''} - (-A - A) * A\|_{L^p(B')} \leq 10^{-4}\sqrt{\alpha\beta}\|(-A - A) * A\|_{L^{p/2}(B')}^{1/2} + 10^{-8}\alpha\beta \leq 10^{-3}\alpha\beta,$$

using again the trivial upper bound $(-A - A)(x) \leq G(x)$. Thus, by Holder's inequality

$$|\mathbb{E}_{x \in B'}((-A - A) * A * m_{B'''}(x) - (-A - A) * A(x))(-A')(x)| \leq 10^{-3}\alpha\beta \cdot (1+c)\alpha \leq \frac{\alpha^2 \beta}{256},$$

where we use that $\mathbb{E}_{x \in B'}(-A')(x) \leq (1+c)\alpha$ since otherwise we would have density increment. Hence,

$$\mathbb{E}_{x \in B'}(-A - A) * A * m_{B'''}(x)(-A')(x)$$

$$\geq \mathbb{E}_{x \in B'}(-A - A) * A(x)(-A')(x) - \frac{\alpha^2 \beta}{256}$$

$$\geq \alpha^2 \beta - \frac{\alpha^2 \beta}{256}$$

$$= \frac{255\alpha^2 \beta}{256}.$$

If $\|A * m_{B'''}\|_\infty \leq (1+c)\alpha$, since $A * m_{B'''}$ is supported on $B + B'$ and $|(-A - A) \cap (B + B' + B')| < \frac{7|B|}{8}$, for $x \in B'$,

$$(-A - A) * A * m_{B'''}(x) = \frac{1}{|G|}\sum_{y \in B+B'}(-A - A)(y)A * m_{B'''}(x - y) \leq \frac{7|B|/8}{|G|}(1+c)\alpha < \frac{15\alpha\beta}{16},$$

49

so
$$\mathbb{E}_{x \in B'}(-A - A) * A * m_{B'''}(x)(-A')(x) \leq \frac{15\alpha\beta}{16} \cdot (1 + c)\alpha < \frac{255\alpha^2\beta}{256},$$

contradicting the above inequality.

We can easily check that in both cases the Bohr sets B'' and B''' satisfy the required condition on the dimension and radius. □

Corollary 48. *The largest subset of $\mathbb{Z}_n$ with no nontrivial solution to $x + y + z = 3w$ has size at most $n \exp(-c(\log n)^{1/5})$.*

Proof. Let $\mathcal{P}$ be the property of containing a nontrivial solution to $x + y + z = 3w$. If $\frac{|A'|}{|B'|} > \frac{9}{8}\alpha$ then we immediately obtain density increment. Otherwise, assume that A contains no nontrivial solution. Then

$$\mathbb{E}_x A * (A + A)(x)(3 \cdot A')(x) = \mathbb{E}_{x,y} A(y)(A + A)(x - y)(3 \cdot A')(x) \leq \frac{|A'|}{|G|^2} \leq \frac{9}{8}\frac{\alpha\beta'}{|G|} < \frac{\alpha^2\beta\beta'}{16},$$

as long as $|B| > 32\alpha^{-1}$. We get the desired bound from Theorem 37. □

7 Upper bounds for almost periodicity

Having seen the applications of L^p almost periodicity, in this section, we turn to discuss the limits on the quantitative bounds on the size of the set of almost periods in Theorem 12, which has direct consequences to the quantitative bounds in Roth's theorem obtained from almost periodicity. There are several parameters of interest here, the dimension p, the set density α and the approximation error ϵ. We will first give a construction showing that the linear dependence on p in the exponent is necessary.

Theorem 49. *Let α be an absolute constant, and q a fixed prime. Let ϵ be a sufficiently small constant depending only on α. There exists a subset $\bar{A}$ of $\mathbb{F}_q^n$ of density α such that the set of elements $x \in \mathbb{F}_q^n$ such that*

$$\|\bar{A} * \bar{A}(x + \cdot) - \bar{A} * \bar{A}(\cdot)\|_p < \epsilon\alpha + \epsilon^2$$

has size at most $|\mathbb{F}_q^n|/(1/\alpha)^{cp}$ for some constant c depending on α and q.

Proof. We first construct a subset $\tilde{A}$ of $\mathbb{F}_q^p$ of density α such that $\|\tilde{A} * \tilde{A}(y + \cdot) - \tilde{A} * \tilde{A}(\cdot)\|_p^p$ is large for all $y \neq 0$. We will construct $\tilde{A}$ as a random set with a special structure. Group each nonzero element of $\mathbb{F}_q^p$ into pairs $\{x, -x\}$, and let B be the set of representatives (one from each pair). We pick a random subset A' of B by choosing each element independently with probability $2\alpha\frac{q^p}{q^p-1}$. For each $a' \in A'$, we flip a fair coin to decide whether we include x or $-x$ (so at most one of them is included). We have $\mathbb{E}\mathbb{E}A = \alpha$. Furthermore, by Hoeffding's inequality, the probability that $|\mathbb{E}A - \alpha| > \epsilon$ is at most $2\exp(-2\epsilon^2(q^p-1)/2)$.

Clearly
$$A * A(0) = 0.$$

For any $y \neq 0$,

$$\mathbb{E}A * A(y) = \frac{\sum_{x \in \mathbb{F}_q^p} \mathbb{E}A(x)A(y - x)}{q^p} = \frac{q^p - 2}{q^p}\frac{\alpha^2 q^{2p}}{(q^p - 1)^2} = \alpha^2 - \frac{\alpha^2}{(q^p - 1)^2},$$

since for nonzero x and $y - x$ (belonging to different pairs $\{u, -u\}, u \in B$), they are included in A independently with probability $\alpha\frac{q^p}{q^p-1}$. Furthermore, for each $x \in \mathbb{F}_q^p$, $A(x)A(y - x)$ is independent of all but at most two other random variables $A(x')A(y - x')$. Thus, we can greedily partition $\mathbb{F}_q^p$ into at most 5 sets S_h, $1 \leq h \leq 5$, such that the random variables $A(x)A(y - x)$ for $x \in S_h$ are independent, and $|S_h| \geq q^p/10$ for each h. By the union bound and Hoeffding's inequality, the probability that $\left|A * A(y) - \left(\alpha^2 - \frac{\alpha^2}{(q^p-1)^2}\right)\right| > \epsilon$ is at most $10\exp(-2\epsilon^2 q^p/10)$.

Choose $\epsilon = q^{-p/3}$, we get by the union bound that with probability at most $12q^p \exp(-2q^{p/4}) < 1/2$ (assuming p is sufficiently large), $|\mathbb{E}A - \alpha| \le q^{-p/3}$ and for all $y \ne 0$, $\left| A * A(y) - \left(\alpha^2 - \frac{\alpha^2}{(q^p-1)^2} \right) \right| \le q^{-p/3}$. By arbitrarily adding or removing at most $q^{2p/3}$ elements from A, we get a set $\tilde{A}$ with $\mathbb{E}\tilde{A} = \alpha$,

$$\tilde{A} * \tilde{A}(0) \le q^{-p/3} + 2q^{-p/3} = 3q^{-p/3},$$

and for all $y \ne 0$,

$$\tilde{A} * \tilde{A}(y) \ge \alpha^2 - \frac{\alpha^2}{(q^p - 1)^2} - q^{-p/3} - 2q^{-p/3} \ge \alpha^2 - 4q^{-p/3}.$$

Here, we used that removing an element decreases $\sum_{x\in\mathbb{F}_q^n} A(x)A(y - x)$ by at most 2, and adding an element increases $\sum_{x\in\mathbb{F}_q^n} A(x)A(y - x)$ by at most 2.

