# Higher-order Fourier analysis with applications to additive combinatorics and theoretical computer science

by

Jonathan Tidor

B.S., Massachusetts Institute of Technology (2017)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author .....
Department of Mathematics
April 20, 2022

Certified by .....
Yufei Zhao
Assistant Professor
Thesis Supervisor

Accepted by .....
Jonathan Kelner
Chairman, Department Committee on Graduate Theses

# Higher-order Fourier analysis with applications to additive combinatorics and theoretical computer science

by

Jonathan Tidor

Submitted to the Department of Mathematics
on April 20, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

Fourier analysis has been used for over one hundred years as a tool to study certain additive patterns. For example, Vinogradov used Fourier-analytic techniques (known in this context as the Hardy-Littlewood circle method) to show that every sufficiently-large odd integer can be written as the sum of three primes, while van der Corput similarly showed that the primes contain infinitely-many three-term arithmetic progressions.

Over the past two decades, a theory of higher-order Fourier analysis has been developed to study additive patterns which are not amenable to classical Fourier-analytic techniques. For example, while three-term arithmetic progressions can be studied with Fourier analysis, all longer arithmetic progressions require higher-order techniques. These techniques have led to a new proof of Szemerédi's theorem in addition to results such as counts of $k$-term arithmetic progressions in the primes.

This thesis contains five results in the field of higher-order Fourier analysis. In the first half, we use these techniques to give applications in additive combinatorics and theoretical computer science. We prove an induced arithmetic removal lemma first in complexity 1 and then for patterns of all complexities. This latter result solves a central problem in property testing known as the classification of testable arithmetic properties. We then study a class of multidimensional patterns and show that many of them satisfy the popular difference property analogously to the one-dimensional case. However there is a surprising spectral condition which we prove necessarily appears in higher dimensions that is not present in the one-dimensional problem.

In the second half of this thesis, we further develop the foundations of higher-order Fourier analysis. We determine the set of higher-order characters necessary over $\mathbb{F}_p^n$, showing that classical polynomials suffice in the inverse theorem for the Gowers $U^k$-norm when $k \leq p+1$, but that non-classical polynomials are necessary whenever $k > p+1$. Finally, we prove the first quantitative bounds on the $U^4$-inverse theorem in the low-characteristic regime $p < 5$.

Thesis Supervisor: Yufei Zhao
Title: Assistant Professor

# Acknowledgments

I thank my advisor Yufei Zhao for showing me how to be a mathematician, for teaching me that clearly and effectively communicating research is just as important as proving new results, and for his guidance, encouragement, and support over these past five years.

This thesis would not have been possible without my coauthors, Aaron Berger, Jacob Fox, Ashwin Sah, Mehtaab Sawhney, and Yufei Zhao. Though their work does not appear in this thesis, I also greatly benefited from my collaborations with David Conlon, Zilin Jiang, Dain Kim, Anqi Li, Yannick Yao, Hans Yu, and Shengtong Zhang.

This research was supported in part by NSF Graduate Research Fellowship Program DGE-1745302. I have also been previously supported by an MIT Presidential Fellowship. Their support allowed me to devote much more of my time and energy to this research.

# Contents

**1 Introduction** **11**

1.1 Gowers norms, patterns, and complexity . . . . . 13

1.2 The inverse theorem . . . . . 15

1.3 Arithmetic regularity . . . . . 16

1.4 Szemerédi-type and popular difference results . . . . . 17

1.5 Arithmetic removal and property testing . . . . . 19

**2 Induced arithmetic removal of complexity 1 patterns over finite fields** **21**

2.1 Definitions and statement of results . . . . . 23

2.1.1 Colored patterns . . . . . 23

2.1.2 Complexity 1 . . . . . 25

2.2 Proof strategy . . . . . 28

2.3 Proof of main theorem . . . . . 32

2.4 Proof of Proposition 2.3.4: a density Ramsey dichotomy . . . . . 39

2.5 Proof of Proposition 2.3.2: arithmetic regularity lemmas . . . . . 43

2.5.1 Proof 1: Strong arithmetic regularity . . . . . 45

2.5.2 Proof 2: An improved bound . . . . . 48

2.6 Extensions . . . . . 54

2.6.1 Infinite removal . . . . . 54

2.6.2 Inhomogeneous patterns . . . . . 56

**3 Popular differences for matrix patterns** **59**

3.1 Summary of results . . . . . 61

3.2 Gowers norms and arithmetic regularity . . . 64
3.3 The $U^3$-arithmetic regularity lemma . . . 66
3.4 Equidistribution and counting lemma . . . 73
3.4.1 Equidistribution results . . . 73
3.4.2 Deriving the linear constraints . . . 77
3.5 Popular differences for 4-point patterns . . . 81
3.6 Counterexample to popular differences for rotated squares in $\mathbb{F}_5^n$ . . . 86
3.6.1 Initial construction . . . 87
3.6.2 Fixing most special directions . . . 89
3.6.3 Finishing the construction . . . 93
3.7 Popular differences for 3-point patterns . . . 94

**4 Testing linear-invariant properties** **101**
4.1 Proof techniques and previous work . . . 104
4.1.1 Graph removal lemmas and property testing . . . 104
4.1.2 Arithmetic analogs . . . 106
4.1.3 Our contributions . . . 108
4.2 Colored patterns and removal lemmas . . . 110
4.3 Preliminaries on higher-order Fourier analysis . . . 113
4.3.1 Gowers norms and complexity . . . 113
4.3.2 Non-classical polynomials and homogeneity . . . 114
4.3.3 Polynomial factors . . . 116
4.3.4 Equidistribution and consistency sets . . . 120
4.3.5 Subatom selection functions . . . 124
4.4 Arithmetic regularity and subatom selection . . . 127
4.4.1 Arithmetic regularity lemmas . . . 127
4.4.2 Subatom selection . . . 136
4.5 Patching . . . 139
4.5.1 A motivating example . . . 139
4.5.2 Colored labeled patterns and canonical colorings . . . 140

4.5.3 Ramsey dichotomy result . . . . . 142
4.5.4 Main patching result . . . . . 153
4.6 Proof of removal lemmas . . . . . 160
4.7 Proof of property testing results . . . . . 172

**5 Non-classical polynomials and the inverse theorem 177**
5.1 Background on non-classical polynomials . . . . . 179
5.2 Classical polynomials for the $U^{p+1}$-inverse theorem . . . . . 181
5.3 Symmetrization techniques . . . . . 182
5.4 Non-classical polynomials are necessary . . . . . 184

**6 Quantitative bounds for the $U^4$-inverse theorem over low characteristic finite fields 195**
6.1 Introduction . . . . . 196
6.1.1 Statement of main result . . . . . 196
6.1.2 Proof strategy . . . . . 197
6.1.3 Low characteristic . . . . . 199
6.2 Preliminaries . . . . . 200
6.3 Non-classical symmetric multilinear forms . . . . . 201
6.4 Symmetrization in low characteristic . . . . . 204
6.5 Proof of main theorem . . . . . 210
6.6 Concluding remarks . . . . . 214

# Chapter 1

# Introduction

It has been known for over one hundred years that Fourier analysis is a useful tool for studying certain *additive patterns*. For example, in the 1930s Vinogradov used Fourier analysis to show that that every sufficiently-large odd integer can be written as the sum of three primes [Vin37] (solutions to $x + y + z = N$) and van der Corput similarly showed that the prime numbers contain infinitely-many three-term arithmetic progressions [vdC39] (solutions to $x - 2y + z = 0$). In addition to these applications in combinatorial number theory, in the 1990s Blum, Luby, and Rubinfeld demonstrated the utility of Fourier analysis in theoretical computer science [BLR93]. They gave a very efficient property testing algorithm, now known as the BLR test, that determines if a function $f\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is close to a globally linear function from local information, namely if $f(x) + f(y) = f(x + y)$ holds for many pairs $x, y$. This local test can be thought of as studying solutions to the equation $x + y = z$.

To explore this connection more formally, we introduce some terminology. For a finite abelian group $G$, we say that a function $f\colon G \to \mathbb{C}$ is *Fourier-pseudorandom* if all Fourier coefficients of $f$ are small in magnitude. We call a set $S \subseteq G$ Fourier-pseudorandom if the normalized indicator function of $S$, the function $f_S(x) = \mathbf{1}(x \in S) - |S|/|G|$, has the same property.

The following is a key property of Fourier-pseudorandomness which is closely tied to why all the aforementioned examples. For a Fourier-pseudorandom set $S \subseteq G$ of density $\alpha$, the density of solutions to $x - 2y + z = 0$ in $S$ is close to $\alpha^3$ (meaning that of all the solutions to $x - 2y + z = 0$ in $G$, the fraction of them that lie in $S$ is close to $\alpha^3$). Thus, with respect to the

equation $x - 2y + z = 0$, a Fourier-pseudorandom set behaves similarly to a randomly-chosen subset. The same property holds for solutions to $x + y + z = N$ and to $x + y = z$.

However, not all additive patterns have this convenient property. As an example, consider four-term arithmetic progressions, $a, a+d, a+2d, a+3d$. Consider the set $S \subseteq \mathbb{F}_5^n$ defined by $S = \{x \in \mathbb{F}_5^n : x_1^2 + x_2^2 + \cdots + x_n^2 \equiv 0 \pmod 5\}$. One can calculate that this set has density $5^{-1} + o(1)$ and is Fourier-pseudorandom. However, if $a$, $a+d$, and $a+2d$ all lie in $S$, this implies that $a+3d$ also lies in $S$. Thus the density of four-term arithmetic progressions in $S$ is the same as the density of three-term arithmetic progressions, $5^{-3} + o(1)$. This is very far from what is expected for a randomly-chosen set with density $5^{-1} + o(1)$ in which the expected density of four-term arithmetic progressions is $5^{-4} + o(1)$. This example shows that Fourier analysis is not the right tool for studying four-term arithmetic progressions (and in fact is not the right tool for studying most additive patterns). To study four-term arithmetic progressions we need a more powerful tool which is capable of detecting the quadratic structure of the set $S$.

The field of higher-order Fourier analysis was developed to solve this deficiency of Fourier analysis. Timothy Gowers defined a sequence of norms, now called the *Gowers uniformity norms* [Gow01]. For each $k \geq 2$, we have a norm called the Gowers $U^k$-norm which, morally speaking, measures structure coming from something resembling a degree $k-1$ polynomial. All additive patterns which satisfy a certain non-degeneracy condition have a *complexity*. Then if a set is pseudorandom with respect to the Gowers $U^{k+1}$-norm, the set has close to the expected number of solutions to to every additive pattern of complexity at most $k$.

Higher-order Fourier analysis consists of a large amount of machinery dedicated to studying the Gowers uniformity norms. This includes the *inverse theorem*, which states that a function which is not $U^{k+1}$-pseudorandom must have global structure. Specifically, it must have some structure resembling a degree $k$ polynomial. Building upon these inverse results are *arithmetic regularity lemmas* which give decompositions of arbitrary functions into structured and pseudorandom pieces. It is an active area of research both to develop further the machinery of higher-order Fourier analysis and to apply it to problems in combinatorial number theory, additive combinatorics, theoretical computer science, and other diverse fields.

This thesis contains five main results. In Chapters 2, 3 and 4 we give three applications

of higher-order Fourier analysis to solve problems in additive combinatorics and theoretical computer science. In Chapters 5 and 6 we further develop the inverse theory of the Gowers uniformity norms, specifically over low-characteristic finite-field vector spaces. This thesis is based on the papers [FTZ22, BSST22, TZ20, BSST21, Tid21].

## 1.1 Gowers norms, patterns, and complexity

For a finite abelian group $G$, a function $f\colon G \to \mathbb{C}$, and $k \geq 2$, the *kth Gowers uniformity norm* is defined by

$$\|f\|_{U^k}^{2^k} = \mathop{\mathbb{E}}_{x,h_1,\ldots,h_k \in G} (\partial_{h_1}\partial_{h_2}\cdots\partial_{h_k} f)(x) \qquad \text{where} \qquad \partial_h f(x) = f(x+h)\overline{f(x)}.$$

Though it is not immediately obvious from the definition, these are well-defined norms which satisfy the additional nesting property $\|f\|_{U^2} \leq \|f\|_{U^3} \leq \|f\|_{U^4} \leq \cdots$ Expanding out the definition, the Gowers $U^k$-norm is an average over $k$-dimensional parallelepipeds, though for the rest of this introduction, the precise definition of the Gowers uniformity norms will not be particularly relevant.

In this thesis we will define an *additive pattern* to be given by a system of homogeneous linear equations (usually with $\mathbb{Z}$-coefficients). We represent such an additive pattern by an $\ell \times k$ matrix $A$ where $\boldsymbol{x} = (x_1, \ldots, x_k) \in G^k$ is an *instance* of $A$ if $A\boldsymbol{x} = \mathbf{0}$. Alternatively, it can be convenient to view such a pattern as a list of $\boldsymbol{L} = (L_1, \ldots, L_k)$ of linear forms in $\ell$ variables. For example, one can view a four-term arithmetic progression either as a solution $(x_1, x_2, x_3, x_4)$ to

$$\begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or as a quadruple $(a, a+d, a+2d, a+3d)$.

For a function $f\colon G \to \mathbb{C}$ and a pattern $A$, we define the *$A$-density* in $f$ to be

$$\Lambda_A(f) = \mathop{\mathbb{E}}_{\substack{\boldsymbol{x}=(x_1,\dots,x_k)\in G^k:\\ A\boldsymbol{x}=\boldsymbol{0}}} f(x_1)f(x_2)\cdots f(x_k).$$

When $f(x) = \mathbf{1}(x \in S)$ then $\Lambda_A(f)$ is the $A$-density in the set $S$. We say that a pattern $A$ has *complexity at most $k$* if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all 1-bounded function $f\colon G \to \mathbb{C}$,

$$|\Lambda_A(f)| \le \epsilon \qquad \text{whenever} \qquad \|f\|_{U^{k+1}} \le \delta.$$

In other words, if a function has small $U^{k+1}$-norm, then that function must have $A$-density close to zero.

We say that a set $S \subseteq G$ is *pseudorandom with respect to the $U^k$-norm* if the normalized indicator function of $S$ has small $U^k$-norm. It then follows shortly from the above definitions that any set which is pseudorandom with respect to the $U^{k+1}$-norm has close to the expected counts of all patterns with complexity at most $k$.

A pattern $A$ is *non-degenerate* if every non-zero vector in the row space of $A$ has at least three non-zero coefficients. Equivalently, we require that the equation $A\boldsymbol{x} = \boldsymbol{0}$ does not force one of the variables to be zero or two of the variables to be multiples of each other. It is known that every non-degenerate pattern has some finite complexity [GT10c, GW10]. Furthermore, there is a simple combinatorial characterization of the minimum complexity of any pattern [GW11a, HHL16] (see also [Man21b] for recent results on the quantitative relationship between $\epsilon$ and $\delta$ in the definition of complexity).

All of this is to say that to study an additive pattern $A$, all we need to understand is the behavior of the $U^{k+1}$-norm where $k$ is the complexity of $A$. If $f$ is pseudorandom with respect to the $U^{k+1}$-norm then we already understand the $A$-density in $f$ as it is close to what is expected for a random function. All that remains is to understand what happens when $f$ is not pseudorandom, i.e., $\|f\|_{U^{k+1}}$ is large.

## 1.2 The inverse theorem

A short calculation shows that $\|f\|_{U^2}$ is large if and only if $f$ has a large Fourier coefficient. Thus the notions of Fourier-pseudorandom and pseudorandom with respect to the $U^2$-norm agree. We wish to generalize this fact to the higher $U^k$-norms.

First note that the aforementioned fact is a local-to-global result. Starting with a local property ($\|f\|_{U^2}$ is large), we deduce a global property ($f$ correlates with a Fourier character, i.e., a linear phase function). The inverse theorem for the Gowers $U^k$-norm is a natural generalization of this fact: $\|f\|_{U^k}$ is large if and only if $f$ correlates with a *higher-order character* which is a sort of degree $k-1$ polynomial phase function [BTZ10, TZ10, TZ12, GTZ11, GTZ12]. As an example, for $G = \mathbb{F}_p^n$, the function $\exp(2\pi i(x_1^2 + x_2^2 + \cdots + x_n^2)/p)$ is a quadratic phase function, one of the higher-order characters for the $U^3$-norm.

Unfortunately much of the rest of the theory of Fourier analysis does not generalize nicely to the higher-order setting. The main issue is that the higher-order characters are not orthogonal (indeed there are way too many of them to form an orthonormal basis). This means that familiar results such as Parseval, Plancherel, and the Fourier inverse formula all fail and there is no useful notion of a Fourier series. In addition, there are further complications simply with the precise statement of the inverse theorem for the Gowers $U^k$-norm.

We restrict to the case of finite field vector spaces for almost all of this thesis. This is for several reasons; first, some, though not all, of the technical difficulties are ameliorated in this case which makes it a useful model problem and second, because this setting is of particular interest for applications in theoretical computer science. Even in the setting of finite-field vector spaces, there are several difficulties even with the statement of the inverse theorem. First, what definition of higher-order character makes the inverse theorem true? In the high-characteristic case of $p \geq k$, it is known that the correct characters are *classical polynomial phase functions*: functions of the form $\exp(2\pi i P/p)$ where $P\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is the usual notion of a polynomial. However even in the case of $p = 2, k = 4$, these functions are not sufficient and objects known as *non-classical polynomial phase functions* necessarily appear. These are functions of the form $\exp(2\pi i Q)$ where $Q\colon \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ is an object known as a *non-classical polynomial*. (We defer the precise definition to Chapter 4.) Prior to this work, it

was known that non-classical polynomial phase functions were a sufficient set of higher-order characters to make the inverse theorem for the Gowers $U^k$-norm over $\mathbb{F}_p^n$ whenever $p < k$ and it was known that non-classical polynomials were necessary in a few sporadic cases (specifically the $U^{p^\ell}$-inverse theorem over $\mathbb{F}_p^n$ for all primes $p$ and $\ell \geq 2$). In Chapter 5 we will give a more satisfactory resolution to this problem by showing that non-classical polynomials are necessary if and only if $p + 1 < k$.

