Title

Structure and Randomness in Complexity Theory and Additive Combinatorics

Author

Hosseini, Seyed Kaave

Publication Date

2019

UNIVERSITY OF CALIFORNIA SAN DIEGO

Structure and Randomness in Complexity Theory and Additive Combinatorics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Computer Science

by

Seyed Kaave Hosseini

Committee in charge:

Professor Shachar Lovett, Chair
Professor Sanjoy Dasgupta
Professor Russell Impagliazzo
Professor Daniele Micciancio
Professor Alireza Salehi Golsefidy

2019

The dissertation of Seyed Kaave Hosseini is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2019

DEDICATION

To my sister, Nasim.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

My sincerest thanks surely must go to my advisor, Shachar Lovett. I am greatly indebted to him for his endless support and kindness and generosity and patience with me. His attitude towards research and life has affected me deeply and forever. He has given me countless problems to work on and has taught me a vast number of ideas, encouraged me to work on good problems and has given me absolute freedom in following my interests, and has quite transparently demonstrated how to think about and learn from problems during hundreds of hours of enjoyable discussions. No acknowledgment would do justice of how great of a mentor Shachar has been and I am quite lucky that I have been able to work with him.

I also thank all my other collaborators for many valuable and illuminating discussions about various projects. It has been a great pleasure knowing and working with Divesh Aggarwal, Boaz Barak, Abhishek Bhrushundi, Eshan Chattopadhyay, Tim Gowers, Hamed Hatami, Omid Hatami, Pooya Hatami, Guy Moshkovitz, Daniel Kane, Pravesh Kothari, Sasha Knop, Sam McGuire, Shay Moran, Sankeerth Rao, Asaf Shapira, Madhur Tulsiani, Gregory Yaroslavstev, and David Zuckerman.

I also thank Noga Alon, Thomas Bloom, Jop Briet, Fan Chung Graham, Jacob Fox, Russell Impagliazzo, Nets Katz, Ryan Martin, Abhishek Methuku, Daniele Miccianco, Ivan Mikhailin, Alireza Salehi Golsefidy, Tom Sanders, Olof Sisask, Julia Wolf, Jiapeng Zhang, and many others for valuable and instructive discussions at various points. I also thank organizers of many programs and workshops at Simons Institute for Theory of Computing, American Mathematical Institute, Center of Mathematical Sciences and Applications, and Joint Mathematics Meetings, for inviting me and giving me the opportunity to meet and learn from amazing people.

I thank all of the CSE family for creating a warm and productive atmosphere, especially members of the theory group for creating an exceptionally friendly environment. I am also grateful to my good friends for their support and many fun adventures we have gone on together.

Finally, I am indebted to my dear family, my Mom and Dad, Homa and Hamed, and my sister, Nasim, for their constant love and encouragement, and for believing in me. I am deeply

sorry that I have not been able to see them for the duration of my PhD because of all sorts of travel bans and visa restrictions. However, hearing their voice has always been giving me energy and hope. I am dedicating this dissertation to my dear sister, Nasim who has always inspired me with her energy and enthusiasm.

Chapter 2 in part, contains a reprint of material from several papers that are being discussed in further detail in upcoming chapters as follows.

Chapter 3 and Section 2.2.2 contain a reprint of material as it appears in Discrete Analysis 2019:10, 14 pp. Kaave Hosseini and Shachar Lovett. "A bilinear Bogolyubov-Ruzsa lemma with poly-logarithmicbounds." The dissertation author was a primary investigator and author of this paper.

Chapter 4 and Section 2.2.3 contain a reprint of material as it appears in *SIAM Journal on Computing, 47, no. 1 (2018): 208-217.* Hamed Hatami, Kaave Hosseini, and Shachar Lovett. "Structure of protocols for XOR functions." The dissertation author was a primary investigator and author of this paper.

Chapter 5 and Section 2.2.4 contain a reprint of material as it appears in Journal ofCombinatorial Theory, Series A 148 (2017): 1-14. Kaave Hosseini and Shachar Lovett. "On the structure of the spectrum of small sets." The dissertation author was a primary investigator and author of this paper.

Chapter 6 and Section 2.3.1 contain a reprint of material as it appears in Mathematical Proceedings of the Cambridge Philosophical Society,vol. 161, no. 2, pp. 193-197. Cambridge University Press, 2016. Kaave Hosseini , Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. "An improved lower bound for arithmetic regularity." The dissertation author was a primary investigator and author of this paper.

Chapter 7 and Section 2.3.2 contain a reprint of material as it appears in 33rd Computational Complexity Conference (CCC2018). Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. "Pseudorandomgenerators from polarizing random walks." The dissertation author was a primary investigator and author of this paper.

VITA

2013	B. S. in Mathematics and in Computer Science, Sharif University of Technology, Tehran
2013-2019	Graduate Research Assistant, University of California San Diego
2019	Ph. D. in Computer Science, University of California San Diego

PUBLICATIONS

Hatami, Hamed, Kaave Hosseini, and Shachar Lovett. "Structure of protocols for XOR functions." SIAM Journal on Computing 47, no. 1 (2018): 208-217.

Chattopadhyay, Eshan, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. "Pseudorandom generators from polarizing random walks." In 33rd Computational Complexity Conference (CCC 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Hosseini, Kaave, Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. "An improved lower bound for arithmetic regularity." In Mathematical Proceedings of the Cambridge Philosophical Society, vol. 161, no. 2, pp. 193-197. Cambridge University Press, 2016.

Hosseini, Kaave, and Shachar Lovett. "On the structure of the spectrum of small sets." Journal of Combinatorial Theory, Series A 148 (2017): 1-14.

Hosseini, Kaave, and Shachar Lovett. "A bilinear Bogolyubov-Ruzsa lemma with poly-logarithmic bounds." Discrete Analysis 2019:10, 14 pp.

ABSTRACT OF THE DISSERTATION

Structure and Randomness in Complexity Theory and Additive Combinatorics

by

Seyed Kaave Hosseini

Doctor of Philosophy in Computer Science

University of California San Diego, 2019

Professor Shachar Lovett, Chair

This dissertation involves the interplay between structure, randomness, and pseudorandomness in theoretical computer science and additive combinatorics. Such interplay in particular materializes when one is extracting algebraic structure in scenarios where only weak combinatorial information is available. We develop new tools to address some problems of this type where the objects are sumsets and its bilinear generalizations, set of large Fourier spectra, and protocols in communication complexity. Later we move on to constructions of objects with certain pseudorandom properties. We construct a highly irregular set showing the limits of regularity lemma in the algebraic setting which is a major tool in pseudorandomness. Moreover, we introduce a new framework to construct pseudorandom generators and give some applications.

Chapter 1

Overview

The contribution of this dissertation is two-fold. The first part involves some problems in additive combinatorics and computer science that have to do with extracting algebraic structure in situations where only weak structural information is available. The second part is about pseudorandomness and explicit constructions in computer science and combinatorics. In the following, we give a summary of the contributions of this dissertation .

First we discuss the problems that involve extracting algebraic structure. A fundamental result in additive combinatorics is Bogolyubov-Ruzsa lemma. In particular the quantitative bounds obtained by Sanders, plays a central role in obtaining effective bounds for the inverse U^3 theorem for the Gowers norms. Recently, Gowers and Milićević [GM17b] applied a bilinear Bogolyubov-Ruzsa lemma as part of a proof of the inverse U^4 theorem with effective bounds. Here we obtain quantitative bounds for the bilinear Bogolyubov-Ruzsa lemma which are similar to those obtained by Sanders for the Bogolyubov-Ruzsa lemma. We show that if a set $A \subset \mathbb{F}^n \times \mathbb{F}^n$ has density α, then after a constant number of horizontal and vertical sums, the set A would contain a bilinear structure of co-dimension $r = \log^{O(1)} \alpha^{-1}$. This improves the results of Gowers and Milićević [GM17a] which obtained similar results with a weaker bound of $r = \exp(\exp(\log^{O(1)} \alpha^{-1}))$ and by Bienvenu and Lê [BL17] which obtained $r = \exp(\exp(\exp(\log^{O(1)} \alpha^{-1})))$. This work is published in [HL19].

Another instance of extracting algebraic structure which has close connections to sumsets is in the study of communication protocols for XOR functions. Let f be a boolean function on n variables. Its associated XOR function is the two-party function $f_{\oplus}(x,y) = f(x \oplus y)$. We show that, up to polynomial factors, the deterministic communication complexity of $f_{\oplus}$ is equal to the parity decision tree complexity of f. we develop a variation of the energy increment technique to study protocols for XOR functions. Most of previous techniques in communication complexity are local in the sense that they work by isolating one combinatorial rectangle. Here we combine tools such as Bogolyubov-Ruzsa theorem with a novel entropy decrements method to study the global structure of the protocols. This work is published in [HHL18].

The final contribution in extracting structure involves study of combinatorial structure of the set of large Fourier coefficients of subsets of abelian groups. Let G be a finite abelian group and A a subset of G. The spectrum of A is the set of its large Fourier coefficients. Known combinatorial results on the structure of spectrum, such as Chang's theorem, become trivial in the regime $|A| = |G|^{\alpha}$ whenever $\alpha \leq c$, where $c \geq 1/2$ is some absolute constant. On the other hand, there are statistical results, which apply only to a noticeable fraction of the elements, which give nontrivial bounds even to much smaller sets. One such theorem (due to Bourgain) goes as follows. For a noticeable fraction of pairs γ_1, γ_2 in the spectrum, $\gamma_1 + \gamma_2$ belongs to the spectrum of the same set with a smaller threshold. Here we show that this result can be made combinatorial by restricting to a large subset. That is, we show that for any set A there exists a large subset A', such that the sumset of the spectrum of A' has bounded size. Our results apply to sets of size $|A| = |G|^{\alpha}$ for any constant $\alpha > 0$, and even in some sub-constant regime. Proving this result also involves advancing the energy increment method. Here we deal with a process involving two energy functions where we want to increase both, but increasing one might result in decreasing the other. We developed some delicate analysis to show that the process needs to stop after a few iterations. This work is published [HL17].

The second part of the dissertation involves two problems related to pseudorandomness. The first one is about regularity lemma in the algebraic setting which is a fundamental tool. Arithmetic regularity lemma due to Green [Gre05b] is an analog of Szemerédi's regularity lemma. Similar to Gowers' tower type lower bound for Szemerédi's regularity lemma, Green proved a lower bound of tower of 2's of height $\log 1/\varepsilon$ for arithmetic regularity lemma [Gre05b]. We show a lower bound of tower of 2's of height $1/\sqrt{\varepsilon}$; So it's similar to Szemerédi's regularity lemma in this sense. This work is published in [HLMS16].

Finally, we consider explicit construction of pseudorandom generators. We propose a new framework for constructing pseudorandom generators for n-variate Boolean functions. It is based on two new notions. First, we introduce fractional pseudorandom generators, which are pseudorandom distributions taking values in $[-1,1]^n$. Next, we use a fractional pseudorandom generator as steps of a random walk in $[-1,1]^n$ that converges to $\{-1,1\}^n$. We prove that this random walk converges fast (in time logarithmic in n) due to polarization. As an application, we construct pseudorandom generators for Boolean functions with bounded Fourier tails. We use this to obtain a pseudorandom generator for functions with sensitivity s, whose seed length is polynomial in s. Other examples include functions computed by branching programs of various sorts or by bounded depth circuits. This work is published in [CHHL18].

1.1 Organization

This dissertation is organized as follows. First in Chapter 2 we provide the necessary background and motivation, and introduce the results of the dissertation. The actual proofs are deferred to chapters 3-7.

Chapter 2

Background and Introduction

In this chapter we provide the necessary background, history, and motivation. Almost all the results of this dissertation have to do with Fourier analysis over finite abelian groups and therefore, before proceeding to the main results, we give a brief introduction to Fourier analysis in section 2.1.

2.1 Basics of Fourier Analysis

Most of the results of this dissertation is concerened with the groups $\mathbb{F}_2^n$ or more generally $\mathbb{F}_q^n$, the n-dimensional vector space over the finite field $\mathbb{F}_q$. However, in some chapters such as chapter 5 we discuss all abelian groups and so we discuss Fourier analysis over all abelian groups in this section.

Let G be a finite abelian group and let $L^2(G)$ be the inner product space of all complex-valued functions $f : G \to \mathbb{C}$ where the inner product of $f, f' : G \to \mathbb{C}$ is defined by

$$\langle f, f' \rangle = \mathop{\mathbb{E}}_{x \in G} f(x)\overline{f'(x)}$$

where $\overline{f'(x)}$ is the complex-conjugate of $f'(x)$. A *linear character* $\gamma : G \to \mathbb{C}^\times$ of G is a multiplicative homomorphism to the group $\mathbb{C}^\times$. The dual group of G, denoted by $\widehat{G}$, is the group of all linear characters of G. The group $\widehat{G}$ has the group structure introduced by $(\gamma_1+\gamma_2)(x)=\gamma_1(x)\gamma_2(x)$ and is isomorphic to G. Moreover, it is easy to see that $\widehat{G}$ forms an orthonormal basis for $L^2(G)$. Therefore given any function $f : G \to \mathbb{C}$, we can write f in its *Fourier basis* as

$$f(x) = \sum_{\gamma\in\widehat{G}} \langle f,\gamma\rangle \gamma(x).$$

Denote the Fourier coefficient $\langle f,\gamma\rangle$ by $\widehat{f}(\gamma)$ and so we have

$$f(x) = \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\gamma(x).$$

In the special case of $f : \mathbb{F}_2^n \to \mathbb{R}$, we can simplify the notation to

$$\widehat{f}(\gamma) = \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n} f(x)(-1)^{\langle x,\gamma\rangle}$$

where $\langle x,\gamma\rangle$ is computing the inner product over $\mathbb{F}_2^n$. The function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by $\gamma \mapsto \langle f,\gamma\rangle$ is called the *Fourier transform* of f. We endow the functions in the physical space with uniform measure, and functions in the Fourier space with counting measure. That is, given $f : G \to \mathbb{C}$ and $p > 0$, let

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

Moreover for $g : \widehat{G} \to \mathbb{C}$,

$$\|g\|_p = \left(\sum_{\gamma\in\widehat{G}} |g(\gamma)|^p\right)^{\frac{1}{p}}.$$

The inner product is also defined for $g, g' : \widehat{G} \to \mathbb{C}$ by

$$\langle g,g'\rangle = \sum_{x\in G} f(x)\overline{f'(x)}.$$

Given this , one can check the *Plancherel identity*,

$$\langle f, f' \rangle = \langle \widehat{f}, \widehat{f'} \rangle$$

for given $f, f' : G \to \mathbb{C}$. In the special case where $f = f'$, this is called *Parseval's identity* which can be restated as

$$\|f\|_2 = \|\widehat{f}\|_2.$$

Convolution A particularly useful operation especially when working with sumsets is *convolution*. The convolution of f and f' is a function denoted by $f * f' : G \to \mathbb{C}$ and is defined by

$$f * f'(x) = \mathop{\mathbb{E}}_{y \in G} f(x-y) f'(y).$$

One can check that Fourier transform turns convolution into multiplication, namely $\widehat{f * f'} = \widehat{f} \cdot \widehat{f'}$. Convolution has a particularly useful meaning when one is dealing with sumsets. Suppose we have a subset $A \subset G$. In this case we make a small abuse of notation and show the indicator function of A by A instead of 1_A. Note that, $A * A(x)$ equals (up to a normalization factor of $|G|$) the number of ways one can write x as the sum $a + a' = x$ for $a, a' \in A$. Particularly, we obtain an analytic description of the combinatorial notion of sumset. Let $A + A = \{a + a' : a, a' \in A\}$. Then we have

$$A + A = \{x : A * A(x) > 0\}.$$

Convolution of sets also may be interpreted as *measure of relative density*. Suppose we have two sets $A, B \subset G$ and define the function $\varphi_B = \frac{|G|}{|B|} 1_B$. Then for every x we have

$$\frac{|A \cap (B + x)|}{|B|} = A * \varphi_B(x).$$

We finally introduce a useful fact, which is *uncertainty principle*.

Lemma 2.1.1 (Uncertainty principle)**.** *For any function $f : G \to \mathbb{C}$ that is not the constant zero* *function, we have*

$$|\mathrm{supp}(f)| \cdot |\widehat{\mathrm{supp}(f)}| \geq |G|.$$

Proof. Using Parseval's identity,

$$\|f\|_2^2 = \|\widehat{f}\|_2^2 \leq \sup_{\gamma \in \widehat{G}}(|\widehat{f}(\gamma)|) \cdot \sum_{\gamma \in \widehat{G}} |\widehat{f}(\gamma)| = \mathop{\mathbb{E}}_{x \in G} |f(x)| \cdot \sum_{\gamma \in \widehat{G}} |\widehat{f}(\gamma)|.$$

Moreover, using Cauchy-Schwartz we have that $\mathbb{E}_{x \in G} |f(x)| \leq \sqrt{\frac{|\mathrm{supp}(f)|}{|G|}} \cdot \|f\|_2$ and $\sum_{\gamma \in \widehat{G}} |\widehat{f}(\gamma)| \leq \sqrt{\mathrm{supp}(\widehat{f})} \cdot \|\widehat{f}\|_2$. This finishes the proof. □

2.2 Approximate Structure

The first part of the dissertation is related to approximate algebraic structure which is a fundamental part of theoretical computer science. Understanding notions of approximate linear maps, groups, polynomials, and so on have deep applications in a wide range of areas. In particular, it is quite useful to be able to extract algebraic information from a given object, where only some weak combinatorial information has been provided. This will open the way to being able to choose from a wide variety of algebraic tools to say develop algorithms, or prove lower bounds for the given object. It turns out, that the notion of *sumsets* in additive combinatorics provides a flexible enough theory to deal with a diverse variety of objects such as approximate groups, linear and multilinear maps, polynomials, communication protocols, linear sketching, and so on. Here we start with the basics of sumsets and its applications. Later on, we consider a generalization of the notion of sumset that has recently played a crucial role in obtaining quantitative bounds for inverse U^4 theorem for Gowers' norm. Then we move on to communication complexity and the applications of structure theory of sumsets in communication complexity of XOR functions. Finally we study the sumset of the set of large Fourier coefficients of sparse sets.

2.2.1 Standard theory of sumsets

We start with introducing the notion of approximate subgroup. Suppose we have an abelian group G, and A is a subset of G. On one hand, if A was a subgroup, then we would know that by definition $\forall x, y \in A$, $x + y \in A$. However, typically we are dealing with a subset A that doesn't satisfy the rigid structure of a subgroup, and only satisfies it approximately, and yet we would like to extract some algebraic conclusion about A. There are two ways to formalize the notion of approximate subgroup which both turn out to be equivalent up to polynomial factors. The first involves the notion of additive energy and is more statistical and weaker compared to the second definition which involves the notion of doubling and is more combinatorial. We start with additive energy.

Definition 2.2.1. *Let $A \subset G$. Then the additive energy of A is defined by*

$$E(A) = |\{(a,b,c,d) : a - b = c - d, a,b,c,d \in A\}|.$$

Observe that if A is a coset of a subgroup of G, then $E(A) = |A|^3$. The second important definition is that of doubling constant. First define the sumset

$$A + A = a + a' : a, a' \in A.$$

The *doubling constant* of A is defined by $\frac{|A+A|}{|A|}$. Moreover, note that if A has doubling constant K, then its additive energy $E(A) \geq \Omega(K^{-1})|A|^3$. The following result initially due to Balog and Szemerédi [BS94] and with polynomial bounds obtained by Gowers [Gow01] establishes an equivalence between the two. The strongest bound is due to Schoen [Sch15] which we state in the following without proof.

Theorem 2.2.2 (BSG theorem[Sch15])**.** *Let $\alpha \in (0,1)$ and A be a subset of an abelian group such* *that* $E(A) = \alpha|A|^3$ *. Then there exists* $A' \subset A$ *with* $|A'| \geq \alpha|A|$ *so that*

$$|A' - A'| \leq \alpha^{-4}|A'|$$

Given the BSG theorem, we can focus on the study of the structure of sets with bounded doubling. Let K be a constant and $H \leq G$ be a subgroup. Observe that if A is a subset of H with $|A| \geq K^{-1}|H|$ then we would have $|A+A| \leq K|A|$. A natural question if we can have an inverse to this. Suppose that we have a arbitrary subset $A \subset G$ satisfying $\frac{|A+A|}{|A|} \leq K$. The so called inverse theorems are formulating that in this case A should have a structure close to a subgroup or some other nice sets such as Arithmetic progressions over the integers. Exactly formalizing such a statement depends on the ambient group G, so here for the sake of simplicity we assume that the ambient group is $G = \mathbb{F}_p^n$, the n-dimensional vector space over $\mathbb{F}_p$. Observe that in $\mathbb{F}_p^n$ the subgroups are the subspaces. The first way to formalize an inverse theorem is to show that the given set A is trapped inside a subspace whose size is not much larger than the size of A itself. This type of result was originally formulated and proved by Freiman [Fre73, Fre87] in the case of integers, (the ambient group being $\mathbb{Z}$). It was later generalized and improved in a long sequence of work [Ruz99, Kon08, Sch11, EZ12, EZL14]. We state the result in $\mathbb{F}_2^n$ due to [EZ12].

Theorem 2.2.3 (Freiman's theorem over $\mathbb{F}_2^n$ [EZ12])**.** *Suppose* $A \subset \mathbb{F}_2^n$ *and* $|A+A| \leq K|A|$*. Then*

$$\frac{|\langle A \rangle|}{|A|} \leq 2^{K^{1+o(1)}}.$$

One drawback of this result is that the bound on $\frac{|\langle A \rangle|}{|A|}$ is exponential on K and this is inevitable. An example showing this is taking A to be a union of a subspace of size K, and K independent vectors. However, in this example, a big portion of the set A namely the subspace part, does not expand, and therefore one may hope that similar thing happens in general. We would like to show that a big portion of the set A is trapped inside a subspace. In this direction, the first breakthrough bound was obtained by Schoen [Sch11] via Fourier analysis. Then Sanders [San12a] employed a fundamentally new tool due to Croot and Sisask [CS10] and obtained the following.

Theorem 2.2.4 (Quasi-polynomial Freiman-Ruzsa theorem [San12a])**.** *Suppose $A \subset \mathbb{F}_2^n$ and also $|A+A| \leq K|A|$. Then there is a subset $A' \subset A$ with $|A'| \geq L^{-1}|A|$ so that*

$$\frac{|\langle A' \rangle|}{|A'|} \leq L$$

where $L = 2^{O(\log^4 K)}$.

Moreover, using a bootstrapping argument due to Konyagin, Sanders [San12b] improved the bound $\log^4 K$ to $\log^{3+o(1)} K$. The famous polynomial Freiman-Ruzsa Conjecture speculates whether one can make the dependence on K to be a polynomial.

Conjecture 2.2.5 (Polynomial Freiman-Ruzsa conjecture)**.** *Suppose $A \subset \mathbb{F}_2^n$ and $|A+A| \leq K|A|$. Then there is a subset $A' \subset A$ with $|A'| \geq L^{-1}|A|$ so that*

$$\frac{|\langle A' \rangle|}{|A'|} \leq L$$

where $L = K^{O(1)}$.

It's known that one can not take the bound better than $L = K^{1.4}$ [GT09]. There is yet another family of results called Bogolyubov-Ruzsa type theorems that only deal with sets that have constant density but give a strong conclusion regarding structure of kA for $k = O(1)$. Sanders in fact proved the following result which by a standard method essentially due to Ruzsa implies the previously mentioned Theorem 2.2.4.

Theorem 2.2.6 (Bogolyubov-Ruzsa lemma [San12a])**.** *Suppose $A \subset \mathbb{F}_p^n$ and $|A| \geq K^{-1}|\mathbb{F}_p^n|$. Then there is a subspace V of co-dimension $O(\log^4 K)$ with $V \subset 4A$.*

It turns out that inverse theorems on the structure of sumsets can be used to study other approximate algebraic structures such as approximate linear maps and polynomials. In particular, one can directly prove an inverse theorem for linear maps with very good bounds. We don't discuss such inverse theorem here, although we will state and use it in section 3.1. On the other hand, inverse theorems for approximate degree-d polynomials is more technical and is the subject matter of the so called area of Higher order Fourier analysis, which was pioneered by Gowers [Gow98]. In general there are no good bounds for polynomials with degree more than 3. Nevertheless, it turns out that the inverse theorems for sumsets play a crucial role here as well. We discuss this in the next section.

2.2.2 Higher order theory of sumsets

One of the key ingredients in the proof of quantitative inverse theorem for Gowers U^3 norm over finite fields, due to Green and Tao [GT08] and Samorodnitsky [Sam07], is an inverse theorem on the structure of sumsets. More concretely, the tool that gives the best bounds is the improved Bogolyubov-Ruzsa lemma which we recall in the following.

Theorem 2.2.6 (Bogolyubov-Ruzsa lemma [San12a])**.** *Suppose $A \subset \mathbb{F}_p^n$ and $|A| \geq K^{-1}|\mathbb{F}_p^n|$. Then there is a subspace V of co-dimension $O(\log^4 K)$ with $V \subset 4A$.*

In fact the link between the inverse U^3 theorem and inverse sumset theorems is deeper. It was shown in [GT10, Lov12] that an inverse U^3 conjecture with polynomial bounds is equivalent to the polynomial Freiman-Ruzsa conjecture, one of the central open problems in additive combinatorics. Given this, one can not help but wonder whether there is a more general inverse sumset phenomena that would naturally correspond to quantitative inverse theorems for U^k norms. In a recent breakthrough, Gowers and Milićević [GM17b] showed that this is indeed the case, at least for the U^4 norm. They used a *bilinear* generalization of Theorem 2.2.6 to obtain a quantitative inverse U^4 theorem.

To be able to explain this result we need to introduce some notation. Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$. Define two operators, capturing subtraction on horizontal and vertical fibers as follows:

$$\phi_{\mathrm{h}}(A) := \{(x_1 - x_2, y) : (x_1, y), (x_2, y) \in A\},$$

$$\phi_{\mathrm{v}}(A) := \{(x, y_1 - y_2) : (x, y_1), (x, y_2) \in A\}.$$

Given a word $w \in \{\mathrm{h}, \mathrm{v}\}^k$ define $\phi_w = \phi_{w_1} \circ \ldots \circ \phi_{w_k}$ to be their composition. A *bilinear variety* $B \subset \mathbb{F}^n \times \mathbb{F}^n$ of co-dimension $r = r_1 + r_2 + r_3$ is a set defined as follows:

$$B = \{(x,y) \in V \times W : b_1(x,y) = \ldots = b_{r_3}(x,y) = 0\},$$

where $V, W \subset \mathbb{F}^n$ are subspaces of co-dimension r_1, r_2, respectively, and $b_1, \ldots, b_{r_3} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ are bilinear forms.

Gowers and Milićević [GM17a] and independently Bienvenu and Lê [BL17] proved the following, although [BL17] obtained a weaker bound of $r = \exp(\exp(\exp(\log^{O(1)} \alpha^{-1})))$.

Theorem 2.2.7 ([GM17a])**.** *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and let* $w = \mathrm{hhvvhh}$. *Then there exists a bilinear variety $B \subset \phi_w(A)$ of co-dimension* $r = \exp(\exp(\log^{O(1)} \alpha^{-1}))$.

To be fair, it was not Theorem 2.2.7 directly but a more analytic variant of it that was used (combined with many other ideas) to prove the inverse U^4 theorem in [GM17b]. However, we will not discuss that analytical variant here.

Here we improve the bound in Theorem 2.2.7 to $r = \log^{O(1)} \alpha^{-1}$. Our proof is arguably simpler and is obtained only by invoking Theorem 2.2.6 a few times, without doing any extra Fourier analysis. The motivation behind this work — other than obtaining the right form of bound — is to employ this result in a more algebraic framework to obtain a modular and simpler proof of an inverse U^4 theorem.

One more remark before explaining the result is that Theorem 2.2.7 generalizes Theorem 2.2.6 because given a set $A \subset \mathbb{F}^n$, one can apply Theorem 2.2.7 to the set $A' = \mathbb{F}^n \times A$ and find $\{x\} \times V \subset \phi_w(A')$ where x is arbitrary, and V a subspace of co-dimension $3r$. This implies $V \subset 2A - 2A$.

Theorem 2.2.8. *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and let* $w = \mathrm{hvvhvvvhh}$. *Then there exists a bilinear variety $B \subset \phi_w(A)$ of co-dimension $r = O(\log^{80} \alpha^{-1})$.*

Note that the choice of the word w in Theorem 2.2.8 is $w = \mathrm{hvvhvvvhh}$ which is slightly longer than in Theorem 2.2.7 being hhvvhh. However, for applications this usually does not matter and any constant length w would do the job. In fact allowing w to be longer is what enables us to obtain a result with a stronger bound.

A robust analog of Theorem 2.2.8

Going back to the theorem of Sanders, there is a more powerful variant of Theorem 2.2.6 which guarantees that V enjoys a stronger property rather than just being a subset of $2A-2A$. The stronger property is that every element $y \in V$ can be written in many ways as $y = a_1 + a_2 - a_3 - a_4$, with $a_1, a_2, a_3, a_4 \in A$. This stronger property of V has a number of applications such as obtaining upper bounds for Roth theorem in four variables. We refer the reader to [SS16] where Theorem 3.2 is similarly obtained from Theorem 2.2.6 and also for the noted application.

Theorem 2.2.9 ([San12a, SS16])**.** *Let $A \subset \mathbb{F}^n$ be a subset of density α. Then there exists a subspace $V \subset 2A-2A$ of co-dimension $O(\log^4 \alpha^{-1})$ such that the following holds. Every $y \in V$ can be expressed as $y = a_1 + a_2 - a_3 - a_4$ with $a_1, a_2, a_3, a_4 \in A$ in at least $\alpha^{O(1)}|\mathbb{F}|^{3n}$ many ways.*

In Theorem 2.2.10 we also state a statistical analog of Theorem 2.2.9 by slightly modifying the proof of Theorem 2.2.8. To explain it, we need just a bit more notation.