Notice that for $y \ne 0$,

$$\|\tilde{A} * \tilde{A}(y + \cdot) - \tilde{A} * \tilde{A}(\cdot)\|_p^p \ge \frac{|\tilde{A} * \tilde{A}(y) - \tilde{A} * \tilde{A}(0)|^p}{q^n} \ge \frac{(\alpha^2 - 7q^{-p/3})^p}{q^p}.$$

We assume that q is a fixed constant (for example, $q = 3$), and choose ϵ a small enough constant depending only on α and q such that

$$\epsilon\alpha + \epsilon^2 < \frac{\alpha^2}{2q}.$$

We now come back to $\mathbb{F}_q^n$. For $x \in \mathbb{F}_q^n$, let $x_{[p]}$ be the first p coordinates of x. Let V be the subspace consisting of elements x for which $x_i = 0$ for all $i > p$. Let $\bar{A} = \{x : x_{[p]} \in \tilde{A}\}$. We have $\mathbb{E}\bar{A} = \alpha$, and

$$\bar{A} * \bar{A}(y) = \tilde{A} * \tilde{A}(y_{[p]}).$$

Thus, for any $y \notin V$, so $y_{[p]} \ne 0$, we have

$$\|\bar{A} * \bar{A}(y + \cdot) - \bar{A} * \bar{A}(\cdot)\|_p > \epsilon\alpha + \epsilon^2.$$

Hence, the set of almost periods is exactly V, which has codimension p and size

$$|V| = \frac{|\mathbb{F}_q^n|}{q^p} = \frac{|\mathbb{F}_q^n|}{(1/\alpha)^{cp}},$$

where c is some constant depending on α and q. $\qquad\square$

The above construction has the additional property that the set of almost periods remain small even if we replace A by a constant density subset of A. The above construction starts to work when $\epsilon \ll \alpha/q$. If we instead change the dependency and force the elements $\{x, -x\}$ to appear in pair (instead of appearing exclusively), we get a lower bound which starts to work when $\epsilon \ll (\alpha/q)^{1/2}$.

This theorem gives the tight asymptotic dependence on p when α and ϵ are treated as fixed constants. The tight dependence when α and ϵ are treated as asymptotic parameters and p is fixed comes from a known construction of Green [18] showing that Chang's lemma is essentially tight.

Theorem 50. *Let* $\alpha \le 1/8$, $\delta \le 1/32$ *and* $\delta^{-2} \log \alpha^{-1} \le \frac{\log N}{\log \log N}$. *There exists a set* A *of density* α *for which the large spectrum* $\Delta_\delta(A)$ *is not contained in* $\{\sum_{s\in S} \epsilon_s s, \epsilon_s \in \{-1, 0, 1\}\}$ *for any set* $|S| \le c\delta^{-2} \log \alpha^{-1}$.

The construction first constructs a weighted set with the required property on the large spectrum, then transform this to a genuine set using discrepancy theory. The construction of the weighted set is a "smooth intersection" of a version of Ruzsa's niveau sets. We omit further details of the proof of Theorem 50, which can be found in [18].

Using Theorem 50, we prove that the dependency on α and ϵ in Theorem 12 is tight.

Theorem 51. *Let α be an absolute constant, $p \geq 2$. Let ϵ be sufficiently small (depending on α). There exists a subset A of $\mathbb{Z}_N$ of density α such that the set of elements $x \in \mathbb{Z}_N$ such that*

$$\|A * A(x + \cdot) - A * A(\cdot)\|_p < \epsilon\alpha + \epsilon^2$$

has size at most $N/(1/\alpha)^{c/\epsilon^2}$ for some constant c.

Proof. Note that

$$\|A * A(x + \cdot) - A * A(\cdot)\|_p \geq \|A * A(x + \cdot) - A * A(\cdot)\|_2,$$

hence it suffices to prove the above for $p = 2$.

Let A be the set given in Theorem 50 with $\alpha \leq 1/8$ and $\delta = C\epsilon$ for a large absolute constant $C > 1000$. In this case

$$\|A * A(x + \cdot) - A * A(\cdot)\|_2^2 = \sum_\chi |\widehat{A * A}(\chi)(\chi(-x) - 1)|^2$$

$$= \sum_\chi |\hat{A}(\chi)|^2 |\chi(-x) - 1|^2$$

$$\geq \delta^2\alpha^2 \sum_{\chi \in \Delta_\delta(A)} |\chi(-x) - 1|^2.$$

In particular, if

$$\|A * A(x + \cdot) - A * A(\cdot)\|_2 < \epsilon\alpha + \epsilon^2 < 2\epsilon\alpha,$$

then

$$\delta^2\alpha^2 \sum_{\chi \in \Delta_\delta(A)} |\chi(-x) - 1|^2 < 4\epsilon^2\alpha^2,$$

so

$$\sum_{\chi \in \Delta_\delta(A)} |\chi(-x) - 1|^2 < 4/C^2.$$

Hence, for all such x, $|\chi(-x) - 1| < 2/C$. Thus, $x \in B(\Delta_\delta(A), 8\pi/C)$.

Finally, we need to prove an upper bound on $|B(\Delta_\delta(A), 8\pi/C)|$. Assume that

$$|B(\Delta_\delta(A), 8\pi/C)|/N > \exp(-\tilde{C}\epsilon^{-2}\log\alpha^{-1}).$$

Notice that for all $\lambda \in \Lambda$, and $B = B(\lambda, \rho)$,

$$\left|\frac{|B|}{N} - \hat{B}(\lambda)\right| \leq \mathbb{E}_x |B(x)(1 - \lambda(x))| \leq \mathbb{E}_x \rho B(x) = \frac{\rho|B|}{N}.$$

Hence,

$$|\hat{B}(\lambda)| \geq \frac{|B|}{N}(1 - \rho).$$

Thus, $\lambda \in \Delta_{1-\rho}(B)$. Hence,

$$\Delta_\delta(A) \subseteq \Delta_{1-8\pi/C}(B(\Delta_\delta(A), 8\pi/C)).$$

Theorem 22 implies that

$$\Delta_{1-8\pi/C}(B(\Delta_\delta(A), 8\pi/C)) \subseteq \{\sum_{t \in T} \epsilon_t t, \epsilon_t \in \{-1, 0, 1\}\}$$

for a set

$$|T| \leq C_{22}(1 - 8\pi/C)^{-2}\tilde{C}\epsilon^{-2}\log\alpha^{-1}.$$

By choosing $\tilde{C}$ to be a sufficiently small positive constant (depending only on C), we get $|T| \le c_{50}\delta^{-2}\log\alpha^{-1}$, which contradicts our assumption on $\Delta_\delta(A)$. Hence,

$$|B(\Delta_\delta(A), 8\pi/C)|/N \le \exp(-\tilde{C}\epsilon^{-2}\log\alpha^{-1}).$$

Thus the set of elements $x \in \mathbb{Z}_N$ such that

$$\|A * A(x + \cdot) - A * A(\cdot)\|_2 < \epsilon\alpha + \epsilon^2$$

has size at most $N/(1/\alpha)^{c/\epsilon^2}$ for some constant c. $\qquad\square$

8 L^∞ almost periodicity for sets with bounded VC dimension

Since we benefit from L^p almost periodicity for very large p, it is natural to wonder if one can get a uniform almost periodicity result, i.e., an L^∞ almost periodicity result. However, by the upper bound in Theorem 49, it is evident that one cannot get an L^∞ almost periodicity result for general sets. Nevertheless, with an extra condition on the set A, it is possible to get a large set of L^∞ almost periods. The notion we need is bounded Vapnik-Chervonenkis dimension (VC dimension).