In addition to this complication related to higher-order characters, the quantitative nature of the inverse theorem is poorly-understood. To be more precise, the inverse theorem states that for all $\epsilon > 0$, there exists $\delta > 0$ such that any 1-bounded function $f \colon G \to \mathbb{C}$ satisfying $\|f\|_{U^k} \geq \epsilon$ has correlation at least $\delta$ with some higher-order character. For $k = 2$, a short Fourier-analytic calculation shows that $\delta$ can be taken to be a polynomial function of $\epsilon$. For $k = 3$, the best-known bounds are quasipolynomial, following from Sanders' quasipolynomial Freiman-Ruzsa theorem [San08]. Furthermore, polynomial bounds on the $U^3$-inverse theorem are known to be equivalent to the celebrated polynomial Freiman-Ruzsa conjecture [GT10b, Lov12]. However for $k \geq 4$ the situation deteriorates rapidly. The original bounds, proved by Bergelson, Tao, and Ziegler, rely heavily on techniques from ergodic theory and thus give no quantitative bounds [BTZ10, TZ10, TZ12]. More recent work of Gowers and Milićević gives an iterated exponential bound in the regime $p \geq k$ [GM17, GM20]. In Chapter 6 we build upon this result to give quantitative bounds in the first unknown case: the inverse theorem for the $U^4$-norm over $\mathbb{F}_p^n$ when $p < 5$. It remains an important open problem to extend this result to give quantitative bounds for the inverse theorem for all $p, k$ and to improve these quantitative bounds.

## 1.3 Arithmetic regularity

We know return to the problem of how to use the inverse theorem to study additive patterns. Given the lack of a "higher-order Fourier series" we have to resort to more involved techniques. First we note that the inverse theorem can be viewed as a *structure-versus-pseudorandomness dichotomy*: either $\|f\|_{U^k}$ is small, meaning that $f$ is pseudorandom with respect to the $U^k$-norm, or $\|f\|_{U^k}$ is large, which implies that $f$ is correlated with a higher-order character,

meaning that $f$ has global structure. We can iteratively apply the inverse theorem to deduce a result known as the *arithmetic regularity lemma*. This states that any 1-bounded function $f\colon G \to \mathbb{C}$ can be decomposed as

$$f = f_{\mathsf{str}} + f_{\mathsf{psr}}$$

where $f_{\mathsf{psr}}$ is pseudorandom with respect to the $U^{k+1}$-norm and $f_{\mathsf{str}}$ has certain $k$-th order structure.

To use the correct terminology, this result is known as a *weak arithmetic regularity lemma*, and there are a variety of *strong arithmetic regularity lemmas* which generally give a decomposition

$$f = f_{\mathsf{str}} + f_{\mathsf{sml}} + f_{\mathsf{psr}}$$

where $f_{\mathsf{str}}$ is structured, $f_{\mathsf{sml}}$ is small with respect to the $L^2$-norm, and $f_{\mathsf{psr}}$ is extremely pseudorandom with respect to the $U^{k+1}$-norm where the amount of pseudorandomness can be made arbitrarily small with respect to the "complexity" of $f_{\mathsf{str}}$. We will delay further details of these regularity lemmas until later. (We give a strong $U^2$-arithmetic regularity lemma in Section 2.5, a strong $U^3$-arithmetic regularity lemma in Section 3.3, and several of $U^k$-arithmetic regularity lemmas of various strengths in Section 4.4.) For now it is enough to know that strong arithmetic regularity lemmas give quite detailed structural information about arbitrary functions which can be quite useful in applications.

## 1.4 Szemerédi-type and popular difference results

So far we have laid the foundation of the machinery of higher-order Fourier analysis that we use to study additive patterns. We now turn to describing some of the applications which motivated the development of this theory.

Szemerédi's theorem is a foundational result in additive combinatorics, stating that any subset of $\mathbb{Z}$ with positive upper density contains arbitrarily-long arithmetic progressions. There are several conceptually-distinct proofs of Szemerédi's theorem. In addition to Szemerédi's original proof, there are proofs using ergodic theory, hypergraph regularity, and higher-order Fourier analysis.

These approaches each have their own benefits and drawbacks. The ergodic theoretic approach is the most general, giving Szemerédi-type results for multidimensional and polynomial patterns, but gives absolutely no quantitative bounds. The hypergraph regularity approach works for multidimensional patterns but not polynomial patterns and gives effective but quite bad (Ackermann-type) bounds. Finally, the higher-order Fourier analytic method has many difficulties applying to either multidimensional or polynomial patterns, but it gives the best-known quantitative bounds on Szemerédi's theorem as well as providing the most detailed structural information.

One of the triumphs of higher-order Fourier analysis is counting solutions to systems of linear equations in the prime numbers [GT10c]. For any additive pattern $A$ of finite complexity, this result gives the density of $A$ in the prime numbers. In a similar direction, one can ask to count pattern densities in arbitrary sets. As an example, for an arbitrary subset $S \subseteq G$ of density $\alpha$, in addition to the Szemerédi-type result that $S$ contains a three-term arithmetic progression, one might ask what is the density of three-term progressions. Unfortunately, it is not true that all such $S$ contain at least $\alpha^3|G|^2$ three-term progressions, the number that would be expected in a random set of density $\alpha$. There are examples due to Behrend that give such sets containing only $\alpha^{c\log(1/\alpha)}|G|^2$ progressions [Beh46]. However, it is known that one can always find a *popular difference* $d \in G$ with $d \neq 0$ such that $S$ always contains at least $(\alpha^3 - o(1))|G|$ three-term progressions of the form $(a, a+d, a+2d)$ [Gre05b].

Green's Fourier-analytic techniques show that all three-point patterns satisfy the *popular difference property*, while further work using quadratic Fourier analysis shows that certain "parallelogram" four-point patterns, including four-term arithmetic progressions, satisfy the popular difference property [Gre07, GT10a]. However, it is somewhat surprising to know that most patterns do not satisfy the popular difference property. In fact, the patterns listed above are the only one-dimensional patterns which satisfy the popular difference property [BHK05, SSZ21].

It is an interesting and active area of research to attempt to use higher-order Fourier analysis to study multidimensional patterns. If successful, this approach would give the first-known reasonable quantitative bounds for such problems. However, there are a number of very fundamental problems that are open in this direction. Indeed it is not even known how to prove

Szemerédi-type results for the square, the pattern $\{(x,y),(x+d,y),(x,y+d),(x+d,y+d)\}$ in a set $S \subseteq G \times G$, using higher-order Fourier analysis. In this thesis we probe the limits of what can be proved about multidimensional patterns using higher-order Fourier analysis. We study multidimensional patterns which satisfy a *full-rank* constraint. This rules out the square, but leaves a large class of multidimensional patterns. For full-rank three- and four-point patterns, it is known how to prove Szemerédi-type results using quadratic Fourier analysis [Pre15]. In Chapter 3 we develop new techniques in quadratic Fourier analysis to study popular difference results for these full-rank multidimensional patterns. We prove that, analogously to the one-dimensional case, that three-point and most "parallelogram" four-point patterns satisfy the popular difference property. However we discover that a new spectral condition appears in higher dimensions. We construct an example that shows, at least in one case, that this spectral condition is necessary.

## 1.5 Arithmetic removal and property testing

As an application in another direction, higher-order Fourier analysis gives more detailed structural information which allows us to prove *arithmetic removal lemmas*. One such result, due to Green, states that for all $\epsilon > 0$ there exists $\delta > 0$ such that any set $S \subseteq G$ with fewer than $\delta|G|^2$ solutions to $x + y = z$ can be made to have no solutions by removing at most $\epsilon|G|$ elements from $S$ [Gre05b]. This result follows from a $U^2$-regularity lemma. Generalizations of this result to other patterns follow easily from an appropriate strong arithmetic regularity lemma. (It was also later noticed that these results can be deduced directly from a removal lemma for directed hypergraphs [KSV12, Sha10].)

Arithmetic removal lemmas are natural generalizations of well-studied graph removal lemmas. Furthermore, in computer science there is much interest in removal lemmas due to their connection to property testing. For example, one can use Green's arithmetic removal lemma to give an algorithm for distinguishing between sets $S \subseteq G$ which have no solutions to $x + y = z$ and those which are far from having no solutions to $x + y = z$ with only a constant number of membership queries to $S$.

From both of these perspectives it is natural to consider results known as *induced*

*arithmetic removal lemmas* which are generalizations of the induced graph removal lemma to the arithmetic setting. These types of result do not appear to follow directly from the $U^2$-arithmetic regularity lemma. (It also does not appear possible to deduce this result directly from the induced graph or hypergraph removal lemma.) In Chapter 2 we state and prove an induced arithmetic removal lemma for complexity 1 patterns over $\mathbb{F}_p^n$. This result is proved via a combination of the $U^2$-arithmetic regularity lemma and a novel combinatorial technique called *patching* which is inspired by Ramsey theory. As a consequence of this result, we immediately deduce that a class of complexity 1 properties are testable with constant query complexity.

After we have warmed up with some $U^3$-theory in Chapter 3, we will return to the induced arithmetic removal lemma in Chapter 4, now for patterns of all complexities. We solve the full problem, using a new $U^k$-arithmetic regularity lemma as well as a stronger version of the Ramsey-theoretic patching technique from Chapter 2 that is suitable for the higher-order setting. This full induced arithmetic removal lemma implies that all the so-called "semi-hereditary" arithmetic properties are testable with constant query complexity. It was previously shown that no other arithmetic property is testable, so the results of Chapter 4 completely classify the arithmetic properties that are testable with constant query complexity.

# Chapter 2

# Induced arithmetic removal of complexity 1 patterns over finite fields

Removal lemmas play a central role in extremal and additive combinatorics (see the survey [CF13]). The *triangle removal lemma*, first proved by Ruzsa and Szemerédi [RS78], states that for every $\epsilon > 0$ there exists $\delta > 0$ such that every $n$-vertex graph with fewer than $\delta n^3$ triangles can be made triangle-free by deleting at most $\epsilon n^2$ edges.

In computer science, there is much interest in removal lemmas due to the connection between removal lemmas and constant query complexity algorithms for property testing [RS96, GGR98]. For example, the triangle removal lemma implies that to distinguish a triangle-free graph from one that is $\epsilon$-far from triangle free, one simply needs to sample a constant number of triples of vertices (this constant depends on $\epsilon$, but not the size of the graph) and check them for triangles. All our discussions and results below can be interpreted using the language of property testing.

An arithmetic analog of the triangle removal lemma was first established by Green [Gre05b]. It states that for every $\epsilon > 0$ there exists $\delta > 0$ such that for every finite abelian group $(G, +)$ and every $A \subset G$ with at most $\delta|G|^2$ triples $(x, y, z) \in A^3$ satisfying $x + y = z$, there exists $A' \subseteq A$ with $|A \setminus A'| \leq \epsilon|G|$ such that $A'$ contains no solution to $x + y = z$.

Green's arithmetic triangle removal lemma also has a corresponding property testing algorithm. This algorithm distinguishes between *triangle-free* maps $f\colon G \to \{0, 1\}$ (i.e., those with no $x, y \in G$ satisfying $f(x) = f(y) = f(x + y) = 1$) and maps that are $\epsilon$-far from

triangle free by sampling a constant number of pairs $(x, y) \in G^2$ and checking if any satisfy $f(x) = f(y) = f(x + y) = 1$.

Generalizing the triangle removal lemma, there is a *graph removal lemma* which states that for every fixed graph $H$ (or a finite collection of graphs), every $n$-vertex graph with $H$-density $o(1)$ can be made $H$-free by removing $o(n^2)$ edges [ADL+94, Für95]. Furthermore, there is an *induced graph removal lemma* [AFKS00] which states that every $n$-vertex graph with induced $H$-density $o(1)$ can be made induced $H$-free by adding or removing $o(n^2)$ edges. Both of these removal lemmas give constant query complexity algorithms for property testing of the corresponding properties.

The graph removal lemma was proved using Szemerédi's graph regularity lemma, an important tool in extremal combinatorics. The arithmetic removal lemma was initially proved by Green via an arithmetic variant of the graph regularity lemma; an alternate proof was later given by Král'–Serra–Vena [KSV09], deducing the arithmetic removal lemma from the graph removal lemma. This technique was extended to deduce a removal lemma for general linear systems from the hypergraph removal lemma [KSV12, Sha10].

A natural common extension of the induced graph (or hypergraph) removal lemma and the arithmetic removal lemma would be an induced arithmetic removal lemma, which is not currently known (e.g., [Sha10, Conjecture 5.3]). In particular, the technique for deducing arithmetic removal from graph removal (from [KSV09]) does not appear to extend to the induced setting.

The goal of this chapter is to establish an induced arithmetic removal lemma for arithmetic patterns of complexity 1 over finite fields. (See Section 2.1.2 for the definition of complexity.) Previous results [BGS15] established this claim over $\mathbb{F}_2$, though the proof requires specific Ramsey-theoretic results that fail for all other finite fields. We overcome this difficulty by introducing a novel recoloring technique that works over all base fields.

It remains an open problem to extend the induced arithmetic removal lemma to more general settings, both to higher complexity patterns and to general groups (even $\mathbb{Z}$ is open). For affine-invariant patterns (which are always translation-invariant), a special case of the general problem that avoids an important difficulty (no regularization near the origin; see discussion in Section 2.2), the problem for bounded complexity patterns over finite fields has

been solved [BFH+13].

We begin with a discussion in Section 2.1 motivating the precise setup of our main theorem. In Section 2.2 we review the background on the regularity proof of the graph removal lemma and its relationship to our proof, and we describe our proof strategy. In Section 2.3 we prove our main theorem, modulo two key tools, proved in Sections 2.4 and 2.5. In Section 2.6 we give two extensions of our main theorem, to an infinite set of patterns and to inhomogeneous patterns.

This chapter is based on the paper [FTZ22].

## 2.1 Definitions and statement of results

### 2.1.1 Colored patterns

Induced subgraphs can be equivalently rephrased as colored patterns in a 2-edge-colored clique, where the two colors correspond to edges and non-edges. We consider the natural generalization allowing an arbitrary fixed number of edge colors.

The proof of the induced graph removal lemma gives the following result about colored patterns. Given $r$-edge-colored cliques $G$ and $H$, the *$H$-density* in $G$ is defined to be the fraction of $|H|$-vertex subgraphs of $G$ that have edge-coloring is isomorphic to $H$. We say that $G$ is *$H$-free* if no such subgraph exists.

**Theorem 2.1.1** (Induced graph removal)**.** *Fix a finite collection $\mathcal{H}$ of $r$-edge-colored cliques. For every $\epsilon > 0$ there exists $\delta > 0$ so that if an $r$-edge-coloring of $K_n$ has $H$-density at most $\delta$ for every $H \in \mathcal{H}$, then one can recolor at most $\epsilon n^2$ edges of $K_n$ (using the original $r$ colors) so that the resulting coloring is $H$-free for every $H \in \mathcal{H}$.*

An analogous statement for hypergraphs is also known to hold [RS07]. In addition, there is even a result allowing an infinite collection of subgraphs $\mathcal{H}$ (with a slightly modified statement); see Section 2.6.1.

We would like to obtain a similar statement for colored arithmetic patterns. However, there must be some caveats, as illustrated by the following two non-examples.

**Non-example 2.1.2.** Consider the 2-coloring of $\mathbb{Z}/N\mathbb{Z}$ with $1, \dots, \lfloor N/2 \rfloor$ colored red and the remaining elements colored blue, and the pattern $x = y + 1$ where $x$ is blue and $y$ is red. There is only one instance of this pattern in the given coloring, but it is impossible to recolor fewer than $\lfloor N/2 \rfloor$ elements to remove this pattern. For an example with a homogeneous equation, a similar situation occurs for $x = 2y$ as long as 2 has large multiplicative order in $(\mathbb{Z}/N\mathbb{Z})^\times$.

The above non-example does not worry us too much since it has infinite complexity and is not over a fixed finite field. Nevertheless, it reminds us that we need to be cautious when formulating a more general conjecture about induced arithmetic removal.

The second non-example is more relevant to our discussion as it is a complexity 1 pattern over a fixed finite field.

**Non-example 2.1.3.** Consider the 4-coloring $\phi\colon \mathbb{F}_5^n \to [4]$ defined as follows: $\phi((0, \dots, 0)) = 1$ and $\phi((0, \dots, 0, x_k, \dots, x_n)) = x_k$ where $x_k \neq 0$ is the first nonzero coordinate. We consider monochromatic solutions to $x + y + z = 0$. The only such solution is $x = y = z = 0$, but it always remains a monochromatic solution no matter how $\phi$ is modified. Thus there is no removal lemma here either.

This non-example shows that we need to treat zero with care. While not all patterns exhibit the above behavior, the simplest way to remove this issue in general seems to be to ignore zero entirely. Thus the statement of our main theorem only removes solutions where none of the variables are equal to zero. An alternative and weaker possible formulation, which is implied by our main theorem, is to only remove *generic* solutions, i.e., solutions whose span has maximum possible dimension.