Fix an arbitrary $(x,y) \in \mathbb{F}^n \times \mathbb{F}^n$, and note that (x,y) can be written as $(x,y) = \phi_{\mathrm{h}}((x+x_1,y),(x_1,y))$ for any $x_1 \in \mathbb{F}^n$. Moreover, for any fixed x_1, each of the points $(x+x_1,y),(x_1,y)$ can be written as $(x+x_1,y) = \phi_{\mathrm{v}}((x+x_1,y+y_1),(x+x_1,y_1))$ and $(x_1,y) = \phi_{\mathrm{v}}((x_1,y+y_2),(x_1,y_2))$ for arbitrary $y_1,y_2 \in \mathbb{F}^n$. So over all, the point (x,y) can be written using the operation ϕ_{vh} in exactly $|\mathbb{F}^n|^3$ many ways, namely, the total number of two-dimensional parallelograms $(x+x_1,y+y_1),(x+x_1,y_1),(x_1,y+y_2),(x_1,y_2)$ where (x,y) is fixed. We can continue this and consider an arbitrary word $w \in \{\mathrm{h},\mathrm{v}\}^k$. Then (x,y) can be written using the operation ϕ_w in exactly $|\mathbb{F}^n|^{2^k-1}$ many ways.

Now, we have a set $A \subset \mathbb{F}^n \times \mathbb{F}^n$ and fix a word $w \in \{\mathrm{h},\mathrm{v}\}^k$. Define $\phi_w^{\varepsilon}(A)$ to be the set of all elements $(x,y) \in \mathbb{F}^n \times \mathbb{F}^n$ that can be obtained in at least $\varepsilon|\mathbb{F}^n|^{2^k-1}$ many ways by applying the operation $\phi_w(A)$.

The following is an extension of Theorem 2.2.8 similar in spirit to Theorem 2.2.9.

Theorem 2.2.10. *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and $w = \mathrm{hvvhvvvhh}$ and $\varepsilon = \exp(-O(\log^{20}\alpha^{-1}))$. Then there exists a bilinear variety $B \subset \phi_w^{\varepsilon}(A)$ of co-dimension $r = O(\log^{80}\alpha^{-1})$.*

As a final comment, we remark that if one keeps track of dependence on the field size in the proofs, then the bound in Theorem 2.2.8 and Theorem 2.2.10 is $r = O(\log^{80}\alpha^{-1} \cdot \log^{O(1)}|\mathbb{F}|)$.

We prove theorems theorem 2.2.8 and theorem 2.2.10 in sections section 3.1 and section 3.2 respectively.

2.2.3 Communication Complexity of XOR functions

Basics of Communication Complexity Communication complexity is a surprisingly flexible theory to prove lower bounds in various uniform and non-uniform models of computation. In fact, communication complexity is used in contexts where there is no explicit computation happening at all, and one is working with a static mathematical structure such as a polytope. Communication complexity, originally introduced by Yao [Yao79] is defined as follows. Informally communication complexity captures or at least lower bounds the amount of information that has to be moved around inside a computation model to do a certain computation. The most basic definition is as follows. Say we have a function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. We have two parties Alice and Bob where Alice has input $x \in \{0,1\}^n$, and Bob has input $y \in \{0,1\}^n$, and they would like to jointly compute $F(x,y)$, by sending bits to each other. We assume that each party has full computational power as we are only concerned with the amount of communication between Alice and Bob. After fixing the function F, the parties can agree on a communication protocol so that they can jointly compute $F(x,y)$ for any input (x,y). The most trivial protocol is to have Alice send her entire input x to Bob and let Bob compute the function $F(x,y)$. However, the question is whether they can do better if the function F is not too complicated.

One may think of the function F, as a matrix, with the rows corresponding to inputs $x \in \{0,1\}^n$, and the columns corresponding to the inputs $y \in \{0,1\}^n$. A protocol Π corresponds to a specific type of binary tree for F as follows. For simplicity suppose that the protocol has the property that Alice and Bob every time send one bit and they alternate speaking. So Alice sends a bit at odd time steps and Bob sends a bit a even time steps. Alice and Bob start at the root of the tree, and Alice (depending her input x) sends a bit to Bob and they both go to the corresponding child of the root. Then Bob sends a bit to Alice and they both go the corresponding node. At the end, the reached leaf is labeled with $0,1$ which should be the output $f(x,y)$. Every leaf corresponds to a *monochromatic combinatorial rectangle* of the matrix F, namely an all-0 or all-1 submatrix of the form $A \times B$ where A,B are subsets of rows and columns respectively. After fixing a protocol Π, the *cost* of the protocol Π is the total number of the bits communicated between Alice and Bob in order to computer $F(x,y)$, maximized over x,y. The *communication complexity* of F denoted by $D(F)$ here, is the minimum cost of a deterministic protocol computing F. Note that $D(F)$ is always between 0 and n, since there is always the trivial protocol of Alice sending all of her input $x \in \{0,1\}^n$ to Bob.

There are several natural questions one may ask about communication complexity.

1. *Which functions have bounded communication complexity?*

2. *Is there a fast algorithm that computes an optimal or close to optimal protocol for a given function F?*

3. *What does such an optimal protocol look like? Can we obtain a rough classification of structure of protocols?*

All of these questions are wide open. A fundamental conjecture in the area is called Log-rank conjecture and is related to these questions. The Log-rank conjecture originally formulated by Lovasz and Saks [LS93], asks whether deterministic communication complexity of a matrix is roughly equivalent to logarithm of its rank over the real numbers. As a first observation, notice that if a matrix has communication complexity k, then it decomposes the matrix into at most 2^k monochromatic combinatorial rectangles and since each such monochromatic rectangle is a rank 1 matrix, then the rank of the given matrix is at most 2^k. The log-rank conjecture is that the reverse of this roughly holds.

Conjecture 2.2.11 (Log-rank conjecture [LS93]). *Is it true that for every boolean function $F : X \times Y \to \{0,1\}$,*

$$D(F) \leq \text{polylog}(rank(F))$$

where $D(\cdot)$ is the deterministic communication complexity and $rank(\cdot)$ is matrix rank over the reals.

Despite a great deal of effort, this conjecture is still wide open. The best known bound in this direction is the following due to Lovett [Lov14b].

Theorem 2.2.12 ([Lov14b]). *For every boolean function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$,*

$$D(F) \leq O(\sqrt{rank(F)} \log rank(F))$$

The purpose here is to make progress towards these questions by restricting to a class of matrices that have an algebraic structure. To do so, we identify $\{0,1\}^n$ with the vector space $\mathbb{F}_2^n$. Moreover suppose we have a function $f : \mathbb{F}_2^n \to 0,1$. We may define the matrix

$$f_{\oplus}(x,y) = f(x+y)$$

where $+$ is the addition operation of $\mathbb{F}_2^n$. Note that one may more generally, consider more general abelian groups instead of $\mathbb{F}_2^n$, however, here we are more concerned with the group $\mathbb{F}_2^n$ to simplify notation. This class of matrices is called XOR functions in the literature. The class of XOR functions has been studied in recent years, see [MO09, ZS10, TWXZ13, Zha14, STlV17]. This class of functions is sufficiently large to capture many interesting examples (e.g., equality and Hamming distance functions), but it is also especially attractive for it allows use of tools from discrete Fourier analysis. This is because the eigenvalues of $f_{\oplus}$ as a matrix are the same as the Fourier coefficients of f; therefore, the rank of $f_{\oplus}$ is equal to the *Fourier sparsity* of f, which is the number of non-zero Fourier coefficients of f. Moreover, if $A \times B \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ is a monochromatic rectangle in $f_{\oplus}$, then f is constant on all of $A+B$, where the sum-set $A+B$ is defined as $\{a+b : a \in A, b \in B\}$. This directly links communication complexity of XOR functions to the structure of sum-sets in additive combinatorics. We will discuss this relation in more detail later.

Going back to the log-rank conjecture for XOR functions, an interesting approach to settle the conjecture is via another complexity measure, called the *parity decision tree* complexity (PDT in short), denoted $\mathrm{pdt}(\cdot)$. A parity decision tree for a boolean function f is an extension of the usual notion of decision trees. While in a regular decision tree, intermediate nodes query variables, in a parity decision tree they are allowed to query an arbitrary linear function of the inputs. A depth-k parity decision tree for a boolean function f can be used to construct a $2k$-bit communication protocol for $f_{\oplus}(x,y)$. Indeed for every linear function L, since $L(x \oplus y) = L(x) \oplus L(y)$, Alice and Bob need to exchange only 2 bits to evaluate $L(x \oplus y)$. Hence they can simulate the PDT by exchanging only $2k$ bits, and thus $D(f_{\oplus}) \leq 2 \cdot \mathrm{pdt}(f)$.

In the opposite direction, since Fourier characters are exponentials of linear functions and f has Fourier sparsity at most $2^{D(f_{\oplus})}$, we have $\mathrm{pdt}(f) \leq 2^{D(f_{\oplus})}$. Our main interest in this work is whether this direction can be made efficient. Namely, is is true that an efficient deterministic protocol for an XOR function implies a polynomial-depth parity decision tree for the corresponding boolean function. Our main result is a polynomial relation between the two.

Theorem 2.2.13 ([HHL16]). *For any $f : \mathbb{F}_2^n \to \{0,1\}$ we have $\mathrm{pdt}(f) \leq O(D(f_{\oplus})^6)$.*

This theorem can be put into a broader context as follows. Consider a function $f : \{0,1\}^m \to \{0,1\}$ and a function $g : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ (called a gadget). Then one can define the composed function $f \circ g^m : \{0,1\}^{km} \times \{0,1\}^{km} \to \{0,1\}$ where $f \circ g^m(x,y) = f(g(x_1,y_1), \cdots, g(x_m,y_m))$, where $x = (x_1, \cdots, x_m) \in (\{0,1\}^k)^m$ and similarly for y. There exists several results in the literature that relate query complexity of f to communication complexity of $f \circ g^m$ when a gadget g (such as the inner product function) is chosen with $k \approx \log m$. This line was initiated by Raz and McKenzie in [RM97]; for example, see also [GPW15, WYY17, GPW17]. In this paper however, the gadget g is $\oplus$ which depends on two bits.

Open problems There are two natural open problems which stem directly from our work. The first is whether our result can be extended to randomized protocols vs randomized parity decision trees. This will be discussed in the next paragraph. The second question asks about what happens if we replace XOR with other gadgets. Sherstov [She11] showed that for many gadgets, including some natural 2-bit gadgets, efficient protocols imply low-degree approximating polynomials, which by the work of Nisan and Szegedy [NS94] imply efficient (standard) decision trees. This however does not hold for 1-bit gadgets. Except for XOR functions, the other class of gadgets that can be considered are AND gadgets (any other 1-bit gadget is either trivial or equivalent to either XOR or AND).

That is, for a boolean function $f : \{0,1\}^n \to \{0,1\}$ define its corresponding AND function as $f_\wedge(x,y) = f(x \wedge y)$, where $\wedge$ is bitwise AND function. An example of an AND function is disjointness. The analog class of decision trees are AND decision trees, where each internal node may query the AND of a subset of the inputs or their negations.

Problem 2.2.14. *Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a function. Assume that $f_\wedge$ has a deterministic protocol with complexity k. Does there exist a deterministic AND decision tree of depth* $\mathrm{poly}(k)$ *which computes f?*

Randomized Communication Complexity All of the questions that were asked in the the beginning of previous section regarding deterministic communication become significantly harder in the case of randomized communication. In particular the following concrete question seems elusive, which is a randomized analog of Theorem 2.2.13.

Problem 2.2.15. *Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a function. Assume that $f_\oplus$ has a randomized protocol with complexity k. Does there exist a randomized parity decision tree of depth* $\mathrm{poly}(k)$ *which computes f?*

One may simplify the model by considering one-way randomized communication complexity where each party speaks once when it is their turn. In this setting, one can classify the structure of one-way protocols for three parties and more. The reader is referred to [HLY18] for more details.

2.2.4 Structure of the Fourier spectrum of sparse sets

The objects that we were dealing with in the previous sections all were involving dense sets, i.e. a set whose size is at least a polynomial of the size of ambient space. However, for many applications one necessarily deals with sets that are much smaller, in particular we have no assumptions on the size of the set or if we do, the size of the set is sub-polynomial in the size of the ambient space. The purpose of this section to obtain some combinatorial information about the structure of Fourier coefficients of arbitrary small sets. The applications we have in mind are in computer science which we will briefly describe afterwards. In the following we give some basic structural information regarding the structure Fourier coefficients of sets and explain how they break down in the sparse setting.

One of the fundamental methods in analyzing several questions regarding subsets of abelian groups is by analyzing the set of large Fourier coefficients the indicator functions of the set. Let G be a finite abelian group, and let A be a subset of G. Fix a parameter $\varepsilon \in [0,1]$. For a character $\gamma \in \widehat{G}$, the corresponding Fourier coefficient of 1_A is

$$\widehat{1_A}(\gamma) = \sum_{x \in A} \gamma(x).$$

The spectrum of A is the set of characters with large Fourier coefficients,

$$\mathrm{Spec}_{\varepsilon}(A) = \{\gamma \in \widehat{G} : |\widehat{1_A}(\gamma)| \geq \varepsilon |A|\}.$$

Note that the spectrum of a set is a symmetric set, that is $\mathrm{Spec}_{\varepsilon}(A) = -\mathrm{Spec}_{\varepsilon}(A)$, where we view $\widehat{G}$ as an additive group (which is isomorphic to G).

Understanding the structure of the spectrum of sets is an important topic in additive combinatorics, with several striking applications discussed below. As we illustrate, there is a gap in our knowledge between *combinatorial* structural results, which apply to all elements in the spectrum, and *statistical* structural results, which apply to most elements in the spectrum. The former results apply only to large sets, typically of the size $|A| \geq |G|^c$ for some absolute constant $c > 0$, where the latter results apply also for smaller sets. The goal here is to bridge this gap.

Our interest in this problem originates from applications of it in computational complexity, where a better understanding of the structure of the spectrum of small sets can help to shed light on some of the main open problems in the area, such as constructions of two source extractors [Bou05b, Rao07, RB11] or the log rank conjecture in communication complexity [BLR12]. We refer the interested reader to a survey on applications of additive combinatorics in theoretical computer science [Lov14a].

We assume from now on that $|A| = |G|^\alpha$ where $\alpha > 0, \varepsilon > 0$ are arbitrarily small constants, which is the regime where current techniques fail. In fact, our results extend to some range of sub-constant parameters, but only mildly. First, we review the current results on the structure of the spectrum, and their limitations.

Size bound The most basic property of the spectrum is that it cannot be too large. Parseval's identity bounds the size of the spectrum by

$$|\mathrm{Spec}_\varepsilon(A)| \leq \frac{|G|}{\varepsilon^2|A|} = \frac{|G|^{1-\alpha}}{\varepsilon^2}.$$

However, this does not reveal any information about the structure of the spectrum, except from a bound on its size.

Dimension bound A combinatorial structural result on the spectrum was obtained by Chang in [Cha02]. She discovered that the spectrum is low dimensional. For a set $\Gamma \subseteq \widehat{G}$, denote its *dimension* as the minimal integer d, such that there exist $\gamma_1, \ldots, \gamma_d \in \widehat{G}$ with the following property: any element $\gamma \in \Gamma$ can be represented as $\gamma = \sum \varepsilon_i \gamma_i$ with $\varepsilon_i \in \{-1, 0, 1\}$. With this definition,

Chang's theorem asserts that

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

Chang [Cha02] used this result to obtain improved bounds for Freiman's theorem on sets with small doubling, and Green [Gre02] used it to find arithmetic progressions in sumsets. Moreover, Green [Gre04] showed that the bound in Chang's theorem cannot in general be improved, at least when A is not too small. Recently, Bloom [Blo] obtained sharper bounds for a large subset of the spectrum. He showed that there exists a subset $\Gamma \subseteq \mathrm{Spec}_{\varepsilon}(A)$ of size $|\Gamma| \geq \varepsilon \cdot |\mathrm{Spec}_{\varepsilon}(A)|$ such that

$$\dim(\Gamma) \leq O(\varepsilon^{-1}\log(|G|/|A|)).$$

He applied these structural results to obtain improved bounds for Roth's theorem and related problems. However, we note that in our regime of interest, where $|A| = |G|^{\alpha}$ with $0 < \alpha < 1$, both results become trivial if ε is a small enough constant. This is because both give a bound on the dimension of the form $O(\varepsilon^{-c}(1-\alpha)) \cdot \log|G|$ with $c \in \{1,2\}$. However, any set $\Gamma \subseteq \widehat{G}$ trivially has dimension at most $\log|G|$. As our interest is in the regime of any arbitrarily small constant $\alpha, \varepsilon > 0$, we need to turn to a different set of techniques.

Statistical doubling Bourgain [Bou05a] showed that for many pairs of elements in the spectrum, their sum lands in a small set. Concretely,

$$\Pr_{\gamma_1,\gamma_2 \in \mathrm{Spec}_{\varepsilon}(A)}[\gamma_1 + \gamma_2 \in \mathrm{Spec}_{\varepsilon^2/2}(A)] \geq \varepsilon^2/2,$$

where we note that by Parseval's identity, $|\mathrm{Spec}_{\varepsilon^2/2}(A)| \leq O(|G|^{1-\alpha}/\varepsilon^4)$. He used these results to obtain improved bounds on exponential sums. Similar bounds can be obtained for linear combinations of more than two elements in the spectrum, for example as done by Shkredov [Shk08]. If we assume that $|\mathrm{Spec}_{\varepsilon^2/2}(A)| \leq K|\mathrm{Spec}_{\varepsilon}(A)|$ and apply the Balog-Szemerédi-Gowers theorem [BS94, Gow98], this implies that there exists a large subset $\Gamma \subseteq \mathrm{Spec}_{\varepsilon}(A)$ such that $|\Gamma + \Gamma| \leq$

$(K/\varepsilon)^{O(1)}|\Gamma|$. However, it does not provide any bounds on the sumset of the entire spectrum, that is on $|\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A)|$. In fact, we will later see an example showing that this sumset could be much large than the spectrum, whenever $\varepsilon \le 1/2$.

Combinatorial doubling The motivating question for the current work is to understand whether the statistical doubling result described above, can be applied for the entire spectrum. That is, can we obtain combinatorial structural results on the sumset of the entire spectrum $\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A)$.

As a first step, we ask for which $\alpha,\varepsilon > 0$ is is true that, for any set A of size $|A| = |G|^\alpha$, the sumset $\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A)$ is much smaller than the entire group. There are two regimes where this is trivially true. First, when $\alpha > 1/2$, it is true since by Parseval's identity, $\mathrm{Spec}_\varepsilon(A)$ is smaller than the square root of the group size, and hence

$$|\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A)| \le |\mathrm{Spec}_\varepsilon(A)|^2 \le \frac{|G|^{2-2\alpha}}{\varepsilon^4}.$$

Also, when $\varepsilon > 1/2$ then $\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A) \subseteq \mathrm{Spec}_{2\varepsilon-1}(A)$ (see, e.g., [TV06] for a proof) and hence again by Parseval's identity, the size of the sumset is bounded by

$$|\mathrm{Spec}_\varepsilon(A)+\mathrm{Spec}_\varepsilon(A)| \le |\mathrm{Spec}_\varepsilon(A)|^2 \le \frac{|G|^{1-\alpha}}{(2\varepsilon-1)^2}.$$

As the following example shows, the thresholds of $\alpha = 1/2, \varepsilon = 1/2$ are tight.

Example 2.2.16. *Let $G = \mathbb{Z}_2^{2n}$ and $A = (\mathbb{Z}_2^n \times \{0^n\}) \cup (\{0^n\} \times \mathbb{Z}_2^n)$. Then $|A| = 2|G|^{1/2} - 1$, $\mathrm{Spec}_{1/2}(A) = A$ and $A + A = G$.*

So, it seems that such structural results are hopeless when $\alpha,\varepsilon < 1/2$. However, there is still hope: in the example, if we restrict to a large subset $A' = \mathbb{Z}_2^n \times \{0^n\} \subseteq A$, then $\mathrm{Spec}_{1/2}(A') = \{0^n\} \times \mathbb{Z}_2^n$ is a subgroup, and specifically the size of $\mathrm{Spec}_{1/2}(A') + \mathrm{Spec}_{1/2}(A')$ is bounded away from the entire group. Our first result is that this is true in general. In fact, the size of the sumset is close to the bound given by Parseval's identity, which is approximately $|G|^{1-\alpha}$.

Theorem 2.2.17. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ such that*

$$\left|\mathrm{Spec}_{\varepsilon}(A') + \mathrm{Spec}_{\varepsilon}(A')\right| \leq (1/\varepsilon)^{O(1/\delta)} \cdot \frac{|G|^{1+\delta}}{|A'|}$$

where $C \leq \exp((1/\varepsilon)^{O(1/\delta)})$.

A more refined notion of structure is that of bounded doubling. Here, we say that a set Γ has a doubling constant K if $|\Gamma+\Gamma| \leq K|\Gamma|$. Note that if $|\mathrm{Spec}_{\varepsilon}(A')|$ has size close to the bound given by Parseval's identity, which is roughly $|G|^{1-\alpha}$, then Theorem 2.2.17 would show that $\mathrm{Spec}_{\varepsilon}(A')$ has a small doubling constant $K = C|G|^{\delta}$. We conjecture that this is always the case. However, we could only show it if we are allowed to change the value of ε somewhat. We state both the theorem and the conjecture below.

Theorem 2.2.18. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ and $\varepsilon' \geq \varepsilon^{2^{1/\delta}}$ such that*

$$|\mathrm{Spec}_{\varepsilon'}(A')| \geq |\mathrm{Spec}_{\varepsilon}(A)|/C$$

and

$$|\mathrm{Spec}_{\varepsilon'}(A') + \mathrm{Spec}_{\varepsilon'}(A')| \leq C|G|^{\delta} \cdot |\mathrm{Spec}_{\varepsilon'}(A')|,$$

where $C \leq \exp\left((1/\varepsilon)^{O(2^{4/\delta})}\right)$.

Conjecture 2.2.19. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ such that*

$$|\mathrm{Spec}_{\varepsilon}(A') + \mathrm{Spec}_{\varepsilon}(A')| \leq C|G|^{\delta} \cdot |\mathrm{Spec}_{\varepsilon}(A')|,$$

where $C = C(\varepsilon, \delta)$.

The original motivation for establishing these results was in pseudorandomness and in particular that the inner produce function over $\mathbb{F}_p^n$ is a certain kind of extractor called affine-malleable extractor. However, the results of this section fall short of achieving this application which requires a stronger conjecture regarding the doubling of the spectrum to be true. We refer the reader to [AHL16] for more details and some partial results.

2.3 Pseudorandomness

In the next two sections we discuss two different aspects of pseudorandomness. The first is the notion of regularity of sets and structure vs randomness dichotomy. The second involves explicit constructions of pseudorandom sets.

2.3.1 Limits of Regularity Lemma

In many applications, obtaining some global information about a given object, say a graph is necessary. For example, suppose we have a big dense graph and would like to estimate the number of triangles in the graph. There are two kinds of graphs for which we know how to easily estimate the number of triangles. One is a complete graph (or a graph with a fixed constant weight on all its edges). Another example is a random graph. A powerful method to handle a given arbitrary graph G is to decompose it into two parts, G_{str} and G_{psd} where G_{str} is a structured part of low complexity, similar to a complete graph, and G_{psd} is pseudorandom. We might possibly have a third error component G_{err}. Then one can estimate the number of triangles in each part separately using known methods and combine them. This decomposition is achieved by Szemerédi's regularity lemma which is a fundamental tool in combinatorics and pseudorandomness. As an analogue of Szemerédi's regularity lemma in graph theory [Sze75], Green [Gre05b] proposed an arithmetic regularity lemma for abelian groups. Given an abelian group G and a bounded function $f : G \to [0,1]$, Green showed that one can find a subgroup $H \leq G$ of bounded index, such that when restricted to most cosets of H, the function f is pseudorandom in the sense that all of its nonzero

Fourier coefficients are small. Quantitatively, the index of H in G is bounded by a tower of twos of height polynomial in the error parameter. The goal of this note is to provide an example showing that these bounds are essentially tight. This strengthens a previous example due to Green [Gre05b] which shows that a tower of height logarithmic in the error parameter is necessary; and makes the lower bounds in the arithmetic case analogous to these obtained in the graph case [Gow97].

We restrict our attention here to the group $G = \mathbb{Z}_2^n$, and note that our construction can be generalized to groups of bounded torsion in an obvious way. We first make some basic definitions. Let A be an affine subspace (that is, a translation of a vector subspace) of $\mathbb{Z}_2^n$ and let $f : A \to [0,1]$ be a function. The Fourier coefficient of f associated with $\eta \in \mathbb{Z}_2^n$ is

$$\widehat{f}(\eta) = \frac{1}{|A|}\sum_{x\in A} f(x)(-1)^{\langle x,\eta\rangle} = \mathop{\mathbb{E}}_{x\in A}\left[f(x)(-1)^{\langle x,\eta\rangle}\right] .$$

Any subspace $H \le \mathbb{Z}_2^n$ naturally determines a partition of $\mathbb{Z}_2^n$ into affine subspaces

$$\mathbb{Z}_2^n/H = \{H + g \, : \, g \in \mathbb{Z}_2^n\} .$$

The number $\left|\mathbb{Z}_2^n/H\right| = 2^{n-\dim H}$ of translations is called the *index* of H.

For an affine subspace $A = H + g$ of $\mathbb{Z}_2^n$, where $H \le \mathbb{Z}_2^n$ and $g \in \mathbb{Z}_2^n$, we say that a function $f : A \to [0,1]$ is *ε-regular* if all its nontrivial Fourier coefficients are bounded by ε, that is,

$$\max_{\eta\notin H^{\perp}} \left|\widehat{f}(\eta)\right| \le \varepsilon .$$

Note that a trivial Fourier coefficient $\eta \in H^{\perp}$ satisfies $|\widehat{f}(\eta)| = |\mathbb{E}_{x\in A} f(x)|$. Henceforth, for any $f : \mathbb{Z}_2^n \to [0,1]$ we denote by $f|_A : A \to [0,1]$ the restriction of f to A.

Definition 2.3.1 (ε-regular subspace)**.** *Let $f : \mathbb{Z}_2^n \to [0,1]$. A subspace $H \le \mathbb{Z}_2^n$ is* ε-regular *for f if $f|_A$ is ε-regular for at least $(1-\varepsilon)\left|\mathbb{Z}_2^n/H\right|$ translations A of H.*

Green [Gre05b] proved that any bounded function has an ε-regular subspace H of bounded index, that is, whose index depends only on ε (equivalently, H has bounded codimension). In the following, $\operatorname{twr}(h)$ is a tower of twos of height h; formally, $\operatorname{twr}(h) := 2^{\operatorname{twr}(h-1)}$ for a positive integer h, and $\operatorname{twr}(0) = 1$.

Theorem 2.3.2 (Arithmetic regularity lemma in $\mathbb{Z}_2^n$, Theorem 2.1 in [Gre05b]). *For every $0 < \varepsilon < \frac{1}{2}$ there is $M(\varepsilon)$ such that every function $f : \mathbb{Z}_2^n \to [0,1]$ has an ε-regular subspace of index at most $M(\varepsilon)$. Moreover, $M(\varepsilon) \leq \operatorname{twr}(\lceil 1/\varepsilon^3 \rceil)$.*

A lower bound on $M(\varepsilon)$ of about $\operatorname{twr}(\log_2(1/\varepsilon))$ was given in the same paper [Gre05b], following the lines of Gowers' lower bound on the order of ε-regular partitions of graphs [Gow97]. While Green's lower bound implies that $M(\varepsilon)$ indeed has a tower-type growth, it is still quite far from the upper bound in Theorem 2.3.2.

Our main result here nearly closes the gap between the lower and upper bounds on $M(\varepsilon)$, showing that $M(\varepsilon)$ is a tower of twos of height at least linear in $1/\varepsilon$. Our construction follows the same initial setup as in [Gre05b], but will diverge from that point on. Our proof is inspired by the recent simplified lower bound proof for the graph regularity lemma in [MS16].

Theorem 2.3.3. *For every $\varepsilon > 0$ it holds that $M(\varepsilon) \geq \operatorname{twr}(\lfloor 1/16\varepsilon \rfloor)$.*

2.3.2 Explicit construction of pseudorandom objects

One of the greatest questions in computer science is whether randomness helps speed up computation. Namely, whether BPP equals P or not. Motivated by this question, a rather more general approach has been pursued over the past 40 years or so via the construction of an object called pseudorandom generator. In the following we give a brief description of pseudorandom generators. and introduce a new approach to construct them.

Pseudorandom generators (PRG) are widely studied in complexity theory. There are several general frameworks used to construct PRGs. One is based on basic building blocks, such as small bias generators [NN93, AGHP92], k-wise independence, or expander graphs [HLW06]. Another approach is based on hardness vs randomness paradigm, which was introduced by Nisan and Wigderson [NW88] and has been very influential. Many of the hardness results used in the latter framework are based on random restrictions, and the analysis of how they simplify the target class of functions. The number of papers in these lines of work is on the order of hundreds, so we do not even attempt to give a comprehensive survey of them all; instead we refer the reader to survey articles [Gol10, Vad12]. The purpose of this work is to introduce a new framework for constructing PRGs based on polarizing random walks. Let us first introduce the standard notion of PRGs.