Definition 52. The VC dimension $\dim_{VC}(\mathcal{S})$ of a collection $\mathcal{S}$ of subsets of a ground set $\mathcal{X}$ is the largest k such that there exists a subset Y of $\mathcal{X}$ of size k such that for every subset Z of Y, there exists $S \in \mathcal{S}$ such that $Z = S \cap Y$. We say that Y is shattered by $\mathcal{S}$ if the above happens.

The VC dimension $\dim_{VC}(A)$ of a subset A of an abelian group G is the VC dimension of the collection $\mathcal{S} = \{A + x, x \in G\}$.

The relative VC dimension of subsets A and X of an abelian group G, denoted by $\dim_{VC}(A, X)$, is the VC dimension of the collection $\mathcal{S} = \{A \cap (x - X), x \in A + X\}$.

We remark that we use a different notation from [39], where they use $\dim_{VC}(A)$ to denote $\dim_{VC}(A, A)$. However, the difference between the two definitions is not crucial, as made clear by the following immediate facts about the VC dimension of subsets of an abelian group.

Lemma 53. *For $A \subseteq G$,*

$$\dim_{VC}(A) = \dim_{VC}(-A) = \dim_{VC}(A + x) = \dim_{VC}(G, A).$$

For $A, X \subseteq G$,

$$\dim_{VC}(A) \ge \dim_{VC}(\{(A + x) \cap X, x \in G\}),$$

$$\dim_{VC}(\{A \cap (x - X), x \in G\}) - 1 \le \dim_{VC}(A, X) \le \dim_{VC}(\{A \cap (x - X), x \in G\}) \le \dim_{VC}(X).$$

For $A \subseteq G$,

$$\dim_{VC}(A) - 1 \le \dim_{VC}(A, A) \le \dim_{VC}(A).$$

Proof. The first statement is immediate from the definition.

For the first part of the second statement, notice that if a set Y of size k is shattered by $\{(A + x) \cap X\}$ then $Y \subseteq X$, and for any $Z \subseteq Y$, there is $x \in G$ such that $Z = Y \cap ((A + x) \cap X)$. Then

$$Z = (A + x) \cap X \cap Y = (A + x) \cap Y.$$

Hence, Y is shattered by the collection $\{A + x\}$. For the second part of the second statement, notice that $A \cap (x - X) = \emptyset$ for all $x \notin X + A$.

The upper bound in the third statement follows from the first part of the second statement. For the lower bound, notice that if a set Y of size k is shattered by $\{A + x\}$, then $Y = (A + x) \cap Y$ for some x, so $Y \subseteq A + x$. For every subset Z of Y, we can find $u \in G$ such that $Z = Y \cap (A + u)$, so

$$Z = Y \cap (A + x) \cap (A + u),$$

and
$$Z - x = (Y - x) \cap ((A + u - x) \cap A).$$

In particular, $Y - x$ is shattered by $\{(A + u) \cap A, u \in G\}$. Hence,

$$\dim_{VC}(A) \leq \dim_{VC}(\{(A + u) \cap A, u \in G\}) \leq \dim_{VC}(A, A) + 1.$$

$\square$

The third property implies that the condition $\dim_{VC}(A) \leq d$ and $\dim_{VC}(A, A) \leq d$ are essentially the same. To motivate our results, we state some examples of sets with bounded VC dimension.

Example 54. A coset of a subgroup of G has VC dimension 1, since the translates of a subgroup are disjoint.

A union of k cosets of a subgroup of G has VC dimension at most k. This is since any $(k+1)$-element set which is shattered by a union of k cosets must be contained in a single translate of the union, so two elements must be in the same coset, which cannot be shattered by translates of unions of cosets.

Example 55. An arithmetic progression in $\mathbb{Z}_n$ or $\mathbb{Z}$ of length at least 3 has VC dimension 2. This is since 2 adjacent elements of the arithmetic progression are shattered. However, assuming that a 3-element set is shattered, this set must be a subset of a translate of the arithmetic progression, and then middle element of the 3-element set cannot be shattered.

A generic generalized arithmetic progression of dimension k, $\{\sum_{i=1}^{k} a_i n_i, 0 \leq a_i \leq L_i\}$, has VC dimension at most $2k$. Here, we assume $n_{i-1} = o(n_i)$. Indeed, assume that there exists a set Y of $2k + 1$ points which is shattered. Without loss of generality, we can assume $Y \subseteq \{\sum_{i=1}^{k} a_i n_i, 0 \leq a_i \leq L_i\}$. By genericity, points $\sum_{i=1}^{k} b_i^j n_i$ are covered in a translate of the generalized arithmetic progression if and only if $x_i \leq b_i^j \leq x_i + L_i$ for some x_i and all j. Thus, if we choose $M_i \in Y$ with maximum i-coordinate, $m_i \in Y$ with minimum i-coordinate. Then if Z is any subset of Y containing all M_i and m_i and a translate of the generalized arithmetic progression covers Z, it must in fact cover Y, since any other point in Y has the i-coordinate between the i-coordinate of m_i and M_i. Since there are at most $2k$ such points m_i and M_i, Y cannot be shattered.

In [39], an L^∞ almost periodicity result is established for sets of bounded VC dimensions. While VC dimension is a combinatorial notion of set systems, it is closely related to the notion of stability in model theory. A set A is d-stable if there does not exist $x_1, x_2, \cdots, x_d$ such that for all $i \leq d$, there is some y_i such that $\{x_1, x_2, \cdots, x_i\} = \{x_1, x_2, \cdots, x_d\} \cap (A - y_i)$. Thus, stability is a much stronger notion than having bounded VC dimension. In particular, d-stable sets have VC dimension at most $d - 1$, since if $\{x_1, x_2, \cdots, x_d\}$ is shattered by $\{A - x\}$ then it also witnesses d-instability. One can find in [2, 39, 42] how assumptions on stability or VC dimension lead to a much more efficient bound for the arithmetic regularity lemma. In fact, L^∞ almost periodicity [39] and a close analog of L^∞ almost periodicity [2] are used to establish this result.