As an example of our main result, Theorem 2.1.15, applied to the above system, we prove:

> If $\phi\colon \mathbb{F}_5^n \setminus \{0\} \to [4]$ has $o(5^{2n})$ monochromatic solutions to $x + y + z = 0$, then one can remove all such solutions by recoloring an $o(1)$-fraction of the space.

**Definition 2.1.4.** An *$r$-colored pattern* $H = (A, \psi)$ is represented by an $\ell \times k$ matrix $A$ with entries in $\mathbb{F}_q$ along with a coloring $\psi\colon [k] \to [r]$ of the columns of $A$. Let $V$ be a finite dimensional $\mathbb{F}_q$-vector space. An *$H$-instance* in a coloring $\phi\colon V \to [r]$ is some

$\boldsymbol{x} = (x_1, \ldots, x_k) \in V^k$ satisfying $A\boldsymbol{x} = \mathbf{0}$ and $\phi(x_i) = \psi(i)$ for all $i \in [k]$. The *$H$-density* is defined to the number of $H$-instances divided by $|V|^{k-\operatorname{rank} A}$, i.e., the probability that a uniform random $\boldsymbol{x} \in V^k$ satisfying $A\boldsymbol{x} = \mathbf{0}$ also satisfies $\phi(x_i) = \psi(i)$ for all $i \in [k]$. We say that $\phi$ is *$H$-free* if there are no $H$-instances $\boldsymbol{x} \in (V \setminus \{0\})^k$.

We generally need to consider a collection of colored patterns.

**Example 2.1.5.** Monochromatic solutions to $x + y + z = 0$ can be encoded by a collection of $r$ separate $r$-colored patterns, $\{(A, \psi_i)\}_{i \in [r]}$, all having the same $A = (1\ 1\ 1)$ where $\psi_i \equiv i$ is the constant function.

**Example 2.1.6.** A rainbow solution to $x + 2y + 3z = 0$ is one where $x$, $y$, and $z$ have distinct colors, and it can be encoded by a collection of $r(r-1)(r-2)$ separate $H$'s, all having the same $A = (1\ 2\ 3)$, and $\psi$ ranging over all injective maps $[3] \to [r]$.

**Example 2.1.7.** The property of having either a blue 3-term arithmetic progression or a red 4-term arithmetic progression can be encoded by two patterns:

$$\left(A = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}, \psi \equiv \text{blue}\right) \quad \text{and} \quad \left(A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}, \psi \equiv \text{red}\right).$$

Note that the second pattern has complexity 2.

**Example 2.1.8.** Suppose that $p$ is prime. We consider maps $\phi\colon \mathbb{F}_p^n \setminus \{0\} \to \mathbb{F}_p$ as $p$-colorings of $\mathbb{F}_p^n$. Consider the set of patterns

$$\mathcal{H} = \{(A, \psi_{a,b,c})\}_{\substack{a,b,c \in \mathbb{F}_p \\ a+b \neq c}}$$

where $A = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$ and $\psi_{a,b,c}$ is defined by $\psi_{a,b,c}(1) = a, \psi_{a,b,c}(2) = b, \psi_{a,b,c}(3) = c$. Then $\phi\colon \mathbb{F}_p^n \setminus \{0\} \to \mathbb{F}_p$ is linear if and only if it is $\mathcal{H}$-free.

### 2.1.2 Complexity 1

Fourier analysis allows us to analyze certain patterns such as $x + y = z$ or 3-term arithmetic progressions, but not others such as 4-term arithmetic progressions. This distinction is

captured by a notion known as the *complexity* of a linear system, introduced by Green and Tao in their work on linear patterns in the primes [GT10c]. Informally, a system has *complexity* $s$ it is "controlled" by the $U^{s+1}$ Gowers uniformity norm (the *true complexity* [GW11a] of the system is the smallest such $s$). We will defer the general definition to Chapter 4 since in this chapter we only address complexity 1 systems.

We consider weighted, or functional, versions of densities of linear patterns. For a finite dimensional $\mathbb{F}_q$-vector space $V$, an $\ell \times k$ matrix $A$ over $\mathbb{F}_q$, and functions $f_1, \dots, f_k \colon V \to \mathbb{C}$, we write

$$\Lambda_A(f_1, \dots, f_k) := \mathbb{E}_{\boldsymbol{x}=(x_1,\dots,x_k)\in V^k : A\boldsymbol{x}=\boldsymbol{0}}[f_1(x_1)\cdots f_k(x_k)].$$

Then the $(A,\psi)$-density in $\phi \colon V \to [r]$ is given by $\Lambda_A(1_{\phi^{-1}(\psi(1))}, \dots, 1_{\phi^{-1}(\psi(k))})$.

**Example 2.1.9.** For the matrix $A = (1 \ -2 \ 1)$ corresponding to the pattern $x - 2y + z = 0$, one has

$$\Lambda_A(f_1, f_2, f_3) = \mathbb{E}_{x,y\in V}[f_1(x)f_2(y)f_3(2y - x)].$$

The above expression illustrates the equivalence between expressing a linear pattern as a linear system of equations (e.g., $x - 2y + z = 0$) and as a collection of linear forms of unconstrained variables (e.g., $(x, y, 2y - x)$). The complexity of a system is often defined in the literature in the latter formulation. It is sometimes useful to keep both versions in mind.

The Fourier transform of a function $f \colon V \to \mathbb{C}$ is $\widehat{f} \colon V^* \to \mathbb{C}$ given by

$$\widehat{f}(\gamma) := \langle f, \gamma\rangle = \mathop{\mathbb{E}}_{x\in V} f(x)\overline{\gamma(x)}$$

where $\gamma \in V^*$ is a character, i.e., $\gamma \colon V \to \mathbb{C}^\times$ satisfying $\gamma(x + y) = \gamma(x)\gamma(y)$ for all $x, y \in V$. For a finite dimensional $\mathbb{F}_q$-vector space $V$ of characteristic $p$, one has $V^* \cong V$ by associating $z \in V$ with the character $\gamma_z$ given by $\gamma_z(x) := \exp(2\pi i \operatorname{tr}(x\cdot z)/p)$ where $x\cdot z := x_1z_1+\cdots+x_nz_n$ is the dot product and $\operatorname{tr} \colon \mathbb{F}_q \to \mathbb{F}_p$ is the trace map. It is fine to think only about the slightly simpler case of prime $q = p$ so that we can write $\widehat{f}(\gamma_z) = \mathbb{E}_{x\in V} f(x)e^{-2\pi i(x\cdot z)/p}$ for $z \in V$.

**Definition 2.1.10.** We say that an $\ell \times k$ matrix $A$ over $\mathbb{F}_q$ has *complexity 1* if for every $\epsilon > 0$ there exists $\delta > 0$ so that for every finite dimensional $\mathbb{F}_q$-vector space $V$ and $f_1, \dots, f_k \colon V \to$

$[-1,1]$, one has

$$|\Lambda_A(f_1,\dots,f_k)| \le \epsilon \quad \text{whenever} \quad \min_{1\le i\le k} \|\widehat{f_i}\|_\infty \le \delta.$$

*Remark* 2.1.11. It is known that a system given by linear forms $L_1,\dots,L_k$ (i.e., generating the null space of $A$) has complexity 1 if and only if the quadratic forms $L_1^2,\dots,L_k^2$ are linearly independent (a technical modification is required if $q$ is even) [GW11b, GW11a, HHL16]. The general dependence of $\delta$ on $\epsilon$ is a wide open problem, though conjecturally $\delta$ can be taken to be of the form $\epsilon^{O_A(1)}$. See [Man18] for some discussion.

**Example 2.1.12.** Any single equation $\sum a_i x_i = 0$ with at least three non-zero coefficients is a complexity 1 system. In matrix form, this says that any matrix $A$ of dimensions $1\times k$ has complexity 1 as long as it has at least three non-zero entries.

**Example 2.1.13.** Let $G=(V,E)$ be a directed graph. Suppose $A$ is a matrix with $|E|$ columns satisfying the property that $\boldsymbol{x}=(x_e)_{e\in E}$ satisfies $A\boldsymbol{x}=\mathbf{0}$ if and only if the signed sum of $x_e$ around each cycle of $G$ is 0. Such an $A$ is complexity 1. This example can be generalized by scaling each column of $A$ by a non-zero constant. In this form Example 2.1.12 is a special case corresponding to the cycle.

**Example 2.1.14.** There are complexity 1 systems that do not arise from the above constructions. One example is given by

$$A = \begin{pmatrix} 2 & 1 & 1 & -1 & 0 & 0 \\ 1 & 2 & 1 & 0 & -1 & 0 \\ 1 & 1 & 2 & 0 & 0 & -1 \end{pmatrix} \quad \text{over } \mathbb{F}_7.$$

For those familiar with the terminology, this system has Cauchy-Schwarz complexity 2, but true complexity 1. All systems arising from the two previous examples have true complexity and Cauchy-Schwarz complexity 1.

Now we are ready to state our main result.

**Theorem 2.1.15** (Induced arithmetic removal)**.** *Fix a finite set $\mathcal{H}$ of $r$-colored complexity 1 patterns over $\mathbb{F}_q$. For every $\epsilon>0$ there exists $\delta=\delta(\epsilon,\mathcal{H})>0$ such that the following holds.*

*Given a finite dimensional $\mathbb{F}_q$-vector space $V$, if a coloring $\phi\colon V \to [r]$ has $H$-density at most $\delta$ for every $H \in \mathcal{H}$, then there exists a recoloring $\phi'\colon V \setminus \{0\} \to [r]$ that is $\mathcal{H}$-free and differs from $\phi$ on at most $\epsilon|V|$ elements of $V$.*

*Remark* 2.1.16. Our proof gives a bound on $1/\delta$ that is tower-type in $1/\epsilon$ provided that the conjecture in Remark 2.1.11 holds for the patterns under consideration. For example, this is the case if the patterns in $\mathcal{H}$ all have Cauchy-Schwarz complexity 1. Although we write $\delta = \delta(\epsilon, \mathcal{H})$ for simplicity, the dependence on $\mathcal{H}$ is not strong. We could instead write $\delta = \delta(\epsilon, q, r, k)$ where $q$ is the size of the finite field, $r$ is the number of colors used, and $k$ is the maximum number of columns in any of the patterns in $\mathcal{H}$.

For those familiar with the terminology, our main theorem states the following in the language of property testing: every linear-invariant, linear-subspace hereditary property of complexity 1 is testable with constant query complexity and one-sided error. (To be precise one needs the generalization to an infinite number of patterns; see Theorem 2.6.2.)

We conjecture that a similar statement holds for arbitrary patterns (with coefficients in $\mathbb{Z}$) in arbitrary abelian groups. However there must be some caveats related to infinite complexity patterns and trivial solutions. As such, we do not state a precise conjecture here.

In Chapter 4 we generalize the main result to patterns of arbitrary complexity (still over finite field vector spaces).

## 2.2 Proof strategy

We begin by recalling the basic strategy for proving the graph removal lemma [RS78] (see also the survey [CF13]). We are given an $n$-vertex graph $G$ with $H$-density $o(1)$ and we wish to make the graph $H$-free by removing $o(n^2)$ edges.

**Strategy for graph removal:**

1. **Regularize.** Apply Szemerédi's regularity lemma to $G$, partitioning the vertex set into a small number of roughly equal-sized parts, $V(G) = V_1 \cup \cdots \cup V_k$, in such a way that $G$ is regular (i.e., quasirandom) between almost all pairs of parts.

2. **Clean up.** Modify $G$ by deleting all edges that lie between a pair of parts $(V_i, V_j)$ that either is irregular or has low edge-density.

The resulting graph, after the edge deletions, should be $H$-free since if there were an $H$-subgraph it must sit among high-density regular pairs. By a counting lemma this would imply the existence of many copies of $H$ in the original graph $G$. This would contradict the assumption that $G$ has $H$-density $o(1)$.

For the induced graph removal lemma, the above strategy needs to be further extended [AFKS00], since removing all the edges between an irregular pair $(V_i, V_j)$ may actually create induced $H$-subgraphs where there were none. The new tool here is to use a strong regularity lemma to find a "regular model" for $G$ in which every pair of parts is regular.

**Strategy for induced graph removal:**

1. **Find a regular model.** Apply a strong regularity lemma to $G$ to produce a partition of the vertex set into a small number of roughly equal-sized parts, $V(G) = V_1 \cup \cdots \cup V_k$, and subsets $U_i \subseteq V_i$ each a positive fraction. This is done in such a way that $U := U_1 \cup \cdots \cup U_k$ forms a "regular model" for $G$ in the sense that $(U_i, U_j)$ is a regular pair for *every* $i, j$ and that the edge densities between pairs of $V_i$'s and the corresponding pairs of $U_i$'s are close on average.

2. **Clean up.** Modify $G$ by deleting all edges that lie between a pair $(V_i, V_j)$ if the edge-density between $(U_i, U_j)$ is close to 0 and by adding all edges that lie between a pair $(V_i, V_j)$ if the edge-density between $(U_i, U_j)$ is close to 1.

The resulting graph, after the edge modifications, should be induced $H$-free for the following reason. Suppose that there were an induced $H$-subgraph sitting among some of the $V_i$'s. Then between the corresponding $U_i$'s, all the pairs are regular and all must have either high edge-density or high non-edge-density as required. By an appropriate counting lemma, this would produce many more induced copies of $H$ in $G$, contradicting the assumption that $G$ has induced $H$-density $o(1)$.

The arithmetic removal lemma is proved analogously to the graph removal lemma [Gre05b]. Here a subset $S \subseteq V$ of a finite field vector space $V$ plays the role of a graph and a regular

partition of $S \subseteq V$ is given by a subspace $V_1 \leq V$ which has bounded codimension. This induces the partition of $V$ into cosets of $V_1$. For this partition to be regular we require $x + V_1$ to be regular (i.e., $S \cap (x + V_1)$ sits inside $x + V_1$ quasirandomly) for almost all $x \in V$.

Suppose $H$ is a complexity 1 system of linear forms over $\mathbb{F}_q$ that has $o(1)$ density in $S \subseteq V$.

**Strategy for arithmetic removal:**

1. **Regularize.** Apply Green's arithmetic regularity lemma to $S \subseteq V$, producing a regular partition determined by a subspace $V_1 \leq V$ of bounded codimension.

2. **Clean up.** Modify $S$ by deleting all elements that lie in a coset $x + V_1$ that either is irregular or has low density of $S$.

The resulting set, after the deletions, should be $H$-free since if there were an $H$-instance it must sit in high-density regular cosets. By a counting lemma, this would imply the existence of many copies of $H$ in the original set $S$. This would contradict the assumption that $S$ has $H$-density $o(1)$.

A natural attempt to prove the induced arithmetic removal lemma would be to invoke a strong arithmetic regularity lemma to find a regular model. In the arithmetic setting a regular model is a subspace $W$ such that $W \cap (x + V_1)$ is regular for each $x$.

Suppose $H$ is a 2-colored pattern of complexity 1 that has $o(1)$ density in $S \subseteq V$. (Here we view $S$ and $V \setminus S$ as the two color classes.)

**(Incorrect) strategy for induced arithmetic removal:**

1. **Find a regular model.** Apply a strong arithmetic regularity lemma to $S \subseteq V$ to produce a partition of $V$ into cosets of $V_1 \leq V$ as well as a bounded codimension subspace $W \leq V$ that forms a "regular model" for $V$ in the following sense. We require that $W \cap (x + V_1)$ is a regular coset for *every* $x$ and the densities of $S$ in the $x + V_1$'s and the corresponding $W \cap (x + V_1)$'s are close on average.

2. **Clean up.** Modify $S$ by deleting all elements that lie in a coset $x + V_1$ if the density of $S$ in $W \cap (x + V_1)$ is close to 0 and by adding all elements that lie in $x + V_1$ if the density of $S$ in $W \cap (x + V_1)$ is close 1.

If it were possible to carry out this strategy, the resulting set would be induced $H$-free. However, an important difficulty arises that was not present in the proof of the induced graph removal lemma. It turns out that it is impossible to perform the "find a regular model" step as the subspace $W \cap V_1$ cannot always be made regular. Green and Sanders [GS15] give an example with the following properties.

> For all $n$, there exists a subset $S \subset \mathbb{F}_3^n$ such that for every positive-dimension subspace $V \leq \mathbb{F}_3^n$ the set $S \cap V$ is *not* an $\epsilon$-regular subset of $V$ for any $\epsilon < \sqrt{3}/6$.

Our strategy for circumventing this difficulty is two-fold. First we get as much out of arithmetic regularity as possible: we say that $W$ is a regular model if $W \cap (x + V_1)$ is regular for every $x \in V \setminus V_1$. Next we develop a new tool to deal with $V_1$. Our new tool is a Ramsey-type dichotomy. It will be convenient in the full proof to have a statement that works with a finite set $\mathcal{H}$ of 2-colored patterns instead of a single pattern. The Ramsey-type dichotomy roughly says the following:

> For every finite set $\mathcal{H}$ of 2-colored patterns over $\mathbb{F}_q$, there exist constants $n_0$ and $\epsilon_0 > 0$ such that either:
>
> (a) for every finite dimensional $\mathbb{F}_q$-vector space $V$ satisfying $\dim V \geq n_0$, every $S \subseteq V$ has $H$-density at least $\epsilon_0$ for some $H \in \mathcal{H}$; or
>
> (b) for every finite dimensional $\mathbb{F}_q$-vector space $V$ there exists $S \subseteq V$ that is $\mathcal{H}$-free (meaning that $S$ is $H$-free for all $H \in \mathcal{H}$).

In the above dichotomy we view $S$ and $V \setminus S$ as the two color classes.

Suppose $H$ is a 2-colored pattern of complexity 1 that has $o(1)$ density in $S \subseteq V$.