Definition 2.3.4. *(PRG) Let $\mathscr{F}$ be a class of Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon > 0$ be an error parameter. A PRG for $\mathscr{F}$ is a random variable $X \in \{-1,1\}^n$ such that*

$$\forall f \in \mathscr{F}, |\mathop{\mathbb{E}}_X[f(X)] - \mathop{\mathbb{E}}_U[f(U)]| \le \varepsilon,$$

where U denotes a random variable with the uniform distribution in $\{-1,1\}^n$. Moreover, X has seed length r if $X = G(U)$ for some function $G : \{-1,1\}^r \to [-1,1]^n$.

We relax this definition by introducing a new object called a *fractional PRG*. To prepare the notation for the definition, identify f with a real multi-linear polynomial, namely its Fourier expansion. This extends $f : \{-1,1\}^n \to \{-1,1\}$ to $f : \mathbb{R}^n \to \mathbb{R}$, although, we would only be interested in inputs from $[-1,1]^n$. Observe that if $x \in [-1,1]^n$ then $f(x) = \mathbb{E}_X[f(X)]$ where $X \in \{-1,1\}^n$ is a random variable sampled as follows: for every $i \in [n]$ sample $X_i \in \{-1,1\}$ independently with $\mathbb{E}[X_i] = x_i$. In particular, f on $[-1,1]^n$ is bounded, namely $f : [-1,1]^n \to [-1,1]$. Also, $f(\overline{0}) = \mathbb{E}_U[f(U)]$. The following is a key definition.

Definition 2.3.5 (Fractional PRG)**.** *Let $\mathscr{F}$ be a class of Boolean functions $f : [-1,1]^n \to [-1,1]$ that are multi-linear and $\varepsilon > 0$ be an error parameter. A fractional PRG for $\mathscr{F}$ is a random variable $X \in [-1,1]^n$ such that*

$$\forall f \in \mathscr{F}, |\mathop{\mathbb{E}}_X[f(X)] - f(\overline{0})| \le \varepsilon.$$

One trivial construction of a fractional PRG is $X \equiv \overline{0}$ but this is not going to be useful for our purpose of constructing PRGs. To disallow such examples, we require each coordinate of X to be far from zero with some noticeable probability. Formally, $X \in [-1,1]^n$ is called p-noticeable if $\mathbb{E}[X_i^2] \geq p$ for all $i = 1, \ldots, n$. A good example to keep in mind is the following. Let $G : \{-1,1\}^r \to \{-1,1\}^n$ be a (Boolean valued) function, and set $X = pG(U)$, where $U \in \{-1,1\}^r$ is uniform. Notice that X is p^2-noticeable.

The main result here is that via doing a certain polarizing random walk we can combine a few independent copies of a fractional PRG and obtain a standard PRG. Let $\mathscr{F}$ be a family of n-variate Boolean functions that is closed under restrictions.

Theorem 2.3.6 (Main theorem, informal version of Theorem 7.1.6)**.** *Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathscr{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1, \ldots, X_t$ be i.i.d. copies of X. There is an explicit random variable $G = G(X_1, \cdots, X_t) \in \{-1,1\}^n$ so that G is a PRG for $\mathscr{F}$ with error $(t+1)\varepsilon$.*

In fact, the function G is simulating a certain polarizing random walk over $[-1,1]^n$ that converges very quickly to the vertices of the hypercube $\{-1,1\}^n$. The details are deferred to chapter 7.

Fractional PRG for functions with bounded Fourier growth

Fractional PRGs are easier to construct than standard PRGs, since they can take values in $[-1,1]^n$. For example, assume that f has Fourier tails bounded in L_1. That is, there exist parameters $a, b \geq 1$ for which

$$\sum_{S \subseteq [n] : |S| = k} |\widehat{f}(S)| \leq a \cdot b^k \qquad \forall k = 1, \ldots, n.$$

We show (in Lemma 7.3.4) that if $X \in \{-1,1\}^n$ is roughly (ε/a)-biased, then pX is a fractional PRG for f with $p \approx 1/b$ and error ε. The reason is that this choice of p controls all the Fourier coefficients of f with large Hamming weight, while X controls the ones with small weight. (In fact, to optimize parameters one can choose X to be almost k-wise independent; see Lemma 7.3.4 for details). In any case, note that pX is p^2-noticeable as pX takes values in $\{-p, p\}^n$.

PRG for functions with bounded Fourier growth

An example of families of Boolean functions that are fooled by our PRG include ones that satisfy the following two properties: (i) being closed under restrictions; (ii) having bounded L_1 Fourier tails.

Theorem 2.3.7 (PRG for functions of bounded L_1 Fourier tail, informal version of Theorem 7.3.5)**.** *Let $\mathscr{F}$ be a family of n-variate Boolean functions closed under restrictions. Assume that there exist $a, b \geq 1$ such that for every $f \in \mathscr{F}$,*

$$\sum_{S \subseteq [n] : |S| = k} |\widehat{f}(S)| \leq a \cdot b^k.$$

Then, for any $\varepsilon < \varepsilon \leq \frac{1}{\mathrm{poly}(b \log n)}$ there exists an explicit PRG $X \in \{-1,1\}^n$ which fools $\mathscr{F}$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log(a/\varepsilon))b^2)$.

It is shown that several major classes of boolean functions such as functions of bounded sensitivity, read-once branching programs of bounded width, and bounded depth circuits have bounded Fourier tails. We refer the reader to [CHHL18] for details of the applications of this result. We also discuss a few technical open problems regarding this result in chapter 7.

Acknowledgement Chapter 2 contains material from several different papers discussed as follows.

Section 2.2.2 contains a reprint of material as it appears in Discrete Analysis 2019:10, 14 pp. Kaave Hosseini and Shachar Lovett. "A bilinear Bogolyubov-Ruzsa lemma with poly-logarithmicbounds." The dissertation author was a primary investigator and author of this paper.

Section 2.2.3 contains a reprint of material as it appears in *SIAM Journal on Computing, 47, no. 1 (2018): 208-217.* Hamed Hatami, Kaave Hosseini, and Shachar Lovett. "Structure of protocols for XOR functions." The dissertation author was a primary investigator and author of this paper.

Section 2.2.4 contains a reprint of material as it appears in Journal of Combinatorial Theory,

Series A 148 (2017): 1-14. Kaave Hosseini and Shachar Lovett. "On the structure of the spectrum of small sets." The dissertation author was a primary investigator and author of this paper.

Section 2.3.1 contains a reprint of material as it appears in Mathematical Proceedings of the Cambridge Philosophical Society,vol. 161, no. 2, pp. 193-197. Cambridge University Press, 2016. Kaave Hosseini , Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. "An improved lower bound for arithmetic regularity." The dissertation author was a primary investigator and author of this paper.

Section 2.3.2 contains a reprint of material as it appears in 33rd Computational Complexity Conference (CCC2018). Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. "Pseudorandomgenerators from polarizing random walks." The dissertation author was a primary investigator and author of this paper.

Part I

Approximate Structure

Chapter 3

Higher order Theory of Sumsets

In this chapter we prove the following two theorems. The reader is referred to section 2.2.2 for an introduction to these results.

Theorem 2.2.8. *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and let $w = \mathrm{hvvhvvvhh}$. Then there exists a bilinear variety $B \subset \phi_w(A)$ of co-dimension $r = O(\log^{80} \alpha^{-1})$.*

Theorem 2.2.10. *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and $w = \mathrm{hvvhvvvhh}$ and $\varepsilon = \exp(-O(\log^{20} \alpha^{-1}))$. Then there exists a bilinear variety $B \subset \phi_w^{\varepsilon}(A)$ of co-dimension $r = O(\log^{80} \alpha^{-1})$.*

We recall necessary definitions. Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$. Define two operators, capturing subtraction on horizontal and vertical fibers as follows:

$$\phi_{\mathrm{h}}(A) := \{(x_1 - x_2, y) : (x_1, y), (x_2, y) \in A\},$$
$$\phi_{\mathrm{v}}(A) := \{(x, y_1 - y_2) : (x, y_1), (x, y_2) \in A\}.$$

Given a word $w \in \{\mathrm{h}, \mathrm{v}\}^k$ define $\phi_w = \phi_{w_1} \circ \ldots \circ \phi_{w_k}$ to be their composition. A *bilinear variety* $B \subset \mathbb{F}^n \times \mathbb{F}^n$ of co-dimension $r = r_1 + r_2 + r_3$ is a set defined as follows:

$$B = \{(x, y) \in V \times W : b_1(x, y) = \ldots = b_{r_3}(x, y) = 0\},$$

where $V, W \subset \mathbb{F}^n$ are subspaces of co-dimension r_1, r_2, respectively, and $b_1, \dots, b_{r_3} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ are bilinear forms. We proceed to the definition of $\phi_w^\varepsilon(\cdot)$. Fix an arbitrary $(x,y) \in \mathbb{F}^n \times \mathbb{F}^n$, and note that (x,y) can be written as $(x,y) = \phi_{\text{h}}((x+x_1,y),(x_1,y))$ for any $x_1 \in \mathbb{F}^n$. Moreover, for any fixed x_1, each of the points $(x+x_1,y),(x_1,y)$ can be written as $(x+x_1,y) = \phi_{\text{v}}((x+x_1,y+y_1),(x+x_1,y_1))$ and $(x_1,y) = \phi_{\text{v}}((x_1,y+y_2),(x_1,y_2))$ for arbitrary $y_1, y_2 \in \mathbb{F}^n$. So over all, the point (x,y) can be written using the operation ϕ_{vh} in exactly $|\mathbb{F}^n|^3$ many ways, namely, the total number of two-dimensional parallelograms $(x+x_1,y+y_1),(x+x_1,y_1),(x_1,y+y_2),(x_1,y_2)$ where (x,y) is fixed. We can continue this and consider an arbitrary word $w \in \{\text{h},\text{v}\}^k$. Then (x,y) can be written using the operation ϕ_w in exactly $|\mathbb{F}^n|^{2^k-1}$ many ways. Now, we have a set $A \subset \mathbb{F}^n \times \mathbb{F}^n$ and fix a word $w \in \{\text{h},\text{v}\}^k$. Define $\phi_w^\varepsilon(A)$ to be the set of all elements $(x,y) \in \mathbb{F}^n \times \mathbb{F}^n$ that can be obtained in at least $\varepsilon|\mathbb{F}^n|^{2^k-1}$ many ways by applying the operation $\phi_w(A)$.

Organization We prove Theorem 2.2.8 in Section 3.1 and Theorem 2.2.10 in Section 3.2.

3.1 Proof of Bilinear Bogolyubov-Ruzsa theorem

We recall theorem 2.2.6 which we will frequently use in the proof:

Theorem 2.2.6 (Bogolyubov-Ruzsa lemma [San12a]). *Suppose $A \subset \mathbb{F}_p^n$ and $|A| \geq K^{-1}|\mathbb{F}_p^n|$. Then there is a subspace V of co-dimension $O(\log^4 K)$ with $V \subset 4A$.*

We prove Theorem 2.2.8 in six steps. It corresponds to applying chain of operators $\phi_{\text{h}} \circ \phi_{\text{vv}} \circ \phi_{\text{h}} \circ \phi_{\text{v}} \circ \phi_{\text{vv}} \circ \phi_{\text{hh}}$ to A. In the proof, we invoke Theorem 2.2.6 (or Theorem 2.2.9, or the Freiman-Ruzsa theorem (theorem 2.2.4) which is a corollary of Theorem 2.2.6), four times in total, in steps 1,2,4, and 5.

We will assume that $A \subset \mathbb{F}^m \times \mathbb{F}^n$, where initially $m = n$ but where throughout the proof we update m, n independently when we restrict x or y to large subspaces. It also helps readability, as we will always have that x and related sets or subspaces are in $\mathbb{F}^m$, while y and related sets or subspace are in $\mathbb{F}^n$.

We use three variables r_1, r_2, r_3 that hold the total number of linear forms on x, linear forms on y, and bilinear forms on (x,y) that are being fixed throughout the proof, respectively. Initially, $r_1 = r_2 = r_3 = 0$, but their values will be updated as we go along and at the end, $r = \max(r_1, r_2, r_3)$ will be the codimension of the final bilinear variety.

Step 1 Decompose $A = \bigcup_{y \in \mathbb{F}^n} A_y \times \{y\}$ with $A_y \subset \mathbb{F}^m$. Define $A^1 := \phi_{\mathrm{hh}}(A)$, so that

$$A^1 = \bigcup_{y \in \mathbb{F}^n} (2A_y - 2A_y) \times \{y\}.$$

Let α_y denote the density of A_y. By Theorem 2.2.6, there exists a linear subspace $V'_y \subset 2A_y - 2A_y$ of co-dimension $O(\log^4 \alpha_y^{-1})$. Let $S := \{y : \alpha_y \geq \alpha/2\}$, where by averaging S has density $\geq \alpha/2$. Note that for every $y \in S$ the co-dimension of each V'_y is $O(\log^4 \alpha^{-1})$. We have

$$B^1 := \bigcup_{y \in S} V'_y \times \{y\} \subset A^1.$$

Step 2 Consider $A^2 := \phi_{\mathrm{vv}}(B^1)$. It satisfies

$$A^2 = \bigcup_{y_1, y_2, y_3, y_4 \in S} \left(V'_{y_1} \cap V'_{y_2} \cap V'_{y_3} \cap V'_{y_4}\right) \times \{y_1 + y_2 - y_3 - y_4\}.$$

By Theorem 2.2.6, there is a subspace $W' \subset 2S - 2S$ of co-dimension $O(\log^4 \alpha^{-1})$. Note that the co-dimension of W', as well as the co-dimension of each $V'_{y_1} \cap V'_{y_2} \cap V'_{y_3} \cap V'_{y_4}$, is at most $O(\log^4 \alpha^{-1})$. We thus have

$$B^2 := \bigcup_{y \in W'} V_y \times \{y\} \subset A^2,$$

where $V_y = V'_{y_1} \cap V'_{y_2} \cap V'_{y_3} \cap V'_{y_4}$ for some $y_1, y_2, y_3, y_4 \in S$ which satisfy $y = y_1 + y_2 - y_3 - y_4$.

Update $r_2 := \mathrm{codim}(W')$, where we restrict $y \in W'$. To simplify notations, identify $W' \cong$

$\mathbb{F}^{n-\mathrm{codim}(W')}$ and update $n := n - \mathrm{codim}(W')$. Thus we assume from now that

$$B^2 := \bigcup_{y \in \mathbb{F}^n} V_y \times \{y\},$$

where each V_y has co-dimension $d = O(\log^4 \alpha^{-1})$.

Step 3 Consider $A^3 := \phi_{\mathrm{v}}(B^2)$. It satisfies

$$A^3 = \bigcup_{y,z \in \mathbb{F}^n} (V_z \cap V_{y+z}) \times \{y\}.$$

Step 4 Consider $A^4 := \phi_{\mathrm{h}}(A^3)$. It satisfies

$$A^4 = \bigcup_{y,z,w \in \mathbb{F}^n} ((V_z \cap V_{y+z}) + (V_w \cap V_{y+w})) \times \{y\}.$$

Define $U_y := V_y^{\perp}$, so that $\dim(U_y) = d$ and

$$A^4 = \bigcup_{y,z,w \in \mathbb{F}^n} ((U_z + U_{y+z}) \cap (U_w + U_{y+w}))^{\perp} \times \{y\}.$$

We pause for a moment to introduce one useful notation. We recall that an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ is $L(y) = My + b$ where $M \in \mathbb{F}^{m \times n}, b \in \mathbb{F}^m$. Given a set of affine maps $\mathscr{L} = \{f_i : \mathbb{F}^n \to \mathbb{F}^m, i \in [k]\}$ and $y \in \mathbb{F}^n$, let $\mathscr{L}(y) = \{f_1(y), \ldots, f_k(y)\} \subset \mathbb{F}^m$, and also let $\overline{\mathscr{L}}$ denote the linear span of $\mathscr{L}$. Our goal in this step is to find a small family of affine maps $\mathscr{L}$ with $|\mathscr{L}| \leq O(d)$, and a fixed choice of z, w, so that

$$\Pr_{y \in \mathbb{F}^n}\left[(U_z + U_{y+z}) \cap (U_w + U_{y+w}) \subset \overline{\mathscr{L}}(y)\right] \gg 1. \tag{3.1}$$

as this will give us a dense set $T \subset \mathbb{F}^n$ so that

$$\bigcup_{y \in T} \overline{\mathscr{L}}(y)^{\perp} \times \{y\} \subset A^4.$$

Now we explain how to get Equation (3.1). For every $a \in \mathbb{F}^n$, let $\mathscr{L}_a$ be a collection of affine maps where initially $\mathscr{L}_a = \{0\}$ for all a's. We keep adding affine maps to some of the $\mathscr{L}_a$'s, while always maintaining $|\mathscr{L}_a| \leq 2^d$ for all $a \in \mathbb{F}^n$, until we satisfy

$$\Pr_{y,z,w\in\mathbb{F}^n}\left[(U_z+U_{y+z})\cap(U_w+U_{y+w}) \subset \overline{\mathscr{L}}_z(z)+\overline{\mathscr{L}}_{y+z}(y+z)+\overline{\mathscr{L}}_w(w)+\overline{\mathscr{L}}_{y+w}(y+w)\right] \geq \frac{1}{2} \tag{3.2}$$

and then we will pick some popular affine maps $\mathscr{L} \subset \cup_{a\in\mathbb{F}^n}\mathscr{L}_a$ with $|\mathscr{L}| = O(d)$ that will give us Equation (3.1). For now, we show how to get Equation (3.2). We need the following lemma.

Lemma 3.1.1. *For each $y \in \mathbb{F}^n$, let $U_y \subset \mathbb{F}^m$ be a subspace of dimension d. Assume that*

$$\Pr_{y,z,w\in\mathbb{F}^n}\left[(U_z+U_{y+z})\cap(U_w+U_{y+w}) \subset \overline{\mathscr{L}}_z(z)+\overline{\mathscr{L}}_{y+z}(y+z)+\overline{\mathscr{L}}_w(w)+\overline{\mathscr{L}}_{y+w}(y+w)\right] \leq \frac{1}{2}.$$

Then there exists an affine function $L : \mathbb{F}^n \to \mathbb{F}^m$ such that

$$\Pr_{y\in\mathbb{F}^n}\left[L(y) \in U_y \setminus \overline{\mathscr{L}}_y(y)\right] \geq \exp(-O(d^4)).$$

In the following we prove Lemma 3.1.1. We will use a modified version of a *function version* of the Freiman-Ruzsa theorem, with the quasi-polynomial bounds obtained by Sanders [San12a]. We first recall the standard version. For details how it is derived from Theorem 2.2.6 we refer the reader to [Gre05a].

Theorem 3.1.2. *(Frieman-Ruzsa theorem; function version). Let $f : \mathbb{F}^n \to \mathbb{F}^m$ be a function. Suppose that*

$$\Pr_{y,z,z'\in\mathbb{F}^n}\left[f(y+z)-f(z) = f(y+z')-f(z')\right] \geq \alpha.$$

Then there exists an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ such that

$$\{z \in \mathbb{F}^n : L(z) = f(z)\} \geq \exp(-O(\log^4(\alpha^{-1})))|\mathbb{F}^n|.$$

Now, this result may be strengthened as follows.

Lemma 3.1.3. *Let $f : \mathbb{F}^n \to \mathbb{F}^m$ be a function and $Z \subset \mathbb{F}^n$ with $|Z| \geq \alpha|\mathbb{F}^n|$. Suppose that*

$$\Pr_{y \in \mathbb{F}^n, z, z' \in Z} \left[f(y+z) - f(z) = f(y+z') - f(z') \right] \geq \alpha.$$

Then there exists an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ such that

$$\{z \in Z : L(z) = f(z)\} \geq \exp(-O(\log^4(\alpha^{-1})))|\mathbb{F}^n|.$$

Proof. Let $\Gamma = \{(x, f(x)) : x \in \mathbb{F}^n\}$ and $\Gamma' = \{(x, f(x)) : x \in Z\}$. The additive energy $E(\Gamma, \Gamma')$ satisfies

$$E(\Gamma, \Gamma') = |\{(a,b,c,d) : a - b = c - d, a, c \in \Gamma, b, d \in \Gamma'\}| \geq \alpha^{O(1)}|\Gamma|^3.$$

Using the Cauchy-Schwartz inequality for additive energy (Corollary 2.10 from [TV06]), we have

$$E(\Gamma, \Gamma') \leq \sqrt{E(\Gamma, \Gamma) \cdot E(\Gamma', \Gamma')}.$$

Using the fact that $|\Gamma'| \geq \alpha|\Gamma|$, we get that $E(\Gamma', \Gamma') \geq \alpha^{O(1)}|\Gamma|^3$. Let $M \geq m$ be large enough, and define a function $f' : \mathbb{F}^n \to \mathbb{F}^M$ by setting $f'(z) = f(z)$ if $z \in Z$, and otherwise f takes random values in $\mathbb{F}^M$. Apply Theorem 3.1.2 to f'. The obtained linear function L has to necessarily agree with f' (and hence with f) on a subset $Z' \subset Z$ of the claimed density. □

Now we may go back to the proof of Lemma 3.1.1.

Proof of Lemma 3.1.1. Consider a choice of y, w, z for which

$$(U_{y+z} + U_z) \cap (U_{y+w} + U_w) \not\subset \overline{\mathscr{L}}_{y+z}(y+z) + \overline{\mathscr{L}}_z(z) + \overline{\mathscr{L}}_{y+w}(y+w) + \overline{\mathscr{L}}_w(w).$$

This directly implies that there is an ordered quadruple (a,b,c,d) so that $a \in U_{y+z}, b \in U_z, c \in$

$U_{y+w}, d \in U_w$ with and $a-b=c-d \neq 0$, and

$$\left(\left[a \notin \overline{\mathscr{L}}_{y+z}(y+z)\right] \text{ OR } \left[b \notin \overline{\mathscr{L}}_{z}(z)\right]\right) \text{ AND } \left(\left[c \notin \overline{\mathscr{L}}_{y+w}(y+w)\right] \text{ OR } \left[d \notin \overline{\mathscr{L}}_{w}(w)\right]\right).$$

Consider all the possible solutions of the above formula, namely:

- $\left[a \notin \overline{\mathscr{L}}_{y+z}(y+z)\right]$ AND $\left[c \notin \overline{\mathscr{L}}_{y+w}(y+w)\right]$
- $\left[b \notin \overline{\mathscr{L}}_{z}(z)\right]$ AND $\left[c \notin \overline{\mathscr{L}}_{y+w}(y+w)\right]$
- $\left[a \notin \overline{\mathscr{L}}_{y+z}(y+z)\right]$ AND $\left[d \notin \overline{\mathscr{L}}_{w}(w)\right]$
- $\left[b \notin \overline{\mathscr{L}}_{z}(z)\right]$ AND $\left[d \notin \overline{\mathscr{L}}_{w}(w)\right]$

One of these cases occur for at least $1/4$ of the choices of y, w, z; assume without loss of generality that it is the last one.

Next, sample a random function $f : \mathbb{F}^n \to \mathbb{F}^m$ by picking $f(x) \in U_x$ uniformly and independently for each $x \in \mathbb{F}^n$. Note that the quadruple a, b, c, d depends on y, w, z, and that for each such choice

$$\Pr_f[f(y+z)=a, f(z)=b, f(y+w)=c, f(w)=d] \geq |\mathbb{F}|^{-4d}.$$

Note that when this event happens, by construction we have $f(y+z)-f(z)=f(y+w)-f(w)$. Combining this with the assumption of the lemma, we get

$$\Pr_{y,z,w\in\mathbb{F}^n, f}\left[f(y+z)-f(z)=f(y+w)-f(w), f(z) \in U_z \setminus \overline{\mathscr{L}_z}(z), f(w) \in U_w \setminus \overline{\mathscr{L}_w}(w)\right] \geq \frac{1}{8}\cdot|\mathbb{F}|^{-4d}.$$

Fix f where the above bound holds. Let $Z = \{z : f(z) \in U_z \setminus \overline{\mathscr{L}_z}(z)\}$. Then, surpassing the dependence on the field size, we have $|Z| \geq \exp(-O(d))|\mathbb{F}|^n$ and

$$\Pr_{y\in\mathbb{F}^n, z,w\in Z}[f(y+z)-f(z)=f(y+w)-f(w)] \geq \exp(-O(d)).$$

By Lemma 3.1.3, there is an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ and a set $Z' \subset Z$ with $|Z'| \geq \exp(-O(d^4))|\mathbb{F}^n|$

such that for all $z' \in Z'$, $f(z') = L(z')$ and hence $L(z') \in U_{z'} \setminus \overline{\mathscr{L}_{z'}}(z')$. □

Next, we proceed as follows. As long as Equation (3.2) is satisfied, apply Lemma 3.1.1 to find an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$. For every x that satisfies $L(x) \in U_x \setminus \overline{\mathscr{L}_x}(x)$, add the map L to $\mathscr{L}_x$. This process needs to stop after $t = \exp(O(d^4))$ many steps. Let $L_1, \dots, L_t : \mathbb{F}^n \to \mathbb{F}^m$ be the affine maps obtained in this process. Using this notation, set $\mathscr{L}' = \cup_{x \in \mathbb{F}^n} \mathscr{L}_x$. For every subspace U_x, there is a set $\mathscr{L}'_x \subset \mathscr{L}'$ of size $|\mathscr{L}'_x| \leq d$ such that

$$\overline{\mathscr{L}_x}(x) \subset \overline{\mathscr{L}'_x}(x).$$

This implies that

$$\Pr_{y,z,w \in \mathbb{F}^n}\left[\Big((U_z + U_{y+z}) \cap (U_w + U_{y+w}) \subset \overline{\mathscr{L}'_z}(z) + \overline{\mathscr{L}'_{y+z}}(y+z) + \overline{\mathscr{L}'_w}(w) + \overline{\mathscr{L}'_{y+w}}(y+w)\Big)\right] \geq \frac{1}{2}.$$

Consider the most popular quadruple $\mathscr{L}'_1, \mathscr{L}'_2, \mathscr{L}'_3, \mathscr{L}'_4 \subset \mathscr{L}'$ so that

$$\Pr_{y,z,w \in \mathbb{F}^n}\left[\Big((U_z + U_{y+z}) \cap (U_w + U_{y+w}) \subset \overline{\mathscr{L}'_1}(z) + \overline{\mathscr{L}'_2}(y+z) + \overline{\mathscr{L}'_3}(w) + \overline{\mathscr{L}'_4}(y+w)\Big)\right] \geq \frac{1}{2} \cdot \binom{t}{d}^{-4}.$$

Let $\mathscr{L} := \mathscr{L}'_1 \cup \mathscr{L}'_2 \cup \mathscr{L}'_3 \cup \mathscr{L}'_4$. Recall that $t = \exp(O(d^4))$ and hence $\binom{t}{d} = \exp(O(d^5))$. We have

$$\Pr_{y,z,w \in \mathbb{F}^n}\left[(U_z + U_{y+z}) \cap (U_w + U_{y+w}) \subset \overline{\mathscr{L}}(z) + \overline{\mathscr{L}}(y+z) + \overline{\mathscr{L}}(w) + \overline{\mathscr{L}}(y+w)\right] \geq \exp(-O(d^5)).$$

By averaging, there is some choice of z, w such that,

$$\Pr_{y \in \mathbb{F}^n}\left[(U_z + U_{y+z}) \cap (U_w + U_{y+w}) \subset \overline{\mathscr{L}}(z) + \overline{\mathscr{L}}(y+z) + \overline{\mathscr{L}}(w) + \overline{\mathscr{L}}(y+w)\right] \geq \exp(-O(d^5)).$$

Recall that each $L \in \mathscr{L}$ is an affine map and that $|\mathscr{L}| \leq 4d$. Thus, $\overline{\mathscr{L}}(z), \overline{\mathscr{L}}(y+z), \overline{\mathscr{L}}(w), \overline{\mathscr{L}}(y+$

$w) \subset \overline{\mathscr{L}}(y) + Q$ where $Q \subset \mathbb{F}^m$ is a linear subspace of dimension $O(d)$. We thus have

$$B^4 := \bigcup_{y \in T} (\overline{\mathscr{L}}(y) + Q)^{\perp} \times \{y\} \subset A^4,$$

where $T \subset \mathbb{F}^n$ has density $\exp(-O(d^5))$.

To simplify the presentation, we would like to assume that the maps in $\mathscr{L}$ are linear maps instead of affine maps, that is, that they do not have a constant term. This can be obtained by restricting x to the subspace orthogonal to Q and to the constant term in the affine maps in $\mathscr{L}$. Correspondingly, we update $r_1 := r_1 + \dim(Q) + |\mathscr{L}| = O(d)$. So, from now we assume that $\mathscr{L}$ is defined by $4d$ linear maps, and that

$$B^4 := \bigcup_{y \in T} \overline{\mathscr{L}}(y)^{\perp} \times \{y\} \subset A^4,$$

where $T \subset \mathbb{F}^n$ has density $\exp(-O(d^5))$.