Theorem 56 (L^∞ almost periodicity). *Let $A, X \subseteq G$ be subsets with $\dim_{VC}(A, X) \leq d$. Assume that $|A + S| \leq K|A|$. There exists a subset T of S of size at least $K^{-Cdk^2/\epsilon^2}|S|$ such that for all $t_1, t_2, \cdots, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_\infty \leq \epsilon \frac{|A|}{|G|}.$$

Theorem 56 is the immediate analog of Theorem 12. All of the machineries we developed, particularly the bootstrapping procedure to transform an arbitrary set of almost periods to a structured set of almost periods (subspace or Bohr set), hold exactly as before. We will give a proof of Theorem 56 in the next subsection, following [39]. This proof directly follows the ideas in the proof of Theorem 12, with the only change being the control of $\mathbf{E} \sup_x |m_{\bar{a}} * X(x) - m_A * X(x)|$ instead of $\mathbf{E}\mathbb{E}_x |m_{\bar{a}} * X(x) - m_A * X(x)|^p$, where

$\bar{a}$ is a sample of n independent elements from A. The control of the supremum is well-known in probability theory, specifically in the study of empirical processes. This perhaps justifies the VC dimension condition, which is exactly what is needed in empirical process theory.

Notice that $\dim_{VC}(A, X) \leq d$ if $\dim_{VC}(X) \leq d$. A much shorter proof of a version of Theorem 56 under the condition $\dim_{VC}(A) \leq d$ is available in [2], which also makes essentially similar use of the VC dimension condition.

Theorem 57 (L^∞ almost periodicity). *Let $A, X \subseteq G$. Assume that $\dim_{VC}(A) \leq d$, and $|A + S| \leq K|A|$. There exists a subset T of S of size at least $\left(\frac{c\epsilon}{Kk}\right)^d |S|$ such that for all $t_1, t_2, \cdots, t_{2k} \in T$,*

$$\|A * X(t_1 - t_2 + \cdots + t_{2k-1} - t_{2k} + \cdot) - A * X(\cdot)\|_\infty \leq \epsilon \frac{|A|}{|G|}.$$

Furthermore, the set T does not depend on X.

In particular, Theorem 57 gives a much stronger quantitative bound when $A = X$ under only a slightly stronger condition. While in [2] and in the remarks in [39], Theorem 57 is only stated for dense A and symmetric convolution $A * A$, we in fact observe that a small generalization of the proof in fact gives the stronger conclusion above. Notice the asymmetric roles of A and X with respect to the bounded VC dimension condition in Theorem 56 and Theorem 57.

We cover the short proof of Theorem 57 in Subsection 8.2, slightly generalizing the result in [2] from the dense case ($|A| \geq \alpha|G|$) to the small expansion case ($|A + S| \leq K|A|$). Even though this proof is shorter, we decide to also give the longer proof via chaining, since it has a direct correspondence with the techniques we developed earlier, and extends the previously observed connection between results in additive combinatorics and results in probability theory. In fact, we believe that this difficulty is understandable since Theorem 56 is slightly more general than Theorem 57. In Theorem 56, the bounded VC dimension condition is placed on X while the small expansion condition is placed on A. In Theorem 57, both conditions are placed on A. Further details of the comparison between two results are made clear at the end of Subsection 8.2, after both proofs are given.

8.1 L^∞ almost periodicity via chaining

In this subsection, we will follow [39] to prove an L^∞ almost periodicity result for sets with bounded VC dimension using techniques from the theory of empirical processes, which perhaps suggests that the condition on the VC dimension is not surprising.

We will use a similar sampling approach to find the almost periods. However, instead of getting an L^p bound, we will get an L^∞ bound under the extra assumption of bounded VC dimension. The main idea in the proof is to bound the expectation of the L^∞ norm of $|m_{\bar{a}} * X(x) - m_A * X(x)|$ where $\bar{a}$ is again a vector of n elements sampled independently at random from A. This is done via a process referred to as chaining in the study of empirical processes. For the reader's convenience, we reproduce the chaining argument, which uses many nice ideas. We follow [7] in this exposition.

Chaining is used to bound empirical processes of the form $S = \sup_j S_j$ for

$$S_j = \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) - (\mathbf{E}U_i)\frac{|X_j|}{|\mathcal{X}|},$$

where X_j are subsets of $\mathcal{X}$ and a_i are sampled independently and uniformly at random from $\mathcal{X}$, and U_i are independent random variables which are further independent of all a_i. If we take U_i to be 1, $\mathcal{X} = A$, $X_j = (j - X) \cap A$, we would arrive back at our setting. The necessity of this generalization is discussed later. The rough idea of chaining is to equip the sets X_j with the metric $d_{\mathcal{X}}(X_j, X_{j'}) = |X_j \Delta X_{j'}|/|\mathcal{X}|$, then show that if $|X_j \Delta X_{j'}|$ is small, we have very good concentration of $S_j - S_{j'}$. Together with a good covering of the sets so that each set is close to at least one set in the covering in the above metric, we

obtain a bound on the L^∞ norm of $S_j - S_{j'}$. We then chain an arbitrary S_j to S_0 by a telescoping sum of sets which are close in the metric $d_\mathcal{X}$, and obtain the required L^∞ bound. A crucial tool, which takes into account the bounded VC dimension condition, is Haussler's packing lemma, which allows us to run efficient union bounds in bounding the L^∞ norms.

Lemma 58 (Haussler's packing lemma). *Let $|A| = n$. Let $\mathcal{S}$ be a collection of subsets of A of VC dimension d. If $\mathcal{D}$ is a subset of $\mathcal{S}$ such that the symmetric difference of any two sets in $\mathcal{D}$ has size at least k, then*

$$|\mathcal{D}| \leq \left(\frac{Cn}{k} \right)^d.$$

We skip the proof of Haussler's packing lemma, which can be found in [24].

In the chaining argument, we link S_j to S_0 using dyadic steps in the metric $d_\mathcal{X}$. In particular, we take a cover D_t of the collection of sets $\{X_j\}$ so that each set is at distance 2^{-t} from some set in the cover D_t. We use exponential concentration and a trivial union bound to bound $\mathbf{E} \sup_{j \in D_t, j' \in D_{t+1}} |S_j - S_{j'}|$. A naive use of the argument takes $\mathcal{X} = A$ and $X_j = (j - X) \cap A$ for $j \in G$, and $U_i = 1$. However, the chaining must then take place with the number of steps depending on $|A|$, while our quantitative bounds should depend only on d and not $|A|$. This occurs since when we localize the chain at very small scale (t much bigger than $\log n$), the union bound starts to blow up while the fluctuations cannot be brought below $1/n$. However, since we sample only n elements, we actually expect no fluctuation at very small scale (large t). We need to be able to take into account our "finite sample size". The idea to do this is a beautiful idea in probability theory, known as symmetrization.

Lemma 59 (Symmetrization). *Let $\mathcal{X}$ be a finite set, $\{X_j\}$ a collection of subsets of $\mathcal{X}$. Let a_i be independently and uniformly chosen elements of $\mathcal{X}$. Then*

$$\mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n X_j(a_i) - \frac{|X_j|}{|\mathcal{X}|} \right| \right) \leq 2 \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) \right| \right),$$

where U_i are independent random variables receiving value 1 with probability $1/2$ and value -1 with probability $1/2$, which are further independent of all a_i.