**(Revised) strategy for induced arithmetic removal:**

1. **Find a regular model.** Apply a strong arithmetic regularity lemma to $S \subseteq V$ to produce a partition of $V$ into cosets of $V_1 \leq V$ and a bounded codimension subspace $W \leq V$ that forms a "regular model" for $V$ in the following sense. We require that $W \cap (x + V_1)$ is a regular coset for *every* $x \notin V_1$ and the densities of $S$ in the $x + V_1$'s and the corresponding $W \cap (x + V_1)$'s are close on average.

2. **Clean up.** Modify $S$ by deleting all elements that lie in a coset $x + V_1$ if the density of $S$ in $W \cap (x + V_1)$ is close to 0 and by adding all elements that lie in $x + V_1$ if the density of $S$ in $W \cap (x + V_1)$ is close to 1.

3. **Patch around the origin.** Since $S \subseteq V$ is a subset with $H$-density $o(1)$, we cannot be in case $(a)$ of the dichotomy above. Therefore there exists a subset $T \subseteq V_1$ that is $H$-free. Replace $S \cap V_1$ with $T$. (One actually needs a slightly more involved application of our Ramsey-type dichotomy.)

Suppose that there is an $H$-instance sitting among some of the $x + V_1$'s. If none of these cosets contains the origin, then all the $W \cap (x + V_1)$ are regular and all either have high $S$-density or high $(V \setminus S)$-density as required. This is impossible since it would imply the existence of many $H$-instances by an appropriate counting lemma. It is also impossible for an $H$-instance to sit entirely in $V_1$ since $V_1$ was patched to be $H$-free. With a little more work we can also rule out the existence of $H$-instances which are contained partially in $V_1$ and partially in $V \setminus V_1$.

## 2.3 Proof of main theorem

In this section we prove the main theorem, modulo two key tools which are proved in Sections 2.4 and 2.5. As described in Section 2.2 the proof has three steps: find a regular model; clean up; patch around the origin. We first give the precise definition of regularity. As is typical in this field we work in the weighted, or functional, setting where functions represent weighted colorings.

**Definition 2.3.1.** A function $f \colon V \to [-1, 1]$ is *$\epsilon$-regular* if $\|\widehat{f - \mathbb{E}f}\|_\infty \leq \epsilon$. A coloring $\phi \colon V \to [r]$ is *$\epsilon$-regular* if $1_{\phi^{-1}(i)}$ is $\epsilon$-regular for all $1 \leq i \leq r$.

The next proposition accomplishes the "find a regular model" step. As before, $V$ is partitioned into cosets as $V = \bigcup_{x \in U}(x + V_1)$. The regular model will consist of $U \oplus V_2$ for some $V_2 \leq V_1$ where $U$ is a complement for $V_1$ in the sense that $U \oplus V_1 = V$. Property (1) states that the model consists of a positive fraction of $V$, property (2) that the colorings of $x + V_1$ and $x + V_2$ have similar densities, and property (3) that $U \oplus V_2$ is regular. This

proposition is proved in Section 2.5 via an application of a strong arithmetic regularity lemma. We give a novel proof the strong arithmetic regularity lemma, which can be viewed as an arithmetic analog of the cylinder regularity argument in [CF12, Theorem 1.3], which achieves tower-type bounds as opposed to wowzer-type.

**Proposition 2.3.2** (Regular model)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, functions $f_1, \dots, f_k \colon V \to [-1,1]$, a subspace $V_0 \leq V$, and a parameter $\epsilon > 0$, there is some $n_{\mathsf{reg}'} = n_{\mathsf{reg}'}(\epsilon, q, k, \operatorname{codim} V_0)$ such that there exist subspaces $V_2 \leq V_1 \leq V_0$ and a choice of complement $U$ satisfying $U \oplus V_1 = V$ such that*

1. *$\operatorname{codim} V_2 \leq n_{\mathsf{reg}'}$;*

2. *for all but an $\epsilon$-fraction of $x \in U$ we have*

$$|\mathbb{E}_{y \in V_1}[f_i(x+y)] - \mathbb{E}_{y \in V_2}[f_i(x+y)]| \leq \epsilon$$

   *for all $1 \leq i \leq k$;*

3. *for each $x \in U \setminus \{0\}$ and $1 \leq i \leq k$, the function $f_i|_{x+V_2}$ is $\epsilon$-regular (meaning that the function $g_i \colon V_2 \to [-1,1]$ defined by $g_i(y) := f_i(x+y)$ is $\epsilon$-regular).*

With Proposition 2.3.2 in hand it is easy to accomplish the "clean up" step.

**Proposition 2.3.3** (Regularity recoloring)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, a coloring $\phi \colon V \to [r]$, and parameters $0 < \epsilon, \epsilon' \leq 1$, there is some $n_{\mathsf{reg}} = n_{\mathsf{reg}}(\epsilon, \epsilon', q, r)$ such that there exist subspaces $V_2 \leq V_1 \leq V$, a complement $U$ satisfying $U \oplus V_1 = V$, and a recoloring $\phi' \colon V \to [r]$ that agrees with $\phi$ on all but an at most $\epsilon|V|$ values satisfying:*

1. *$1/\epsilon \leq \operatorname{codim} V_1 \leq \operatorname{codim} V_2 \leq n_{\mathsf{reg}}$;*

2. *if a color appears in some coset $x + V_1$ under $\phi'$, then at least an $\epsilon/(2r)$-fraction of $x + V_2$ is that color under $\phi$;*

3. *for each $x \in U \setminus \{0\}$, the original coloring $\phi|_{x+V_2}$ is $\epsilon'$-regular.*

*Proof.* For $1 \le i \le r$, define $f_i \colon V \to [-1,1]$ by $f_i = 1_{\phi^{-1}(i)}$. Let $V_0$ be an arbitrary subspace of codimension $\lceil 1/\epsilon \rceil$. Applying Proposition 2.3.2 with parameter $\epsilon'' = \min(\epsilon/(4r), \epsilon')$ gives subspaces $V_2 \le V_1 \le V_0$ and a choice of complement $U$ satisfying $U \oplus V_1 = V$ with the following properties:

(1′) $\operatorname{codim} V_2 \le n_{\mathsf{reg}'}(\epsilon'', q, r, \lceil 1/\epsilon \rceil)$;

(2′) for all but an $\epsilon/(4r)$-fraction of $x \in U$ we have

$$|\mathbb{E}_{y \in V_1}[f_i(x+y)] - \mathbb{E}_{y \in V_2}[f_i(x+y)]| \le \frac{\epsilon}{4r}$$

for all $1 \le i \le r$;

(3′) for each $x \in U \setminus \{0\}$ and $1 \le i \le r$, the function $f_i|_{x+V_2}$ is $\epsilon'$-regular.

Since $V_1 \le V_0$, clearly $\operatorname{codim} V_1 \ge \operatorname{codim} V_0 \ge 1/\epsilon$. This proves part (1). Part (3) follows from (3′). Next we produce a recoloring $\phi' \colon V \to [r]$ that satisfies part (2). For each $x \in U$, there is some color $i_x$ that appears with density at least $\epsilon/(2r)$ in $x + V_2$. For each color $i$ which appears with density less than $\epsilon/(2r)$ in $x + V_2$, recolor every element of color $i$ in $x + V_1$ to color $i_x$. This recoloring satisfies part (2).

All that remains is to show that $\phi$ and $\phi'$ agree on all but an $\epsilon$-fraction of $V$. At least a $(1 - \epsilon/(4r))$-fraction of $x \in U$ satisfy

$$|\mathbb{E}_{y \in V_1}[f_i(x+y)] - \mathbb{E}_{y \in V_2}[f_i(x+y)]| \le \frac{\epsilon}{4r} \quad \text{for all } 1 \le i \le r. \tag{$*$}$$

For the $x \in U$ failing to satisfy $(*)$, we use the trivial bound: the number of elements of $x + V_1$ that we recolor is at most all $|V_1|$ of them. For the rest of the $x \in U$, for each color $i$ which occurs with density at most $\epsilon/(2r)$ in $x + V_2$, color $i$ has density at most $\epsilon/(2r) + \epsilon/(4r)$ in $x + V_1$. Thus for each $x \in U$ satisfying $(*)$ we recolor at most a $3\epsilon/4$-fraction of $x + V_1$. Combining these two cases, $\phi'$ and $\phi$ agree on all but at most a $3\epsilon/4 + \epsilon/(4r) \le \epsilon$-fraction of $V$. □

To accomplish the "patch around the origin" step we need the following dichotomy which is proved in Section 2.4.

**Proposition 2.3.4** (Ramsey dichotomy)**.** *Fix a finite set $\mathcal{H}$ of $r$-colored patterns over $\mathbb{F}_q$. There exist $n_{\mathsf{rado}} = n_{\mathsf{rado}}(\mathcal{H})$ and $\epsilon_{\mathsf{rado}} = \epsilon_{\mathsf{rado}}(\mathcal{H}) > 0$ such that either:*

*(a) every coloring $\phi\colon V \to [r]$ of $V$, a finite dimensional $\mathbb{F}_q$-vector space with $\dim V \geq n_{\mathsf{rado}}$, has $H$-density at least $\epsilon_{\mathsf{rado}}$ for some $H \in \mathcal{H}$; or*

*(b) for every finite dimensional $\mathbb{F}_q$-vector space $V$ there exists a coloring $\phi\colon V \setminus \{0\} \to [r]$ that is $\mathcal{H}$-free.*

As we saw in the previous discussion, Proposition 2.3.2 and Proposition 2.3.3 allow us to recolor $V \setminus V_1$ to be $\mathcal{H}$-free. Furthermore, applying Proposition 2.3.4 to $\mathcal{H}$ allows us to recolor $V_1 \setminus \{0\}$ to be $\mathcal{H}$-free. However this does not rule out $H$-instances which are contained partially in $V_1 \setminus \{0\}$ and partially in $V \setminus V_1$.

To deal with this problem we apply Proposition 2.3.4 not only to $\mathcal{H}$ but also to some of the *subpatterns* of the patterns $H \in \mathcal{H}$. The subpatterns of an arithmetic pattern $H$ are analogous to the subgraphs of a graph $H$.

**Definition 2.3.5.** Let $H = (A, \psi)$ be an $r$-colored pattern where $A$ is an $\ell \times k$ matrix with entries in $\mathbb{F}_q$. For $I \subseteq [k]$ say that $H' = (A', \psi|_I)$ is a *subpattern of $H$ restricted to the variables $I$* if for all $V$, a finite dimensional $\mathbb{F}_q$-vector space, it holds that for $\boldsymbol{x} \in V^I$ we have $A'\boldsymbol{x} = \mathbf{0}$ if and only if there exists $\boldsymbol{y} \in V^k$ satisfying $A\boldsymbol{y} = \mathbf{0}$ and $x_i = y_i$ for all $i \in I$. (Here $\psi|_I\colon I \to [r]$ is the restriction of $\psi$ to $I$.)

**Example 2.3.6.** Consider the following pattern which represents a red 4-term arithmetic progression:

$$H = \left( A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}, \psi \equiv \text{red} \right).$$

The subpattern of $H$ corresponding to the first three variables (i.e., $I = \{1, 2, 3\}$) is a red 3-term arithmetic progression. One possible representation is as:

$$H' = \left( A' = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}, \psi \equiv \text{red} \right).$$

**Example 2.3.7.** Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

The equation $A\boldsymbol{x} = \mathbf{0}$ imposes the constraints $x_1 + x_2 + x_3 = 0$ and $x_3 + x_4 + x_5 = 0$. Restricting to variables $x_1, x_2, x_3$ (i.e., $I = \{1, 2, 3\}$), the only constraint is $x_1 + x_2 + x_3 = 0$ which may be represented by the matrix

$$A' = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

Similarly, restricting to variables $x_1, x_2, x_4, x_5$ (i.e., $I = \{1, 2, 4, 5\}$), the only constraint is $x_1 + x_2 - x_4 - x_5 = 0$ which may be represented by the matrix

$$A' = \begin{pmatrix} 1 & 1 & -1 & -1 \end{pmatrix}.$$

Let $H = (A, \psi)$ be an $r$-colored pattern where $A$ is an $\ell \times k$ matrix with entries in $\mathbb{F}_q$. Suppose $H' = (A', \psi|_{[j]})$ is a subpattern of $H$ restricted to the first $j$ variables. Our definitions are such that for every finite dimensional $\mathbb{F}_q$-vector space $V$ and $f_1, \ldots, f_k \colon V \to [-1, 1]$, one has

$$\Lambda_A(f_1, \ldots, f_j, 1, \ldots, 1) = \Lambda_{A'}(f_1, \ldots, f_j).$$

*Proof of Theorem 2.1.15.* We are given $\mathcal{H}$, a finite set of $r$-colored complexity 1 patterns over $\mathbb{F}_q$, and $\phi \colon V \to [r]$, an $r$-coloring of a finite dimensional $\mathbb{F}_q$-vector space.

Let $\tilde{\mathcal{H}}$ be a finite set containing one representative of each subpattern of each $H \in \mathcal{H}$. More precisely, for each $H = (A, \psi) \in \mathcal{H}$, a pattern where $A$ is an $\ell \times k$ matrix, and each $I \subseteq [k]$, the set $\tilde{\mathcal{H}}$ contains some $H' = (A', \psi|_I)$, a subpattern of $H$ restricted to variables $I$. We define the following constants. (Recall that $n_{\mathsf{rado}}$ and $\epsilon_{\mathsf{rado}}$ were defined in Proposition 2.3.4 while $n_{\mathsf{reg}}$ was defined in Proposition 2.3.3.)

$$\epsilon_{\mathsf{rado}} = \min_{\mathcal{H}' \subseteq \tilde{\mathcal{H}}} \epsilon_{\mathsf{rado}}(\mathcal{H}') \quad \text{and} \quad n_{\mathsf{rado}} = \max_{\mathcal{H}' \subseteq \tilde{\mathcal{H}}} n_{\mathsf{rado}}(\mathcal{H}').$$

Write $k_{\mathsf{max}}$ for the maximum number of columns in any of the patterns $H \in \mathcal{H}$. With foresight, we define

$$\epsilon_{\mathsf{count}} = \frac{1}{2k_{\mathsf{max}}} \left(\frac{\epsilon}{4r}\right)^{k_{\mathsf{max}}} \epsilon_{\mathsf{rado}}.$$

Define $\epsilon_{\mathsf{reg}}$ such that for all $H = (A, \psi) \in \tilde{\mathcal{H}}$ and for all $f_1, \ldots, f_k \colon V \to [-1, 1]$ we have

$$|\Lambda_A(f_1, \ldots, f_k)| \leq \epsilon_{\mathsf{count}} \quad \text{whenever} \quad \min_{1 \leq i \leq k} \|\widehat{f_i}\|_\infty \leq \epsilon_{\mathsf{reg}}. \tag{$\dagger$}$$

We can do this since all $H \in \mathcal{H}$ are complexity 1 (and this implies that all $H \in \tilde{\mathcal{H}}$ are complexity 1). Define $n_0 = n_{\mathsf{rado}} + n_{\mathsf{reg}}(\epsilon/2, \epsilon_{\mathsf{reg}}, q, r)$. Finally define

$$\delta = \min\left(\frac{1}{4} q^{-k_{\mathsf{max}} \cdot n_{\mathsf{reg}}(\epsilon/2, \epsilon_{\mathsf{reg}}, q, r)} \left(\frac{\epsilon}{4r}\right)^{k_{\mathsf{max}}} \epsilon_{\mathsf{rado}}, q^{-n_0 k_{\mathsf{max}}}\right).$$

First note that the desired result is trivially true for $\dim V < n_0$; if a pattern $H$ with at most $k_{\mathsf{max}}$ columns has density at most $\delta \leq q^{-n_0 k_{\mathsf{max}}}$ in $\phi \colon V \to [r]$, then $\phi$ must have fewer than one $H$-instance, i.e., be $H$-free.

Now assume that $\dim V \geq n_0$. With these constants chosen we move on to the recoloring algorithm. First, we apply Proposition 2.3.3 to $\phi$ with parameters $\epsilon/2, \epsilon_{\mathsf{reg}}$. This produces $V_2 \leq V_1 \leq V$ and $U$ satisfying $U \oplus V_1 = V$ and a recoloring $\phi' \colon V \setminus V_1 \to [r]$ that differs from $\phi$ on at most $(\epsilon/2)|V|$ elements that satisfies:

1. $2/\epsilon \leq \operatorname{codim} V_1 \leq \operatorname{codim} V_2 \leq n_{\mathsf{reg}}(\epsilon/2, \epsilon_{\mathsf{reg}}, q, r)$;
2. if a color appears in some coset $x + V_1$ under $\phi'$, then at least an $\epsilon/(4r)$-fraction of $x + V_2$ is that color under $\phi$;
3. for each $x \in U \setminus \{0\}$, the original coloring $\phi|_{x+V_2}$ is $\epsilon_{\mathsf{reg}}$-regular.

Second, define $\mathcal{H}' \subseteq \tilde{\mathcal{H}}$ to be the set of all patterns $H \in \tilde{\mathcal{H}}$ which appear with density less than $\epsilon_{\mathsf{rado}}$ in $\phi|_{V_2}$. Note that we have $\dim V_2 \geq n_{\mathsf{rado}}$. We now apply Proposition 2.3.4; by the definition of $\mathcal{H}'$, we cannot be in case (a), since $\phi|_{V_2}$ is a counterexample. Thus we are in case (b) which implies the existence coloring $\phi' \colon V_1 \setminus \{0\} \to [r]$ that is $\mathcal{H}'$-free.