Step 5 Consider $A^5 := \phi_{\mathrm{vv}}(B^4)$ so that

$$A^5 = \bigcup_{y_1,y_2,y_3,y_4 \in T} \left(\overline{\mathscr{L}}(y_1)^{\perp} \cap \overline{\mathscr{L}}(y_2)^{\perp} \cap \overline{\mathscr{L}}(y_3)^{\perp} \cap \overline{\mathscr{L}}(y_4)^{\perp}\right) \times \{y_1 + y_2 - y_3 - y_4\}.$$

By Theorem 2.2.4 there exists a subspace $W \subset 2T - 2T$ with co-dimension $O(d^{20})$. However, this time, this is not enough for us. We need to use Theorem 2.2.9 instead. The following equivalent formulation of Theorem 2.2.9 will be more convenient for us: there is a subspace $W \subset \mathbb{F}^n$ of co-dimension $O(\log^4 \alpha^{-1})$ such that, for each $y \in W$ there is a set $S_y \subset (\mathbb{F}^n)^3$ of density $\alpha^{O(1)}$, for which

$$\forall (a_1, a_2, a_3) \in S_y : \quad a_1, a_2, a_3, a_1 + a_2 - a_3 - y \in A.$$

Apply Theorem 2.2.9 to the set T to obtain the subspace W and the sets S_y. We have

$$B^5 := \bigcup_{y\in W}\left(\bigcup_{(y_1,y_2,y_3)\in S_y}\left(\overline{\mathscr{L}}(y_1)+\overline{\mathscr{L}}(y_2)+\overline{\mathscr{L}}(y_3)+\overline{\mathscr{L}}(y_1+y_2-y_3-y)\right)^{\perp}\right)\times\{y\}\subset A^5.$$

To simplify the presentation we introduce the notation $\overline{\mathscr{L}}(y_1,y_2,y_3) := \overline{\mathscr{L}}(y_1)+\overline{\mathscr{L}}(y_2)+\overline{\mathscr{L}}(y_3)$. Next, observe that for any $y,y'\in\mathbb{F}^n$, $\overline{\mathscr{L}}(y')+\overline{\mathscr{L}}(y+y') = \overline{\mathscr{L}}(y')+\overline{\mathscr{L}}(y)$. Thus we can simplify the expression of B^5 to

$$B^5 = \bigcup_{y\in W}\left(\bigcup_{(y_1,y_2,y_3)\in S_y}\left(\overline{\mathscr{L}}(y_1,y_2,y_3)+\overline{\mathscr{L}}(y)\right)^{\perp}\right)\times\{y\},$$

which can be re-written as

$$B^5 = \bigcup_{y\in W}\left(\bigcup_{(y_1,y_2,y_3)\in S_y}\overline{\mathscr{L}}(y_1,y_2,y_3)^{\perp}\cap\overline{\mathscr{L}}(y)^{\perp}\right)\times\{y\}.$$

Step 6 Consider $A^6 := \phi_{\mathrm{h}}(B^5)$. It satisfies

$$A^6 = \bigcup_{y\in W}\left(\bigcup_{\substack{(y_1,y_2,y_3)\in S_y\\(y'_1,y'_2,y'_3)\in S_y}}\overline{\mathscr{L}}(y_1,y_2,y_3)^{\perp}\cap\overline{\mathscr{L}}(y)^{\perp}+\overline{\mathscr{L}}(y'_1,y'_2,y'_3)^{\perp}\cap\overline{\mathscr{L}}(y)^{\perp}\right)\times\{y\}.$$

In order to complete the proof, we will find a large subspace V such that for every $y\in W$,

$$V\cap\overline{\mathscr{L}}(y)^{\perp}\subset\bigcup_{\substack{(y_1,y_2,y_3)\in S_y\\(y'_1,y'_2,y'_3)\in S_y}}\overline{\mathscr{L}}(y_1,y_2,y_3)^{\perp}\cap\overline{\mathscr{L}}(y)^{\perp}+\overline{\mathscr{L}}(y'_1,y'_2,y'_3)^{\perp}\cap\overline{\mathscr{L}}(y)^{\perp}.$$

In fact, we will prove something stronger: there is a large subspace V such that for each $y\in W$, there is a choice of $(y_1,y_2,y_3),(y'_1,y'_2,y'_3) \in S_y$ for which

$$V \cap \overline{\mathscr{L}}(y)^{\perp} \subset \overline{\mathscr{L}}(y_1,y_2,y_3)^{\perp} \cap \overline{\mathscr{L}}(y)^{\perp} + \overline{\mathscr{L}}(y'_1,y'_2,y'_3)^{\perp} \cap \overline{\mathscr{L}}(y)^{\perp}.$$

The following lemma is key. Given a set $\mathscr{L}$ of linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$, let $\dim(\overline{\mathscr{L}})$ denote the dimension of linear span of $\mathscr{L}$ as a vector space over $\mathbb{F}$.

Lemma 3.1.4. *Fix $\delta > 0$. Let $\mathscr{L}$ be a set of linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ with $\dim(\overline{\mathscr{L}}) = k$. Then there is a subspace $Z \subset \mathbb{F}^m$ of dimension at most $k(2k + \log\delta^{-1} + 3)$ such that the following holds. For every subset $S \subset \mathbb{F}^n$ of density at least δ, and arbitrary $y \in \mathbb{F}^n$, at least half the pairs $s,s' \in S$ satisfy that*

$$(\overline{\mathscr{L}}(s) + \overline{\mathscr{L}}(y)) \cap (\overline{\mathscr{L}}(s') + \overline{\mathscr{L}}(y)) \subset Z + \overline{\mathscr{L}}(y).$$

Proof. The proof is by induction on $\dim(\overline{\mathscr{L}})$. Consider first the base case of $\dim(\overline{\mathscr{L}}) = 1$ and suppose $\overline{\mathscr{L}} = \langle L \rangle$ for some map L. We consider two cases based on minimum rank of maps in $\overline{\mathscr{L}}$. First suppose that rank of every non-zero map in $\overline{\mathscr{L}}$ (which is the same as rank of L) is bigger than $\log\delta^{-1} + 5$. Fix arbitrary $L_1, L_3 \in \overline{\mathscr{L}} \setminus \{0\}$ and $L_2, L_4 \in \overline{\mathscr{L}}$ and $s, y \in \mathbb{F}^n$ and observe that

$$\Pr_{s' \in S}\left[L_1(s) + L_2(y) = L_3(s') + L_4(y)\right] < \frac{|\mathbb{F}|^{-(\log\delta^{-1}+5)}}{\Pr_{s' \in \mathbb{F}^n}\left[s' \in S\right]} \leq |\mathbb{F}|^{-(\log\delta^{-1}+5)}\delta^{-1}.$$

By applying the union bound over all quadruples $L_1, \cdots, L_4 \in \overline{\mathscr{L}}$, we obtain that

$$\Pr_{s,s' \in S}\left[\left(\overline{\mathscr{L}}(s) + \overline{\mathscr{L}}(y)\right) \cap \left(\overline{\mathscr{L}}(s') + \overline{\mathscr{L}}(y)\right) \neq \overline{\mathscr{L}}(y)\right] \leq |\mathbb{F}|^4 |\mathbb{F}|^{-(\log\delta^{-1}+5)}\delta^{-1} \leq \frac{1}{2}.$$

Therefore, we can safely choose $Z = \{0\}$ in the lemma. Now, for the second case, suppose that $\mathrm{rank}(L) \leq \log\delta^{-1} + 5$. Let $Z = \mathrm{Im}(L)$. Then for all $s \in \mathbb{F}^n$, $\overline{\mathscr{L}}(s) \subset \mathrm{Im}(L) = Z$, and so $(\overline{\mathscr{L}}(s) + \overline{\mathscr{L}}(y)) \cap (\overline{\mathscr{L}}(s') + \overline{\mathscr{L}}(y)) \subset Z \subset Z + \overline{\mathscr{L}}(y)$.

Now let $\dim(\overline{\mathscr{L}}) = k$. First, suppose that $\forall L \in \overline{\mathscr{L}}$, $\mathrm{rank}(L) > 4k + \log\delta^{-1} + 1$. Then similar to the base case, for all $y \in \mathbb{F}^n$,

$$\Pr_{s,s' \in S}\left[\left(\overline{\mathscr{L}}(s) + \overline{\mathscr{L}}(y)\right) \cap \left(\overline{\mathscr{L}}(s') + \overline{\mathscr{L}}(y)\right) \neq \overline{\mathscr{L}}(y)\right] \leq |\mathbb{F}|^{4k}|\mathbb{F}|^{-(4k+\log\delta^{-1}+1)}\delta^{-1} \leq \frac{1}{2}.$$

Otherwise, suppose there is some $L \in \overline{\mathscr{L}} \setminus \{0\}$ with rank at most $4k + \log\delta^{-1} + 1$. Let Y be a subspace so that $Y \oplus \mathrm{Im}(L) = \mathbb{F}^m$. Let $\mathrm{Proj}_Y : \mathbb{F}^n \to Y$ be the projection map along $\mathrm{Im}(L)$ with $\mathrm{Proj}_Y(\mathrm{Im}(L)) = 0$. Consider the new family of maps

$$\mathscr{L}' = \{\mathrm{Proj}_Y \circ M : M \in \mathscr{L}\}.$$

Note that $\overline{\mathscr{L}'}$ has dimension $\leq k-1$ because $\mathrm{Proj}_Y \circ L \equiv 0$ and so by induction hypothesis, there exists a subspace Z' of dimension at most $(k-1)(2(k-1)+\log\delta^{-1}+3)$ such that, for all $y \in \mathbb{F}^n$, for least half the pairs $s, s' \in S$ it holds that

$$(\overline{\mathscr{L}'}(s) + \overline{\mathscr{L}'}(y)) \cap (\overline{\mathscr{L}'}(s') + \overline{\mathscr{L}'}(y)) \subset Z' + \overline{\mathscr{L}'}(y).$$

The above implies that

$$\mathrm{Proj}_Y((\overline{\mathscr{L}}(s) + \overline{\mathscr{L}}(y)) \cap (\overline{\mathscr{L}}(s') + \overline{\mathscr{L}}(y))) \subset Z' + \mathrm{Proj}_Y(\overline{\mathscr{L}}(y)) \subset Z' + \overline{\mathscr{L}}(y) + \mathrm{Im}(L).$$

So we can take $Z = Z' + \mathrm{Im}(L)$.

□

We note that for Theorem 2.2.8 we only need a weaker form of Lemma 3.1.4, which states that at least one pair $y, y' \in S$ exists; however, we would need the stronger version for Theorem 2.2.10.

We apply Lemma 3.1.4 as follows. Define a new family of linear maps $\mathscr{L}^*$ from $\mathbb{F}^{3n}$ to $\mathbb{F}^m$

as follows. For each $L \in \mathscr{L}$ define three linear maps L_i, $i \in \{1,2,3\}$ by:

$$L_i : (\mathbb{F}^n)^3 \to \mathbb{F}^m, L_i(y_1,y_2,y_3) = L(y_i)$$

and let

$$\mathscr{L}^* := \{L_i : L \in \mathscr{L}, i \in [3]\}.$$

Apply Lemma 3.1.4 to the family $\mathscr{L}^*$ with $\delta = \exp(-O(d^5))$ and obtain a subspace $V \subset \mathbb{F}^m$ of codimension $O(d^2 \log(\exp(-O(d^5))) = O(d^7)$ so that, for every $S_y \subset (\mathbb{F}^n)^3$ with $y \in W$, there exist $(y_1,y_2,y_3),(y_1',y_2',y_3') \in S_y$ for which

$$V \cap \overline{\mathscr{L}^*}((y,y,y))^\perp \subset (\overline{\mathscr{L}^*}((y_1,y_2,y_3))^\perp \cap \overline{\mathscr{L}^*}((y,y,y))^\perp) + (\overline{\mathscr{L}^*}((y_1',y_2',y_3'))^\perp \cap \overline{\mathscr{L}^*}((y,y,y))^\perp).$$

This directly implies that

$$V \cap \overline{\mathscr{L}}(y)^\perp \subset (\overline{\mathscr{L}}(y_1,y_2,y_3)^\perp \cap \overline{\mathscr{L}}(y)^\perp) + (\overline{\mathscr{L}}(y_1',y_2',y_3')^\perp \cap \overline{\mathscr{L}}(y)^\perp).$$

Define

$$B^6 := \bigcup_{y \in W} \left(V \cap \overline{\mathscr{L}}(y)^\perp\right) \times \{y\} \subset A^6.$$

Observe that B^6 is a bilinear variety defined by $\operatorname{codim}(V)$ many linear equations on x, $\operatorname{codim}(W)$ linear equations on y and $|\mathscr{L}|$ bilinear equations on (x,y).

To complete the proof we calculate the quantitative bounds obtained. We have $d = O(\log^4 \alpha^{-1})$ where α was the density of the original set A, and

$$\begin{aligned} r_1 &= O(d) + \operatorname{codim}(V) = O(d^7), \\ r_2 &= O(d) + \operatorname{codim}(W) = O(d^{20}), \\ r_3 &= |\mathscr{L}| = O(d). \end{aligned}$$

Together these give the final bound of $r = \max(r_1, r_2, r_3) = O(\log^{80} \alpha^{-1})$.

3.2 Proof of a robust version of Bilinear Bogolyubov-Ruzsa theorem

In this section we prove Theorem 2.2.10 by slightly modifying the proof of Theorem 2.2.8. We point out the necessary modifications to proof of Theorem 2.2.8. In this proof we use the Theorem 2.2.9 which we recall in the following.

Theorem 2.2.9 ([San12a, SS16])**.** *Let $A \subset \mathbb{F}^n$ be a subset of density α. Then there exists a subspace $V \subset 2A - 2A$ of co-dimension $O(\log^4 \alpha^{-1})$ such that the following holds. Every $y \in V$ can be expressed as $y = a_1 + a_2 - a_3 - a_4$ with $a_1, a_2, a_3, a_4 \in A$ in at least $\alpha^{O(1)}|\mathbb{F}|^{3n}$ many ways.*

Step 1 In this step, we use Theorem 2.2.9 instead of Theorem 2.2.6, and directly obtain

$$B^1 \subset \phi_{\mathrm{hh}}^{\varepsilon_1}(A) \tag{3.3}$$

for $\varepsilon_1 = \alpha^{O(1)}$.

Step 2 Similarly in this step as well, using Theorem 2.2.9 instead of Theorem 2.2.6 gives

$$B^2 \subset \phi_{\mathrm{vv}}^{\varepsilon_2}(B^1) \tag{3.4}$$

with $\varepsilon_2 = \alpha^{O(1)}$. To recall, we assume for simplicity of exposition from now on that $B^2 = \bigcup_{y \in \mathbb{F}^n} V_y \times \{y\}$.

Steps 3 and 4 This step is slightly different than steps 1 and 2. Here, we are not able to directly produce some set B^4 that would satisfy $B^4 \subset \phi_{\mathrm{hv}}^{\varepsilon_4}(B^2)$. But what we can do is to apply the remaining operation ϕ_{hvvhv} altogether to B^2 and obtain the final bilinear structure B^6 that satisfies what we want, which is

$$B^6 \subset \phi_{\mathrm{hvvhv}}^{\varepsilon_6}(B^2) \tag{3.5}$$

for $\varepsilon_6 = \exp(-\operatorname{poly}\log \alpha^{-1})$. Combining Equations (3.3) to (3.5) gives

$$B^6 \subset \phi^{\varepsilon}_{\mathrm{hvvhvvvhh}}(A)$$

for $\varepsilon = \exp(-\operatorname{poly}\log \alpha^{-1})$.

We establish Equation (3.5) in the rest of the proof. Recall that previously we showed that the following holds: there is a set of affine maps $\mathscr{L}$, with $|\mathscr{L}| = O(d)$, such that

$$\Pr_{y,w,z\in\mathbb{F}^n}\left[\left(\overline{\mathscr{L}}(z)+\overline{\mathscr{L}}(y+z)+\overline{\mathscr{L}}(w)+\overline{\mathscr{L}}(y+w)\right)^{\perp} \subset \left(V_z^{\perp}\cap V_{y+z}^{\perp}\right)+\left(V_w^{\perp}\cap V_{y+w}^{\perp}\right)\right] \geq \exp(-O(d^5))$$

and consequently

$$\Pr_{y,w,z\in\mathbb{F}^n}\left[\left(\overline{\mathscr{L}}(y)+\overline{\mathscr{L}}(z)+\overline{\mathscr{L}}(w)\right)^{\perp} \subset \left(V_z^{\perp}\cap V_{y+z}^{\perp}\right)+\left(V_w^{\perp}\cap V_{y+w}^{\perp}\right)\right] \geq \exp(-O(d^5)).$$

Remember that $d = O(\log^4 \alpha^{-1})$. Furthermore, we may assume the maps in $\mathscr{L}$ are linear (instead of affine) after we update $r_1 := r_1 + |\mathscr{L}| = O(d)$.

Then what we did in the proof of Theorem 2.2.8 was to fix one popular choice of w,z. However, here we can't do that, as we need many pairs of w,z. Let T be the set of y's that satisfy

$$\Pr_{w,z\in\mathbb{F}^n}\left[\left(\overline{\mathscr{L}}(y)+\overline{\mathscr{L}}(z)+\overline{\mathscr{L}}(w)\right)^{\perp} \subset \left(V_z^{\perp}\cap V_{y+z}^{\perp}\right)+\left(V_w^{\perp}\cap V_{y+w}^{\perp}\right)\right] \geq \exp(-O(d^5)), \tag{3.6}$$

and so T has density $\exp(-O(d^5))$. We deduce something stronger from Equation (3.6) but we need to introduce some notation first.

For $A,B \subset \mathbb{F}^n$ let $A -_{\eta} B$ denote the set of all elements $c \in A - B$ that can be written in at least $\eta|\mathbb{F}^n|$ many ways as $c = a - b$ for $a \in A, b \in B$. To use this notation, note that if A,B are two subspaces of co-dimension k, then $A - B = A -_{\eta} B$ for $\eta = \exp(-O(k))$. This is because every element $c \in A - B$ can be written as $c = (a+v) - (b+v)$ where v is an arbitrary element in the subspace $A \cap B$ of codimension at most $2k$. So we can improve the Equation (3.6) to

$$\Pr_{w,z \in \mathbb{F}^n}\left[\left(\overline{\mathscr{L}}(y)+\overline{\mathscr{L}}(z)+\overline{\mathscr{L}}(w)\right)^{\perp} \subset \left(V_z^{\perp} \cap V_{y+z}^{\perp}\right) -_{\eta} \left(V_w^{\perp} \cap V_{y+w}^{\perp}\right)\right] \geq \exp(-O(d^5)), \quad (3.7)$$

for $\eta = \exp(-O(d))$

Step 5 Similar to before, consider the subspace $W \subset 2T - 2T$ of co-dimension $O(d^{20})$ that is given by Theorem 2.2.9. This subspace W has the following property: fix arbitrary $y \in W$. Sample $y_1, y_2, y_3 \in \mathbb{F}^n$ uniformly and independently, and set $y_4 = -y + y_1 + y_2 - y_3$. Then with probability at least $\exp(-O(d^5))$ we have $y_1, y_2, y_3, y_4 \in T$. This means that if we furthermore sample $w_1, w_2, w_3, w_4, z_1, z_2, z_3, z_4 \in \mathbb{F}^n$ uniformly and independently, then, with probability at least $\exp(-O(d^5))$, the following four equations simultaneously hold:

$$\left(\overline{\mathscr{L}}(y_i)+\overline{\mathscr{L}}(z_i)+\overline{\mathscr{L}}(w_i)\right)^{\perp} \subset \left(V_{z_i}^{\perp} \cap V_{y_i+z_i}^{\perp}\right) -_{\eta} \left(V_{w_i}^{\perp} \cap V_{y_i+w_i}^{\perp}\right) \qquad \forall i = 1, \ldots, 4.$$

By computing the intersection of the left hand sides and the right hand sides we obtain that with probability at least $\exp(-O(d^5))$, it holds that

$$\left(\overline{\mathscr{L}}(y)+\sum_{i=1}^{3}\overline{\mathscr{L}}(y_i)+\sum_{i=1}^{4}\overline{\mathscr{L}}(z_i)+\sum_{i=1}^{4}\overline{\mathscr{L}}(w_i)\right)^{\perp} \subset \bigcap_{i=1}^{4}\left(\left(V_{z_i}^{\perp} \cap V_{y_i+z_i}^{\perp}\right) -_{\eta} \left(V_{w_i}^{\perp} \cap V_{y_i+w_i}^{\perp}\right)\right). \quad (3.8)$$

For a given $y \in \mathbb{F}^n, \boldsymbol{s} = (y_1, y_2, y_3, w_1, w_2, w_3, w_4, z_1, z_2, z_3, z_4) \in (\mathbb{F}^n)^{11}$, let

$$\mathscr{V}_{y,\boldsymbol{s}} = \bigcap_{i=1}^{4}\left(\left(V_{z_i}^{\perp} \cap V_{y_i+z_i}^{\perp}\right) -_{\eta} \left(V_{w_i}^{\perp} \cap V_{y_i+w_i}^{\perp}\right)\right),$$

where to recall $y_4 = -y + y_1 + y_2 - y_3$. Observe that for any $\boldsymbol{s}$,

$$\bigcup_{y \in W} \mathscr{V}_{y,\boldsymbol{s}} \times \{y\} \subset \phi_{\mathrm{vvhv}}(B^2).$$

We rewrite Equation (3.8) more compactly as

$$\Pr_{\boldsymbol{s}}\left[\left(\overline{\mathscr{L}}(y)+\overline{\mathscr{L}}(\boldsymbol{s})\right)^{\perp}\subset\mathscr{V}_{y,\boldsymbol{s}}\right]\geq\exp(-O(d^5)),\tag{3.9}$$

where we use the notation $\overline{\mathscr{L}}(\boldsymbol{s})=\sum_{i=1}^{3}\overline{\mathscr{L}}(y_i)+\sum_{i=1}^{4}\overline{\mathscr{L}}(z_i)+\sum_{i=1}^{4}\overline{\mathscr{L}}(w_i)$.

Step 6 Now we consider applying the operation hvvhv altogether to B^2. Only the last operation h remains to be applied, which after doing so, we will find a subspace $V\subset\mathbb{F}^m$ of co-dimension $O(d^7)$ that satisfies the following: for any $y\in W$, choose $\mathbf{s_1},\mathbf{s_2}\in(\mathbb{F}^n)^{11}$ uniformly and randomly. Then with probability $\exp(-O(d^5))$,

$$V\cap\overline{\mathscr{L}}(y)^{\perp}\subset\mathscr{V}_{y,\mathbf{s_1}}-_{\eta}\mathscr{V}_{y,\mathbf{s_2}}.$$

where to recall $\eta=\exp(-O(d))$.

Fix $y\in W$ and let S_y be the set of all tuples $\boldsymbol{s}=(y_1,y_2,y_3,w_1,w_2,w_3,w_4,z_1,z_2,z_3,z_4)\in(\mathbb{F}^n)^{11}$ that satisfy Equation (3.9). Note that the density of each S_y is at least $\exp(-O(d^5))$. To simplify notation denote $\boldsymbol{s}=(s_1,\dots,s_{11})$. We call up Lemma 3.1.4 in a similar way as we did before. Define a family $\mathscr{L}^*$ of linear maps, containing linear maps L_i for each $L\in\mathscr{L}$ and $i=1,\dots,11$, where

$$L_i:(\mathbb{F}^n)^{11}\to\mathbb{F}^m, L_i(\boldsymbol{s})=L(s_i).$$

Apply Lemma 3.1.4 to $\mathscr{L}^*$ and density parameter $\exp(-O(d^5))$. So, we obtain a subspace $V\subset\mathbb{F}^m$ of co-dimension $O(d^7)$ such that for each $y\in W$,

$$\Pr_{\mathbf{s_1},\mathbf{s_2}\in S_y}\left[V\cap\overline{\mathscr{L}}(y)^{\perp}\subset(\overline{\mathscr{L}}(\mathbf{s_1})+\overline{\mathscr{L}}(y))^{\perp}+(\overline{\mathscr{L}}(\mathbf{s_2})+\overline{\mathscr{L}}(y))^{\perp}\right]\geq\frac{1}{2},\tag{3.10}$$

which implies

$$\Pr_{\mathbf{s_1},\mathbf{s_2}\in(\mathbb{F}^n)^{11}}\left[V\cap\overline{\mathscr{L}}(y)^{\perp}\subset\mathscr{V}_{y,\mathbf{s_1}}-_{\eta}\mathscr{V}_{y,\mathbf{s_2}}\right]\geq\exp(-O(d^5)).\tag{3.11}$$

Define the final bilinear structure as

$$B^6 := \bigcup_{y \in W} \left(V \cap \overline{\mathscr{L}}(y)^\perp \right) \times \{y\}.$$

It satisfies

$$B^6 \subset \phi^{\varepsilon_6}_{\mathrm{hvvhv}}(B^2)$$

for $\varepsilon_6 = \exp(-O(d^5))$ and so over all

$$B^6 \subset \phi^{\varepsilon}_{\mathrm{hvvhvvvhh}}(A)$$

for $\varepsilon = \exp(-O(d^5))$.

Acknowledgement Chapter 3 is a reprint of material as it appears in Discrete Analysis 2019:10, 14 pp. Kaave Hosseini and Shachar Lovett. "A bilinear Bogolyubov-Ruzsa lemma with poly-logarithmicbounds." The dissertation author was a primary investigator and author of this paper.

Chapter 4

Sumsets and Communication Complexity of XOR functions

The purpose of this chapter is to prove the following result.

Theorem 2.2.13 ([HHL16]). *For any* $f : \mathbb{F}_2^n \to \{0,1\}$ *we have* $\mathrm{pdt}(f) \leq O(D(f_\oplus)^6)$.

Recall that $f_\oplus(x,y) = f(x+y)$, and $\mathrm{pdt}(\cdot)$ is the parity-decision-tree complexity and $D(f_\oplus)$ is deterministic communication complexity.

Organization We start with giving a proof overview in section 4.1. Then we give some preliminary definitions in Section 4.2. We establish the key steps required in the proof of theorem 2.2.13 in Sections 4.3.1, 4.3.2, 4.3.3, and we apply them in Section 4.3.4 to prove of Theorem 2.2.13.

4.1 Proof overview

Fix $f : \mathbb{F}_2^n \to \{0,1\}$, where we assume that $f_\oplus$ has an efficient deterministic protocol. Our goal is to design a low-depth PDT for f.

Reduction to monochromatic subspaces Note that if f has a PDT of depth k, then in particular, the leaves of the PDT determine affine subspaces of co-dimension at most k on which f

is constant. We call such subspaces *monochromatic subspaces* for f. From here onwards, we use “subspace" as a shorthand for “affine subspace".

It turns out that in order to design a PDT for f, it suffices to show that there exists a large monochromatic subspace for f. This follows from [TWXZ13] who showed (among other things) that if f is constant on a subspace V, then the Fourier sparsity of f restricted to any coset of V reduces by at least a factor of two. This is sufficient for our application, as the existence of an efficient deterministic protocol for $f_{\oplus}$ implies in particular that f has low Fourier sparsity. This reduces Theorem 2.2.13 to the following question, which is the main problem we investigate in this paper.

Question 4.1.1. *Let $f : \mathbb{F}_2^n \to \{0,1\}$ with $D(f_{\oplus}) \le k$. Is there a subspace V of co-dimension* $\mathrm{poly}(k)$ *on which f is constant?*

In the next few paragraphs we give a brief discussion of how to find such a subspace. We first describe a natural approach, which only tries to exploit the existence of a large monochromatic rectangle for $f_{\oplus}$ (many techniques in communication complexity follow this approach; in the randomized settings, one needs to replace “monochromatic rectangle" with “biased rectangle"). However, as we discuss below, a direct application of this technique fails, and a more careful application requires unproven conjectures in additive combinatorics. As such, we follow a different route, which exploits the entire structure of the protocol. This is uncommon in communication complexity, and we view this is as a conceptual contribution of this work.

Using a large monochromatic rectangle, and why it fails The existence of an efficient deterministic protocol for $f_{\oplus}$ implies that it is constant on a large rectangle $A \times B$, and consequently f is constant on $A + B$. As a first attempt, one may hope that if $A, B \subseteq \mathbb{F}_2^n$ are large sets, then $A + B$ must contain a large subspace. This would directly imply that f is constant on this subspace. Unfortunately this is false, as the following example of Green [Gre04] shows.

Example 4.1.2. *Let $A = B = \mathscr{B}(n/2 - \sqrt{n})$ where $\mathscr{B}(r) \subset \{0,1\}^n$ is the hamming ball of radius r. Then $|A| = |B| = \Omega(2^n)$, $A + B = \mathscr{B}(n - 2\sqrt{n})$ but the largest subspace contained in $A + B$ has* *co-dimension $2\sqrt{n}$. For example, such a subspace can be obtained by fixing the first $2\sqrt{n}$ bits to zero.*

The situation improves for sum-sets involving more than two sets. Sanders [San12a] showed that for a set $A \subset \mathbb{F}_2^n$ with $|A| \geq \varepsilon 2^n$, $4A = A + A + A + A$ contains a subspace of co-dimension $O(\log^4(1/\varepsilon))$. As Yao showed [Yao16], it follows directly from this result that a k-bit deterministic protocol for the 4-party function $F(x,y,z,w) = f(x \oplus y \oplus z \oplus w)$ implies a parity decision tree of depth $O(k^5)$ for f.

Going back to two-fold sum-sets, we note that despite Example 4.1.2, for our application one might still be able to use other properties of f to find a large monochromatic subspace in $A+B$. For example, since f has low Fourier sparsity, if we find a subspace V on which f is nearly constant, then f will be in fact constant on this subspace. More precisely, since the Fourier sparsity of f is at most 2^k, using uncertainty principle, lemma 2.1.1, we deduce the following: $\mathbb{E}[f|_V] < 2^{-k}$ implies $f|_V \equiv 0$, and $\mathbb{E}[f|_V] > 1 - 2^{-k}$ implies $f|_V \equiv 1$. Therefore, given large sets $A, B \subseteq \mathbb{F}_2^n$, rather than showing the existence of a large subspace in $A+B$, it suffices to show that $A+B$ contains most of a large subspace, and then the Fourier sparsity of f implies that f is constant on this subspace. Working out the details, it turns out that we would need the following conjecture:

Conjecture 4.1.3. *Let $A \subset \mathbb{F}_2^n$ be of size $|A| \geq \varepsilon 2^n$. Then for any $\delta > 0$ there exists a subspace V such that $|2A \cap V| \geq (1-\delta)|V|$, where the co-dimension of V is at most* $\operatorname{polylog}(1/\varepsilon\delta)$.