Proof. Let a_i' be independently and uniformly chosen elements of $\mathcal{X}$, which are further independent of all a_i and U_i. Then

$$
\begin{aligned}
\mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n X_j(a_i) - \frac{|X_j|}{|\mathcal{X}|} \right| \right) &= \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n [X_j(a_i) - \mathbf{E}(X_j(a_i'))] \right| \right) \\
&\leq \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n [X_j(a_i) - X_j(a_i')] \right| \right) \\
&= \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i [X_j(a_i) - X_j(a_i')] \right| \right) \\
&\leq \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) \right| \right) + \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i') \right| \right) \\
&= 2 \mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) \right| \right),
\end{aligned}
$$

where the first inequality follows from Jensen's inequality applied to the convex function $|x|$, the second equality follows since $\{U_i(X_j(a_i) - X_j(a_i'))\}_j$ has the same distribution as $\{X_j(a_i) - X_j(a_i')\}_j$, and the final equality follows since a_i and a_i' are identically distributed. $\square$

The symmetrization inequality is extremely useful, since it allows us to turn the problem of bounding the supremum of the empirical averages of the set system X_j of a possibly large $\mathcal{X}$ to bounding

$$\sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) \right| \right)$$

(thinking of a_i as being fixed and U_i being random). This allows us to localize the problem, reducing the problem over $\mathcal{X}$ to an "effective" problem over an n-element set. One can imagine that chaining over the original universe $\mathcal{X}$ would lead to many steps that localize beyond $1/n$, the union bound over such steps is very inefficient since most atoms would contain no point from the n sampled points. Of course, reducing the union bound is not possible since that requires knowledge of the position of the points. However, by symmetrizing with the n Rademacher variables one essentially reduces the randomness to that around the finite sample of points. This localization allows us to apply the previously described chaining idea to $\frac{1}{n} \sum_{i=1}^n U_i X_j(a_i)$.

Lemma 60 (Chaining). *Let $a_1, a_2, \cdots, a_n$ be fixed elements in $\mathcal{X}$. Let $\{X_j\}$ be a collection of subsets of $\mathcal{X}$ with VC dimension at most d. Then*

$$\mathbf{E} \sup_j \left(\left| \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i) \right| \right) \leq \frac{C\sqrt{d}}{\sqrt{n}}.$$

Proof. For $\bar{a} = \{a_1, a_2, \cdots, a_n\}$, $\{\bar{a} \cap X_j\}$ (as a collection of subsets of $\bar{a}$) has VC dimension at most d, by Lemma 53.

For $t = 0, 1, \cdots, \lceil \log_2 n \rceil$, let D_t be a maximal collection of sets in $\{\bar{a} \cap X_j\}$ such that for any $D_{t,1}, D_{t,2} \in D_t$, $|D_{t,1} \Delta D_{t,2}|/n \geq 2^{-t}$. By Lemma 58, $|D_t| \leq 2^{(t+C_0)d}$ for some constant C_0. Furthermore, by maximality, for any $\bar{a} \cap X_j$, there must exist $D_t(j) \in D_t$ for which $|D_t(j) \Delta (\bar{a} \cap X_j)| < 2^{-t}n$.

Let

$$f(\bar{a} \cap X_j) = \frac{1}{n} \sum_{i=1}^n U_i X_j(a_i).$$

For each t, let $D_t(j)$ be a point in D_t such that $|D_t(j) \Delta (\bar{a} \cap X_j)| \leq 2^{-t}n$. Define $y_t(j)$ inductively by

$$y_{\lceil \log_2 n \rceil}(j) = \bar{a} \cap X_j,$$

and for $1 \leq t \leq \lceil \log_2 n \rceil - 1$,

$$y_{t-1} = D_{t-1}(y_t).$$

We can write

$$f(\bar{a} \cap X_j) = (f(y_{\lceil \log_2 n \rceil}) - f(y_{\lceil \log_2 n \rceil - 1})) + (f(y_{\lceil \log_2 n \rceil - 1}) - f(y_{\lceil \log_2 n \rceil - 2})) + \cdots + (f(y_1) - f(y_0)) + f(y_0).$$

Note the crucial property that $|y_{t+1} \Delta y_t| \leq 2^{-t}n$, and $y_t \in D_t$. We will bound $\mathbf{E} \sup_{y_t \in D_t, y_{t+1} \in D_{t+1}} |f(y_{t+1}) - f(y_t)|$, from which we can derive a bound on $\mathbf{E} \sup |f(\bar{a} \cap X_j)|$.

We have

$$f(y_{t+1}) - f(y_t) = \frac{1}{n} \sum_{i \in y_{t+1} \setminus y_t} U_i - \frac{1}{n} \sum_{i \in y_t \setminus y_{t+1}} U_i,$$

so by independence,

$$\mathbf{E} \exp(\lambda(f(y_{t+1}) - f(y_t))) = \left(\frac{\exp(\lambda/n)}{2} + \frac{\exp(-\lambda/n)}{2} \right)^{|y_{t+1} \Delta y_t|}.$$

Using that $\exp(x) \leq 1 + x + x^2$ for $|x| \leq 1$, assuming that $\lambda \leq n$, we have

$$\mathbf{E} \exp(\lambda(f(y_{t+1}) - f(y_t))) \leq (1 + \lambda^2/n^2)^{2^{-t}n} \leq \exp(2^{-t}\lambda^2/n).$$

By the trivial bound

$$\mathbf{E}\exp\left(\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}\lambda(f(y_{t+1})-f(y_t))\right)\leq\sum_{y_t\in D_t, y_{t+1}\in D_{t+1}}\mathbf{E}\exp(\lambda(f(y_{t+1})-f(y_t)))$$

$$\leq 2^{1+2(t+C_0)d}\exp(2^{-t}\lambda^2/n).$$

By Jensen's inequality,

$$\exp\left(\mathbf{E}\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}\lambda(f(y_{t+1})-f(y_t))\right)\leq\mathbf{E}\exp\left(\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}\lambda(f(y_{t+1})-f(y_t))\right),$$

so

$$\mathbf{E}\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}(f(y_{t+1})-f(y_t))\leq\frac{C_1 d(1+t)+2^{-t}\lambda^2/n}{\lambda}$$

for an absolute constant C_1.