Combining these two results gives a recoloring $\phi' \colon V \setminus \{0\} \to [r]$. First note that $\phi'$ differs from $\phi$ on at most $(\epsilon/2)|V| + q^{-\operatorname{codim} V_1}|V| \leq \epsilon|V|$ elements of $V$. Now suppose that $\boldsymbol{x}$ is

an $H$-instance in $\phi'$ for some $(A, \psi) = H \in \mathcal{H}$. Write $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$ with $\boldsymbol{u} \in U^k$ and $\boldsymbol{v} \in V_1^k$. We will prove that this implies that the $H$-density is large in $\phi$; we do this by counting $H$-instances of the form $\boldsymbol{u} + \boldsymbol{y}$ with $\boldsymbol{u}$ fixed as above and $\boldsymbol{y} \in V_2^k$.

Relabeling variables as needed, assume that $u_1, \dots, u_j = 0$ and $u_{j+1}, \dots, u_k \neq 0$. Let $H' = (A', \psi|_{[j]}) \in \tilde{\mathcal{H}}$ be a subpattern of $H$ restricted to the first $j$ variables. The existence of $\boldsymbol{x}$ means that for $i = j+1, \dots, k$, the color $\psi(i)$ appears at least once in $u_i + V_1$ under $\phi'$ and it means that there exists an $H'$-instance in $\phi'|_{V_1}$, namely $(x_1, \dots, x_j)$. By property (2) and the $\mathcal{H}'$-freeness of $\phi'|_{V_1}$, this implies that for $i = j+1, \dots, k$, the density of color $\psi(i)$ in $u_i + V_2$ under $\phi$ is at least $\epsilon/(4r)$ and that the $H'$-density in $\phi|_{V_2}$ is at least $\epsilon_{\mathsf{rado}}$.

Define $f_i \colon V \to [0,1]$ by $f_i(x) = 1_{\phi^{-1}(\psi(i))}(x)$ and $g_i \colon V_2 \to [0,1]$ by $g_i(x) = 1_{\phi^{-1}(\psi(i))}(x + u_i)$. In this notation, the conclusions of the last paragraph are that $\mathbb{E} g_i \geq \epsilon/(4r)$ for $i = j+1, \dots, k$ and $\Lambda_{A'}(g_1, \dots, g_j) \geq \epsilon_{\mathsf{rado}}$. Finally we compute

$$\begin{aligned}
\Lambda_A(f_1, \dots, f_k) &\geq q^{-(k - \operatorname{rank} A)\cdot \operatorname{codim} V_2} \Lambda_A(g_1, \dots, g_k) \\
&\geq q^{-(k - \operatorname{rank} A)\cdot \operatorname{codim} V_2} \left(\Lambda_A(g_1, \dots, g_j, \mathbb{E} g_{j+1}, \dots, \mathbb{E} g_k) - (k-j)\epsilon_{\mathsf{count}}\right) \\
&\geq q^{-(k - \operatorname{rank} A)\cdot \operatorname{codim} V_2} \left(\left(\frac{\epsilon}{4r}\right)^{k-j} \Lambda_{A'}(g_1, \dots, g_j) - (k-j)\epsilon_{\mathsf{count}}\right) \\
&\geq q^{-k\cdot \operatorname{codim} V_2} \left(\left(\frac{\epsilon}{4r}\right)^{k} \epsilon_{\mathsf{rado}} - k\epsilon_{\mathsf{count}}\right) \\
&> \delta.
\end{aligned}$$

The first line represents restricting from arbitrary $H$-instances to $H$-instances of the form $\boldsymbol{u} + \boldsymbol{y}$ with $\boldsymbol{y} \in V_2^k$ (note that the $f_i$ are non-negative). The second line follows by iterating (†). The last line follows from our choice of $\epsilon_{\mathsf{count}}$ and $\delta$.

Thus if the above recoloring procedure does not produce an $\mathcal{H}$-free coloring, then $\Lambda_A(f_1, \dots, f_k) > \delta$, meaning that the original coloring had $H$-density more than $\delta$ for some $H \in \mathcal{H}$. This completes the proof. □

## 2.4 Proof of Proposition 2.3.4: a density Ramsey dichotomy

In this section we prove Proposition 2.3.4. The proof is based on the following trivial dichotomy:

> For any $\mathcal{H}$, a finite set of $r$-colored patterns over $\mathbb{F}_q$, there exists a constant $n_0$ such that either:
>
> (a) every $r$-coloring of every $V$ with $\dim V \geq n_0$ contains an $H$-instance for some $H \in \mathcal{H}$; or
>
> (b) for every finite dimensional $\mathbb{F}_q$-vector space $V$, there exists an $r$-coloring of $V$ that is $\mathcal{H}$-free.

To boost this result to the full strength of Proposition 2.3.4 we use a sampling argument. The idea of the sampling is simple. Suppose we have some $\mathcal{H}$ that is in case (a). Then for an $r$-coloring of $V$ we know that there is not just one $H$-instance, but one $H$-instance in every dimension $n_0$ subspace of $V$. Adding up all of these gives many $H$-instances in $V$. Unfortunately we have overcounted each $H$-instance many times. Double counting more carefully gives us the desired result, but only if the original $H$-instances are *generic* in the following sense.

**Definition 2.4.1.** For $A$ an $\ell \times k$ matrix with entries in $\mathbb{F}_q$ and $\boldsymbol{x} = (x_1, \ldots, x_k) \in V^k$ satisfying $A\boldsymbol{x} = \mathbf{0}$, say that $\boldsymbol{x}$ is a *generic solution* if $\dim(\operatorname{span}\{x_1, \ldots, x_k\}) = k - \operatorname{rank} A$. For $H = (A, \psi)$ a colored pattern, define a *generic $H$-instance* to be an $H$-instance $\boldsymbol{x}$ that is a generic solution to $A\boldsymbol{x} = \mathbf{0}$.

To prove Proposition 2.3.4 we first show that we can always turn an $H$-instance into a generic $H$-instance, then we perform the sampling argument.

The first step relies on an argument which is inspired by elements of the proof of Rado's theorem (see [Deu75]). We reduce the question of whether $\mathcal{H}$ lies in case (a) of Proposition 2.3.4 to a finite check. Instead of checking if there exists an $\mathcal{H}$-instance in every coloring we only

need to check the so-called *canonical colorings*. (With some manipulations this check can be made to resemble the "columns condition" of Rado's theorem.)

**Definition 2.4.2.** For $\chi\colon \mathbb{F}_q \setminus \{0\} \to [r]$ and $n > 0$, define the *$\chi$-canonical coloring* $\Phi_{n,\chi}\colon \mathbb{F}_q^n \setminus \{0\} \to [r]$ by $\Phi_{n,\chi}(x) := \chi(\mathrm{fnz}(x))$ where $\mathrm{fnz}\colon \mathbb{F}_q^n \setminus \{0\} \to \mathbb{F}_q \setminus \{0\}$ maps a vector to its first non-zero coordinate.

**Lemma 2.4.3.** *For $\mathcal{H}$, a finite set of $r$-colored patterns over $\mathbb{F}_q$, write $k$ for the maximum number of columns in any of the patterns $H \in \mathcal{H}$. There exists some $n_{\mathsf{rado}} = n_{\mathsf{rado}}(\mathcal{H})$ such that for any $n_0 \geq n_{\mathsf{rado}}$ the following are equivalent:*

*(a) For every finite dimensional $\mathbb{F}_q$-vector space $V$ satisfying $\dim V \geq n_0$ and every coloring $\phi\colon V \setminus \{0\} \to [r]$, there is an $H$-instance in $\phi$ for some $H \in \mathcal{H}$.*

*(b) For every $\chi\colon \mathbb{F}_q \setminus \{0\} \to [r]$, there is an $H$-instance in $\Phi_{k,\chi}$ for some $H \in \mathcal{H}$.*

*(c) For every finite dimensional $\mathbb{F}_q$-vector space $V$ satisfying $\dim V \geq n_{\mathsf{rado}}$ and every coloring $\phi\colon V \setminus \{0\} \to [r]$, there is a generic $H$-instance in $\phi$ for some $H \in \mathcal{H}$.*

Note that in the above lemma statement, an $H$-instance always refers to $\boldsymbol{x}$ where none of the vectors $x_i$ are 0 since all of the colorings considered leave 0 uncolored. We will define the function $n_{\mathsf{rado}}(\mathcal{H})$ in the statement of Proposition 2.3.4 to be the same as the $n_{\mathsf{rado}}(\mathcal{H})$ in this lemma.

We use the following vector space Ramsey theorem of Graham and Rothschild which was proved with primitive-recursive bounds by Shelah.

**Theorem 2.4.4** (Graham-Rothschild [GR71], Shelah [She88][1])**.** *There exists $n_{\mathsf{GR}} = n_{\mathsf{GR}}(q, r, k)$ such that for a finite dimensional $\mathbb{F}_q$-vector space $V$ with $\dim V \geq n_{\mathsf{GR}}$ and $\phi\colon V \setminus \{0\} \to [r]$, an $r$-coloring that is constant on one-dimensional subspaces (with zero removed), there exists a $k$-dimensional subspace $U \leq V$ such that $U \setminus \{0\}$ is monochromatic under $\phi$.*

The best-known upper bound on $n_{\mathsf{GR}}$ as a function of $k$ is broadly comparable to the third function in the Ackermann hierarchy.

[1]Note that both of these references prove a slight variant of this result, the "affine Ramsey theorem". Spencer [Spe79] gives an easy deduction of this result, referred to as the "vector space Ramsey theorem", from the affine Ramsey theorem.

*Proof of Lemma 2.4.3.* Define $n_{\mathsf{rado}}(\mathcal{H}) = n_{\mathsf{GR}}(q, r^{q-1}, 2k)$ and fix any $n_0 \geq n_{\mathsf{rado}}$. With these choices we show that $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

$(a) \Rightarrow (b)$: Fix $\chi \colon \mathbb{F}_q \setminus \{0\} \to [r]$ and take $n \geq n_0$. Assuming (a), there is an $H$-instance in $\Phi_{n,\chi}$ for some $H \in \mathcal{H}$. Say that $H = (A, \psi)$ where $A$ is an $\ell \times k'$ matrix and call the $H$-instance $\boldsymbol{x} \in (\mathbb{F}_q^n \setminus \{0\})^{k'}$. We use this $\boldsymbol{x}$ to find an $H$-instance in $\Phi_{k,\chi}$.

For each $1 \leq j \leq k'$, let $b_j$ be the position of the first non-zero coordinate of $x_j$. Relabeling the variables as necessary, assume that $b_1 \leq b_2 \leq \cdots \leq b_{k'}$. Consider the map $\Pi \colon \mathbb{F}_q^n \to \mathbb{F}_q^k$ which sends $v = (v_1, \ldots, v_n) \in \mathbb{F}_q^n$ to $(v_{b_1}, v_{b_2}, \ldots, v_{b_{k'}}, 0, \ldots, 0) \in \mathbb{F}_q^k$, where there are $k - k'$ zeros added at the end. One can easily check that each $x_j$ satisfies $\Phi_{n,\chi}(x_j) = \Phi_{k,\chi}(\Pi(x_j))$. Thus $\Pi(\boldsymbol{x})$ is an $H$-instance in $\Phi_{k,\chi}$, as desired.

$(b) \Rightarrow (c)$: Take $n \geq n_{\mathsf{rado}}$ and a coloring $\phi \colon V \setminus \{0\} \to [r]$ of a finite dimensional $\mathbb{F}_q$-vector space $V$ satisfying $\dim V = n$. We define an $r^{q-1}$-coloring $\overline{\phi} \colon V \setminus \{0\} \to [r]^{q-1}$ that is constant on one-dimensional subspaces as follows. Pick an isomorphism $V \cong \mathbb{F}_q^n$. For $x \in V \setminus \{0\}$ there exists a unique $a \in \mathbb{F}_q \setminus \{0\}$ such that the first non-zero coordinate of $ax$ is 1. Then define

$$\overline{\phi}(x) := (\phi(abx))_{b \in \mathbb{F}_q \setminus \{0\}} \in [r]^{q-1}.$$

By Theorem 2.4.4 and our choice of $n_{\mathsf{rado}}$, there exists $U \leq V$, a $2k$-dimensional subspace such that $\overline{\phi}$ is constant on $U \setminus \{0\}$. We claim that this implies that there is a coloring $\chi \colon \mathbb{F}_q \setminus \{0\} \to [r]$ and an isomorphism $\iota \colon U \xrightarrow{\sim} \mathbb{F}_q^{2k}$ such that $\phi$ agrees with the canonical coloring $\Phi_{2k,\chi}$ on $U \setminus \{0\}$.

The color of $U \setminus \{0\}$ under $\overline{\phi}$ is some vector in $[r]^{q-1}$, in other words, a function $\mathbb{F}_q \setminus \{0\} \to [r]$. This is our $\chi$. Now pick a basis for $U \leq \mathbb{F}_q^n$. Performing Gaussian elimination on this basis transforms it into a basis $u_1, \ldots, u_{2k} \in U$ such that the first non-zero coordinate of each of these vectors is 1 and the location of the first non-zero coordinates is strictly increasing from $u_1$ to $u_{2k}$. Then define $\iota \colon U \xrightarrow{\sim} \mathbb{F}_q^{2k}$ to send $c_1 u_1 + \cdots + c_{2k} u_{2k} \mapsto (c_1, \ldots, c_{2k}) \in \mathbb{F}_q^{2k}$. These $\chi, \iota$ have the desired properties since $\mathrm{fnz}(c_1 u_1 + \cdots + c_{2k} u_{2k}) = \mathrm{fnz}((c_1, \ldots, c_{2k}))$.

Consider $\mathbb{F}_q^k \oplus \{0\} \subset \mathbb{F}_q^{2k}$, the subspace of vectors of the form $(*, \ldots, *, 0, \ldots, 0)$. Assuming (b), there exists an $H$-instance $\boldsymbol{x} \in ((\mathbb{F}_q^k \setminus \{0\}) \oplus \{0\})^{k'}$ in $\Phi_{2k,\chi}$. Now for any $\boldsymbol{y} \in \left(\{0\} \oplus \mathbb{F}_q^k\right)^{k'}$ we have that $x_j$ and $x_j + y_j$ are the same color under $\Phi_{2k,\chi}$ for all $j$. Pick $\boldsymbol{y}$ any generic

solution to $A\boldsymbol{y} = \mathbf{0}$ in $\{0\} \oplus \mathbb{F}_q^k$. Then $\boldsymbol{x} + \boldsymbol{y}$ is a generic $H$-instance in $\Phi_{2k,\chi}$. Since $\phi|_U$ agrees with $\Phi_{2k,\chi}$, we have found a generic $H$-instance in $\phi$, as desired.

$(c) \Rightarrow (a)$: Obvious. □

*Proof of Proposition 2.3.4.* Fix a finite set $\mathcal{H}$ of $r$-colored patterns over $\mathbb{F}_q$. Write $k_{\mathsf{max}}$ for the maximum number of columns in any of the patterns $H \in \mathcal{H}$. Define $n_{\mathsf{rado}} = n_{\mathsf{rado}}(\mathcal{H})$ to be the same as in Lemma 2.4.3 and, with foresight, define

$$\epsilon_{\mathsf{rado}} = \frac{1}{10|\mathcal{H}|} q^{-n_{\mathsf{rado}} k_{\mathsf{max}}}.$$

Suppose that we are not in case (b) of Proposition 2.3.4; namely, that there is some finite dimensional $\mathbb{F}_q$-vector space $V_0$ such that every coloring $\phi\colon V_0 \setminus \{0\} \to [r]$ contains an $H$-instance for some $H \in \mathcal{H}$. We now apply Lemma 2.4.3 to $\mathcal{H}$ with parameter $n_0 := \max\{\dim V_0, n_{\mathsf{rado}}\}$. By assumption, statement (a) of Lemma 2.4.3 is true, so we conclude that statement (c) is also true: for every finite dimensional $\mathbb{F}_q$-vector space $V$ satisfying $\dim V \geq n_{\mathsf{rado}}$ and every coloring $\phi\colon V \setminus \{0\} \to [r]$, there is a generic $H$-instance in $\phi$ for some $H \in \mathcal{H}$.

Now we prove we are in case (a) of Proposition 2.3.4. Fix $V$, a finite dimensional $\mathbb{F}_q$-vector space with $\dim V = n \geq n_{\mathsf{rado}}$, and a coloring $\phi\colon V \to [r]$. We count the total number of generic $\mathcal{H}$-instances in $\phi$. For ease of notation, given $H = (A, \psi) \in \mathcal{H}$, write $k_H$ for the number of columns in $A$.