For this and related conjectures see [SS16] (in particular Section 9, the paragraph on correlations of $2A, 3A, 4A$). We note that two partial results towards Conjecture 4.1.3 are known, both due to Sanders:

- [San10] proves the existence of a subspace with co-dimension $O((1/\varepsilon)\log(1/\delta))$.
- [San12a] proves the existence of a subspace with co-dimension $O((1/\delta^2)\log^4(1/\varepsilon))$.

Unfortunately, neither of these two bounds is strong enough for our application. If $f_\oplus$ has a k-bit deterministic protocol, then the largest monochromatic rectangle satisfies $|A|, |B| \geq 2^{n-k}$. We thus have $\varepsilon = 2^{-k}$. Furthermore, $f_\oplus$ has at most 2^k nonzero Fourier coefficients, which means that we need a subspace which is 2^{-k} close to being monochromatic, and thus we need to set $\delta < 2^{-k}$. Hence to achieve our goal of finding a subspace of co-dimension $\mathrm{poly}(k)$, we need poly-logarithmic dependency on both ε and δ.

Our approach: utilizing the entire protocol We circumvent the need to use unproven conjectures by devising an alternative route based on information theory, which exploits the entire structure of the protocol. Fix a deterministic protocol for $f_\oplus$ which sends k bits, and let $K = 2^k$. Let $A_i \times B_i$ for $i \in [K]$ be the partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ induced by the protocol. For an input (x,y), let $\Pi_{xy} \in [K]$ denote the index of the unique rectangle that contains (x,y). By our assumption $f_\oplus$ is constant on each $A_i \times B_i$ (or equivalently the value of $f_\oplus(x,y)$ is determined by Π_{xy}), which means that f is constant on each $A_i + B_i$.

Let $\mu = \mathbb{E}[f]$ be the average of f on the entire space, and assume without loss of generality that $\mu \geq 1/2$. We may use the existence of a large monochromatic rectangle to find a large subspace V on which the average of f is far from the global average. Concretely, let $A \times B$ be the largest rectangle on which f equals to zero. It can be shown that $|A|, |B| \geq 2^{n-2k}$. The result of [San12a] implies the existence of a subspace V such that $|V \cap (A+B)| \geq (3/4)|V|$, where the co-dimension of V is $O(k^4)$. This implies that $\mathbb{E}[f|_V] \leq 1/4$. For $x \in \mathbb{F}_2^n$, let $\tilde{x}$ be the unique element in $\mathbb{F}_2^n/V$ satisfying $x \in V + \tilde{x}$. Note that $x + y \in V$ if and only if $\tilde{x} = \tilde{y}$. Let X, Y be random variables, independently and uniformly sampled from $\mathbb{F}_2^n$. As is conventional, here and throughout the rest of the paper we use the notation XY for the random variable (X,Y). We have

$$\mathbb{E}[f(X+Y)] - \mathbb{E}[f(X+Y)|\tilde{X} = \tilde{Y}] \geq \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \tag{4.1}$$

This shows that Π_{XY} is not independent from $\tilde{X}\tilde{Y}$. However, we need to quantify this, and to this end, in Lemma 4.3.4 we show that (4.1) implies that the mutual information between Π_{XY} and $\tilde{X}\tilde{Y}$ is large:

$$I(\Pi_{XY}; \tilde{X}\tilde{Y}) = H(\Pi_{XY}) - H(\Pi_{XY}|\tilde{X}\tilde{Y}) \geq 2^{-8}.$$

In other words, knowing which shifts of V, X and Y belong to, decreases the entropy of Π_{XY} significantly on average. In particular, there exists a coset $(V+w_1)\times(V+w_2)$ on which the entropy decreases by at least 2^{-8}. We may now iterate this process. As originally we have $H(\Pi_{XY})\leq k$ (since the partition $\mathscr{P}$ is to $K=2^k$ rectangles), after $O(k)$ iterations we will reach a constant function on a subspace of co-dimension $O(k^5)$.

4.2 Preliminaries

Combinatorial Rectangles and Partitions The Cartesian product of two sets $A,B\subseteq\mathbb{F}_2^n$ is called a combinatorial rectangle. It is well-known that the inputs that lead to a particular leaf in a deterministic communication protocol form a combinatorial rectangle, and thus every protocol taking input from $\mathbb{F}_2^n\times\mathbb{F}_2^n$ provides a partition of $\mathbb{F}_2^n\times\mathbb{F}_2^n$ into combinatorial rectangles.

We will use functions $\Pi:\mathbb{F}_2^n\times\mathbb{F}_2^n\to[K]$ to denote partitions of $\mathbb{F}_2^n\times\mathbb{F}_2^n$. Here Π maps every input to the index of the unique rectangle that contains it. For every vector space V over $\mathbb{F}_2$ we extend these definitions to $V\times V$ by identifying $V\cong\mathbb{F}_2^n$ for $n=\dim(V)$.

Entropy, Mutual Information, and Divergence The *entropy* of a discrete random variable X is defined as

$$H(X)=\sum_{a\in\operatorname{supp}(X)}\Pr[X=a]\log\frac{1}{\Pr[X=a]},$$

where here and throughout the paper, logarithms are in base two. The entropy of X conditioned on a random variable Y is defined as

$$H(X|Y)=\sum_y\Pr[Y=y]H(X|Y=y)=H(XY)-H(Y),$$

and corresponds to the amount of information that is left in X after knowing Y. Here and throughout the paper, as is customary in information theory, we use XY to denote (X,Y).

The *mutual information* between X and Y is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(XY) - H(X) - H(Y).$$

Mutual information is symmetric, it is always non-negative, and it measures the amount of the information shared between two random variables. Let μ and ν be two probability distributions on the same space. The *Kullback-Leibler divergence* (or KL-divergence, or simply divergence) of ν from μ is defined as

$$D(\mu\|\nu) = \mathop{\mathbb{E}}_{a\sim\mu}\left[\log\frac{\mu(a)}{\nu(a)}\right].$$

The divergence $D(\mu\|\nu)$ is non-negative, and it is not symmetric in μ and ν. It is equal to $+\infty$ if $\mathrm{supp}(\mu) \not\subseteq \mathrm{supp}(\nu)$. The so called *Pinsker's inequality* states that divergence can be used to bound the distance between the two probability measures:

$$\sum_a |\mu(a) - \nu(a)| \leq \sqrt{2D(\mu\|\nu)}. \tag{4.2}$$

Mutual information can be expressed using divergence. Indeed if $p(x,y)$ denotes the joint distribution of (X,Y), then

$$I(X;Y) = D(p(x,y)\|p_1(x)p_2(y)), \tag{4.3}$$

where $p_1(x)$ is the marginal distribution of X and $p_2(y)$ is the marginal distribution of Y.

4.3 Proof of Main theorem

As we have discussed in the introduction, the proof of Theorem 2.2.13 can be divided into the following three steps:

- *Step I:* Applying Sanders's result [San12a] together with Fourier sparsity of f to find a large subspace V such that

$$|\mathbb{E}[f]-\mathbb{E}[f|V]| \geq \frac{1}{4}.$$

- *Step II:* Applying information theoretic techniques to deduce from Step I that there exist $w', w'' \in \mathbb{F}_2^n$ with

$$H(\Pi | XY \in (V+w') \times (V+w'')) \leq H(\Pi) - 2^{-8}.$$

  Repeated application of Steps I and II will show the existence of a large subspace V such that $f|_V$ is constant; this answers Question 4.1.1.

- *Step III:* Using Fourier sparsity of f to deduce from Step II that f can be computed by a parity decision tree of low depth.

Next we will show how these three steps are carried out.

4.3.1 Step I: A large subspace on which the average changes significantly

We use the following result of Sanders [San12a] (see also [CS10] and [CLS13]).

Theorem 4.3.1. *Let $A, B \subseteq \mathbb{F}_2^n$ be sets of size $|A|, |B| \geq 2^n/K$. For any $\eta > 0$, there exists an affine subspace V of co-dimension $d \leq O(\log(K)^4/\eta)$ such that*

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

We also recall a corollary of the uncertainty principle, lemma 2.1.1, which implies that Fourier-sparse boolean functions cannot be too close to constant without actually being constant.

Lemma 4.3.2. *Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a function which has at most 2^s nonzero Fourier coefficients. If $\mathbb{E}[f] < 2^{-s}$ then $f \equiv 0$, and if $\mathbb{E}[f] > 1-2^{-s}$ then $f \equiv 1$.*

The following corollary establishes Step I of the proof.

Corollary 4.3.3. *Let $\Pi : \mathbb{F}_2^n \times \mathbb{F}_2^n \to [2^k]$ be a partition into $f_\oplus$-monochromatic rectangles. There exists a subspace $V \subseteq \mathbb{F}_2^n$ of co-dimension $O(k^4)$ such that*

$$|\mathbb{E}[f] - \mathbb{E}[f|V]| \geq \frac{1}{4}.$$

Proof. Assume without loss of generality that $\mathbb{E}[f] \geq 1/2$ (otherwise replace f with $1-f$). By Lemma 4.3.2, we have $\mathbb{E}[f_\oplus] = \mathbb{E}[f] \leq 1 - 2^{-k}$. Considering all the 0-rectangles in the partition, there must exist a rectangle $A \times B$ in the partition such that $f(A+B) = 0$ and $|A \times B| \geq 2^{2n-2k}$. In particular, $|A|, |B| \geq 2^{n-2k}$. Applying Theorem 4.3.1 to A, B with $K = 2^{2k}, \eta = 1/4$, we deduce the existence of an affine subspace V of co-dimension $O(k^4)$ such that $|(A+B) \cap V| \geq (3/4)|V|$. In particular, $\mathbb{E}[f|V] \leq 1/4$. □

4.3.2 Step II: Decreasing the entropy of the partition

Consider $f : \mathbb{F}_2^n \to \{0,1\}$, a partition $\Pi : \mathbb{F}_2^n \times \mathbb{F}_2^n \to [K]$ into rectangles such that $f_\oplus$ is constant on each rectangle, and a subspace V of $\mathbb{F}_2^n$. For $x \in \mathbb{F}_2^n$, let $\tilde{x}$ be the unique element in $\mathbb{F}_2^n/V$ satisfying $x \in V + \tilde{x}$.

Lemma 4.3.4. *If $|\mathbb{E}[f] - \mathbb{E}[f|V]| \geq \varepsilon$ and (X,Y) takes values in $\mathbb{F}_2^n \times \mathbb{F}_2^n$ uniformly at random, then for $\Pi_{XY} = \Pi(X,Y)$, we have*

$$I(\Pi_{XY}; \tilde{X}\tilde{Y}) \geq \varepsilon^2/16.$$

Proof. Denote $W = \mathbb{F}_2^n/V$, and for every $t \in [K]$ and $w \in W$, let $p_t = \Pr[\Pi_{XY} = t]$ and $p_{t|w,w} = \Pr[\Pi_{XY} = t | \tilde{X} = \tilde{Y} = w]$. It follows from the assumption

$$|\mathbb{E}[f] - \mathbb{E}[f|V]| = |\mathbb{E}[f(X+Y)] - \mathbb{E}[f(X+Y)|\tilde{X} = \tilde{Y}]| \geq \varepsilon$$

that $\sum_t |p_t - \mathbb{E}_{w\in W}\left[p_{t|w,w}\right]| \geq \varepsilon$. In particular

$$\sum_t \mathop{\mathbb{E}}_{w\in W}\left[|p_t - p_{t|w,w}|\right] \geq \varepsilon.$$

Since the function Π gives a partition into *rectangles*, for every $w \in W$ and $t \in [K]$, we have

$$p_{t|w,w} = \frac{\Pr[\tilde{X}=\tilde{Y}=w|\Pi_{XY}=t]\times p_t}{\Pr[\tilde{X}=\tilde{Y}=w]} = p_t \times \frac{\Pr[\tilde{X}=w|\Pi_{XY}=t]}{\Pr[\tilde{X}=w]} \times \frac{\Pr[\tilde{Y}=w|\Pi_{XY}=t]}{\Pr[\tilde{Y}=w]}.$$

Consequently, using $\max(0, 1-ab) \leq |1-a|+|1-b|$ and Pinsker's inequality (4.2) we have

$$\begin{aligned}
\varepsilon &\leq \sum_t \mathop{\mathbb{E}}_{w\in W}\left[|p_t - p_{t|w,w}|\right] = 2\sum_t \mathop{\mathbb{E}}_{w\in W}\left[\max(0, p_t - p_{t|w,w})\right] \\
&= 2\sum_t p_t \mathop{\mathbb{E}}_{w\in W}\left[\max\left(0, 1-\frac{\Pr[\tilde{X}=w|\Pi_{XY}=t]}{\Pr[\tilde{X}=w]}\times\frac{\Pr[\tilde{Y}=w|\Pi_{XY}=t]}{\Pr[\tilde{Y}=w]}\right)\right] \\
&\leq 2\sum_t p_t \mathop{\mathbb{E}}_{w\in W}\left[\left|1-\frac{\Pr[\tilde{X}=w|\Pi_{XY}=t]}{\Pr[\tilde{X}=w]}\right| + \left|1-\frac{\Pr[\tilde{Y}=w|\Pi_{XY}=t]}{\Pr[\tilde{Y}=w]}\right|\right] \\
&\leq 2\sqrt{2I(\Pi_{XY};\tilde{X})} + \sqrt{2I(\Pi_{XY};\tilde{Y})} \leq 4\sqrt{I(\Pi_{XY};\tilde{X}) + I(\Pi_{XY};\tilde{Y})}.
\end{aligned}$$

where we used (4.3) to show that $I(\Pi_{XY};\tilde{X}) = D(p_{t,w} \parallel p_t q_w)$ with $p_{t,w} = \Pr[\Pi_{XY}=t, \tilde{X}=w]$ and $q_w = \Pr[\tilde{X}=w]$, and the similar identity for $I(\Pi_{XY};\tilde{Y})$. Finally since $\tilde{X}$ and $\tilde{Y}$ are independent (even after conditioning on $\Pi_{XY}=t$), we have $I(\Pi_{XY};\tilde{X}\tilde{Y}) = I(\Pi_{XY};\tilde{X}) + I(\Pi_{XY};\tilde{Y})$. □

Note that $\tilde{X}=\tilde{Y}$ can be a very small-probability event (this is the case when V is a small subspace), and thus in the first glance it might be surprising that it is possible to use $|\mathbb{E}[f]-\mathbb{E}[f|V]|$ to obtain a lower bound for $I(\Pi_{XY};\tilde{X}\tilde{Y})$, independent of the size of V. Indeed Lemma 4.3.4 exploits the assumption that Π_{XY} is defined by a partition into combinatorial rectangles and as the following example shows this is not true for partitions into generic sets.

Example 4.3.5. *Let $V = \{x \in \mathbb{F}_2^n : x_1 = 0\}$ so that $\tilde{x} = (x_2, \ldots, x_n)$ for $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$. Con-*

sider the following partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ *into three sets*

$$\Pi_{xy} = \begin{cases} 1 & \tilde{x} = \tilde{y} \\ 2 & \tilde{x} \neq \tilde{y}, x_1 = 0 \\ 3 & \tilde{x} \neq \tilde{y}, x_1 = 1 \end{cases},$$

and let $f(x,y) = 1$ *if* $\Pi_{xy} = 2$ *and* $f(x,y) = 0$ *otherwise. Then* $\mathbb{E}[f] \approx \frac{1}{2}$ *while* $\mathbb{E}[f|\tilde{X} = \tilde{Y}] = 0$. *However* $I(\Pi_{XY}; \tilde{X}\tilde{Y}) = o(1)$. *Similarly it is easy to construct examples showing that it is essential that* X *and* Y *are independent.*

Remark 4.3.6. *Note that the proof of Lemma 4.3.4 shows that the following general statement is true. Let* μ *and* ν *be two distributions on* $\mathbb{F}_2^n$, *and let* A *and* B *be two functions on* $\mathbb{F}_2^n$ *such that* $A(X)$ *and* $B(Y)$ *have the same distribution if* $(X,Y) \sim \mu \times \nu$. *If* $\Pi : \mathbb{F}_2^n \times \mathbb{F}_2^n \to [K]$ *is a partition into rectangles, and* $g : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \{0,1\}$ *is constant on each rectangle, then*

$$|\mathbb{E}[g(X,Y)] - \mathbb{E}[g(X,Y)|A(X) = B(Y)]| \leq 4\sqrt{I(\Pi(X,Y); A(X)B(Y))}.$$

4.3.3 Step III: Constructing the PDT

Tsang et al. [TWXZ13] showed that in order to design a parity decision tree, it suffices to find a large subspace on which the function is constant; and then recurse. For completeness, we reproduce their argument. Let $\mathrm{rank}(f)$ denote the rank of the real matrix $M_{x,y} = f(x+y)$. It equals the number of nonzero Fourier coefficients of f. Note that $\log \mathrm{rank}(f) \leq D(f_\oplus)$.

Lemma 4.3.7. *Let* $T : \mathbb{N} \to \mathbb{N}$ *be a function for which the following holds. For any function* $f : \mathbb{F}_2^n \to \{0,1\}$, *if* $D(f_\oplus) = k$ *then there exists an affine subspace* V *of co-dimension* $T(k)$ *on which* f *is constant. Then for any function* $f : \mathbb{F}_2^n \to \{0,1\}$, $\mathrm{pdt}(f) \leq T(D(f_\oplus)) \cdot (D(f_\oplus) + 1)$.

Proof. The main idea is that if f is constant on V, then its rank on any coset of V reduces by at least a factor of two, which then allows for induction. To see that, assume that $\mathrm{rank}(f) = r$. Then

$$f(x) = \sum_{i=1}^{r} \widehat{f}(\alpha_i)(-1)^{\langle x, \alpha_i \rangle},$$

for some $\alpha_1, \dots, \alpha_r \in \mathbb{F}_2^n$. We know by assumption that f is constant on an affine subspace V of co-dimension $t = T(D(f_\oplus))$. We may assume that V is linear subspace, by replacing $f(x)$ with $f(x+v)$ for some $v \in V$ (note that this does not change $D(f_\oplus)$ or $\mathrm{rank}(f)$). Let W be the quotient subspace $\mathbb{F}_2^n/V$ so that $\dim(W) = t$ and $\mathbb{F}_2^n = V + W$. Note that any $x \in \mathbb{F}_2^n$ can be uniquely decomposed as $x = v + w$ with $v \in V, w \in W$. Let $\pi_V : \mathbb{F}_2^n \to V$ and $\pi_W : \mathbb{F}_2^n \to W$ be the projection maps to V and W, respectively, mapping $x = v + w$ to $\pi_V(x) = v$ and $\pi_W(x) = w$. Then

$$f|_V(v) = \sum_{i=1}^{r} \widehat{f}(\alpha_i)(-1)^{\langle v, \pi_V(\alpha_i) \rangle},$$

In particular, as f is constant on V, it must be the case that for every non-zero α_i there exists some α_j such that $\pi_V(\alpha_i) = \pi_V(\alpha_j)$, or equivalently $\alpha_i + \alpha_j \in W$. Thus

$$|\{\pi_V(\alpha_i) : i \in [r]\}| \leq \frac{r+1}{2}.$$

Let $V + w$ be any coset of V. Then

$$f|_{V+w}(v+w) = \sum_{i=1}^{r} \widehat{f}(\alpha_i)(-1)^{\langle w, \pi_W(\alpha_i) \rangle}(-1)^{\langle v, \pi_V(\alpha_i) \rangle}.$$

In particular, $\mathrm{rank}(f|_{V+w}) \leq |\{\pi_V(\alpha_i) : i \in [r]\}| \leq \frac{\mathrm{rank}(f)+1}{2}$.

We now construct the parity decision tree for f. We first query $w = \pi_W(x)$, which requires depth $\dim(W) = T(D(f_\oplus))$. Each restricted function $f|_{V+w}$ has $D((f|_{V+w})_\oplus) \leq D(f_\oplus)$ and $\mathrm{rank}(f|_{V+w}) \leq \frac{\mathrm{rank}(f)+1}{2}$, and hence by induction can be computed by a parity decision tree of depth at most $T(D(f_\oplus)) \cdot (\log(\mathrm{rank}(f)) + 1) \leq T(D(f_\oplus)) \cdot (D(f_\oplus) + 1)$. The lemma follows. □

4.3.4 Proof of Theorem 2.2.13

Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a boolean function. The associated XOR function is $f_\oplus(x,y) = f(x+y)$. Let $D(f_\oplus)$ denote the minimum complexity of a deterministic protocol which computes $f_\oplus$. We restate Theorem 2.2.13, which we prove in this section, for the convenience of the reader.

Theorem 2.2.13 ([HHL16])**.** *For any* $f : \mathbb{F}_2^n \to \{0,1\}$ *we have* $\mathrm{pdt}(f) \leq O(D(f_\oplus)^6)$.

Proof. Let $k = D(f_\oplus)$. By Corollary 4.3.3 there exists an affine subspace V of co-dimension $O(k^4)$ such that

$$|\mathbb{E}[f] - \mathbb{E}[f|V]| \geq \frac{1}{4}.$$

Let $W = \mathbb{F}_2^n/V$ so that $\mathbb{F}_2^n = V + W$. Applying Lemma 4.3.4, we obtain

$$I(\Pi;\tilde{X}\tilde{Y}) \geq 2^{-8}.$$

In particular, there exists a choice of $w', w'' \in W$ such that

$$H(\Pi|\tilde{X} = w', \tilde{Y} = w'') \leq H(\Pi) - 2^{-8}.$$

By restricting the rectangles of Π to $(V+w') \times (V+w'')$, we obtain a partition $\Pi|_{(V+w')\times(V+w'')}$ of $(V+w') \times (V+w'')$ into $f|_{V+w'+w''}$-monochromatic rectangles with

$$H(\Pi|_{(V+w')\times(V+w'')}) = H(\Pi|\tilde{X} = w', \tilde{Y} = w'') \leq H(\Pi) - 2^{-8}.$$

Since $H(\Pi) \leq k$, iterating this procedure at most $2^8 k$ times, we find an affine subspace V such that $f|_V$ is constant. Furthermore since each iteration increases the co-dimension by at most $O(k^4)$, the subspace V will have co-dimension $O(k^5)$. Finally, we can apply Lemma 4.3.7 to conclude the theorem. □

Acknowledgement Chapter 4 is a reprint of material as it appears in *SIAM Journal on Computing, 47, no. 1 (2018): 208-217.* Hamed Hatami, Kaave Hosseini, and Shachar Lovett. "Structure of protocols for XOR functions." The dissertation author was a primary investigator and author of this paper.

Chapter 5

Fourier structure of sparse sets

The purpose of this chapter is to prove the following two results. The reader is referred to Section 2.2.4 for the background regarding these results. In the following G is a finite abelian group.

Theorem 2.2.17. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ such that*

$$\left|\mathrm{Spec}_{\varepsilon}(A') + \mathrm{Spec}_{\varepsilon}(A')\right| \leq (1/\varepsilon)^{O(1/\delta)} \cdot \frac{|G|^{1+\delta}}{|A'|}$$

where $C \leq \exp((1/\varepsilon)^{O(1/\delta)})$.

Theorem 2.2.18. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ and $\varepsilon' \geq \varepsilon^{2^{1/\delta}}$ such that*

$$|\mathrm{Spec}_{\varepsilon'}(A')| \geq |\mathrm{Spec}_{\varepsilon}(A)|/C$$

and

$$|\mathrm{Spec}_{\varepsilon'}(A') + \mathrm{Spec}_{\varepsilon'}(A')| \leq C|G|^{\delta} \cdot |\mathrm{Spec}_{\varepsilon'}(A')|,$$

where $C \leq \exp\left((1/\varepsilon)^{O(2^{4/\delta})}\right)$.

Also we remind the reader of the following conjecture.

Conjecture 2.2.19. *Fix $0 < \delta < \alpha < 1/2$ and $0 < \varepsilon < 1/2$. Let $A \subseteq G$ of size $|A| \geq |G|^{\alpha}$. Then there exists a subset $A' \subseteq A$ of size $|A'| \geq |A|/C$ such that*

$$|\mathrm{Spec}_{\varepsilon}(A') + \mathrm{Spec}_{\varepsilon}(A')| \leq C|G|^{\delta} \cdot |\mathrm{Spec}_{\varepsilon}(A')|,$$

where $C = C(\varepsilon, \delta)$.

Organization We prove Theorem 2.2.17 in Section 5.1 and Theorem 2.2.18 in Section 5.2.

5.1 Proof of Theorem 2.2.17

We begin by introducing some notation. For $A \subseteq G$ and $\Gamma \subseteq \widehat{G}$, define an $|A| \times |\Gamma|$ complex matrix $M = M(A, \Gamma)$, with rows indexed by A and columns by Γ, as follows. First, denote by $\gamma(A) := \mathbb{E}_{a \in A}[\gamma(a)]$ the average value of the character γ on A. Define

$$M_{a,\gamma} := \gamma(a) \frac{\overline{\gamma(A)}}{|\gamma(A)|}.$$

With this definition, we have that for any $\Gamma \subseteq \mathrm{Spec}_{\varepsilon}(A)$,

$$\left|\mathbf{1}_A^T M(A,\Gamma) \mathbf{1}_{\Gamma}\right| = \sum_{\gamma \in \Gamma} \left| \sum_{a \in A} \gamma(a) \right| \geq \varepsilon |A||\Gamma|. \tag{5.1}$$

We next define a notion of regularity for $M(A, \Gamma)$.

Definition 5.1.1 (Regularity for $M(A,\Gamma)$)**.** *Let $A \subseteq G, \Gamma \subseteq \widehat{G}$. The matrix $M = M(A,\Gamma)$ is called λ-regular if for every pair of functions $f : A \to \mathbb{C}$, $g : \Gamma \to \mathbb{C}$ such that $\langle f, \mathbf{1}_A \rangle = 0$ or $\langle g, \mathbf{1}_{\Gamma} \rangle = 0$ or both, it holds that*

$$|f^T M g| < \lambda \|f\|_{\infty} \|g\|_{\infty} |A||\Gamma|.$$

It is conventional to use the L_2-norm in definition of regularity, however in our case, the use of L_∞-norm makes the argument more straightforward and gives better bounds.

The argument informally goes as follows. We divide into two cases. First, we show if $M = M(A, \mathrm{Spec}_\varepsilon(A))$ is λ-regular for a suitable choice of λ, then $\mathrm{Spec}_\varepsilon(A)$ has bounded doubling. Otherwise, if M is not λ-regular, we find large subsets $A' \subseteq A, \Gamma' \subseteq \mathrm{Spec}_\varepsilon(A)$ such that $M(A', \Gamma')$ has higher average. This allows us to revert to study $M(A', \mathrm{Spec}_{\varepsilon'}(A'))$ where $\varepsilon' = \varepsilon + \lambda^{O(1)}$ and iterate.

First, we analyze the case where M is regular.

Lemma 5.1.2. *Fix some* $0 < \varepsilon, \rho < 1$ *and* $\Gamma \subseteq \mathrm{Spec}_\rho(A)$. *If* $M = M(A, \Gamma)$ *is* $\varepsilon\rho/150$*-regular, then for any* $\gamma \in \mathrm{Spec}_\varepsilon(A)$*, there is a subset* $\Gamma_\gamma \subseteq \Gamma$, $|\Gamma_\gamma| \geq 0.9|\Gamma|$ *such that*

$$\gamma + \Gamma_\gamma \subset \mathrm{Spec}_{\varepsilon\rho/2}(A).$$

Proof. Suppose towards contradiction that there is some $\gamma_\circ \in \mathrm{Spec}_\varepsilon(A)$ for which the claim does not hold. That is, there exists a subset $\Gamma' \subseteq \Gamma$ of size $|\Gamma'| > 0.1|\Gamma|$ such that $\forall \gamma' \in \Gamma'$,

$$\gamma_\circ + \gamma' \notin \mathrm{Spec}_{\varepsilon\rho/2}(A).$$

Define a pair of functions $f : A \to \mathbb{C}$ and $g : \Gamma \to \mathbb{C}$ by

$$\begin{aligned} f(a) &= \gamma_\circ(a), \\ g(\gamma) &= \frac{|\Gamma|}{|\Gamma'|}\mathbf{1}_{\Gamma'}(\gamma). \end{aligned}$$

We have

$$\begin{aligned}
f^T Mg &= \sum_{\gamma\in\Gamma}\left[\sum_{a\in A}\gamma_\circ(a)\gamma(a)\frac{\overline{\gamma(A)}}{|\gamma(A)|}\right]\left[\frac{|\Gamma|}{|\Gamma'|}\mathbf{1}_{\Gamma'}(\gamma)\right]\\
&= \frac{|\Gamma|}{|\Gamma'|}\sum_{\gamma\in\Gamma}\frac{\overline{\gamma(A)}}{|\gamma(A)|}\sum_{a\in A}\gamma_\circ(a)\gamma(a)\mathbf{1}_{\Gamma'}(\gamma)\\
&= \frac{|\Gamma|}{|\Gamma'|}\sum_{\gamma'\in\Gamma'}\frac{\overline{\gamma'(A)}}{|\gamma'(A)|}\sum_{a\in A}\left(\gamma_\circ+\gamma'\right)(a).
\end{aligned}$$

By our assumption, $\forall\gamma'\in\Gamma', \gamma_\circ+\gamma'\notin \mathrm{Spec}_{\varepsilon\rho/2}(A)$. Therefore

$$\left|f^T Mg\right| \le (\varepsilon\rho/2)\cdot|\Gamma||A|.$$

Decompose f as $f=f_1+f_2$ with $f_1=\mathbb{E}_{a\in A}[f(a)]\cdot\mathbf{1}_A$ and g as $g=g_1+g_2$ with $g_1=\mathbb{E}_{\gamma\in\Gamma}[g(\gamma)]\cdot\mathbf{1}_\Gamma=\mathbf{1}_\Gamma$. Then

$$f^T Mg = f_1^T Mg_1 + f_2^T Mg_1 + f_1^T Mg_2 + f_2^T Mg_2. \tag{5.2}$$

We have that $\langle f_2,\mathbf{1}_A\rangle=0$, $\langle g_2,\mathbf{1}_\Gamma\rangle=0$ and

$$\left|f_1^T Mg_1\right| = \left|\mathop{\mathbb{E}}_{a\in A} f(a)\cdot\left(\mathbf{1}_A^T M\mathbf{1}_\Gamma\right)\right| \ge \left|\mathop{\mathbb{E}}_{a\in A}[\gamma_\circ(a)]\right|\cdot\rho|\Gamma||A| \ge \varepsilon\rho|\Gamma||A|.$$

We show that the other terms in Equation (5.2) are too small to cancel out the contribution of $f_1^T Mg_1$. Consequently, we reach a contradiction.