Choose $\lambda=\sqrt{d2^t n}$. For $t\leq\log_2(\sqrt{n/d})$, $\lambda\leq n$, and

$$\mathbf{E}\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}(f(y_{t+1})-f(y_t))\leq\frac{C_2(1+t)\sqrt{d}}{\sqrt{2^t n}}.$$

If $t>\log_2(\sqrt{n/d})$, we instead use the deterministic bound

$$f(y_{t+1})-f(y_t)\leq\frac{|y_{t+1}\Delta y_t|}{n}\leq 2^{-t},$$

since $|U_i|\leq 1$. Hence,

$$\mathbf{E}\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}(f(y_{t+1})-f(y_t))\leq 2^{-t}.$$

Thus,

$$\mathbf{E}\sum_{t=0}^{\lceil\log_2 n\rceil}\sup_{y_t\in D_t, y_{t+1}\in D_{t+1}}(f(y_{t+1})-f(y_t))\leq\sum_{t=0}^{\lceil\log_2 n\rceil}\frac{C_2(1+t)\sqrt{d}}{2^{t/2}\sqrt{n}}+\sum_{t>\log_2(\sqrt{n/d})}^{\lceil\log_2 n\rceil}2^{-t}$$

$$<\frac{C_3\sqrt{d}}{\sqrt{n}}+2^{-\log_2(\sqrt{n/d})}=\frac{C_4\sqrt{d}}{\sqrt{n}}.$$

Furthermore,

$$\mathbf{E}f(y_0)=0.$$

Hence,

$$\mathbf{E}\sup_j f(\bar{a}\cap X_j)<\frac{C_4\sqrt{d}}{\sqrt{n}}.$$

Similarly,

$$\mathbf{E}\sup_j(-f(\bar{a}\cap X_j))<\frac{C_4\sqrt{d}}{\sqrt{n}}.$$

Hence,

$$\mathbf{E}\sup_j|f(\bar{a}\cap X_j)|<\frac{C\sqrt{d}}{\sqrt{n}}.$$

$\square$

We next give the proof of Theorem 56.

Proof of Theorem 56. Let $\mathcal{X} = A$, $X_j = (j - X) \cap A$ for $j \in X + A$. Then, $\dim_{VC}(\{X_j\}) \leq d$. Note that if $j \notin X + A$,

$$m_{\bar{a}} * X(j) - m_A * X(j) = 0.$$

By Theorem 59 and Theorem 60,

$$
\begin{aligned}
\mathbf{E} \sup_{j \in G} (|m_{\bar{a}} * X(j) - m_A * X(j)|) &= \mathbf{E} \sup_{j \in B+A} \left(\left| \frac{1}{n} \sum_{i=1}^{n} X_j(a_i) - \frac{|X_j|}{|\mathcal{X}|} \right| \right) \\
&\leq 2\mathbf{E} \sup_{j \in B+A} \left(\left| \frac{1}{n} \sum_{i=1}^{n} U_i X_j(a_i) \right| \right) \\
&= 2\mathbf{E} \left[\mathbf{E} \left[\sup_{j \in B+A} \left(\left| \frac{1}{n} \sum_{i=1}^{n} U_i X_j(a_i) \right| \right) \Big| \bar{a} \right] \right] \\
&\leq 2 \frac{C\sqrt{d}}{\sqrt{n}}.
\end{aligned}
$$

Thus, with probability at least $1/2$,

$$\|m_{\bar{a}} * X - m_A * X\|_\infty < 4 \frac{C\sqrt{d}}{\sqrt{n}}.$$

The proof then proceeds identically to the proof of Theorem 13. □

8.2 L^∞ almost periodicity for sets with bounded VC dimension and small expansion

We give in this subsection a proof of Theorem 57, following [2]. We slightly generalize their results to the setting of sets with small expansion $|A + S| \leq K|A|$ instead of dense A. In this proof, we use Haussler's packing lemma, Lemma 58, to directly deduce that for many y, $|(A + y)\Delta A|$ is small, which is much stronger, and in particular, implies that $\|A * X(y + \cdot) - A * X(\cdot)\|_\infty$ is small.

Lemma 61. *For at least $(\delta/C)^d|S|$ elements y in a translate of S, $|(A + y)\Delta A| \leq \delta|A + S|$.*

Proof. Let $X \subseteq S$ be such that $\{x + A, x \in X\}$ forms a maximal collection of translates of A such that $|(A + x)\Delta(A + x')| > \delta|A + S|$ for all $x, x' \in X$. By Lemma 58, $|X| \leq (C/\delta)^d$.

For any $g \in S$, $|(A + g)\Delta(A + x)| \leq \delta|A + S|$ for some $x \in X$ by maximality of X. Hence, $|A\Delta(A + g - x)| \leq \delta|A + S|$. By the pigeonhole principle, for some $x \in X$, there must be at least $|S|(\delta/C)^d$ elements g for which $|(A + g)\Delta(A + x)| \leq \delta|A + S|$. Then $|A\Delta(A + y)| \leq \delta|A + S|$ for all $y = g - x$, which holds for at least $|S|(\delta/C)^d$ elements y in a translate of S. □

Lemma 61 readily implies Theorem 57.

Proof of Theorem 57. Notice that for all y given by Lemma 61,

$$|A * X(x + y) - A * X(x)| = \frac{||(A + y) \cap (x - X)| - |A \cap (x - X)||}{|G|} \leq \frac{|(A + y)\Delta A|}{|G|} \leq \delta \frac{|A + S|}{|G|} \leq \delta K \frac{|A|}{|G|}.$$

Hence,

$$\|A * X(y + \cdot) - A * X(\cdot)\|_\infty \leq \delta K \frac{|A|}{|G|}.$$

This implies Theorem 57 via the standard machinery. □

We remark that this proof instead uses the VC dimension condition on A instead of X. One can think of the result of Lemma 61 as a "first order" almost periodicity result, which applies directly to the set A. In the L^∞ almost periodicity result in the previous subsection and the other L^p almost periodicity results, we get a "second order" almost periodicity, which applies to convolutions of two sets. As above, the first order condition is much stronger than the second order condition. One can think that the bounded VC dimension, coupled with the small expansion condition, is actually much stronger, giving a clean "first order" condition. For Theorem 56, the additively structured condition is placed on A while the bounded VC dimension condition is placed on X, leading to a more complicated situation where we cannot obtain a "first order" condition but only a second order one, by combining the nice properties of A and X. Thus, one can see the conditions in Theorem 56 as being slightly more general, separating the contribution of the additive structure and bounded VC dimension. This should compensate for the longer proof of Theorem 56.

The proof given in this subsection heavily depends on efficient packing (which is in fact the only necessary ingredient), and a nice translation invariant property of the set system considered in the additive setting. In fact, one can roughly imagine that taking into full account this nice translation invariant property of the set system (together with efficient packing and small expansion) essentially allows us to sample a single element in the proof of Lemma 61 instead of a large number of elements as in the proof of Theorem 56, leading also to the improved quantitative bound. It is natural to ask if the weaker conditions in Theorem 56 are useful in any application, or if the proof via chaining allows for deducing uniform continuity of more general objects.

Finally, we remark that the quantitative bound in Lemma 61 is in fact sufficiently good that we can avoid the previous spectral bootstrapping procedure to get a structured set of almost periods via saturation, which is employed in [2]. In particular, by choosing the constants appropriately, we can get some k such that $|2kT| \le K|kT|$ where T is the set of δ-almost periods. Here, K can be taken to be $\exp((\log \delta^{-1})^c)$. Then we can apply directly the quasipolynomial Bogolyubov-Ruzsa lemma [37] to get the structured set of almost periods. Applying this procedure to our previous applications on Roth's theorem would unfortunately lead to some loss on the power of the $\log n$ factor in the final quantitative bound.