Consider the following random process: pick a uniform random $H = (A, \psi) \in \mathcal{H}$ and then pick a uniform random generic solution $A\boldsymbol{x} = 0$ with $\boldsymbol{x} \in V^{k_H}$. Write $N_H(\phi)$ for the number of generic $H$-instances in $\phi$. Then the probability that this process produces a pair $(H, \boldsymbol{x})$ where $\boldsymbol{x}$ is a generic $H$-instance is

$$\frac{1}{|\mathcal{H}|} \sum_{H=(A,\psi)\in\mathcal{H}} \frac{N_H(\phi)}{(q^n-1)(q^n-q)\cdots(q^n-q^{k_H-\operatorname{rank} A-1})} \leq \frac{10}{|\mathcal{H}|} \sum_{H=(A,\psi)\in\mathcal{H}} N_H(\phi) q^{-n(k_H-\operatorname{rank} A)}.$$

Now consider the following random process: pick $U \leq V$, a uniform random subspace of dimension $n_{\mathsf{rado}}$, pick a uniform random $H = (A, \psi) \in \mathcal{H}$, and then pick a uniform random generic solution $A\boldsymbol{x} = 0$ with $\boldsymbol{x} \in U^{k_H}$. Since, by assumption, there is at least one generic

$\mathcal{H}$-instance in $\phi|_U$, the probability that this process finds a pair $(H, \boldsymbol{x})$ where $\boldsymbol{x}$ is a generic $H$-instance is at least

$$\frac{1}{|\mathcal{H}|} \frac{1}{(q^{n_{\mathsf{rado}}} - 1)(q^{n_{\mathsf{rado}}} - q) \cdots (q^{n_{\mathsf{rado}}} - q^{k_{\mathsf{max}}-1})} \geq \frac{1}{|\mathcal{H}|} q^{-n_{\mathsf{rado}} k_{\mathsf{max}}}.$$

Note that these two random processes actually produce the same distribution on pairs $(H, \boldsymbol{x})$. (This is the key place where we use the assumption that the $\boldsymbol{x}$ are generic solutions.) Combining these two bounds, we conclude that

$$\sum_{H=(A,\psi)\in\mathcal{H}} N_H(\phi) q^{-n(k_H - \operatorname{rank} A)} \geq \frac{1}{10} q^{-n_{\mathsf{rado}} k_{\mathsf{max}}}.$$

Thus by the pigeonhole principle, there exists some pattern $H = (A, \psi) \in \mathcal{H}$ such that $N_H(\phi) q^{-n(k_H - \operatorname{rank} A)}$, i.e., the $H$-density in $\phi$, is at least $\epsilon_{\mathsf{rado}}$, as desired. □

## 2.5 Proof of Proposition 2.3.2: arithmetic regularity lemmas

We give two proofs of Proposition 2.3.2, giving wowzer- and tower-type bounds respectively. The first proof can be considered the arithmetic analog of the strong regularity argument in [AFKS00] and the second bears some resemblance to an arithmetic analog of the cylinder regularity argument in [CF12, Theorem 1.3].

Our main technique for proving arithmetic regularity lemmas is an energy increment argument.

**Definition 2.5.1.** Fix a finite dimensional $\mathbb{F}_q$-vector space $V$ and functions $f_1, \ldots, f_k \colon V \to [-1, 1]$. Given $\mathcal{P}$ a partition of $S \subseteq V$, define $(f_i)_{\mathcal{P}} \colon S \to [-1, 1]$ to be the projection of $f_i|_S$ onto the $\sigma$-algebra generated by $\mathcal{P}$. In other words, $(f_i)_{\mathcal{P}}(x)$ is defined to be the average of $f_i$ over the part of $\mathcal{P}$ containing $x$. The *energy* of the partition is defined to be

$$\mathcal{E}(\mathcal{P}) := \sum_{i=1}^{k} \|(f_i)_{\mathcal{P}}\|_{L^2(S)}^2 .$$

One case will be of particular interest to us. Let $V_1 \leq V$ be a linear subspace and $S \subseteq V$ any set which can be written as the union of cosets of $V_1$. Then write $\mathcal{P}(V_1|S)$ for the partition of $S$ into cosets of $V_1$. When $S$ is not given, it is assumed to be all of $V$.

We record the following three properties of energy.

**Proposition 2.5.2.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$ and functions $f_1, \ldots, f_k \colon V \to [-1, 1]$, we have:*

1. *for any partition $\mathcal{P}$ of a set $S \subseteq V$,*

$$0 \leq \mathcal{E}(\mathcal{P}) \leq k;$$

2. *for any partitions $\mathcal{P}, \mathcal{Q}$ of a set $S \subseteq V$ with $\mathcal{Q}$ refining $\mathcal{P}$ (written $\mathcal{Q} \succeq \mathcal{P}$),*

$$\mathcal{E}(\mathcal{Q}) - \mathcal{E}(\mathcal{P}) = \sum_{i=1}^{k} \|(f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}}\|_{L^2(S)}^2 \geq 0.$$

3. *for $V_1 \leq V$, if there exists $1 \leq i \leq k$ and $x \in V$ such that $f_i|_{x+V_1}$ is not $\epsilon$-regular, there exists $V_2 \leq V_1$ of codimension 1 such that*

$$\mathcal{E}(\mathcal{P}(V_2|x + V_1)) - \mathcal{E}(\mathcal{P}(V_1|x + V_1)) > \epsilon^2.$$

*Proof.* The first result follows since trivially $0 \leq \|(f_i)_{\mathcal{P}}\|_{L^2(S)}^2 \leq 1$ for all $1 \leq i \leq k$.

The second result follows from the Pythagorean theorem. For all $1 \leq i \leq k$,

$$\|(f_i)_{\mathcal{P}}\|_{L^2(S)}^2 + \|(f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}}\|_{L^2(S)}^2 = \|(f_i)_{\mathcal{Q}}\|_{L^2(S)}^2$$

as long as $(f_i)_{\mathcal{P}}$ and $(f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}}$ are orthogonal. We compute

$$\langle (f_i)_{\mathcal{P}}, (f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}} \rangle_{L^2(S)} = \langle f_i, ((f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}})_{\mathcal{P}} \rangle_{L^2(S)} = \langle f_i, 0 \rangle_{L^2(S)} = 0.$$

To prove the third result, suppose that $|\widehat{f_i|_{x+V_1}}(\gamma_z)| > \epsilon$ for some $\gamma_z \in V_1^* \setminus \{0\}$ and some $1 \leq i \leq k$. Recall that $\gamma_z$ is the character $\gamma_z(x) := \exp(2\pi i \operatorname{tr}(x \cdot z)/p)$. Define $V_2 \leq V_1$ to

be $V_2 = \{y \in V_1 : y \cdot z = 0\}$. For ease of notation, write $\mathcal{P}$ for $\mathcal{P}(V_1|x+V_1)$ and $\mathcal{Q}$ for $\mathcal{P}(V_2|x+V_1)$ for the rest of the proof. Note that $\gamma_z$ is constant on parts of $\mathcal{Q}$. Thus we can write

$$\begin{aligned}
\epsilon^2 &< \left|\widehat{f_i|_{x+V_1}}(\gamma_z)\right|^2 \\
&= \left|\langle f_i, \gamma_z\rangle_{L^2(x+V_1)}\right|^2 \\
&= \left|\langle (f_i)_{\mathcal{Q}}, \gamma_z\rangle_{L^2(x+V_1)}\right|^2 \\
&= \left|\langle (f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}}, \gamma_z\rangle_{L^2(x+V_1)}\right|^2 && \text{(since } \mathbb{E}\gamma_z = 0 \text{ and } (f_i)_{\mathcal{P}} \text{ is constant)} \\
&\leq \|(f_i)_{\mathcal{Q}} - (f_i)_{\mathcal{P}}\|^2_{L^2(x+V_1)} && \text{(Cauchy-Schwarz inequality)} \\
&\leq \mathcal{E}(\mathcal{Q}) - \mathcal{E}(\mathcal{P}). && \square
\end{aligned}$$

### 2.5.1 Proof 1: Strong arithmetic regularity

We first prove Green's arithmetic regularity lemma in finite field vector spaces by the standard energy increment method. We then iterate this arithmetic regularity lemma to prove a strong arithmetic regularity lemma. Given functions $f_1, \ldots, f_k \colon V \to [-1,1]$, a subspace $V_0 \leq V$, and $\epsilon > 0$, this strong arithmetic regularity lemma gives subspaces $V_2 \leq V_1 \leq V_0$ with several desirable properties. To complete the proof of Proposition 2.3.2 we choose a random complement $U$ satisfying $U \oplus V_1 = V$ and show that $U$ has the desired properties with positive probability.

For completeness we include the standard proof of Green's arithmetic regularity lemma.

**Theorem 2.5.3** (Green's arithmetic regularity)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, functions $f_1, \ldots, f_k \colon V \to [-1,1]$, a subspace $V_0 \leq V$, and a parameter $\epsilon > 0$, there exists $n_{\mathsf{green-reg}} = n_{\mathsf{green-reg}}(\epsilon, q, k, \operatorname{codim} V_0)$ such that there exists a subspace $V_1 \leq V_0$ such that*

1. $\operatorname{codim} V_1 \leq n_{\mathsf{green-reg}}$*;*
2. *for each $1 \leq i \leq k$, for all but an $\epsilon$-fraction of $x \in V$ the function $f_i|_{x+V_1}$ is $\epsilon$-regular.*

*Proof.* We iterate to produce a sequence of subspaces $V_0 = V^{(0)} \geq V^{(1)} \geq \cdots \geq V^{(M)}$ such that setting $V_1 = V^{(M)}$ proves the theorem. This sequence will satisfy $\operatorname{codim} V^{(m+1)} \leq$

$\operatorname{codim} V^{(m)} + q^{\operatorname{codim} V^{(m)}}$ and $\mathcal{E}(\mathcal{P}(V^{(m+1)})) > \mathcal{E}(\mathcal{P}(V^{(m)})) + \epsilon^3$.

Suppose that we have produced some $V^{(m)}$ that does not satisfy property (2) above. Pick some complement $U$ satisfying $U \oplus V^{(m)} = V$. Then there exist some $1 \leq i \leq k$ and $\epsilon|U|$ values of $x \in U$, say $U' \subset U$, such that $f_i|_{x+V^{(m)}}$ is not $\epsilon$-regular for each $x \in U'$.

Fix $x \in U'$. Since $f_i|_{x+V^{(m)}}$ is not $\epsilon$-regular, part (3) of Proposition 2.5.2 gives a subspace $V(x) \leq V^{(m)}$ of codimension 1 such that $\mathcal{E}(\mathcal{P}(V(x)|x+V^{(m)})) > \mathcal{E}(\mathcal{P}(V^{(m)}|x+V^{(m)})) + \epsilon^2$. Define $V^{(m+1)}$ to be the intersection of all of the $V(x)$ for all $x \in U'$. Then $\operatorname{codim} V^{(m+1)} \leq \operatorname{codim} V^{(m)} + |U'| \leq \operatorname{codim} V^{(m)} + q^{\operatorname{codim} V^{(m)}}$, as desired.

Note that for $W \leq V^{(m)}$, we have $\mathcal{E}(\mathcal{P}(W)) = \mathbb{E}_{x \in U}[\mathcal{E}(\mathcal{P}(W|x+V^{(m)}))]$. Therefore

$$\begin{aligned}
\mathcal{E}(\mathcal{P}(V^{(m+1)})) - \mathcal{E}(\mathcal{P}(V^{(m)})) &= \mathbb{E}_{x \in U}\left[\mathcal{E}(\mathcal{P}(V^{(m+1)}|x+V^{(m)})) - \mathcal{E}(\mathcal{P}(V^{(m)}|x+V^{(m)}))\right] \\
&\geq \frac{|U'|}{|U|}\mathbb{E}_{x \in U'}\left[\mathcal{E}(\mathcal{P}(V(x)|x+V^{(m)})) - \mathcal{E}(\mathcal{P}(V^{(m)}|x+V^{(m)}))\right] \\
&> \epsilon^3.
\end{aligned}$$

Since $0 \leq \mathcal{E}(\mathcal{P}) \leq k$ for all $\mathcal{P}$, we conclude that this procedure halts after $M \leq k\epsilon^{-3}$ iterations. From $\operatorname{codim} V^{(m+1)} \leq \operatorname{codim} V^{(m)} + q^{\operatorname{codim} V^{(m)}}$, we deduce that $V^{(M)}$ satisfies the desired properties and $\operatorname{codim} V^{(M)}$ is bounded by function of $\epsilon$, $q$, $k$, and $\operatorname{codim} V_0$. □

**Theorem 2.5.4** (Strong arithmetic regularity)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, functions $f_1, \ldots, f_k \colon V \to [-1,1]$, a subspace $V_0 \leq V$, and parameters $\delta > 0$ and $\boldsymbol{\epsilon} = (\epsilon_0, \epsilon_1, \ldots)$ satisfying $\epsilon_0 \geq \epsilon_1 \geq \cdots > 0$, there exists $n_{\mathsf{strong-reg}} = n_{\mathsf{strong-reg}}(\delta, \boldsymbol{\epsilon}, q, k, \operatorname{codim} V_0)$ such that there exist subspaces $V_2 \leq V_1 \leq V_0$ such that*

1. $\operatorname{codim} V_2 \leq n_{\mathsf{strong-reg}}$*;*
2. $\mathcal{E}(\mathcal{P}(V_2)) \leq \mathcal{E}(\mathcal{P}(V_1)) + \delta$*;*
3. *for each $1 \leq i \leq k$, for all but an $\epsilon_{\operatorname{codim} V_1}$-fraction of $x \in V$ the function $f_i|_{x+V_2}$ is $\epsilon_{\operatorname{codim} V_1}$-regular.*

*Proof.* We iterate Green's arithmetic regularity lemma to produce a sequence of subspaces $V_0 = V^{(0)} \geq V^{(1)} \geq \cdots \geq V^{(M)}$ such that setting $V_1 = V^{(M-1)}$ and $V_2 = V^{(M)}$ proves the theorem. Let $V^{(m+1)}$ be the result of Theorem 2.5.3 to $V^{(m)}$ with parameter $\epsilon_{\operatorname{codim} V_m}$.

Note that $0 \le \mathcal{E}(\mathcal{P}(V^{(0)})) \le \mathcal{E}(\mathcal{P}(V^{(1)})) \le \cdots \le \mathcal{E}(\mathcal{P}(V^{(M)})) \le k$. Therefore we can stop the iteration at $M \le k\delta^{-1}$ such that $\mathcal{E}(\mathcal{P}(V^{(M)})) - \mathcal{E}(\mathcal{P}(V^{(M-1)})) \le \delta$, as desired. □

*Proof of Proposition 2.3.2.* We can assume that $|V_0| \le (\epsilon/4)|V|$. (If this does not hold, simply replace $V_0$ with a subspace of codimension at most $\lceil \log_q(4\epsilon^{-1}) \rceil$; if this is impossible, then the desired result is trivially true by setting $V_1 = V_2 = \{0\}$.)

Apply Theorem 2.5.4 to $f_1, \dots, f_k \colon V \to [-1, 1]$ and $V_0 \le V$ with parameters $\delta = \epsilon^3/4$ and $\epsilon_m = \min(\epsilon, q^{-m}/(2k))$. Write $V_2 \le V_1 \le V_0$ for the subspaces produced. We will pick a uniform random subspace $U \le V$ subject to the constraint that $U \oplus V_1 = V$. In other words, $U$ consists of one representative element from each $V_1$-coset in such a way the elements of $U$ form a subspace. We claim that with positive probability, this choice of $U$ satisfies the desired properties.

For each $i$, pick $x \in V$ uniformly at random. Theorem 2.5.4(3) implies that with probability at least $1 - \epsilon_{\operatorname{codim} V_1}$, the function $f_i|_{x+V_2}$ is $\epsilon$-regular. By the union bound, with probability at least $1 - \epsilon_{\operatorname{codim} V_1} \cdot k \cdot (|U| - 1) > 1/2$, for every $x \in U \setminus \{0\}$ and every $1 \le i \le k$, the function $f_i|_{x+V_2}$ is $\epsilon$-regular.

By Theorem 2.5.4(2) and then Proposition 2.5.2(2), we have

$$\begin{aligned}
\epsilon^3/4 &\ge \mathcal{E}(\mathcal{P}(V_2)) - \mathcal{E}(\mathcal{P}(V_1)) \\
&= \sum_{i=1}^{k} \left\| (f_i)_{\mathcal{P}(V_2)} - (f_i)_{\mathcal{P}(V_1)} \right\|_{L^2(V)}^2 \\
&= \mathbb{E}_{x \in V} \left[ \sum_{i=1}^{k} \left( \mathbb{E}_{y \in V_2}[f_i(x+y)] - \mathbb{E}_{y \in V_1}[f_i(x+y)] \right)^2 \right].
\end{aligned}$$

This implies that for a uniform random choice of $x \in V$, with probability at least $1 - \epsilon/4$, this $x$ satisfies

$$|\mathbb{E}_{y \in V_2}[f_i(x+y)] - \mathbb{E}_{y \in V_1}[f_i(x+y)]| \le \epsilon \quad \text{for all} \quad 1 \le i \le k. \tag{‡}$$

Therefore choosing $U$ randomly, the expected number of $x \in U$ failing to satisfy (‡) is at most $1 + (\epsilon/4)(|U| - 1) < (\epsilon/2)|U|$. By Markov's inequality, with probability at least $1/2$, at most an $\epsilon$-fraction of $x \in U$ fail to satisfy (‡). Thus by the union bound, $U$ has the desired

properties with positive probability. □

### 2.5.2 Proof 2: An improved bound

We prove a weak arithmetic regularity lemma with exponential bounds and then iterate this lemma to produce a strong arithmetic regularity lemma with tower-type bounds that is just strong enough to prove Proposition 2.3.2.

**Definition 2.5.5.** For a finite dimensional $\mathbb{F}_q$-vector space $V$ and $U \leq V$, a *decomposition of $V$ with respect to $U$* is a set of subspaces $\mathcal{D} = \{W_1, \ldots, W_\ell\}$ all of the same codimension and all transverse to $U$ (i.e., $W_i \cup U$ spans $V$ for each $i$) such that $W_1 \setminus U, \ldots, W_\ell \setminus U$ are disjoint sets that together partition $V \setminus U$. We write $\operatorname{codim} \mathcal{D} := \operatorname{codim} W_1 = \cdots = \operatorname{codim} W_\ell$.

One can easily see that every decomposition $\mathcal{D}$ has size

$$|\mathcal{D}| = \frac{q^{\dim V} - q^{\dim U}}{q^{\dim V - \operatorname{codim} \mathcal{D}} - q^{\dim U - \operatorname{codim} \mathcal{D}}} = q^{\operatorname{codim} \mathcal{D}}.$$

To build intuition, we give several constructions of decompositions of increasing complexity. We start with a very simple example.