In each one of the terms $f_1^T Mg_2, f_2^T Mg_1, f_2^T Mg_2$ at least one of the functions are orthogonal to the identity function. Therefore, we can bound the size of these terms using the $\frac{\varepsilon\rho}{150}$-regularity assumption. We have $\|f_1\|_\infty\le 1, \|f_2\|_\infty\le 2, \|g_1\|_\infty\le 1, \|g_2\|_\infty\le 10$, and hence

$$\left|f_2^T Mg_1 + f_1^T Mg_2 + f_2^T Mg_2\right| \le (20+10+20)\cdot(\varepsilon\rho/150)|A||\Gamma| = (\varepsilon\rho/3)|A||\Gamma|.$$

This implies that $\left|f^T Mg\right|\ge\frac{2}{3}\varepsilon\rho|A||\Gamma|$, which is a contradiction. □

Next, we show how to use Lemma 5.1.2 to infer that if $M = M(A, \mathrm{Spec}_\rho(A))$ is $\frac{\varepsilon\rho}{150}$-regular then $|\mathrm{Spec}_\varepsilon(A) - \mathrm{Spec}_\varepsilon(A)|$ is small as long as $|\mathrm{Spec}_{\varepsilon\rho/2}(A)| \approx |\mathrm{Spec}_\rho(A)|$.

Lemma 5.1.3. *If $M = M(A, \mathrm{Spec}_\rho(A))$ is $\frac{\varepsilon\rho}{150}$-regular, then*

$$|\mathrm{Spec}_\varepsilon(A) - \mathrm{Spec}_\varepsilon(A)| \leq 2\frac{\left|\mathrm{Spec}_{\varepsilon\rho/2}(A)\right|^2}{\left|\mathrm{Spec}_\rho(A)\right|}.$$

Proof. Fix arbitrary $\gamma_1, \gamma_2 \in \mathrm{Spec}_\varepsilon(A)$. By Lemma 5.1.2 there exist sets $\Gamma_1, \Gamma_2 \subseteq \mathrm{Spec}_\rho(A)$ of size $|\Gamma_1|, |\Gamma_2| \geq 0.9|\mathrm{Spec}_\rho(A)|$ such that $\gamma_1 + \Gamma_1, \gamma_2 + \Gamma_2 \subseteq \mathrm{Spec}_{\varepsilon\rho/2}(A)$. For any $\gamma \in \Gamma_1 \cap \Gamma_2$ we can then write

$$\gamma_1 - \gamma_2 = (\gamma_1 + \gamma) - (\gamma_2 + \gamma)$$

where $\gamma_1 + \gamma, \gamma_2 + \gamma \in \mathrm{Spec}_{\varepsilon\rho/2}(A)$. This gives $|\Gamma_1 \cap \Gamma_2| \geq 0.8|\mathrm{Spec}_\rho(A)|$ distinct ways to write $\gamma_1 - \gamma_2$ as the difference of a pair of elements in $\mathrm{Spec}_{\varepsilon\rho/2}(A)$. Consequently

$$|\mathrm{Spec}_\varepsilon(A) - \mathrm{Spec}_\varepsilon(A)| \leq \frac{\left|\mathrm{Spec}_{\varepsilon\rho/2}(A)\right|^2}{|\Gamma_1 \cap \Gamma_2|} \leq \frac{\left|\mathrm{Spec}_{\varepsilon\rho/2}(A)\right|^2}{0.8\left|\mathrm{Spec}_\rho(A)\right|}.$$

□

Next, we consider the case that the matrix M is not λ-regular for $\lambda = \varepsilon\rho/150$. In the following we denote $\mathbb{E}[M] := \mathbb{E}_{a,\gamma}[M_{a,\gamma}]$.

Lemma 5.1.4. *If $M = M(A, \Gamma)$ is not λ-regular, then there exist subsets $A' \subseteq A$, $\Gamma' \subseteq \Gamma$ such that*

$$\left|\mathbb{E}\left[M(A', \Gamma')\right]\right| \geq |\mathbb{E}\left[M(A, \Gamma)\right]| + c\lambda^{15},$$

where $|A'| \geq c\lambda^{15}|A|, |\Gamma'| \geq c\lambda^{15}|\Gamma|$, and $c > 0$ is an absolute constant.

Assuming that $M = M(A, \Gamma)$ is not λ-regular, there are functions $f : A \to \mathbb{C}$ and $g : \Gamma \to \mathbb{C}$ with $\|f\|_\infty = \|g\|_\infty = 1$, at least one of which is orthogonal to the identity function, such that

$\left|f^T Mg\right| \geq \lambda|A||\Gamma|$. As a first step towards proving Lemma 5.1.4, we approximate f, g by step functions $\widetilde{f}$ and $\widetilde{g}$, respectively.

Claim 5.1.5. *Fix $\eta > 0$. Let $f : A \to \mathscr{C}$ be a function with $\|f\|_\infty = 1$. Then there exists a function $\widetilde{f} : A \to \mathscr{C}$ such that*

$$\|f - \widetilde{f}\|_\infty \leq \eta$$

with $\widetilde{f} = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}$, where $A_i \subseteq A$ are disjoint subsets and $\alpha_i \in \mathbb{C}$ with $|\alpha_i| \leq 1$. Moreover, $k \leq \frac{100}{\eta^2}$.

Proof. We partition A based on the phase and magnitude of f. For $r = \lceil 10/\eta \rceil$ define

$$A_{j,k} = \{a \in A : j/r < |f(a)| \leq (j+1)/r \text{ and } 2\pi k/r < \arg f(a) \leq 2\pi(k+1)/r\}.$$

We partition A to subsets $A_{j,k}$ for $j, k \in \{0, \ldots, r-1\}$. Define the step function $\widetilde{f}$ as

$$\widetilde{f} = \sum_{j,k=0}^{r-1} j/r \cdot e^{(2\pi i)k/r} \cdot \mathbf{1}_{A_{j,k}}.$$

It is easy to verify that for all $a \in A$, $|f(a) - \widetilde{f}(a)| \leq \eta$ as claimed. □

We proceed with the proof of Lemma 5.1.4.

Proof of Lemma 5.1.4. Let $\rho := \mathbb{E}[M]$ be the average of M, and define a matrix M' by $M'_{a,\gamma} = M_{a,\gamma} - \rho$, so that $\mathbb{E}[M'] = 0$. Note that $|M'_{a,\gamma}| \leq 2$ for all $a \in A, \gamma \in \Gamma$. We may assume for simplicity that ρ is real and non-negative, by multiplying all entries of M by an appropriate phase $e^{i\theta}$, as this does not change any of the properties at hand.

As we assume M is not λ-regular, there exist functions $f : A \to \mathscr{C}, g : \Gamma \to \mathscr{C}$ with $\|f\|_\infty, \|g\|_\infty = 1$, one of which at least sums to zero, such that $|f^T Mg| \geq \lambda|A||\Gamma|$. Note that $f^T M' g = f^T Mg$. Let $\widetilde{f}, \widetilde{g}$ be their step function approximations given by Claim 5.1.5 for $\eta = \lambda/8$, where $\widetilde{f} = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}, \widetilde{g} = \sum_{i=1}^{k} \beta_i \mathbf{1}_{\Gamma_i}$ and $k \leq \frac{100}{\eta^2}$. Moreover

$$\left|\widetilde{f}^T M' \widetilde{g}\right| \geq |f^T M' g| - |(f - \widetilde{f})^T M' g| - |\widetilde{f}^T M'(g - \widetilde{g})| \geq \lambda/2 \cdot |A||\Gamma|.$$

That is,

$$\left|\sum_{i,j=1}^{k} \alpha_i \beta_j \mathbf{1}_{A_i}^T M' \mathbf{1}_{\Gamma_j}\right| \geq \lambda/2 \cdot |A||\Gamma|.$$

In particular, there must exist A_i, Γ_j such that

$$\left|\mathbf{1}_{A_i}^T M' \mathbf{1}_{\Gamma_j}\right| \geq (\lambda/2k^2) \cdot |A||\Gamma| \geq c_1 \lambda^5 \cdot |A||\Gamma|,$$

where $c_1 > 0$ is an absolute constant.

If we knew that $\mathbf{1}_{A_i}^T M' \mathbf{1}_{\Gamma_j}$ is real and non-negative, say, then we would be done by choosing $A' = A_i, \Gamma' = \Gamma_j$ as then $\mathbb{E}[M(A', \Gamma')] \geq \rho + c_1 \lambda^5$. However, it may be that its real part is negative, cancelling the average. To overcome this, we consider choosing $A' \in \{A_i, A_i^c\}, \Gamma' \in \{\Gamma_j, \Gamma_j^c\}$ (where $A_i^c = A \setminus A_i, \Gamma_j^c = \Gamma \setminus \Gamma_j$) and show that one of the choices satisfies the required properties. Set

$$\alpha_1 := 1_{A_i}^T M' 1_{\Gamma_j}, \alpha_2 := 1_{A_i^c}^T M' 1_{\Gamma_j}, \alpha_3 := 1_{A_i}^T M' 1_{\Gamma_j^c}, \alpha_4 := 1_{A_i^c}^T M' 1_{\Gamma_j^c}$$

and

$$\beta_1 := |A_i||\Gamma_j|, \beta_2 := |A_i^c||\Gamma_j|, \beta_3 := |A_i||\Gamma_j^c|, \beta_4 := |A_i^c||\Gamma_j^c|.$$

Fix $\delta = c\lambda^{15}$ for an absolute constant $c > 0$ to be chosen later. We will show that for some $i \in \{1,2,3,4\}$, we have $|\beta_i| \geq \delta |A||\Gamma|$ and $|\alpha_i + \rho \beta_i| \geq (\rho + \delta)\beta_i$. This implies that if we take A', Γ' to be the corresponding sets, then $|A'| \geq \delta |A|, |\Gamma'| \geq \delta |\Gamma|$ and $|1_{A'} M 1_{\Gamma'}| = |\alpha_i + \rho \beta_i| \geq (\rho + \delta)|A'||\Gamma'|$.

In order to show that, let us note that $\sum \alpha_i = 0$, $|\alpha_1| \geq c_1 \lambda^5 |A||\Gamma|$, $\beta_1 \geq c_1 \lambda^5 |A||\Gamma|$, and the β_i are real non-negative numbers with $\sum \beta_i = |A||\Gamma|$. If for some i we have $\mathrm{Re}(\alpha_i) \geq \delta |A||\Gamma|$ then $|\alpha_i + \rho \beta_i| \geq \mathrm{Re}(\alpha_i + \rho \beta_i) \geq \delta |A||\Gamma| + \rho \beta_i \geq (\rho + \delta)\beta_i$ and we are done. If $\mathrm{Re}(\alpha_i) \leq -\delta |A||\Gamma|$ then, since $\sum \alpha_i = 0$, there exists some $j \neq i$ for which $\mathrm{Re}(\alpha_j) \geq \delta/3 \cdot |A||\Gamma|$, and we are done by the previous argument. So, we may assume that $|\mathrm{Re}(\alpha_i)| \leq \delta |A||\Gamma|$ for all i. In particular

$|\mathrm{Re}(\alpha_1)| \leq (\delta/c_1\lambda^5)\beta_1$. Hence

$$\begin{aligned}|\alpha_1 + \rho\beta_1|^2 &= |\rho\beta_1 + \mathrm{Re}(\alpha_1)|^2 + \mathrm{Im}(\alpha_1)^2 \\ &\geq \rho^2\beta_1^2 + |\alpha_1|^2 - 2\rho\beta_1|\mathrm{Re}(\alpha_1)| \\ &\geq \beta_1^2(\rho^2 + c_1^2\lambda^{10} - 2\delta/c_1\lambda^5). \\ &\geq \beta_1^2(\rho^2 + (c_1^2 - 2c/c_1)\lambda^{10}),\end{aligned}$$

where we used our choice of $\delta = c\lambda^{15}$. If we choose $c > 0$ small enough, we conclude that also in this case, $|\alpha_1 + \rho\beta_1| \geq (\rho + \delta)\beta_1$. Note that the condition $\beta_i \geq c_1\lambda^5|A||\Gamma|$ is automatically satisfied for all i, by making sure, let's say, $|A_i| \leq |A|/2$ and $|\Gamma_j| \leq |\Gamma|/2$. □

We now combine Lemma 5.1.3 and Lemma 5.1.4 in order to prove Theorem 2.2.17. The high level idea is the following. Initialize $\rho = \varepsilon, \Gamma = \mathrm{Spec}_\varepsilon(A)$. If $M(A,\Gamma)$ is λ-regular for $\lambda = \varepsilon\rho/150$, and $|\mathrm{Spec}_{\varepsilon\rho/2}(A)| \approx |\Gamma|$, then the proof follows from Lemma 5.1.3 and Parseval's identity. Otherwise, one of two cases must occur. The first case that could occur is that $M(A,\Gamma)$ is not λ-regular. Then by Lemma 5.1.4 we can replace A,Γ with A',Γ' and increase ρ by a noticeable amount. This cannot occur too many times, as $\rho \leq 1$. The second case that could occur is that $|\mathrm{Spec}_{\varepsilon\rho/2}(A)| \gg |\Gamma| \approx \mathrm{Spec}_\rho(A)$. In such a case, we set $\rho \to \varepsilon\rho/2$ and increase the spectrum of A by a noticeable amount. As the spectrum is bounded by $|G|$, this again cannot happen too many times. Combining these steps together requires a somewhat delicate balance act.

Let $K = K(\varepsilon,\delta)$ be a parameter to be optimized later. We define a sequence of sets $A_i \subseteq A$ and parameters $\rho_i \in [0,1]$ for $i \geq 1$, where initially $A_0 = A, \rho_0 = \varepsilon$. Given A_i, ρ_i set $\lambda_i = \varepsilon\rho_i/150$ and run the following procedure:

(i) If $M(A_i, \mathrm{Spec}_{\rho_i}(A_i))$ is λ_i-regular and $|\mathrm{Spec}_{\varepsilon\rho_i/2}(A_i)| \leq K|\mathrm{Spec}_{\rho_i}(A_i)|$, then set $A^* = A_i$ and finish.

(ii) If $M(A_i, \mathrm{Spec}_{\rho_i}(A_i))$ is not λ_i-regular then apply Lemma 5.1.4 to A_i and $\mathrm{Spec}_{\rho_i}(A_i)$. Let

$A' \subseteq A_i, \Gamma' \subseteq \mathrm{Spec}_{\rho_i}(A_i)$ be the resulting sets such that $|A'| \geq c\lambda_i^{15}|A_i|$, $|\Gamma'| \geq c\lambda_i^{15}|\Gamma_i|$ and $|\mathbb{E}[M(A',\Gamma')]| \geq \rho_i + c\lambda_i^{15}$. Set $A_{i+1} = A'$ and $\rho_{i+1} = \rho_i + (c/2)\lambda_i^{15}$. Return to step (i).

(iii) If $|\mathrm{Spec}_{\varepsilon\rho_i/2}(A_i)| > K|\mathrm{Spec}_{\rho_i}(A_i)|$ then set $A_{i+1} = A_i$ and $\rho_{i+1} = \varepsilon\rho_i/2$. Return to step (i).

Next, we analyze this procedure. First, note that if the procedure ends with $A^* = A_i$ then by Lemma 5.1.3 and Parseval's identity we have that

$$|\mathrm{Spec}_\varepsilon(A^*) - \mathrm{Spec}_\varepsilon(A^*)| \leq 2K|\mathrm{Spec}_{\varepsilon\rho_i/2}(A_i)| \leq \frac{8K|G|}{\varepsilon^2\rho_i^2|A_i|}. \tag{5.3}$$

So, we need to show that $\rho_i, |A_i|$ are never too small. Suppose that stages (ii) and (iii) occur k_1 and k_2 times, respectively. Let $\eta : \{1,\ldots,k_2\} \to \{1,\ldots,k_1+k_2\}$ be the ordered indices of occurrences of stage (iii). We first bound k_1.

Claim 5.1.6. *If $i < \eta(j)$ then $\rho_i \geq (\varepsilon/2)^j$.*

Proof. The value of ρ_i increases in step (ii), and decreases in step (iii) by a factor of $\varepsilon/2$. If $i < \eta(j)$ then we applied step (iii) at most $j-1$ times, hence $\rho_i \geq (\varepsilon/2)^{j-1}\rho_0 \geq (\varepsilon/2)^j$. □

Claim 5.1.7. *For $\forall j \in \{1,\ldots,k_2-1\}$, $|\eta(j+1) - \eta(j)| \leq (1/\varepsilon)^{O(j)}$.*

Proof. Consider a step i for $\eta(j) \leq i \leq \eta(j+1)$. We have that $\rho_{i+1} \geq \rho_i + (c/2)(\rho_i\varepsilon/150)^{15} \geq \rho_i + c'\varepsilon^{15(j+2)}$, where $c, c' > 0$ are absolute constants. As ρ_i never exceeds 1 for all i, this process cannot repeat more than $(1/c')(1/\varepsilon)^{15(j+2)}$ times. As we assume $\varepsilon < 1/2$, this is bounded by $(1/\varepsilon)^{c'j}$ for a large enough $c' > 0$. □

Corollary 5.1.8. $k_1 \leq (1/\varepsilon)^{O(k_2)}$.

Proof. By claim 5.1.7, $k_1 \leq \sum_{j=1}^{k_2}(1/\varepsilon)^{O(j)} \leq (1/\varepsilon)^{O(k_2)}$. □

We next upper bound k_2. To do so, we will show that in step (ii) we have that $\mathrm{Spec}_{\rho_{i+1}}(A_{i+1})$ is not much smaller than $\mathrm{Spec}_{\rho_i}(A_i)$.

Claim 5.1.9. *Assume that we run step (ii) in iteration i. Then*

$$|A_{i+1}| \geq c\lambda_i^{15}|A_i|$$

and

$$|\mathrm{Spec}_{\rho_{i+1}}(A_{i+1})| \geq c\lambda_i^{30}|\mathrm{Spec}_{\rho_i}(A_i)|,$$

where $c > 0$ is an absolute constant.

Proof. We apply in step (ii) Lemma 5.1.4 to $A_i, \mathrm{Spec}_{\rho_i}(A_i)$. We get subsets $A_{i+1} \subseteq A_i, \Gamma' \subseteq \mathrm{Spec}_{\rho_i}(A_i)$ such that $|A_{i+1}| \geq c\lambda_i^{15}|A_i|$, $|\Gamma'| \geq c\lambda_i^{15}|\mathrm{Spec}_{\rho_i}(A_i)|$ and $\rho_{i+1} \leq |\mathbb{E}[M(A_{i+1},\Gamma')]| - (c/2)\lambda_i^{15}$. Let $S = \Gamma' \cap \mathrm{Spec}_{\rho_{i+1}}(A_{i+1})$. Then

$$|\mathbb{E}[M(A_{i+1},\Gamma')]| \leq \frac{|S|}{|\Gamma'|} + \left(1 - \frac{|S|}{|\Gamma'|}\right)\rho_{i+1}.$$

Hence $|\mathrm{Spec}_{\rho_{i+1}}(A_{i+1})| \geq |S| \geq (c/2)\lambda_i^{15}|\Gamma'|$ and the claim follows. □

Combining Claim 5.1.7 and Claim 5.1.9, we deduce that, for any $j \in \{1,\dots,k_2-1\}$, the ratio in the size of the spectrums immediately after the j-th application of step (iii), and immediately before the $j+1$ application of step (iii), is lower bounded by

$$\begin{aligned} T_j := \frac{|\mathrm{Spec}_{\rho_{\eta(j)}}(A_{\eta(j)})|}{|\mathrm{Spec}_{\rho_{\eta(j+1)-1}}(A_{\eta(j+1)-1})|} &\leq \prod_{i=\eta(j)}^{\eta(j+1)-2} \frac{1}{c\lambda_i^{30}} \leq \left(\frac{1}{c}\left(\frac{150\cdot 2^j}{\varepsilon^{j+1}}\right)^{30}\right)^{\eta(j+1)-\eta(j)} \\ &\leq (1/\varepsilon)^{O(j\cdot(1/\varepsilon)^{O(j)})} \leq \exp\left((1/\varepsilon)^{O(j)}\right). \end{aligned}$$

We will choose K large enough so that $T_j \leq K^{1/2}$ for all $j < k_2$, and hence

$$|\mathrm{Spec}_{\rho_{\eta(j+1)}}(A_{\eta(j+1)})| \geq K \cdot |\mathrm{Spec}_{\rho_{\eta(j+1)-1}}(A_{\eta(j+1)-1})| \geq K^{1/2} \cdot |\mathrm{Spec}_{\rho_{\eta(j)}}(A_{\eta(j)})|.$$

Fix $K = |G|^{\delta}$ and $C = \exp((1/\varepsilon)^{O(1/\delta)})$. We may assume that $|G| \geq C$, as otherwise our bounds are trivial. Then, we must have $k_2 \leq 2/\delta$ and hence $k_1 \leq (1/\varepsilon)^{O(1/\delta)}$. We conclude that

$$\frac{|A|}{|A^*|} \leq \prod_{i=1}^{k_1+k_2} \frac{1}{c\lambda_i^{15}} \leq \exp\left((1/\varepsilon)^{O(1/\delta)}\right)$$

and that plugging these estimates into Equation (5.3) implies that

$$|\mathrm{Spec}_\varepsilon(A^*) - \mathrm{Spec}_\varepsilon(A^*)| \leq (1/\varepsilon)^{O(1/\delta)} \cdot |G|^{1+\delta}/|A^*|.$$

Since the definition of the spectrum is symmetric, $\mathrm{Spec}_\varepsilon(A^*) = -\mathrm{Spec}_\varepsilon(A^*)$, this implies the same bounds on $|\mathrm{Spec}_\varepsilon(A^*) + \mathrm{Spec}_\varepsilon(A^*)|$.

5.2 Proof of Theorem 2.2.18

The proof of theorem 2.2.18 is very similar to the proof of theorem 2.2.17, with a few small tweaks. First, we use Lemma 5.1.2 and Lemma 5.1.3 in the special case of $\rho = \varepsilon$. We restate Lemma 5.1.3 in this special case.

Lemma 5.2.1. *If $M = M(A, \mathrm{Spec}_\varepsilon(A))$ is $\frac{\varepsilon^2}{150}$-regular, then*

$$|\mathrm{Spec}_\varepsilon(A) - \mathrm{Spec}_\varepsilon(A)| \leq 2\frac{\left|\mathrm{Spec}_{\varepsilon^2/2}(A)\right|^2}{|\mathrm{Spec}_\varepsilon(A)|}.$$

We combine Lemma 5.2.1 with Lemma 5.1.4 to prove Theorem 2.2.18. The difference is in the iterative refinement process. Here, instead of setting $\lambda_i = \varepsilon\rho_i/150$, we instead set $\lambda_i = \rho_i^2/150$. To be more precise, initialize $\Gamma = \mathrm{Spec}_\varepsilon(A)$. If $M(A,\Gamma)$ is λ-regular for $\lambda = \varepsilon^2/150$, and $|\mathrm{Spec}_{\varepsilon^2/2}(A)| \approx |\Gamma|$, then the proof follows from Lemma 5.2.1 and Parseval's identity. Otherwise, one of the following two cases must occur. The first case that could occur is that $M(A,\Gamma)$ is not λ-regular. In this case, by Lemma 5.1.4 we can replace A, Γ with A',Γ' and increase ε by a noticeable amount. This can not occur many times as $\varepsilon \leq 1$. The other case that can occur is that $|\mathrm{Spec}_{\varepsilon^2/2}(A)| \gg |\Gamma| \approx \mathrm{Spec}_\varepsilon(A)$. In this case, we set $\varepsilon = \varepsilon^2/2$ and increase the spectrum of A.

Since the spectrum is bounded by $|G|$, this also can not occur too many times. In the following we formalize this high level argument.

Let $K = K(\varepsilon, \delta)$ be a parameter to be optimized later. Define a sequence of sets $A_i \subseteq A$ and parameters $\rho_i \in [0,1]$ for $i \geq 1$, and initialize $A_0 = A$ and $\rho_0 = \varepsilon$. Recall that δ is a parameter, chosen so that the final doubling constant is bounded by $|G|^{\delta}$. Given A_i, ρ_i set $\lambda_i = \rho_i^2/150$ and run the following procedure:

(i) If $M(A_i, \mathrm{Spec}_{\rho_i}(A_i))$ is λ_i-regular and $|\mathrm{Spec}_{\rho_i^2/2}(A_i)| \leq K|\mathrm{Spec}_{\rho_i}(A_i)|$, then set $A^* = A_i$ and finish.

(ii) If $M(A_i, \mathrm{Spec}_{\rho_i}(A_i))$ is not λ_i-regular then apply Lemma 5.1.4 to $A_i, \mathrm{Spec}_{\rho_i}(A_i)$. Let $A' \subseteq A_i, \Gamma' \subseteq \mathrm{Spec}_{\rho_i}(A_i)$ be sets such that $|A'| \geq c\lambda_i^{15}|A_i|$, $|\Gamma'| \geq c\lambda_i^{15}|\Gamma_i|$ and $|\mathbb{E}[M(A', \Gamma')]| \geq \rho_i + c\lambda_i^{15}$. Set $A_{i+1} = A'$ and $\rho_{i+1} = \rho_i + (c/2)\lambda_i^{15}$.

(iii) If $|\mathrm{Spec}_{\rho_i^2/2}(A_i)| > K|\mathrm{Spec}_{\rho_i}(A_i)|$ then set $A_{i+1} = A_i$ and $\rho_{i+1} = \rho_i^2/2$.

The analysis of this procedure is similar to the analysis of the procedure in the proof of Theorem 2.2.17. First note that if the procedure ends with $A^* = A_i$ and $\varepsilon^* = \rho_i$ then by Lemma 5.2.1 we have that

$$|\mathrm{Spec}_{\varepsilon^*}(A^*) - \mathrm{Spec}_{\varepsilon^*}(A^*)| \leq 2K|\mathrm{Spec}_{\varepsilon^{*2}/2}(A^*)| \leq 2K^2|\mathrm{Spec}_{\varepsilon^*}(A^*)|. \tag{5.4}$$

Therefore, we need to show that ε^* and $|A^*|$ are not too small. Suppose that stages (ii) and (iii) occur k_1 and k_2 times, respectively. Let $\eta : \{1, \cdots, k_2\} \to \{1, \cdots, k_1 + k_2\}$ be the ordered indices of occurrences of stage (iii). We first bound k_1.

Claim 5.2.2. *If $i < \eta(j)$ then $\rho_i \geq (\varepsilon/2)^{2^j}$.*

Proof. The value of ρ_i increases in step (ii), and decreases in step (iii). If $i < \eta(j)$ then we applied step (iii) at most $j-1$ times, hence $\rho_i \geq (\varepsilon/2)^{2^j}$. □

Claim 5.2.3. *For $\forall j \in \{1, \ldots, k_2 - 1\}$, $|\eta(j+1) - \eta(j)| \leq (1/\varepsilon)^{O(2^j)}$.*

Proof. Consider a step i for $\eta(j) \leq i \leq \eta(j+1)$. We have that $\rho_{i+1} \geq \rho_i + c(\rho_i^2)^{15} \geq \rho_i + c((\varepsilon/2)^{30\cdot 2^j})$. As ρ_i never exceeds 1 for all i, this process cannot repeat more than $(1/c)(2/\varepsilon)^{30\cdot 2^j}$ times. □

Corollary 5.2.4. $k_1 \leq (1/\varepsilon)^{O(2^{k_2})}$.

Proof. By claim 5.2.3, $k_1 \leq \sum_{j=1}^{k_2}(1/\varepsilon)^{O(2^j)} \leq (1/\varepsilon)^{O(2^{k_2})}$. □

We next upper bound k_2. To do so, we will show that in step (ii) we have that $\mathrm{Spec}_{\rho_{i+1}}(A_{i+1})$ is not much smaller than $\mathrm{Spec}_{\rho_i}(A_i)$. We restate Claim 5.1.9 which was proved before.