9 Concluding remarks

Almost periodicity results show that any dense set, or more generally, sets with small expansion under the addition of an other set, exhibit an L^p periodic behavior under two-fold convolutions. The proof follows from the simple idea of sampling a finite-size subset, which happens entirely in the physical space. We also develop the spectral machineries to bootstrap the set of almost periods to a structured set of almost periods (a subspace of bounded codimension, or a large Bohr set). We use this to obtain almost logarithmic bounds for Roth's theorem over general abelian groups, and Behrend-type bounds for Roth's theorem in four or more variables.

Almost periodicity has found many other exciting applications in additive combinatorics that are not included in this essay. We describe several applications that can be deduced in quite straightforward manner from the almost periodicity result. We refer the interested reader to the listed references for more details.

Arithmetic progression in sumsets.

Bourgain [8] showed that for two subsets A, B of $\mathbb{Z}_n$ with density α, β, $A + B$ contains an arithmetic progression of length at least $\exp(c(\alpha\beta \log n)^{1/3} - \log \log n)$. Using Fourier analytic techniques, Green [19] and Sanders [34] improved the bound to $\exp(c(\alpha\beta \log n)^{1/2} - \log \log n)$.

Since the support of $A * B$ is exactly $A + B$, it is not surprising that that almost periodicity can be applied to this application. Using almost periodicity, Croot-Laba-Sisask [14] further improved the bound

to

$$\exp\left(c\left(\frac{\alpha \log n}{(\log 2\beta^{-1})^3}\right)^{1/2} - \log(\beta^{-1}\log n)\right).$$

An analog with the tri-fold sumset $A+A+A$ is also available in [38], where it is shown that $A+A+A$ contains an arithmetic progression of length at least

$$\alpha \exp\left(\left(\frac{c\log n}{\log^3(2/\alpha)}\right)^{1/2}\right).$$

Quasipolynomial Bogolyubov-Ruzsa lemma.

The Bogolyubov-Ruzsa lemma states that for a set A with density at least α, $2A-2A$ contains a large subspace or Bohr set. It is known that the Bogolyubov-Ruzsa lemma implies the Freiman-Ruzsa theorem on the structure of sets with small expansion. In particular, a polynomial bound for Bogolyubov-Ruzsa lemma implies the polynomial Freiman-Ruzsa conjecture. Using almost periodicity, Sanders [37] shows a quasipolynomial Bogolyubov-Ruzsa lemma, stating that for A with small doubling ($|A+A| \le K|A|$), $2A-2A$ contains a coset progression of dimension at most $\log^{O(1)}(2K)$ and size at least $\exp(-\log^{O(1)}(2K))|A|$. Here, a coset progression is a generalized progression of cosets of a subgroup. A nice exposition of this result over finite field vector space can be found in [29], and a deduction using an exponential tail version of almost periodicity instead of L^p norms together with an algorithmic version can be found in [4].

Arithmetic regularity lemma for sets with bounded VC dimension.

Green's arithmetic regularity lemma [20] states that for a dense subset A of $\mathbb{F}_p^n$, there exists a subspace V of $\mathbb{F}_p^n$ of codimension $M(\epsilon)$ such that in all but an ϵ fraction of translates of V, A is uniform in the translate, in the sense that all nontrivial Fourier coefficients of A restricted to the translate are small. Green's proof uses density increment and leads to a bound on $M(\epsilon)$ of tower type, which is in fact necessary [26]. However, for sets with bounded VC dimension, and in particular, for stable sets, $M(\epsilon)$ can in fact be taken to be polynomial in ϵ^{-1}, as shown in [42] using ideas inspired from model theory. Using L^∞ almost periodicity, in [39], a similar result is shown, which can also be generalized to the case of sets with small doubling. Independently, in [2], the polynomial arithmetic regularity lemma for sets with bounded VC dimension in groups with bounded exponent is obtained from Lemma 61, using a bootstrapping procedure involving saturation and the Bogolyubov-Ruzsa lemma. In fact, it is quite straightforward to get a version of the polynomial arithmetic regularity lemma from the L^∞ almost periodicity results that we covered.

Almost periods and subset sums.

In a recent work of the author in a different setting [31], in trying to understand the expansion property of the set of subset sums $\Sigma(A) = \{\sum_{s\in S} s, S \subseteq A\}$, the central object is the set of "first order" almost periods, $P_d = \{x : |(A+x)\Delta A| \le d\}$. P_d is exactly the set of elements whose inclusion does not expand $\Sigma(A)$ significantly. Instead of showing P_d is large, we show that P_d cannot be too large. However, by using the stability of P_d under iterated sumsets, we transform the problem of understanding the much more complicated subset sums to iterated sumsets. In this way, we obtain very useful information on P_d, which allows us to show good expansion properties of $\Sigma(A)$, resolving a previous conjecture of Alon-Erdős [1] and several other questions on subset sums. It is also quite interesting that our bootstrapping procedure is quite similar to that in [2], using iterated sumsets to achieve a certain level of saturation. This procedure happens entirely in physical space instead of in the spectral domain as the bootstrapping procedure in Section 4.

Finally, we end the essay with several interesting future directions. Many of the questions are communicated to the author by Professor Julia Wolf and Dr. Thomas Bloom.

Upper bounds for almost periodicity.

We discussed in Theorem 49 and Theorem 51 the upper bounds on the set of almost periods, which are tight separately in p and in α, ϵ. The construction for the dependency on α and ϵ uses spectral information,

and therefore cannot be improved as p grows. It would be very interesting to get a joint dependency in p, α and ϵ, as in many applications, almost periodicity is used with p quite large.

Refining almost periodicity by passing to a dense subset.

Chang's lemma is essentially tight by Theorem 50. However, Bloom [5], by a sampling approach in the spectral domain, shows that one can improve Chang's lemma by passing to a large subset of the large spectrum. Bloom's version of Chang's lemma is used to deduce the best known bound on Roth's theorem. One can ask if the quantitative bound for almost periodicity can be significantly improved by passing to a subset of A. We note, however, that the construction of the lower bound in Theorem 49 showing the tight dependency on p yields an example that cannot be improved by passing to a constant density subset. We do not know if such an example exists when one is concerned with the dependency on α and ϵ. One can try sampling a random constant density subset of A, however, since this sampling procedure does not necessarily lead to a smoothing effect (as opposed to sampling in the spectral domain), it is not immediately clear that this would help.

Iterated almost periodicity for multiple convolutions.

As seen in the applications of almost periodicity, L^p almost periodicity on two-fold convolutions leads to L^∞ almost periodicity on three-fold convolutions by Holder's inequality. However, even by taking further convolutions, this procedure does not enlarge the set of almost periods. It is interesting to find a method that takes into account multiple convolutions in finding the set of almost periods, which hopefully leads to larger sets of almost periods. One possible approach is an iterative method that enlarges the set of almost periods, i.e., an almost periodicity result that improves as the sets involved carry more periodic structure. Corresponding upper bounds on the set of almost periods of multi-fold convolutions would also be very important.

A physical argument for higher uniformity norms.

As in the applications of almost periodicity to Roth's theorem, almost periodicity can be used to directly show that a function with large U^2 norm has density increment on a subspace or Bohr set, leading to essentially a physical proof of such result. It would be very interesting to know if there is an analog of almost periodicity that applies to higher uniformity norms.