**Example 2.5.6.** Take $V = \mathbb{F}_q^2$ and $U = \{(0, x) : x \in \mathbb{F}_q\}$. We define $\mathcal{D}$, a decomposition of $V$ with respect to $U$, to be the set of all 1-dimensional subspaces of $V$ other than $U$ itself. More concretely, $\mathcal{D} = \{W_a\}_{a \in \mathbb{F}_q}$ where $W_a = \{(x, ax) : x \in \mathbb{F}_q\}$.

Next we give a generalization of this example.

**Example 2.5.7.** Take $V = \mathbb{F}_q^{2n}$ and $U = \{0\} \oplus \mathbb{F}_q^n$. We view $V \cong \mathbb{F}_{q^n} \oplus \mathbb{F}_{q^n}$, an isomorphism of $\mathbb{F}_q$-vector spaces. Then we take the same construction as the previous example: $\mathcal{D} = \{W_a\}_{a \in \mathbb{F}_{q^n}}$ where $W_a = \{(x, ax) : x \in \mathbb{F}_{q^n}\}$.

Both these examples are very special: they both have the property that $W \cap U = \{0\}$ for each $W \in \mathcal{D}$. We can make our examples slightly more interesting as follows.

**Example 2.5.8.** Take any $U' \leq U \leq V$ with $\dim V - \dim U = \dim U - \dim U'$. The previous example gives a decomposition $\mathcal{D}$ of $V/U'$ with respect to $U/U'$. Then $\mathcal{D}' = \{W \oplus U'\}_{W \in \mathcal{D}}$ is a decomposition of $V$ with respect to $U$.

The decompositions $\mathcal{D}$ produced by Example 2.5.8 are slightly more interesting, but they satisfy $W \cap U = U'$ for each $W \in \mathcal{D}$. To prove our decomposition regularity lemma, we will need to consider more complicated decompositions.

For two decompositions $\mathcal{D}, \mathcal{D}'$ of $V$ with respect to $U$, say that $\mathcal{D}'$ refines $\mathcal{D}$ (written as $\mathcal{D}' \succeq \mathcal{D}$) if the partition $\bigcup_{W \in \mathcal{D}'}(W \setminus U)$ refines the partition $\bigcup_{W \in \mathcal{D}}(W \setminus U)$.

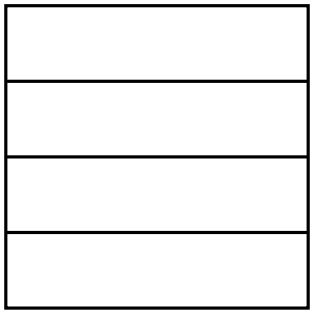
We can create more complicated examples of decompositions of $V$ with respect to $U$ as follows. We construct a sequence of decompositions $\{V\} = \mathcal{D}_0 \preceq \mathcal{D}_1 \preceq \cdots$ iteratively. To get from $\mathcal{D}_m$ to $\mathcal{D}_{m+1}$, we refine each part of $\mathcal{D}_m$ as follows. For each $W \in \mathcal{D}_m$, pick a subspace $U' \leq U \cap W$ satisfying $\dim W - \dim(U \cap W) = \dim(U \cap W) - \dim U'$. Apply Example 2.5.8 with $U' \leq U \leq W$ to find a decomposition of $W$ with respect to $U$. Replacing each $W \in \mathcal{D}_m$ with the subspaces produced in this way gives a new decomposition $\mathcal{D}_{m+1}$.

For a decomposition $\mathcal{D}$ of $V$ with respect to $U$, we associate to it the partition $\mathcal{P}(\mathcal{D})$ whose parts are, for each $W \in \mathcal{D}$, the cosets of $W \cap U$ contained in $W \setminus U$. For functions $f_1, \ldots, f_k \colon V \to [-1, 1]$ and a subspace $U \leq V$, our goal will be to produce a decomposition $\mathcal{D}$ of $V$ with respect to $U$ such that the functions $f_1, \ldots, f_k$ are regular on most parts of the partition $\mathcal{P}(\mathcal{D})$. As in the previous section, we will consider the energy

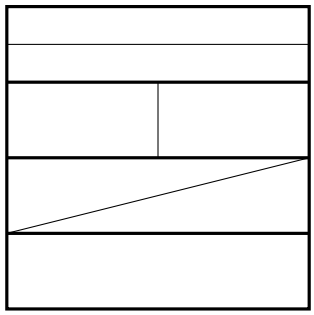
$$\mathcal{E}(\mathcal{P}(\mathcal{D})) = \sum_{i=1}^{k} \left\| (f_i)_{\mathcal{P}(\mathcal{D})} \right\|_{L^2(V \setminus U)}^2 .$$

Before diving into the proof, we give some intuition why using decompositions can help prove a more efficient regularity lemma. Both the standard proof and the decomposition proof start with a partition and, if it is not regular, refine the partition to increase its energy. Consider an initial partition, schematically depicted in Figure 2-1(a). If the partition is not regular, we find a refinement, depicted in Figure 2-1(b), that increases its energy.

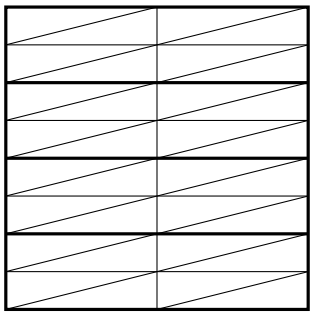
In the standard proof, we require that all parts of the partition must be cosets of the same subspace. The refinement depicted in Figure 2-1(b) does not have this structure, so we must refine the partition further to create the structure. This further refinement is schematically depicted in Figure 2-1(c). This step of the process increases the number of parts of the partition exponentially. The inefficiency in this step is the reason that the standard argument produces wowzer-type bounds instead of tower-type bounds.



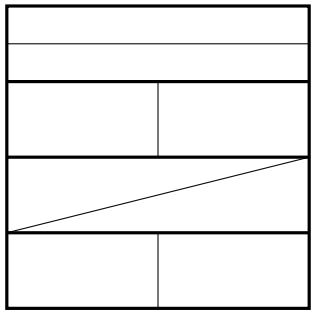
(a) An initial partition.



(b) Refining some parts of the initial partition increases the energy.



(c) Taking the common refinement produces a partition with many parts.



(d) A minimal decomposition which contains (b) has many fewer parts. (The bottom part is partitioned arbitrarily.)

Figure 2-1: An example of refining partitions and decompositions.

Decompositions are a more flexible class of partitions. In the decomposition proof we take the refinement depicted in Figure 2-1(b) and again we further refine the partition so that it is a decomposition. Such a further refinement is schematically depicted in Figure 2-1(d). Because of the flexibility in the definition of decompositions, we can do this step without increasing the number of parts too much. This allows us to prove a more efficient regularity lemma. We now give the details.

**Theorem 2.5.9** (Weak decomposition regularity)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, functions $f_1, \dots, f_k \colon V \to [-1, 1]$, a subspace $U \leq V$, and a parameter $\epsilon > 0$, if* $\dim V \geq \operatorname{codim} U \cdot q^{\operatorname{codim} U} k\epsilon^{-3}$*, then there exists a decomposition $\mathcal{D}$ such that*

1. $\operatorname{codim} \mathcal{D} \leq \operatorname{codim} U \cdot q^{\operatorname{codim} U} k\epsilon^{-3}$*;*

2. *for each $1 \leq i \leq k$, for all but an $\epsilon$-fraction of $W \in \mathcal{D}$, for all $x \in W \setminus U$ the function $f_i|_{x+W\cap U}$ is $\epsilon$-regular.*

*Proof.* We iterate to produce a sequence of decompositions $\{V\} = \mathcal{D}_0 \preceq \mathcal{D}_1 \preceq \cdots \preceq \mathcal{D}_M$ such that $\mathcal{D}_M$ has the desired properties. The sequence will satisfy $\operatorname{codim} \mathcal{D}_{m+1} = \operatorname{codim} \mathcal{D}_m + \operatorname{codim} U$ and $\mathcal{E}(\mathcal{P}(\mathcal{D}_{m+1})) > \mathcal{E}(\mathcal{P}(\mathcal{D}_m)) + \epsilon^3 q^{-\operatorname{codim} U}$.

Suppose that we have produced some $\mathcal{D}_m$ that does not satisfy property (2) above. Then there is some $1 \leq i \leq k$ and $\epsilon|\mathcal{D}_m|$ choices of $W \in \mathcal{D}_m$, say $\mathcal{D}' \subset \mathcal{D}_m$, such that for each $W \in \mathcal{D}'$, there exists $x \in W \setminus U$ such that $f_i|_{x+W\cap U}$ is not $\epsilon$-regular.

Fix $W \in \mathcal{D}'$ with associated $x \in W \setminus U$. Since $f_i|_{x+W\cap U}$ is not $\epsilon$-regular, by part (3) of Proposition 2.5.2, there exists some $W' \leq W \cap U$ of codimension 1 such that $\mathcal{E}(\mathcal{P}(W'|x+W\cap U)) > \mathcal{E}(\mathcal{P}(W\cap U|x+W\cap U))+\epsilon^2$. Now we pick $U' \leq W'$ such that $\dim W - \dim(W\cap U) = \dim(W\cap U) - \dim U'$. Applying Example 2.5.8 to $U' \leq (W\cap U) \leq W$ gives a decomposition $\mathcal{D}_W$ of $W$ with respect to $W\cap U$ such that the partition $\mathcal{P}(\mathcal{D}_W)$ of $W \setminus U$ refines the partition $\mathcal{P}(W'|W \setminus U)$ (the partition of $W \setminus U$ into cosets of $W'$). This implies that

$$\mathcal{E}(\mathcal{P}(\mathcal{D}_W)) > \mathcal{E}(\mathcal{P}(W\cap U|W \setminus U)) + \frac{\epsilon^2}{q^{\operatorname{codim} U} - 1}.$$

Let $\mathcal{D}_{m+1}$ be the decomposition of $V$ with respect to $U$ defined as follows. For each $W \in \mathcal{D}'$, replace $W$ with $\mathcal{D}_W$, the decomposition defined in the previous paragraph. For the remaining $W \in \mathcal{D}_m \setminus \mathcal{D}'$, replace $W$ with an arbitrary instance of Example 2.5.8. This refinement satisfies $\operatorname{codim} \mathcal{D}_{m+1} = \operatorname{codim} \mathcal{D}_m + \operatorname{codim} U$.

Now we compute the energy increment

$$\begin{aligned}
\mathcal{E}(\mathcal{P}(\mathcal{D}_{m+1})) - \mathcal{E}(\mathcal{P}(\mathcal{D}_m)) &= \mathbb{E}_{W\in\mathcal{D}_m}\left[\mathcal{E}(\mathcal{P}(\mathcal{D}_{m+1})|_{W\setminus U}) - \mathcal{E}(\mathcal{P}(\mathcal{D}_m)|_{W\setminus U})\right] \\
&\geq \frac{|\mathcal{D}'|}{|\mathcal{D}_m|}\mathbb{E}_{W\in\mathcal{D}'}\left[\mathcal{E}(\mathcal{P}(\mathcal{D}_{m+1})|_{W\setminus U}) - \mathcal{E}(\mathcal{P}(\mathcal{D}_m)|_{W\setminus U})\right] \\
&\geq \epsilon^3 q^{-\operatorname{codim} U}.
\end{aligned}$$

Since $0 \leq \mathcal{E}(\mathcal{P}(\mathcal{D})) \leq k$ for any decomposition $\mathcal{D}$, this process halts after $M \leq q^{\operatorname{codim} U} k\epsilon^{-3}$ iterations, giving the desired result. □

Given a decomposition $\mathcal{D}$ of $V$ with respect to $U$, define

$$V(\mathcal{D}) := \bigcap_{W \in \mathcal{D}} W.$$

Note that this satisfies $\operatorname{codim} V(\mathcal{D}) \leq |\mathcal{D}| \cdot \operatorname{codim} \mathcal{D} = \operatorname{codim} \mathcal{D} \cdot q^{\operatorname{codim} \mathcal{D}}$. Furthermore, note that if $\mathcal{D}$ is non-trivial, i.e., $\operatorname{codim} \mathcal{D} > 0$, then $V(\mathcal{D}) \leq U$.

**Theorem 2.5.10** (Strong decomposition regularity)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, functions $f_1, \ldots, f_k \colon V \to [-1, 1]$, a subspace $V_0 \leq V$, and a parameter $\epsilon > 0$, there exists $n_{\mathsf{improved-reg}} = n_{\mathsf{improved-reg}}(\epsilon, q, k, \operatorname{codim} V_0)$ such that there exists a subspace $V_1 \leq V_0$ and a decomposition $\mathcal{D}$ of $V$ with respect to $V_1$ such that*

1. $\operatorname{codim} V_1, \operatorname{codim} \mathcal{D} \leq n_{\mathsf{improved-reg}}$;
2. $\mathcal{E}(\mathcal{P}(V_1 | V \setminus V_1)) \leq \mathcal{E}(\mathcal{P}(\mathcal{D})) + \epsilon$;
3. *for each $1 \leq i \leq k$, for all but an $\epsilon$-fraction of $W \in \mathcal{D}$, for all $x \in W \setminus V_1$ the function $f_i|_{x+W \cap V_1}$ is $\epsilon$-regular.*

*Proof.* We can assume that $|V_0| \leq (\epsilon/4)|V|$.

Let $\mathcal{D}_0$ be a decomposition of $V$ with respect to $V_0$ formed by applying Example 2.5.8. We iterate our weak arithmetic regularity lemma to produce a sequence of decompositions $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_M$ such that $V(\mathcal{D}_{M-1}), \mathcal{D}_M$ has the desired properties. Let $V_m = V(\mathcal{D}_m)$ and let $\mathcal{D}_{m+1}$ be the result of applying Theorem 2.5.9 to $V_m$ with parameter $\epsilon$. (Note that we must have $\dim V \geq \operatorname{codim} V_m \cdot q^{\operatorname{codim} V_m} \cdot k \cdot \epsilon^{-3}$ to apply Theorem 2.5.9. However, if this ever fails the desired result is trivially true.)

Note that $V_0 \geq V_1 \geq \cdots \geq V_M$ so $0 \leq \mathcal{E}(\mathcal{P}(V_0)) \leq \mathcal{E}(\mathcal{P}(V_1)) \leq \cdots \leq \mathcal{E}(\mathcal{P}(V_M)) \leq k$. Therefore we can stop the iteration at $M \leq 2k\epsilon^{-1}$ such that $\mathcal{E}(\mathcal{P}(V_M)) \leq \mathcal{E}(\mathcal{P}(V_{M-1})) + \epsilon/2$.

Now since $|V_{M-1}| \leq (\epsilon/4)|V|$, note that $\mathcal{E}(\mathcal{P}(V_M | V \setminus V_{M-1})) \leq \mathcal{E}(\mathcal{P}(V_{M-1} | V \setminus V_{M-1})) + \epsilon$. Furthermore $\mathcal{P}(V_M | V \setminus V_{M-1}) \succeq \mathcal{P}(\mathcal{D}_M) \succeq \mathcal{P}(V_{M-1} | V \setminus V_{M-1})$. This implies property (2), completing the proof. □

*Proof of Proposition 2.3.2.* We can assume that $|V_0| \leq (\epsilon/4)|V|$.

Apply Theorem 2.5.10 to the functions $f_1, \ldots, f_k \colon V \to [-1,1]$ and the subspace $V_0 \leq V$ with parameter $\min(\epsilon^3/4, 1/(4k))$. Write $V_1 \leq V_0$ for the subspace produced and $\mathcal{D}$ for the decomposition produced. For a uniform random choice of $W \in \mathcal{D}$, write $V_2 = V_1 \cap W$ and pick $U$ satisfying $U \oplus V_2 = W$ arbitrarily. We claim that with positive probability, this choice of $U, V_2, V_1$ satisfies the desired properties.

First note that with probability at least $1 - k(1/(4k)) = 3/4$ we have that for all $1 \leq i \leq k$ and for all $x \in W \setminus V_1$ the function $f_i|_{x+V_2}$ is $\epsilon$-regular.

Now by part (2) of Proposition 2.5.2, we write

$$\begin{aligned}
\epsilon^3/4 &\geq \mathcal{E}(\mathcal{P}(\mathcal{D})) - \mathcal{E}(\mathcal{P}(V_1|V \setminus V_1)) \\
&= \sum_{i=1}^{k} \left\| (f_i)_{\mathcal{P}(V_1|V\setminus V_1)} - (f_i)_{\mathcal{P}(\mathcal{D})} \right\|_{L^2(V\setminus V_1)}^2 \\
&= \mathbb{E}_{x \in V \setminus V_1} \left[ \sum_{i=1}^{k} \left( \mathbb{E}_{y \in W(x) \cap V_1}[f_i(x+y)] - \mathbb{E}_{y \in V_1}[f_i(x+y)] \right)^2 \right].
\end{aligned}$$

In the last line above we write $W(x) \in \mathcal{D}$ for the unique part of the decomposition containing $x$.