Claim 5.2.5. *Assume that we run step (ii) in iteration i. Then*

$$|A_{i+1}| \geq c\lambda_i^{15} \cdot |A_i|$$

and

$$|\mathrm{Spec}_{\rho_{i+1}}(A_{i+1})| \geq c\lambda_i^{30} \cdot |\mathrm{Spec}_{\rho_i}(A_i)|.$$

As in the proof of Theorem 2.2.17, if we combine Claim 5.2.3 and Claim 5.2.5, then for any $j \in \{1,\ldots,k_2-1\}$, the ratio in the size of the spectrums immediately after the j-th application of step (iii), and immediately before the $j+1$ application of step (iii), is lower bounded by

$$T_j := \frac{|\mathrm{Spec}_{\rho_{\eta(j)}}(A_{\eta(j)})|}{|\mathrm{Spec}_{\rho_{\eta(j+1)-1}}(A_{\eta(j+1)-1})|} \leq \exp\left((1/\varepsilon)^{O(2^j)}\right).$$

We will choose K large enough so that $T_j \leq K^{1/2}$ for all $j < k_2$, and hence

$$|\mathrm{Spec}_{\rho_{\eta(j+1)}}(A_{\eta(j+1)})| \geq K \cdot |\mathrm{Spec}_{\rho_{\eta(j+1)-1}}(A_{\eta(j+1)-1})| \geq K^{1/2} \cdot |\mathrm{Spec}_{\rho_{\eta(j)}}(A_{\eta(j)})|.$$

Fix $K = |G|^{\delta/2}$ and $C = \exp((1/\varepsilon)^{O(2^{4/\delta})})$. We may assume that $|G| \geq C$, as otherwise our bounds are trivial. Then we deduce that $k_2 \leq 4/\delta$, $k_1 \leq (2/\varepsilon)^{O(2^{4/\delta})}$. We get that

$$\frac{|A|}{|A^*|} \leq \prod_{i=1}^{k_1+k_2} \frac{1}{c(\lambda_i)^{15}} = \exp\left((1/\varepsilon)^{O(2^{4/\delta})}\right)$$

and then by plugging these estimates into Equation (5.4) we conclude that

$$|\mathrm{Spec}_{\varepsilon^*}(A^*) - \mathrm{Spec}_{\varepsilon^*}(A^*)| \leq \exp\left((1/\varepsilon)^{O(2^{4/\delta})}\right) |G|^{\delta} \cdot |\mathrm{Spec}_{\varepsilon^*}(A^*)| \, .$$

Since the definition of the spectrum is symmetric, $\mathrm{Spec}_{\varepsilon^*}(A^*) = -\mathrm{Spec}_{\varepsilon^*}(A^*)$, this implies the same bounds on $|\mathrm{Spec}_{\varepsilon^*}(A^*) + \mathrm{Spec}_{\varepsilon^*}(A^*)|$.

Acknowledgement Chapter 5 is a reprint of material as it appears in Journal ofCombinatorial Theory, Series A 148 (2017): 1-14. Kaave Hosseini and Shachar Lovett. "On the structure of the spectrum of small sets." The dissertation author was a primary investigator and author of this paper.

Part II

Pseudorandomness

Chapter 6

Limits of regularity lemma

In this chapter we construct a function with a tower type lower bound on its regularity. Namely, we prove the following theorem that we discussed in Section 2.3.1.

Theorem 2.3.3. *For every $\varepsilon > 0$ it holds that $M(\varepsilon) \geq \operatorname{twr}(\lfloor 1/16\varepsilon \rfloor)$.*

Organization We recall the basic definitions in Section 6.1. Then we prove Theorem 2.3.3 in Section 6.2. Furthermore, in Section 6.2.3 we show how the construction can be modified to give a binary valued function.

6.1 Preliminaries.

We recall the necessary definitions. Let A be an affine subspace (that is, a translation of a vector subspace) of $\mathbb{Z}_2^n$ and let $f : A \to [0,1]$ be a function. The Fourier coefficient of f associated with $\eta \in \mathbb{Z}_2^n$ is

$$\widehat{f}(\eta) = \frac{1}{|A|} \sum_{x \in A} f(x)(-1)^{\langle x, \eta \rangle} = \mathop{\mathbb{E}}_{x \in A} \left[f(x)(-1)^{\langle x, \eta \rangle} \right] .$$

Any subspace $H \leq \mathbb{Z}_2^n$ naturally determines a partition of $\mathbb{Z}_2^n$ into affine subspaces

$$\mathbb{Z}_2^n / H = \{ H + g \, : \, g \in \mathbb{Z}_2^n \} .$$

The number $\left|\mathbb{Z}_2^n/H\right| = 2^{n-\dim H}$ of translations is called the *index* of H.

For an affine subspace $A = H + g$ of $\mathbb{Z}_2^n$, where $H \le \mathbb{Z}_2^n$ and $g \in \mathbb{Z}_2^n$, we say that a function $f : A \to [0,1]$ is *ε-regular* if all its nontrivial Fourier coefficients are bounded by ε, that is,

$$\max_{\eta \notin H^\perp} \left|\widehat{f}(\eta)\right| \le \varepsilon \, .$$

Note that a trivial Fourier coefficient $\eta \in H^\perp$ satisfies $|\widehat{f}(\eta)| = |\mathbb{E}_{x\in A} f(x)|$. Henceforth, for any $f : \mathbb{Z}_2^n \to [0,1]$ we denote by $f|_A : A \to [0,1]$ the restriction of f to A.

Definition 2.3.1 (ε-regular subspace)**.** *Let $f : \mathbb{Z}_2^n \to [0,1]$. A subspace $H \le \mathbb{Z}_2^n$ is* ε-regular *for f if $f|_A$ is ε-regular for at least $(1-\varepsilon)\left|\mathbb{Z}_2^n/H\right|$ translations A of H.*

Theorem 2.3.2 (Arithmetic regularity lemma in $\mathbb{Z}_2^n$, Theorem 2.1 in [Gre05b])**.** *For every $0 < \varepsilon < \frac{1}{2}$ there is $M(\varepsilon)$ such that every function $f : \mathbb{Z}_2^n \to [0,1]$ has an ε-regular subspace of index at most $M(\varepsilon)$. Moreover, $M(\varepsilon) \le \mathrm{twr}(\lceil 1/\varepsilon^3 \rceil)$.*

6.2 Proof of Theorem 2.3.3

6.2.1 The Construction

To construct a function witnessing the lower bound in Theorem 2.3.3 we will use pseudorandom spanning sets.

Claim 6.2.1. *Let V be a vector space over $\mathbb{Z}_2$ of dimension d. Then there is a set of $8d$ nonzero vectors in V such that any subset of $\frac{3}{4}$ of them spans V.*

Proof. Choose random vectors $v_1, \dots, v_{8d} \in V \setminus \{0\}$ independently and uniformly. Let U be a subspace of V of dimension $d-1$. The probability that a given v_i lies in U is at most $\frac{1}{2}$. By Chernoff's bound, the probability that more than $6d$ of our vectors v_i lie in U is smaller than $\exp(-2(2d)^2/8d) = \exp(-d)$. By the union bound, the probability that there exists a subspace

U of dimension $d-1$ for which the above holds is at most $2^d \exp(-d) < 1$. This completes the proof. □

We now describe a function $f : \mathbb{Z}_2^n \to [0,1]$ which, as we will later prove, has no ε-regular subspace of small index. Henceforth set $s = \lfloor 1/16\varepsilon \rfloor$. Furthermore, let d_i be the following sequence of integers of tower-type growth:

$$d_{i+1} = \begin{cases} 2^{D_i} & \text{if } i = 1,2,3 \\ 2^{D_i - 3} & \text{if } i > 3 \end{cases} \qquad \text{where } D_i = \sum_{j=1}^{i} d_j \text{ and } D_0 = 0 \,.$$

Note that the first values of d_i for $i \geq 1$ are $1, 2, 8, 2^8, 2^{264}$, etc., and it is not hard to see that $d_i \geq \operatorname{twr}(i-1)$ for every $i \geq 1$. Set $n = D_s$ ($\geq \operatorname{twr}(s-1)$). For $x \in \mathbb{Z}_2^n$, partition its coordinates into s blocks of sizes $d_1, \ldots, d_s$, and identify $x = (x^1, \ldots, x^s) \in \mathbb{Z}_2^{d_1 + \cdots + d_s} = \mathbb{Z}_2^n$.

Let $1 \leq i \leq s$. Bijectively associate with each $v \in \mathbb{Z}_2^{D_{i-1}} = \mathbb{Z}_2^{d_1 + \cdots + d_{i-1}}$ a nonzero vector $\xi_i(v) \in \mathbb{Z}_2^{d_i}$ such that the set of vectors $\{\xi_i(v) : v \in \mathbb{Z}_2^{D_{i-1}}\}$ has the property that any subset of $\frac{3}{4}$ of its elements spans $\mathbb{Z}_2^{d_i}$. The existence of such a set, which is a subset of size $2^{D_{i-1}}$ in a vector space of dimension d_i, follows from Claim 6.2.1 when $i > 3$, since then $2^{D_{i-1}} = 8d_i$. When $i \leq 3$ the existence of such a set is trivial since $\lceil (3/4)i \rceil = i$, hence any basis would do (and we take $2^{D_{i-1}} = d_i$). With a slight abuse of notation, if $x \in \mathbb{Z}_2^n$ we write $\xi_i(x)$ for $\xi_i((x^1, \ldots, x^{i-1}))$.

We define our function $f : \mathbb{Z}_2^n \to [0,1]$ as

$$f(x) = \frac{\left|\{1 \leq i \leq s : \langle x^i, \xi_i(x) \rangle = 0\}\right|}{s} \,.$$

The following is our main technical lemma, from which Theorem 2.3.3 immediately follows.

Lemma 6.2.2. *The only ε-regular subspace for f is the zero subspace $\{0\}$.*

Proof of Theorem 2.3.3. The index of $\{0\}$ is $\left|\mathbb{Z}_2^n / \{0\}\right| = 2^n \geq \operatorname{twr}(s) = \operatorname{twr}(\lfloor 1/16\varepsilon \rfloor)$. □

6.2.2 Proof of Lemma 6.2.2

Let H be an ε-regular subspace for f, and assume towards contradiction that $H \neq \{0\}$. Let $1 \leq i \leq s$ be minimal such that there exists $v \in H$ for which $v^i \neq 0$. For any $g \in \mathbb{Z}_2^n$, let

$$\gamma_g = (0, \ldots, 0, \xi_i(g), 0, \ldots, 0) \in \mathbb{Z}_2^n$$

where only the i-th component is nonzero. We will show that for more than an ε-fraction of the translations $H + g$ of H it holds that $\gamma_g \notin H^\perp$ and

$$\widehat{f|_{H+g}}(\gamma_g) > \varepsilon \; .$$

This will imply that H is not ε-regular for f, thus completing the proof.

First, we argue that $\gamma_g \notin H^\perp$ for a noticeable fraction of $g \in \mathbb{Z}_2^n$. We henceforth let $B = \{g \in \mathbb{Z}_2^n : \gamma_g \in H^\perp\}$ be the set of "bad" elements.

Claim 6.2.3. $|B| \leq \frac{3}{4} \left|\mathbb{Z}_2^n\right|$.

Proof. If $g \in B$ then $\langle \xi_i(g), v^i \rangle = 0$. Hence, $\{\xi_i(g) : g \in B\}$ does not span $\mathbb{Z}_2^{d_i}$. By the construction of ξ_i, this means that $\{(g^1, \ldots, g^{i-1}) : g \in B\}$ accounts to at most $\frac{3}{4}$ of the elements in $\mathbb{Z}_2^{D_{i-1}}$, and hence $|B| \leq \frac{3}{4} \left|\mathbb{Z}_2^n\right|$. □

Next, we argue that typically $\widehat{f|_{H+g}}(\gamma_g)$ is large. Let $W \leq \mathbb{Z}_2^n$ be the subspace spanned by the last $s - i$ blocks, that is, $W = \{w \in \mathbb{Z}_2^n : w^1 = \ldots = w^i = 0\}$. Note that for any $g \in \mathbb{Z}_2^n, w \in W$ we have $\gamma_{g+w} = \gamma_g$. In particular, $g + w \in B$ if and only if $g \in B$.

Claim 6.2.4. *Fix $g \in \mathbb{Z}_2^n$ such that $\gamma_g \notin H^\perp$. Then*

$$\mathop{\mathbb{E}}_{w \in W} \left[\widehat{f|_{H+g+w}}(\gamma_g) \right] = \frac{1}{2s} \; .$$

Proof. Write $f(x) = \frac{1}{s} \sum_{j=1}^{s} B_j(x)$ where $B_j(x) : \mathbb{Z}_2^n \to \{0, 1\}$ is the characteristic function for the set of vectors x satisfying $\langle x^j, \xi_j(x)\rangle = 0$. Hence, for any affine subspace A in $\mathbb{Z}_2^n$,

$$\widehat{f|_A}(\gamma_g) = \frac{1}{s}\sum_{j=1}^{s} \widehat{B_j|_A}(\gamma_g)\ . \tag{6.1}$$

Set $A = H + g + w$ for an arbitrary $w \in W$. We next analyze the Fourier coefficient $\widehat{B_j|_A}(\gamma_g)$ for each $j \le i$, and note that in these cases we have $\xi_j(x) = \xi_j(g)$ for any $x \in A$. First, if $j < i$ then for every $x \in A$ we have $x^j = g^j$, which implies that $B_j|_A$ is constant. Since a nontrivial Fourier coefficient of a constant function equals 0, we have

$$\widehat{B_j|_A}(\gamma_g) = 0, \qquad \forall j < i. \tag{6.2}$$

Next, for $j = i$, write $B_i|_A(x) = \frac{1}{2}((-1)^{\langle x^i,\xi_i(x)\rangle} + 1)$. Since $\langle x, \gamma_g\rangle = \langle x^i, \xi_i(x)\rangle$, we have

$$\widehat{B_i|_A}(\gamma_g) = \mathop{\mathbb{E}}_{x\in A}\left[\frac{1}{2}((-1)^{\langle x^i,\xi_i(x)\rangle} + 1)\cdot(-1)^{\langle x^i,\xi_i(x)\rangle}\right] = \mathop{\mathbb{E}}_{x\in A}[B_i(x)] = \frac{1}{2}\ . \tag{6.3}$$

Finally, for $j > i$ we average over all $w \in W$. Let $H + W$ be the subspace spanned by H, W. Writing $B_j(x) = \frac{1}{2}((-1)^{\langle x^j,\xi_j(x)\rangle} + 1)$, the average Fourier coefficient is

$$\mathop{\mathbb{E}}_{w\in W}\mathop{\mathbb{E}}_{x\in H+g+w}\left[B_j(x)(-1)^{\langle x^i,\xi_i(x)\rangle}\right] = \frac{1}{2}\mathop{\mathbb{E}}_{x\in H+W+g}\left[(-1)^{\langle x^i,\xi_i(g)\rangle + \langle x^j,\xi_j(x)\rangle}\right].$$

Note that for every fixing of $x^1, \dots, x^{j-1}$, we have that x^j is uniformly distributed in $\mathbb{Z}_2^{d_j}$ (due to W), and that $(-1)^{\langle x^i,\xi_i(g)\rangle}$ is constant. Since $\xi_j(x) \neq 0$, we conclude that

$$\mathop{\mathbb{E}}_{w\in W}\left[\widehat{B_j|_{H+g+w}}(\gamma_g)\right] = 0, \qquad \forall j > i. \tag{6.4}$$

The proof now follows by substituting (6.2), (6.3) and (6.4) into (6.1). □

Since $\widehat{f|_{H+g+w}}(\gamma_g) \le 1$, we infer (via a simple averaging argument) the following corollary.

Corollary 6.2.5. *Fix $g \in \mathbb{Z}_2^n$ such that $\gamma_g \notin H^\perp$. Then for more than $1/4s$ fraction of the elements*

$w \in W$,

$$\widehat{f|_{H+g+w}}(\gamma_g) > \frac{1}{4s} \,.$$

We are now ready to conclude the proof of Lemma 6.2.2. Partition $\mathbb{Z}_2^n$ into translations of W, and recall that γ_g depends just on the translation $g+W$. By Claim 6.2.3, for at least $\frac{1}{4}$ of the translations, $\gamma_g \notin H^{\perp}$. By Corollary 6.2.5, in each such translation, more than $1/4s$-fraction of the elements $g+w$ satisfy $\widehat{f|_{H+g+w}}(\gamma_g) > 1/4s$. Since $1/16s \geq \varepsilon$, the subspace H cannot be ε-regular for f.

6.2.3 A variant of Theorem 2.3.3 for binary functions

One can also deduce from Theorem 2.3.3 a similar bound for ε-regular *sets*, that is, for binary functions $f : \mathbb{Z}_2^n \to \{0,1\}$. For this, all we need is the following easy probabilistic argument.

Claim 6.2.6. *Let $\tau > 0$ and $f : \mathbb{Z}_2^n \to [0,1]$. There exists a binary function $S : \mathbb{Z}_2^n \to \{0,1\}$ satisfying, for every affine subspace A of $\mathbb{Z}_2^n$ of size $|A| \geq 4n^2/\tau^2$ and any vector $\eta \in \mathbb{Z}_2^n$, that*

$$\left|\widehat{S|_A}(\eta) - \widehat{f|_A}(\eta)\right| \leq \tau.$$

Proof. Choose $S : \mathbb{Z}_2^n \to \{0,1\}$ randomly by setting $S(x) = 1$ with probability $f(x)$, independently for each $x \in \mathbb{Z}_2^n$, Let A, η be as in the statement. The random variable

$$\widehat{S|_A}(\eta) = \frac{1}{|A|} \sum_{x \in A} S(x)(-1)^{\langle x,\eta \rangle}$$

is an average of $|A|$ mutually independent random variables taking values in $[-1,1]$, and its expectation is $\widehat{f|_A}(\eta)$. By Hoeffding's bound, the probability that $\left|\widehat{S|_A}(\eta) - \widehat{f|_A}(\eta)\right| > \tau$ is smaller than

$$2\exp(-\tau^2 \,|A|/2) \leq 2^{-2n^2+1} \,.$$

The number of vector subspaces over $\mathbb{Z}_2^n$ can be trivially bounded by 2^{n^2}, the number of sequences of n vectors in $\mathbb{Z}_2^n$. Hence, the number of pairs (A,η) is bounded by 2^{n^2+n}. The claim follows by the union bound. □

Applying Claim 6.2.6 with $\tau = \varepsilon/2$ (say) implies that if $f : \mathbb{Z}_2^n \to [0,1]$ has no ε-regular subspace of index smaller than $\mathrm{twr}(\lfloor 1/16\varepsilon \rfloor)$ then, provided n is sufficiently large in terms of ε, there is $S : \mathbb{Z}_2^n \to \{0,1\}$ that has no $\varepsilon/2$-regular subspace of index smaller than $\mathrm{twr}(\lfloor 1/16\varepsilon \rfloor)$.

Acknowledgement Chapter 6 is a reprint of material as it appears in Mathematical Proceedings of the Cambridge Philosophical Society,vol. 161, no. 2, pp. 193-197. Cambridge University Press, 2016. Kaave Hosseini , Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. "An improved lower bound for arithmetic regularity." The dissertation author was a primary investigator and author of this paper.

Chapter 7

Pseudorandom generators via polarizing random walks

In this chapter we describe the details of framework we discussed in section 2.3.2, to construct pseudorandom generators based on the notion of fractional pseudorandom generator. In particular we develop the theory to prove the formal analog of the following statement that we described in section 2.3.2.

Theorem 7.1.4 (Main theorem, informal version of Theorem 7.1.6)**.** *Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathscr{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1, \dots, X_t$ be i.i.d. copies of X. There is an explicit random variable $G = G(X_1, \cdots, X_t) \in \{-1,1\}^n$ so that G is a PRG for $\mathscr{F}$ with error $(t+1)\varepsilon$.*

As an application, we obtain the following unified PRG for all classes of boolean functions with bounded Fourier growth in L_1 norm.

Theorem 2.3.7 (PRG for functions of bounded L_1 Fourier tail, informal version of Theorem 7.3.5)**.** *Let $\mathscr{F}$ be a family of n-variate Boolean functions closed under restrictions. Assume that there exist*

$a, b \geq 1$ such that for every $f \in \mathscr{F}$,

$$\sum_{S \subseteq [n] : |S| = k} |\widehat{f}(S)| \leq a \cdot b^k.$$

Then, for any $\varepsilon < \varepsilon \leq \frac{1}{\mathrm{poly}(b \log n)}$ there exists an explicit PRG $X \in \{-1,1\}^n$ which fools $\mathscr{F}$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log(a/\varepsilon))b^2)$.

Organization We describe the general framework in detail in Section 7.1. We prove Theorem 7.1.6 in Section 7.2. We describe applications in Section 7.3.

7.1 General framework

7.1.1 Boolean functions

Let $f : \{-1,1\}^n \to [-1,1]$ be an n-variate Boolean function, identified with its multilinear extension, also known as its Fourier expansion. For $x \in [-1,1]^n$ define $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x^S$ where $x^S = \prod_{i \in S} x_i$. As f is multilinear, a convenient viewpoint is to view $f(x)$ as computing the expected value of f on a product distribution on $\{-1,1\}^n$. That is, let $W = W(x) \in \{-1,1\}^n$ be a random variable, where $W_1, \ldots, W_n$ are independently chosen so that $\mathbb{E}[W_i] = x_i$. Then $f(x) = \mathbb{E} f(W)$. In particular, $f(\overline{0}) = \mathbb{E} f(U)$, where $U \in \{-1,1\}^n$ is uniformly chosen.

A family $\mathscr{F}$ of n-variate Boolean functions is said to be *closed under restrictions* if for any $f \in \mathscr{F}$ and any function $f' : \{-1,1\}^n \to \{-1,1\}$ obtained from f by fixing some of its inputs to $\{-1,1\}$ it holds that also $f' \in \mathscr{F}$.

7.1.2 Pseudorandom generators

Let $\mathscr{F}$ be a family of n-variate Boolean functions. The following is the standard definition of a pseudorandom generator (PRG) for $\mathscr{F}$, adapted to our notation.

Definition 7.1.1 (PRG)**.** *A random variable $X \in \{-1,1\}^n$ is a PRG for $\mathscr{F}$ with error ε, if for any $f \in \mathscr{F}$ it holds that $\left|f(\overline{0}) - \mathbb{E} f(X)\right| \leq \varepsilon$.*

We introduce the notion of a *fractional PRG*. It is the same as a PRG, except that the random variable is allowed to take values in $[-1,1]^n$, instead of only Boolean values. We assume that X has finite support.

Definition 7.1.2 (Fractional PRG)**.** *A random variable $X \in [-1,1]^n$ with finite support, is a fractional PRG for $\mathscr{F}$ with error ε, if for any $f \in \mathscr{F}$ it holds that $\left|f(\overline{0}) - \mathbb{E} f(X)\right| \leq \varepsilon$.*

Our main goal will be to "amplify" fractional PRGs for $\mathscr{F}$ in order to obtain PRGs for $\mathscr{F}$. To that end, we need to enforce some non-triviality conditions on the fractional PRG. For example, $X = \overline{0}$ is a fractional PRG for any function. We require that for any coordinate $i \in [n]$, the value of X_i is far from zero with noticeable probability. Formally, we require a noticeable second moment.

Definition 7.1.3 (p-noticeable random variable)**.** *A random variable $X \in [-1,1]^n$ is p-noticeable if for every $i \in [n]$, $\mathbb{E}[X_i^2] \geq p$.*

For technical reasons, we would also need X to be *symmetric*, which means that the distribution of $-X$ is the same as the distribution of X. This is easy to achieve, for example by multiplying all elements of X with a uniformly chosen sign.

Fractional PRG as steps in a random walk

Let $X \in [-1,1]^n$ be a fractional PRG for f with error ε. That is,

$$|\mathop{\mathbb{E}}_{X}[f(X)] - f(\overline{0})| \leq \varepsilon.$$

The goal is to construct a random variable $Y \in \{-1,1\}^n$ such that $\mathbb{E}_Y[f(Y)] \approx f(\overline{0})$, where the fractional PRG X provides a "small step" towards this approximation. If we can combine these small steps in a way that they converge fast to $\{-1,1\}^n$, then we would be done. To be a bit more precise, consider a random walk starting at $\overline{0}$ with the following properties:

1. The value of f at each step on average does not change by too much.

2. The random walk converges fast to $\{-1,1\}^n$.

Observe that if we take X as the first step, then property 1 is satisfied for the first step. Considering later steps leads to the following question: given a point $y \in [-1,1]^n$, can we find a random variable $A = A(y,X)$ such that

$$|\mathbb{E}[f(A)] - f(y)| \leq \varepsilon,$$

and such that A takes values closer to Boolean values? We show that this is indeed the case if we assume that X not only fools f, but also fools any possible restriction of f.

To formalize this, let $\mathscr{F}$ be a family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$. We say that $\mathscr{F}$ is closed under restrictions if for any $f \in \mathscr{F}$, if we fix some inputs of f to constants $\{-1,1\}$, then the new restricted function is still in $\mathscr{F}$. Most natural families of Boolean functions studied satisfy this condition. Some examples are functions computed by small-depth circuits, functions computed by bounded width branching programs, and functions of low sensitivity.

We show that if X is a fractional PRG for such $\mathscr{F}$, then it can be used to approximate $f(y)$ for any $y \in [-1,1]^n$. Define $\delta_y \in [0,1]^n$ by $(\delta_y)_i = 1 - |y_i|$. For $x, x' \in [-1,1]^n$ define $x \circ x' \in [-1,1]^n$ to be their coordinate-wise product, $(x \circ x')_i = x_i x'_i$. Note that under this definition, the sub-cube $\{y + \delta_y \circ x : x \in [-1,1]^n\}$ is the largest symmetric sub-cube of $[-1,1]^n$ centered at y.

We show (Claim 7.2.3) that if $X \in [-1,1]^n$ is a fractional PRG for $\mathscr{F}$ which is closed under restrictions, then for any $f \in \mathscr{F}$ and any $y \in [-1,1]^n$ it holds that

$$|\mathbb{E}[f(y + \delta_y \circ X)] - f(y)| \leq \varepsilon.$$

Technically, we need to also assume that X is *symmetric*, which means that $\Pr[X = x] = \Pr[X = -x]$ for all x. This is easy to achieve from any X which is not symmetric, for example by multiplying X with a uniform bit (thus, increasing its seed length by 1 bit).

Polarization and fast convergence

Our next goal is to show fast convergence of the random walk to $\{-1,1\}^n$. To that end, we need to analyze the following martingale:

$$Y_1 = X_1$$
$$Y_i = Y_{i-1} + \delta_{Y_{i-1}} \circ X_i$$

where $X_1, X_2, \ldots$ are independent copies of a fractional PRG. We show that for some t not too large, Y_t is close to a point in $\{-1,1\}^n$. But why would that be true? This turns out to be the result of *polarization* in the random walk. It suffices to show this for every coordinate individually.

So, let $Z_1, Z_2, \ldots \in [-1,1]$ be independent random variables (which are the i-th coordinate of $X_1, X_2, \ldots$ for some fixed i), and define the following one-dimensional martingale:

$$W_1 = Z_1$$
$$W_i = W_{i-1} + (1 - |W_{i-1}|)Z_i.$$

Claim 7.2.5 shows that if (i) Z_i is symmetric, and (ii) $\mathbb{E}[Z_i^2] \geq p$ (which follows from our assumption that the fractional PRG is p-noticeable), then it holds that

$$\Pr[|W_t| \geq 1 - \delta] \geq 1 - \delta$$

for $t = O(\log(1/\delta)/p)$. Setting $\delta = \varepsilon/n$ guarantees that with probability $1 - \varepsilon$ all the coordinates of Y_t are ε/n close to $\{-1,1\}$. Then a simple argument shows that rounding the coordinates gives a PRG with error $O(\varepsilon)$, as desired.

We now state our main theorem.

Theorem 7.1.4 (Main theorem, informal version of Theorem 7.1.6)**.** *Let $\mathscr{F}$ be a family of n-variate Boolean functions that is closed under restrictions. Let $X \in [-1,1]^n$ be a symmetric p-noticeable* *fractional PRG for $\mathscr{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1,\dots,X_t$ be i.i.d. copies of X. Define the following random variables taking values in $[-1,1]^n$:*

$$Y_0 = \overline{0}; \qquad Y_i = Y_{i-1} + \delta_{Y_{i-1}} \circ X_i \qquad i = 1,\dots,t.$$

Let $G = sign(Y_t) \in \{-1,1\}^n$ obtained by taking the sign of the coordinates in Y_t. Then G is a PRG for $\mathscr{F}$ with error $(t+1)\varepsilon$.

Note that computing this PRG only involves basic operations such as addition and multiplication over the reals with bounded error.

7.1.3 Polarizing random walks

The main idea is to view a fractional PRG as steps in a random walk in $[-1,1]^n$ that converges to $\{-1,1\}^n$. To that end, we define a gadget that implements the random walk; and moreover, that allows for fast convergence. As we will see later, the fast convergence is an effect of polarization.

Definition 7.1.5 (Random walk gadget)**.** *For any $t \geq 1$ define the random walk gadget $g_t : [-1,1]^t \to [-1,1]$ as follows. Let $a_1,\dots,a_t \in [-1,1]$. Define $g_1(a_1) := a_1$ and for $t > 1$,*

$$g_t(a_1,\dots,a_t) := g_{t-1}(a_1,\dots,a_{t-1}) + (1 - |g_{t-1}(a_1,\dots,a_{t-1})|)a_t.$$

We extend the definition to act on bit-vectors. Define $g_t^n : ([-1,1]^n)^t \to [-1,1]^n$ as follows. For $x_1,\dots,x_t \in [-1,1]^n$ define

$$g_t^n(x_1,\dots,x_t) = (g_t(x_{1,1},\dots,x_{t,1}),\dots,g_t(x_{1,n},\dots,x_{t,n}))\,.$$

Equivalently, we can view g_t^n as follows: construct a $t \times n$ matrix whose rows are $x_1,\dots,x_t$; and then apply g_t to each column of the matrix to obtain a resulting vector in $[-1,1]^n$.

The following theorem shows how to “amplify" fractional PRGs using the random walk gadget to obtain a PRG. Below, for $x \in [-1,1]^n$ we denote by $\operatorname{sign}(x) \in \{-1,1\}^n$ the Boolean vector obtained by taking the sign of each coordinate (the sign of 0 can be chosen arbitrarily).