References

[1] Alon, N. and P. Erdős. Sure monochromatic subset sums. *Acta Arith.* 74 (1976): 269–272. 9

[2] Alon, N., J. Fox and Y. Zhao. Efficient arithmetic regularity and removal lemmas for induced bipartite patterns, preprint. 1, 1, 8, 8, 8, 8.2, 8.2, 9

[3] Behrend, F. A. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.* (1946) 32:331–332. 1

[4] Ben-Sasson, E., N. Ron-Zewi, M. Tulsiani, and J. Wolf. Sampling-Based Proofs of Almost-Periodicity Results and Algorithmic Applications. *Proceedings of 41st International Colloquium on Automata, Languages and Programming (ICALP)* (2014): 955–966. 3, 3, 9

[5] Bloom, T. F. A quantitative improvement for Roth's theorem on arithmetic progressions. *J. Lond. Math. Soc. (2)* 93 (2016): 643–663. 1, 1, 4.1.1, 9

[6] Bloom, T. F. and O. Sisask. Logarithmic bounds for Roth's theorem via almost-periodicity. *Discrete analysis* (to appear). 1, 1, 1, 1, 3, 4, 5.1, 6.1, 6.2, 6.2.1

[7] Boucheron, S., O. Bousquet, and G. Lugosi. Concentration inequalities. In *Advanced Lectures in Machine Learning* (2004). Springer-Verlag, 1st edition. 8.1

[8] Bourgain, J. On arithmetic progressions in sums of sets of integers. In *A tribute to Paul Erdős* (1990). Cambridge Univ. Press, Cambridge, Cambridge. 4.1.1, 9

[9] Bourgain, J. On triples in arithmetic progression. *Geom. Funct. Anal.* 9 (1999): 968–984. 1, 1

[10] Bourgain, J. Roth's theorem on progressions revisited. *J. Anal. Math.* 104 (2008): 155–192. 1, 1

[11] Chang, M. C. *A polynomial bound in Freiman's theorem. Duke Math. J.* 113(3) (2002): 399–419. 1, 4.1.1

[12] Croot, E., V. Lev, and P. Pach. Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Ann. of Math.* 185 (2017): 331–337. 1, 5.1

[13] Croot, E. and O. Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geom. Funct. Anal.* 20 (2010): 1367–1396. 1, 1, 3

[14] Croot, E., I. Laba, and O. Sisask. Arithmetic progression in sumsets and L^p-almost-periodicity. *Combin. Probab. Comput.* 22(3) (2013): 351–365. 9

[15] Elkin, M. An improved construction of progression-free sets. In *Symposium on Discrete Algorithms* (2010): 886–905, arXiv:0801.4310. 1

[16] Ellenberg, J. S. and D. Gijswijt. On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. of Math.* 185 (2017): 339–343. 1, 5.1

[17] Fox, J., H. T. Pham and Y. Zhao. Tower-type bounds for Roth's theorem with popular differences, preprint. 1, 2.3

[18] Green, B. J. Some constructions in the inverse spectral theory of cyclic groups. Combin. Probab. Comput. 12(2) (2003): 127–138. 1, 4.1.1, 7, 7

[19] Green, B. J. Spectral structure of sets of integers. In *Fourier analysis and convexity* (2004): 83–96. Birkhauser Boston, Boston, MA. 4.1.1, 9

[20] Green, B. J. A Szemeredi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.* 15 (2005): 340–376. 9

[21] Green, B. J. Finite field models in additive combinatorics. In Bridget S Webb, editor, *Surveys in combinatorics* (2005): 1–27. Cambridge Univ. Press, Cambridge, Cambridge. 1

[22] Green, B. J. and T. Sanders. Boolean functions with small spectral norm. *Geom. Funct. Anal.* 18(1) (2008): 144–162. 4.1.1

[23] Green, B. J. and J. Wolf. A note on Elkin's improvement of Behrend's construction. In *Additive number theory: Festschrift in honor of the sixtieth birthday of Melvyn B. Nathanson* (2010): 141–144. Springer-Verlag, 1st edition. 1

[24] Haussler, D. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *J. Combin. Theory Ser. A* 69 (1995): 217–232. 8.1

[25] Heath-Brown, D. R. Integer sets containing no arithmetic progressions. *J. London Math. Soc.* 35 (1987): 385–394. 1

[26] Hosseini, K., S. Lovett, G. Moshkovitz, and A. Shapira. An improved lower bound for arithmetic regularity. *Math. Proc. Camb. Phil. Soc.* 161 (2016): 193–197. 9

[27] Impagliazzo, R., C. Moore, and A. Russell. An entropic proof of Chang's inequality. *SIAM J. Discrete Math.* 28(1) (2014): 173–176. 4.1.1

[28] Lee, J. Covering the large spectrum and generalized Riesz products. *SIAM J. Discrete Math.* 31(1) (2017): 562–572. 4.1.1

[29] Lovett, S. An exposition of Sanders quasi-polynomial Freiman-Ruzsa theorem. *Electronic Colloquium on Computational Complexity* 19 (2012). 9

[30] Meshulam, R. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A* 71(1) (1995):168–172. 1, 5.1

[31] Pham, H. T. Subset sums and a conjecture of Alon and Erdős, preprint. 9

[32] Roth, K. F. On certain sets of integers. *J. London Math. Soc.* 28 (1953): 104–109. 1, 5.1

[33] Rudin, W. Fourier analysis on groups. Wiley (1990). 4.1.1

[34] Sanders, T. Additive structures in sumsets. *Math. Proc. Cambridge Philos. Soc.* 144 (2008): 289–316. 9

[35] Sanders, T. On Roth's theorem on progressions. *Ann. of Math.* 174 (2011): 619–636. 1, 6.2, 6.2.2, 6.2.3, 6.2.3

[36] Sanders, T. On certain other sets of integers. *J. Anal. Math.* 116 (2012): 53–82. 1, 1, 1, 4, 4.1.1, 4.2.1

[37] Sanders, T. On the Bogolyubov–Ruzsa lemma, *Analysis & PDE* 5 (2012): 627–655. 8.2, 9

[38] Schoen, T. and O. Sisask. Roth's theorem for four variables and additive structures in sums of sparse sets. *Forum Math. Sigma* 4 (2016): e5 (28 pages). 1, 1, 1, 1, 5.1, 9

[39] Sisask, O. Convolutions of sets with bounded VC-dimension are uniformly continuous, preprint. 1, 1, 8, 8, 8, 8, 8.1, 9

[40] Szemeredi, E. Integer sets containing no arithmetic progressions. *Acta Math. Hungar.* 56 (1990): 155–158. 1

[41] Tao, T. and V. Vu. *Additive Combinatorics*, 1st ed. Cambridge University Press, 2006. 2.3, 4, 4.1.1

[42] Terry, C. and J. Wolf. Stable arithmetic regularity in the finite-field model, preprint. 8, 9

[43] Wolf, J. Finite field models in arithmetic combinatorics–ten years on. *Finite Fields Appl.* Vol. 32 (2015): 233–274. 1

[44] Wolf, J. Some applications of relative entropy in additive combinatorics. *Surveys in Combinatorics, London Mathematical Society Lecture Note Series* (2017): 409–442. Cambridge Univ. Press, Cambridge, Cambridge.

4.1.1