Theorem 7.1.6 (Amplification Theorem)**.** *Let $\mathscr{F}$ be a family of n-variate Boolean functions which is closed under restrictions. Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathscr{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1,\ldots,X_t$ be iid copies of X. Define a random variable $G \in \{-1,1\}^n$ as follows:*

$$G := G(X_1,\ldots,X_t) = sign(g_t^n(X_1,\ldots,X_t)).$$

Then G is a PRG for $\mathscr{F}$ with error $(t+1)\varepsilon$.

7.2 Proof of Amplification Theorem

We prove Theorem 7.1.6 in this section. From here onwards, we fix a family $\mathscr{F}$ of n-variate Boolean functions which is closed under restrictions. The proof is based on the following two lemmas. The first lemma amplifies a p-noticeable fractional PRG to a $(1-q)$-noticeable fractional PRG. The second lemma shows that setting $q = \varepsilon/n$, the latter fractional PRG can be rounded to a Boolean-valued PRG without incurring too much error.

Lemma 7.2.1 (Amplification lemma)**.** *Let $X_1,\ldots,X_t \in [-1,1]^n$ be independent symmetric p-noticeable fractional PRGs for $\mathscr{F}$ with error ε. Define a random variable $Y \in [-1,1]^n$ as*

$$Y := g_t^n(X_1,\ldots,X_t).$$

Then Y is a $(1-q)$-noticeable fractional PRG for $\mathscr{F}$ with error $t\varepsilon$, where $q = 2^{-\Omega(pt)}$.

Lemma 7.2.2 (Rounding lemma)**.** *Let $Y \in [-1,1]^n$ be a $(1-q)$-noticeable fractional PRG for $\mathscr{F}$ with error ε. Then $sign(Y) \in \{-1,1\}^n$ is a PRG for $\mathscr{F}$ with error $\varepsilon + qn$.*

Theorem 7.1.6 follows directly by applying Lemma 7.2.1 with $t = O(\log(n/\varepsilon)/p)$ to obtain $q = \varepsilon/n$ and then applying Lemma 7.2.2.

7.2.1 Proof of Lemma 7.2.1

We prove Lemma 7.2.1 in this section. We need to prove two claims: that $g_t^n(X_1, \dots, X_t)$ is a fractional PRG for $\mathscr{F}$ with error εt, and that it is $(1-q)$-noticeable. This is achieved in the following sequence of claims.

First we need some notations. For $y \in [-1,1]^n$ define $\delta_y \in [-1,1]^n$ by $(\delta_y)_i := 1 - |y_i|$. For two vectors $x, y \in [-1,1]^n$ define $x \circ y \in [-1,1]^n$ to be their point-wise product, namely $(x \circ y)_i := x_i y_i$. Observe that $\{y + \delta_y \circ x : x \in [-1,1]^n\}$ is the largest symmetric sub-cube in $[-1,1]^n$ centered at y.

Claim 7.2.3. *Let $X \in [-1,1]^n$ be a fractional PRG for $\mathscr{F}$ with error ε. Then for any $f \in \mathscr{F}$ and any $y \in [-1,1]^n$,*

$$\left| f(y) - \mathbb{E} f(y + \delta_y \circ X) \right| \leq \varepsilon.$$

Proof. Consider a distribution over $F \in \mathscr{F}$ obtained from f by fixing the i-th input to $\mathrm{sign}(y_i)$ with probability $|y_i|$, independently for each i. That is,

$$F(x) := f(R(x)),$$

where $R(x) \in [-1,1]^n$ is a random variable obtained by sampling $R_1, \dots, R_n$ independently where each R_i is chosen as follows. Pick $R_i(x) = \mathrm{sign}(y_i)$ with probability $|y_i|$ and with probability $1 - |y_i|$ do as follows: pick $R_i(x) = 1$ with probability $(x_i + 1)/2$ and pick $R_i(x) = -1$ otherwise. It's easy to check that $\mathbb{E}_R(R(x)) = y + \delta_y \circ x$. By multi-linearity of f, and as $R(x)$ is a product distribution, for all $x \in [-1,1]^n$,

$$\mathop{\mathbb{E}}_F[F(x)] = \mathop{\mathbb{E}}_R[f(R(x))] = f(\mathop{\mathbb{E}}_R[R(x)]) = f(y + \delta_y \circ x).$$

Setting $x = X$ and averaging over X gives

$$\left| f(y) - \mathop{\mathbb{E}}_{X}[f(y + \delta_y \circ X)] \right| = \left| \mathop{\mathbb{E}}_{F}[F(\overline{0})] - \mathop{\mathbb{E}}_{F,X}[F(X)] \right| \leq \mathop{\mathbb{E}}_{F} \left| F(\overline{0}) - \mathop{\mathbb{E}}_{X}[F(X)] \right| \leq \varepsilon,$$

since $F \in \mathscr{F}$ with probability one and X is a fractional PRG for $\mathscr{F}$ with error ε. □

Claim 7.2.4. *Let $X_1, \ldots, X_t \in [-1,1]^n$ be independent fractional PRGs for $\mathscr{F}$ with error ε. Then for any $f \in \mathscr{F}$,*

$$\left| f(\overline{0}) - \mathop{\mathbb{E}}_{X_1,\ldots,X_t}[f(g_t^n(X_1, \ldots, X_t))] \right| \leq t\varepsilon.$$

Proof. The proof is by induction on t. The base case $t = 1$ follows by definition as $g_1^n(X_1) = X_1$. For $t > 1$ we will show that

$$\left| \mathbb{E}[f(g_{t-1}^n(X_1, \ldots, X_{t-1}))] - \mathbb{E}[f(g_t^n(X_1, \ldots, X_t))] \right| \leq \varepsilon,$$

from which the claim follows by the triangle inequality. In fact, we will show a stronger inequality: for any fixing of $x_1, \ldots, x_{t-1} \in [-1,1]^n$, it holds that

$$\left| f(g_{t-1}^n(x_1, \ldots, x_{t-1})) - \mathop{\mathbb{E}}_{X_t}[f(g_t^n(x_1, \ldots, x_{t-1}, X_t))] \right| \leq \varepsilon.$$

The first inequality then follows by averaging over $x_1 = X_1, \ldots, x_{t-1} = X_{t-1}$. To see why this latter inequality holds, set $y = g_{t-1}^n(x_1, \ldots, x_{t-1})$. Then by definition,

$$g_t^n(x_1, \ldots, x_{t-1}, X_t) = y + \delta_y \circ X_t.$$

The claim now follows from Claim 7.2.3. □

We have so far proved that $g_t^n(X_1, \ldots, X_t)$ is a fractional PRG for $\mathscr{F}$ with slightly worse error. Although we do not need it, it is worth noting that it is symmetric since $X_1, \ldots, X_t$ are symmetric and $-g_t^n(X_1, \ldots, X_t) = g_t^n(-X_1, \ldots, -X_t)$. To conclude, we show that it converges fast to a value close to $\{-1,1\}^n$. This is the effect of *polarization*. It will be enough to analyze this for one-dimensional random variables.

Claim 7.2.5. *Let $A_1,\dots,A_t \in [-1,1]$ be independent symmetric random variables with $\mathbb{E}[A_i^2] \geq p$. For $i = 1,\dots,t$ define*

$$B_i := g_i(A_1,\dots,A_i) = B_{i-1} + (1-|B_{i-1}|)A_i.$$

Then $\mathbb{E}[B_t^2] \geq 1-q$ where $q = 3\exp(-tp/16)$.

Proof. Let $C_i := 1-|B_i|$ be the distance to $\{-1,1\}$ at step i. We show that C_i converges to 0 exponentially fast. Observe that C_i satisfies the following recursive definition:

$$C_i = \begin{cases} C_{i-1}(1-A_i\cdot\mathrm{sign}(B_{i-1})) & \text{if } C_{i-1}(1-A_i\cdot\mathrm{sign}(B_{i-1})) \leq 1 \\ 2-C_{i-1}(1-A_i\cdot\mathrm{sign}(B_{i-1})) & \text{if } C_{i-1}(1-A_i\cdot\mathrm{sign}(B_{i-1})) > 1 \end{cases}.$$

In either case one can verify that $C_i \in [0,1]$ and that

$$C_i \leq C_{i-1}(1-A_i\cdot\mathrm{sign}(B_{i-1})).$$

Now observe that C_{i-1} and $A_i\cdot\mathrm{sign}(B_{i-1})$ are independent. This is because B_{i-1} is symmetric(because A_j's are symmetric), and so $|B_{i-1}|$ and $\mathrm{sign}(B_{i-1})$ are independent. So we can write,

$$\mathbb{E}\left[\sqrt{C_i}\right] \leq \mathbb{E}\left[\sqrt{C_{i-1}}\right]\mathbb{E}\left[\sqrt{1-A_i\cdot\mathrm{sign}(B_{i-1})}\right].$$

The Taylor expansion of $\sqrt{1-x}$ in $[-1,1]$ is

$$\sqrt{1-x} = 1-\frac{x}{2}-\frac{x^2}{8}-\frac{x^3}{16}-\dots$$

In particular, all the coefficients except for the constant term are negative. As $A_i\cdot\mathrm{sign}(B_{i-1})$ is symmetric, $\mathbb{E}[(A_i \cdot \mathrm{sign}(B_{i-1}))^k] = 0$ for any odd k, so

$$\mathbb{E}\left[\sqrt{1 - A_i \cdot \mathrm{sign}(B_{i-1})}\right] \leq 1 - \frac{\mathbb{E}[A_i^2]}{8} \leq 1 - \frac{p}{8} \leq \exp(-p/8).$$

Thus

$$\mathbb{E}\left[\sqrt{C_t}\right] \leq \prod_{i=1}^{t} \mathbb{E}\left[\sqrt{1 - A_i \cdot \mathrm{sign}(B_{i-1})}\right] \leq \exp(-tp/8).$$

Now we use Markov's inequality. We know $\Pr[\sqrt{C_t} \geq \lambda\, \mathbb{E}\left[\sqrt{C_t}\right]] \leq \lambda^{-1}$. By choosing $\lambda = \exp(tp/16)$ we get $\Pr[C_t \geq \exp(-tp/8)] \leq \exp(-tp/16)$. If $C_t \leq \exp(-tp/8)$ then $1 - B_t^2 \leq 2\exp(-tp/8)$. If not, then we can trivially bound $1 - B_t^2 \leq 1$. Putting these together gives

$$\mathbb{E}[1 - B_t^2] \leq 2\exp(-tp/8) + \exp(-tp/16) \leq 3\exp(-tp/16).$$

□

To provide a piece of intuition explaining the fast convergence of this random walk, notice that once C_i becomes sufficiently small, it gets more and more difficult to increase the value of C_i again. This could be best explained with an example. Suppose all A_i's take value in $\{-0.5, 0.5\}$. We start at $B_0 = 0$ and take a step, say $A_1 = 0.5$, and therefore $B_1 = 0.5$. Now observe that the length of the next step would be only $(1 - |B_1|)|A_2| = 0.25$. So even if $A_2 = -0.5$, we get $B_2 = 0.25$, which means we still need to take one more step to become less than 0. In other words, once we get close to the boundary $\{-1, 1\}$, the random walk converges faster as it gets more difficult to move away from the boundary.

Corollary 7.2.6. *Let $X_1, \ldots, X_t \in [-1, 1]^n$ be independent symmetric p-noticeable random variables. Define $Y = g_t^n(X_1, \ldots, X_t)$. Then Y is $(1-q)$-noticeable for $q = 3\exp(-tp/16)$.*

Proof. Apply Claim 7.2.5 to each coordinate of Y. □

Lemma 7.2.1follows by combining Claim 7.2.4 and Corollary 7.2.6.

7.2.2 Proof of Lemma 7.2.2

We prove Lemma 7.2.2 in this section. Let $x \in [-1,1]^n$ be a possible outcome of X. Let $W := W(x) \in \{-1,1\}^n$ be a random variable, where $W_1, \dots, W_n$ are independent and $\mathbb{E}[W_i] = x_i$. Then $\mathbb{E}_W[f(W)] = f(x)$. As f takes values in $[-1,1]$, we can upper bound $|f(x) - f(\text{sign}(x))|$ by

$$|f(x) - f(\text{sign}(x))| = |\mathop{\mathbb{E}}_{W}[f(W)] - f(\text{sign}(x))| \le 2\Pr[W \ne \text{sign}(x)].$$

The last term can be bounded by the union bound,

$$2\Pr[W \ne \text{sign}(x)] \le 2\sum_{i=1}^{n} \Pr[W_i \ne \text{sign}(x_i)] = \sum_{i=1}^{n} 1 - |x_i|.$$

Setting $x = X$ and averaging over X gives

$$|\mathop{\mathbb{E}}_{X}[f(X)] - \mathop{\mathbb{E}}_{X}[f(\text{sign}(X))]| \le \mathop{\mathbb{E}}_{X}|f(X) - f(\text{sign}(X))| \le \sum_{i=1}^{n} \mathbb{E}[1 - |X_i|].$$

As X is $(1-q)$-noticeable it satisfies $\mathbb{E}[X_i^2] \ge 1 - q$ for all i. As $1 - z \le 1 - z^2$ for all $z \in [0,1]$ we have

$$\mathbb{E}[1 - |X_i|] \le \mathbb{E}[1 - X_i^2] \le q.$$

This concludes the proof as

$$|f(\bar{0}) - \mathop{\mathbb{E}}_{X}[f(\text{sign}(X))]| \le |f(\bar{0}) - \mathop{\mathbb{E}}_{X}[f(X)]| + |\mathop{\mathbb{E}}_{X}[f(X)] - \mathop{\mathbb{E}}_{X}[f(\text{sign}(X))]| \le \varepsilon + qn,$$

where the first inequality follows as X is a fractional PRG with error ε, and the second by the discussion above.

7.3 PRGs for functions with bounded Fourier tails

Several natural families of Boolean functions have bounded Fourier tails, such as: AC^0 circuits [LMN93, Man95]; functions with bounded sensitivity [GSW16, LTZ18]; and functions computed by branching programs of various forms [RSV13, CHRT18]. Our goal is to construct a universal PRG which fools any such function. We consider two variants: L_1 bounds and L_2 bounds.

Definition 7.3.1 (L_1 bounds)**.** *For $a,b \geq 1$, we denote by $\mathscr{L}_1^n(a,b)$ the family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ which satisfy*

$$\sum_{\substack{S\subseteq[n]\\|S|=k}} |\widehat{f}(S)| \leq a \cdot b^k \qquad \forall k = 1,\ldots,n.$$

Definition 7.3.2 (L_2 bounds)**.** *For $a,b \geq 1$, we denote by $\mathscr{L}_2^n(a,b)$ the family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ which satisfy*

$$\sum_{\substack{S\subseteq[n]\\|S|\geq k}} \widehat{f}(S)^2 \leq a \cdot 2^{-k/b} \qquad \forall k = 1,\ldots,n.$$

Tal [Tal17] showed that L_2 bounds imply L_1 bounds: if $f \in \mathscr{L}_2(a,b)$ then $f \in \mathscr{L}_1(a,b')$ for $b' = O(b)$. The reverse direction is false, as can be witnessed by the PARITY function. So, the class of functions with L_1 bounded Fourier tails is richer, and we focus on it.

In the following lemma, we construct a fractional PRG for this class, which we will then amplify to a PRG. We note that this lemma holds also for bounded functions, not just Boolean functions. The construction is based on a scaling of almost d-wise independent random variables, whose definition we now recall.

Definition 7.3.3 (Almost d-wise independence)**.** *A random variable $Z \in \{-1,1\}^n$ is ε-almost d-wise independent if, for any restriction of Z to d coordinates, the marginal distribution has statistical distance at most ε from the uniform distribution on $\{-1,1\}^d$.*

Naor and Naor [NN93] gave an explicit construction of an ε-almost d-wise random variable $Z \in \{-1,1\}^n$ with seed length $O(\log\log n + d + \log(1/\varepsilon))$. We note that this seed length is optimal, up to the hidden constants.

Lemma 7.3.4. *Fix $n, a, b \geq 1$ and $\varepsilon > 0$. There exists a fractional PRG $X \in [-1,1]^n$ that fools $\mathscr{L}_1^n(a,b)$ with error ε, such that*

(i) X is p-noticeable for $p = \frac{1}{4b^2}$.

(ii) The seed length of X is $O(\log\log n + \log(a/\varepsilon))$.

Proof. Fix $f \in \mathscr{L}_1^n(a,b)$. Set $d = \lceil \log 2a/\varepsilon \rceil, \delta = \varepsilon/2a, \beta = 1/2b$. Let $Z \in \{-1,1\}^n$ be a δ-almost d-wise independent random variable, and set $X = \beta Z$ which takes values in $\{-\beta, \beta\}^n$. We claim that X satisfies the requirements of the lemma. Claim (i) clearly holds, and claim (ii) holds by the Naor-Naor construction. We thus focus on proving that X fools $\mathscr{F}$ with error ε.

Fix $f \in \mathscr{F}$ and consider its Fourier expansion:

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x^S.$$

We need to show that $\mathbb{E}[f(X)]$ is close to $f(\overline{0})$. Averaging over X gives

$$|\mathbb{E}[f(X)] - f(\overline{0})| \leq \sum_{|S|>0} |\widehat{f}(S)| \cdot |\mathbb{E}[X^S]| = \sum_{|S|>0} |\widehat{f}(S)| \cdot \beta^{|S|} |\mathbb{E}[Z^S]|.$$

We next bound $|\mathbb{E}[Z^S]|$. If $|S| \leq d$ then by the definition of Z we have $|\mathbb{E}[Z^S]| \leq \delta$. If $|S| > d$ we bound trivially $|\mathbb{E}[Z^S]| \leq 1$. Let $W_k = \sum_{S:|S|=k} |\widehat{f}(S)|$, where by assumption $W_k \leq a \cdot b^k$. Thus

$$|\mathbb{E}[f(X)] - f(\overline{0})| \leq \delta \sum_{k=1}^{d} W_k \beta^k + \sum_{k>d} W_k \beta^k \leq \delta a \sum_{k=1}^{d} (\beta b)^k + a \sum_{k>d} (\beta b)^k \leq \delta a + 2^{-d} a$$

where we used the choice of $\beta = 1/2b$. The claim follows as we set $\delta = \varepsilon/2a$ and $2^{-d} \leq \varepsilon/2a$. □

Applying Theorem 7.1.6 using the fractional PRG constructed in Lemma 7.3.4 gives the following PRG construction. Note that we still need to require that $\mathscr{F}$ is closed under restrictions.

Theorem 7.3.5. *Let $\mathscr{F}$ be a family of n-variate Boolean functions closed under restrictions. Assume that $\mathscr{F} \subset \mathscr{L}_1^n(a,b)$ or that $\mathscr{F} \subset \mathscr{L}_2^n(a,b)$. Then, for any $\varepsilon \leq \frac{1}{\mathrm{poly}(b \log n)}$ there exists an explicit PRG $X \in \{-1,1\}^n$ which fools $\mathscr{F}$ with error $\varepsilon > 0$, whose seed length is* $O(\log(n/\varepsilon)(\log\log n + \log(a/\varepsilon))b^2)$.

7.3.1 Open problems

We discuss a couple of open problems in the following.

Early termination

Our analysis requires a random walk with $t = O(\log(n/\varepsilon)/p)$ steps, each coming from a p-noticeable fractional PRG. We believe that for some natural families of functions shorter random walks might also suffice.

Open problem 7.3.6. *Find conditions on classes of Boolean functions so that short random walks can be used to construct PRGs. In particular, are there nontrivial classes where the number of steps is independent of n?*

Less independence

Our analysis of Theorem 7.1.6 currently requires to assume t independent copies of a fractional PRG X. It might be possible that these copies can be chosen in a less independent form, where the analysis still holds.

Open problem 7.3.7. *Can the fractional PRGs $X_1, \dots, X_t$ in Theorem 7.1.6 be chosen not independently, such that the conclusion still holds? Concrete examples to consider are k-wise independence for $k \ll t$, or using an expander random walk.*

Gadgets

We can view the random walk as a “gadget construction". Given independent p-noticeable fractional PRGs $X_1, \ldots, X_t \in [-1,1]^n$, view them as the rows of a $t \times n$ matrix, and then apply a gadget $g : [-1,1]^t \to \{-1,1\}$ to each column to obtain the outcome in $\{-1,1\}^n$. We show that the random walk gives such a gadget which converges for $t = O(\log(n/\varepsilon)/p)$. Many constructions of PRGs can be viewed in this framework, where typically $X_i \in \{-1,1\}^n$. Ours is the first construction which allows X_i to take non-Boolean values. It is interesting whether other gadgets can be used instead of the random walk gadget, and whether there are general properties of gadgets that would suffice.

Acknowledgement Chapter 7 is a reprint of material as it appears in 33rd Computational Complexity Conference (CCC2018). Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. "Pseudorandomgenerators from polarizing random walks." The dissertation author was a primary investigator and author of this paper.

Bibliography

[AGHP92] Noga Alon, Oded Goldreich, Johan Hastad, and Rene Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AHL16] Divesh Aggarwal, Kaave Hosseini, and Shachar Lovett. Affine-malleable extractors, spectrum doubling, and application to privacy amplification. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2913–2917. Ieee, 2016.

[BL17] Pierre-Yves Bienvenu and Thái Hoàng Lê. A bilinear Bogolyubov theorem. *arXiv preprint arXiv:1711.05349*, 2017.

[Blo] Thomas F. Bloom. A quantitative improvement for roth's theorem on arithmetic progressions.

[BLR12] Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 177–186. IEEE, 2012.

[Bou05a] J. Bourgain. Mordell's exponential sum estimate revisited. *J. Amer. Math. Soc.*, 18(2):477–499, 2005.

[Bou05b] J. Bourgain. More on the sum-product phenomenon in prime fields and its applictaions. *International Journal of Number Theory*, 01(01):1–32, 2005.

[BS94] Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[Cha02] Mei-Chu Chang. A polynomial bound in freiman's theorem. *Duke mathematical journal*, 113(3):399–419, 2002.

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. In *33rd Computational Complexity Conference (CCC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. To appear in STOC, 2018.

[CLS13] Ernie Croot, Izabella Laba, and Olof Sisask. Arithmetic progressions in sumsets and L_p-almost-periodicity. *Combinatorics, Probability and Computing*, 22(03):351–365, 2013.

[CS10] Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geometric and functional analysis*, 20(6):1367–1396, 2010.

[EZ12] Chaim Even-Zohar. On sums of generating sets in 2 n. *Combinatorics, probability and computing*, 21(6):916–941, 2012.

[EZL14] Chaim Even-Zohar and Shachar Lovett. The freiman–ruzsa theorem over finite fields. *Journal of Combinatorial Theory, Series A*, 125:333–341, 2014.

[Fre73] Gregory A Freiman. Foundations of a structual theory of set addition. *Translation of Math. Monographs*, 37, 1973.

[Fre87] Gregory A Freiman. What is the structure of k if k+ k is small? In *Number Theory*, pages 109–134. Springer, 1987.

[GM17a] WT Gowers and Luka Milićević. A bilinear version of Bogolyubov's theorem. *arXiv preprint arXiv:1712.00248*, 2017.

[GM17b] WT Gowers and Luka Milićević. A quantitative inverse theorem for the U^4 norm over finite fields. *arXiv preprint arXiv:1712.00241*, 2017.

[Gol10] Oded Goldreich. *A primer on pseudorandom generators*, volume 55. American Mathematical Soc., 2010.

[Gow97] William T Gowers. Lower bounds of tower type for szemerédi's uniformity lemma. *Geometric & Functional Analysis GAFA*, 7(2):322–337, 1997.

[Gow98] William Timothy Gowers. A new proof of szemerédi's theorem for arithmetic progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

[Gow01] William T Gowers. A new proof of szemerédi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1077–1088, 2015.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium*, 2017.

[Gre02] Ben Green. Arithmetic progressions in sumsets. *Geometric & Functional Analysis GAFA*, 12(3):584–597, 2002.

[Gre04] Ben Green. Spectral structure of sets of integers. In *Fourier analysis and convexity*, pages 83–96. Springer, 2004.

[Gre05a] B Green. Notes on the polynomial Freiman-Ruzsa conjecture. *preprint*, 2005. http://people.maths.ox.ac.uk/greenbj/papers/PFR.pdf.

[Gre05b] Ben Green. A szemerédi-type regularity lemma in abelian groups, with applications. *Geometric & Functional Analysis GAFA*, 15(2):340–376, 2005.

[GSW16] Parikshit Gopalan, Rocco A Servedio, and Avi Wigderson. Degree and sensitivity: tails of two distributions. In *Proceedings of the 31st Conference on Computational Complexity*, page 13. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.

[GT08] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3 norm. *Proceedings of the Edinburgh Mathematical Society*, 51(1):73–153, 2008.

[GT09] Ben Green and Terence Tao. Freiman's theorem in finite fields via extremal set theory. *Combinatorics, Probability and Computing*, 18(3):335–355, 2009.

[GT10] Ben Green and Terence Tao. An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 149, pages 1–19. Cambridge University Press, 2010.

[HHL16] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 282–288, 2016.

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for xor functions. *SIAM Journal on Computing*, 47(1):208–217, 2018.

[HL17] Kaave Hosseini and Shachar Lovett. On the structure of the spectrum of small sets. *Journal of Combinatorial Theory, Series A*, 148:1–14, 2017.

[HL19] Kaave Hosseini and Shachar Lovett. A bilinear bogolyubov-ruzsa lemma with polylogarithmic bounds. *Discrete analysis 10.19086/da.8867*, 2019.

[HLMS16] Kaave Hosseini, Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. An improved lower bound for arithmetic regularity. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 161, pages 193–197. Cambridge University Press, 2016.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HLY18] Kaave Hosseini, Shachar Lovett, and Grigory Yaroslavtsev. Optimality of linear sketching under modular updates. In *CCC 2019 (34th Conference on Computational Complexity)*, 2018.

[Kon08] Sergei Vladimirovich Konyagin. On the freiman theorem in finite fields. *Mathematical Notes*, 84(3):435–438, 2008.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

[Lov12] Shachar Lovett. Equivalence of polynomial conjectures in additive combinatorics. *Combinatorica*, 32(5):607–618, 2012.

[Lov14a] Shachar Lovett. Additive combinatorics and its applications in theoretical computer science, 2014. http://cseweb.ucsd.edu/~slovett/files/addcomb-survey.pdf.

[Lov14b] Shachar Lovett. Communication is bounded by root of rank. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 842–846. ACM, 2014.

[LS93] László Lovăsz and Michael Saks. Communication complexity and combinatorial lattice theory. *Journal of Computer and System Sciences*, 47(2):322–349, 1993.

[LTZ18] Shachar Lovett, Avishay Tal, and Jiapeng Zhang. The robust sensitivity of boolean functions. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1822–1833, 2018.

[Man95] Yishay Mansour. An $n^{O(\log\log n)}$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995.

[MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *arXiv preprint arXiv:0909.3392*, 2009.

[MS16] Guy Moshkovitz and Asaf Shapira. A short proof of gowers' lower bound for the regularity lemma. *Combinatorica*, 36(2):187–194, 2016.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational complexity*, 4(4):301–313, 1994.

[NW88] Noam Nisan and Avi Wigderson. Hardness vs. randomness. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 2–11. IEEE, 1988.

[Rao07] Anup Rao. An exposition of bourgainâs 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

[RB11] Noga Ron-Zewi and Eli Ben-Sasson. From affine to two-source extractors via approximate duality. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 177–186. ACM, 2011.

[RM97] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Foundations of Computer Science (FOCS), 1997 IEEE 38th Annual Symposium on*, pages 234–243, 1997.

[RSV13] Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 655–670. Springer, 2013.

[Ruz99] Imre Ruzsa. An analog of freiman's theorem in groups. *Astérisque*, 258(199):323–326, 1999.

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 506–515. ACM, 2007.

[San10] Tom Sanders. Green's sumset problem at density one half. *arXiv preprint arXiv:1003.5649*, 2010.

[San12a] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[San12b] Tom Sanders. The structure theory of set addition revisited. *arXiv preprint arXiv:1212.0458*, 2012.

[Sch11] Tomasz Schoen. Near optimal bounds in freiman's theorem. *Duke Mathematical Journal*, 158(1):1–12, 2011.

[Sch15] Tomasz Schoen. New bounds in balog-szemerédi-gowers theorem. *Combinatorica*, 35(6):695–701, 2015.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[Shk08] I. D. Shkredov. On sets of large trigonometric sums. *Izv. Ross. Akad. Nauk Ser. Mat.*, 72(1):161–182, 2008.

[SS16] Tomasz Schoen and Olof Sisask. Roth's theorem for four variables and additive structures in sums of sparse sets. In *Forum of Mathematics, Sigma*, volume 4. Cambridge University Press, 2016.

[STlV17] Amir Shpilka, Avishay Tal, and Ben lee Volk. On the structure of boolean functions with small spectral norm. *computational complexity*, 26(1):229–273, 2017.

[Sze75] Endre Szemerédi. Regular partitions of graphs. Technical report, STANFORD UNIV CALIF DEPT OF COMPUTER SCIENCE, 1975.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of AC0. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 79. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[TV06] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Foundations of Computer Science (FOCS) , 2013 IEEE 54th Annual Symposium on*, pages 658–667. IEEE, 2013.

[Vad12] Salil P Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[WYY17] Xiaodi Wu, Penghui Yao, and Henry S Yuen. Raz-Mckenzie simulation with the inner product gadget. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:10, 2017.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213. ACM, 1979.

[Yao16] Penghui Yao. Parity decision tree complexity and 4-party communication complexity of XOR-functions are polynomially equivalent. *Chicago Journal of Theoretical Computer Science*, 2016(12), August 2016.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR functions. *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1878–1885, 2014.

[ZS10] Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of boolean functions. *Theoretical Computer Science*, 411(26), 2010.