This implies that for a uniform random choice of $x \in V \setminus V_1$, with probability at least $1 - \epsilon/4$, this $x$ satisfies

$$|\mathbb{E}_{y \in W(x) \cap V_1}[f_i(x+y)] - \mathbb{E}_{y \in V_1}[f_i(x+y)]| \leq \epsilon \quad \text{for all} \quad 1 \leq i \leq k. \tag{§}$$

Therefore choosing $W \in \mathcal{D}$ uniformly at random, the expected number of $x \in W$ failing to satisfy (§) is at most $|W \cap V_1| + (\epsilon/4)(|W| - |W \cap V_1|) \leq (\epsilon/2)|W|$. By Markov's inequality, with probability at least $1/2$, at most an $\epsilon$-fraction of $x \in W$ fail to satisfy (§). Thus by the union bound, with positive probability this choice of $U, V_2, V_1$ satisfies all the desired properties. $\square$

## 2.6 Extensions

### 2.6.1 Infinite removal

First let us recall the infinite graph removal lemma [AS08].

**Theorem 2.6.1.** *For every (possibly infinite) set of graphs $\mathcal{H}$ and $\epsilon > 0$, there are $h_0$ and $\delta > 0$ such that the following holds. If a graph $G$ has induced $H$-density at most $\delta$ for each $H \in \mathcal{H}$ on at most $h_0$ vertices, then $G$ can be made induced $\mathcal{H}$-free by changing at most an $\epsilon$-fraction of the edges.*

This result is proved by a short modification of the usual proof of induced graph removal. By performing an analogous modification to the proof of induced arithmetic removal it is possible to deduce the following infinite arithmetic removal lemma.

**Theorem 2.6.2.** *Fix $\mathcal{H}$ a (possibly infinite) set of $r$-colored complexity 1 patterns over $\mathbb{F}_q$. For every $\epsilon > 0$ there are $k_0$ and $\delta > 0$ such that the following holds. Given a finite dimensional $\mathbb{F}_q$-vector space $V$, if a coloring $\phi\colon V \to [r]$ has $H$-density at most $\delta$ for each $H \in \mathcal{H}$ with at most $k_0$ columns, then $\phi$ can be made $\mathcal{H}$-free by changing the color of at most an $\epsilon$-fraction of the elements.*

We now give a sketch of the proof of this infinite arithmetic removal lemma. The proof uses the same recoloring algorithm as in the proof of Theorem 2.1.15 with one modification. We replace the regularity recoloring result (Proposition 2.3.3) with the following strong variant.

**Proposition 2.6.3** (Strong variant of Proposition 2.3.3)**.** *For a finite dimensional $\mathbb{F}_q$-vector space $V$, a coloring $\phi\colon V \to [r]$, and parameters $0 < \epsilon \leq 1$ and $\boldsymbol{\epsilon}' = (\epsilon'_0, \epsilon'_1, \ldots)$ satisfying $\epsilon'_0 \geq \epsilon'_1 \geq \cdots > 0$, there is some $n_{\mathsf{reg}} = n_{\mathsf{reg}}(\epsilon, \boldsymbol{\epsilon}', q, r)$ such that there exist subspaces $V_2 \leq V_1 \leq V$, a complement $U$ satisfying $U \oplus V_1 = V$, and a recoloring $\phi'\colon V \to [r]$ that agrees with $\phi$ on all but at most $\epsilon|V|$ values satisfying:*

1. *$1/\epsilon \leq \operatorname{codim} V_1 \leq \operatorname{codim} V_2 \leq n_{\mathsf{reg}}$;*
2. *if a color appears in some coset $x + V_1$ under $\phi'$, then at least an $\epsilon/(2r)$-fraction of $x + V_2$ is that color under $\phi$;*

*3. for each $x \in U \setminus \{0\}$, the original coloring $\phi|_{x+V_2}$ is $\epsilon'_{\operatorname{codim} V_1}$-regular.*

The only difference between Proposition 2.3.3 and Proposition 2.6.3 is that the degree of regularity in part (3) is now allowed to depend on $\operatorname{codim} V_1$. The proof strategy given in Section 2.5 also proves this strong variant with an appropriate choice of parameters.

Now suppose we perform our recoloring procedure on an $r$-coloring $\phi\colon V \to [r]$. The output is a recoloring $\phi'\colon V \setminus \{0\} \to [r]$. Suppose that along the way the application of regularity produces $V_2 \leq V_1 \leq V$ where $\operatorname{codim} V_1 = d$. We now define a "summary" of the recoloring $\phi \rightsquigarrow \phi'$ as follows. Let $\Phi\colon \mathbb{F}_q^d \setminus \{0\} \to 2^{[r]} \setminus \{\emptyset\}$ be the map which sends a coset $x + V_1$ to the set of colors which appear under $\phi'$ in $x + V_1$. The recoloring $\phi'|_{V_1}$ is a canonical coloring and thus is determined by a function $\chi\colon \mathbb{F}_q \setminus \{0\} \to [r]$. We will call $(\Phi, \chi)$ the summary of the recoloring procedure $\phi \rightsquigarrow \phi'$. Note that the number of possible summaries is bounded in terms of $d$ (and $q, r$).

**Definition 2.6.4.** Let $H = (A, \psi)$ be an $r$-colored pattern over $\mathbb{F}_q$ where $A$ has dimensions $\ell \times k$. Given $\Phi\colon \mathbb{F}_q^d \setminus \{0\} \to 2^{[r]} \setminus \{\emptyset\}$ and $\chi\colon \mathbb{F}_q \setminus \{0\} \to [r]$, say that $(\Phi, \chi)$ *partially induces an $H$-instance* if there exists $\boldsymbol{x} \in (\mathbb{F}_q^d)^k$ such that:

1. $A\boldsymbol{x} = \mathbf{0}$;
2. for each $i \in [k]$ such that $x_i \neq 0$, it holds that $\psi(i) \in \Phi(x_i)$;
3. setting $I \subset [k]$ to be the set of $i \in [k]$ such that $x_i = 0$, for some $H' = (A', \psi|_I)$, a subpattern of $H$ restricted to the variables $I$, the $\chi$-canonical coloring of $\mathbb{F}_q^n \setminus \{0\}$ contains an $H'$-instance for all large enough $n$.

This definition is chosen so that if $\phi'$ contains an $H$-instance, then the summary $(\Phi, \chi)$ partially induces an $H$-instance. A careful reading of the proof of the main theorem shows that it proves the following:

**Proposition 2.6.5.** *Let $H$ be an $r$-colored complexity 1 pattern and fix $\epsilon > 0$. There exist $\epsilon_{\mathsf{strong}} = \epsilon_{\mathsf{strong}}(\epsilon, H) > 0$ and $\delta = \delta(\epsilon, H) > 0$ such that the following holds. Suppose the recoloring procedure of Proposition 2.6.3 applied to $\phi$ with parameters $\epsilon$ and $\boldsymbol{\epsilon}' = (\epsilon_{\mathsf{strong}}, \epsilon_{\mathsf{strong}}, \ldots)$ produces $\phi'$. Let $(\Phi, \chi)$ be the summary of this recoloring. If $(\Phi, \chi)$ partially induces an $H$-instance, then the $H$-density in $\phi$ is more than $\delta$.*

Now we give the idea of the proof of Theorem 2.6.2. Fix $\mathcal{H}$, a possibly infinite set of $r$-colored complexity 1 patterns over $\mathbb{F}_q$. We define finite subsets $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \cdots \subseteq \mathcal{H}$ as follows. Fix some $d$. Consider the finite set of all possible pairs $\Phi\colon \mathbb{F}_q^d \setminus \{0\} \to 2^{[r]} \setminus \{\emptyset\}$ and $\chi\colon \mathbb{F}_q \setminus \{0\} \to [r]$. For each $(\Phi, \chi)$, if there exists any $H \in \mathcal{H}$ such that $(\Phi, \chi)$ contains an $H$-instance add one such $H$ to $\mathcal{H}_d$. This ensures that $\mathcal{H}_d$ is finite.

Given $\epsilon > 0$, we set

$$\epsilon'_d = \min_{H \in \mathcal{H}_d} \epsilon_{\mathsf{strong}}(\epsilon, H) \quad \text{and} \quad \delta_d = \min_{H \in \mathcal{H}_d} \delta(\epsilon, H).$$

We apply Proposition 2.6.3 with $\boldsymbol{\epsilon}' = (\epsilon'_0, \epsilon'_1, \ldots)$; all other parts of the recoloring produces remain the same. At the end we obtain a coloring that we claim is $\mathcal{H}$-free. However, by the definitions we made in the previous paragraph, it suffices to check that it is $\mathcal{H}_{\operatorname{codim} V_1}$-free. Taking $\delta = \delta_{n_{\mathsf{reg}}(\epsilon, \boldsymbol{\epsilon}', q, r)}$, Proposition 2.6.5 implies that if the recoloring is not $\mathcal{H}_{\operatorname{codim} V_1}$-free, then it has $H$-density more than $\delta_{\operatorname{codim} V_1} \geq \delta$ for some $H \in \mathcal{H}_{\operatorname{codim} V_1} \subseteq \mathcal{H}$. Finally taking $k_0$ to be the maximum number of columns in any of the patterns of $\mathcal{H}_{n_{\mathsf{reg}}(\epsilon, \boldsymbol{\epsilon}', q, r)}$ completes the argument.

### 2.6.2 Inhomogeneous patterns

**Definition 2.6.6.** Let $H = (A, \psi)$ be an $r$-colored pattern over $\mathbb{F}_q$ where $A$ is an $\ell \times k$ matrix. For a finite dimensional $\mathbb{F}_q$-vector space $V$, a coloring $\phi\colon V \to [r]$, and an *offset* $\boldsymbol{b} \in V^\ell$, say that $\boldsymbol{x} \in V^k$ is an $(H, \boldsymbol{b})$-*instance* if $A\boldsymbol{x} = \boldsymbol{b}$ and $\phi(x_i) = \psi(i)$.

Our main theorem can be generalized to remove instances of inhomogeneous patterns. As before, there are some issues relating to the zero vector and these are magnified by the presence of offsets. The simplest statement, presented below, ignores all instances that contain vectors in the span of the offsets.

**Theorem 2.6.7.** *Fix a finite set $\mathcal{H}$ of $r$-colored complexity 1 patterns over $\mathbb{F}_q$. Suppose that each $H \in \mathcal{H}$ is defined by an $\ell_H \times k_H$ matrix. For every $\epsilon > 0$ there exists $\delta = \delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. Let $V$ be a finite dimensional $\mathbb{F}_q$-vector space $V$, let $\phi\colon V \to [r]$ be a coloring, and let $(\boldsymbol{b}_H)_{H \in \mathcal{H}}$ be offsets (each $\boldsymbol{b}_H$ is a tuple of $\ell_H$ elements of $V$). Suppose*

*that $\phi$ has $(H, \boldsymbol{b}_H)$-density at most $\delta$ for every $H \in \mathcal{H}$. Define the subspace $B$ to be the span of $(\boldsymbol{b}_H)_i$ over all choices of $H \in \mathcal{H}$ and $i \in [\ell_H]$. Then there exists a recoloring $\phi' \colon V \setminus B \to [r]$ that is $(H, \boldsymbol{b}_H)$-free for all $H \in \mathcal{H}$ and differs from $\phi$ on at most $\epsilon|V|$ elements.*

We prove this by reduction to the main theorem. We have a subspace $B \leq V$ where $B$ is of bounded size. Pick a complement $\tilde{V}$ satisfying $B \oplus \tilde{V} = V$. Now we convert $\phi \colon V \to [r]$ to $\tilde{\phi} \colon \tilde{V} \to [r]^B$ in the obvious manner: $\tilde{\phi}(x) := (\phi(x+u))_{u \in B}$. One can check that given a pattern $H = (A, \psi)$ where $A$ is an $\ell \times k$ matrix and an offset $\boldsymbol{b} \in B^\ell$ we can create a set of $N = |B|^{k - \operatorname{rank} A} r^{k(|B|-1)}$ patterns $H_1, \ldots, H_N$ so that every $(H, \boldsymbol{b})$-instance in $\phi$ becomes an $H_i$-instance in $\tilde{\phi}$ for some $1 \leq i \leq N$. Then applying the main theorem to this finite set of homogeneous patterns, we deduce the desired result.

# Chapter 3

# Popular differences for matrix patterns

Using an argument of Varnavides [Var59], it is well-known that Roth's theorem [Rot53] on three-term arithmetic progressions can be strengthened to guarantee at least $c_\alpha N^2$ arithmetic progressions in a set $A \subseteq [N]$ of size $\alpha N$. The constant $c_\alpha$ is known not to be polynomial in $\alpha$; in particular, modifying a well-known construction of Behrend [Beh46] allows one to construct sets with $\alpha^{c \log(1/\alpha)} N^2$ three-term arithmetic progressions. However, Green [Gre05b], showed that one has a "popular" common difference $d \neq 0$, i.e., a value $d \in [N]$ such that

$$\#\{a : a, a+d, a+2d \in A\} \geq (\alpha^3 - o(1))N.$$

That is, the set behaves like a random set along certain structured differences, if not all of them. Green's proof involves an arithmetic regularity lemma, which is essentially equivalent to arithmetic regularity for the Gowers $U^2$-norm.

One can ask if this phenomenon holds for longer arithmetic progressions. The analogous result for four-term arithmetic progressions with $\alpha^4 - o(1)$ on the right-hand side was proved by Green and Tao [GT10a] relying on a remarkable "positivity" identity (see [Gre07] for a version over $\mathbb{F}_p^n$ for $p \geq 5$) in combination with the $U^3$-arithmetic regularity results of Green and Tao [GT10a]. However, surprisingly, an $\alpha^k - o(1)$ (or any polynomial) bound does not hold for $k$-term arithmetic progressions for $k \geq 5$ due to a construction of Ruzsa [BHK05, Appendix]. These results were motivated by corresponding ergodic results of Bergelson, Host, and Kra [BHK05], although the theorems do not directly transfer when studying popular

differences (as opposed to Furstenberg's correspondence theorem for Szemerédi's theorem).

One may ask about popularity of more general patterns, for example $\{0, 1, 2, 4\}$. (We use a set to refer to the pattern consisting of homothetic copies of that set; in this case, the pattern is $(a, a + d, a + 2d, a + 4d)$.) The proof of Green and Tao [GT10a] for four-term arithmetic progressions (and [Gre07] over finite fields) immediately extends to patterns of the form $\{0, k_1, k_2, k_1 + k_2\}$ for $k_1k_2(k_1 + k_2) \neq 0$. Work of Sah, Sawhney, and Zhao [SSZ21] shows that two-point, three-point, and these specific "parallelogram" four-point patterns are the only popular patterns over $\mathbb{Z}$.

Popularity of higher-dimensional patterns such as corners, $\{(0, 0), (1, 0), (0, 1)\}$, was first studied by Mandache [Man21a] in the combinatorial setting (see [Chu11, DS16] for related work in the ergodic theory setting), who showed over $\mathbb{F}_p^n$ that they are not $\alpha^3$-popular but do satisfy a weakened bound with $\alpha^4$ instead. Fox, Sah, Sawhney, Stoner, and Zhao [FSS+20] showed that the tight bound is of the form $\alpha^4\tau(\alpha)$, where $\tau$ grows as $\alpha \to 0$, but is of the form $\alpha^{o(1)}$. Finally, Berger [Ber21] showed the same behavior over $\mathbb{Z}^2$. Sah, Sawhney, and Zhao [SSZ21] studied higher-dimensional patterns which are homothetic copies of a set and provide a nearly comprehensive classification.

## Our contributions

The standard toolset of arithmetic regularity in higher-order Fourier analysis, which can prove popular difference results for three and four-point single-dimensional patterns, necessarily breaks to some extent when handling higher-dimensional corners (as pointed out in [SSZ21]), and has not yet been successfully applied to four-point patterns such as squares, $\{(0, 0), (1, 0), (0, 1), (1, 1)\}$, for which the question remains open. However, it was noted by Prendiville [Pre15] that classic single-dimensional techniques extend if one considers *full-rank* matrix patterns (a collection which excludes corners and squares but includes a wide class of multidimensional configurations such as "rotated corners" – also known as "right isosceles triangles" – and "rotated squares"), and he achieves versions of Szemerédi's theorem (for $k \leq 4$ points) with good quantitative bounds in this setting. We continue in this line of work, achieving popular difference results of strength equal to the single-dimensional case, illustrating by comparison the suitability of these methods for studying full-rank matrix

patterns.

The main novelty of this chapter lies in the popular difference results for four-point patterns, where we exhibit further behavior that does not appear even in Prendiville's work. In order to properly handle popularity of four-point patterns, we show that one must apply the method of arithmetic regularity in a manner that sees the spectral properties of the matrices defining the pattern. In particular, the counting lemma (which for four-point patterns over $\mathbb{F}_p^n$ relies heavily on equidistribution over parts in quadratic factors) becomes qualitatively distinct depending on the spectral structure of the matrices in the pattern (see Theorem 3.4.4). This subtlety is not present in earlier counting lemmas for scalar-valued patterns. This also translates concretely to an additional restriction that no pair of eigenvalues of an associated matrix can be negatives of each other for our method to produce a popular difference result (see Theorem 3.1.2). To confirm that this behavior is genuine and not an artifact of the proof, in Theorem 3.1.3 we exhibit a full-rank matrix pattern which does not satisfy the additional spectral condition imposed by Theorem 3.1.2 and for which the conclusion of the theorem is false. In particular, we show that rotated squares in $\mathbb{F}_5^n$ do not satisfy a popular difference result, at least with popularity $\alpha^4$.

This chapter is based on the paper [BSST22].

## 3.1 Summary of results

Our first result is a popular differences result for all full-rank three-point patterns. A three-point pattern is full rank if it can be expressed in the form $\vec{x}, \vec{x} + M_1\vec{d}, \vec{x} + M_2\vec{d}$ where $M_1, M_2, M_1 - M_2$ are invertible. One such example is "rotated corners," which are of the form $\{(x, y), (x+a, y+b), (x+b, y-a)\}$. (By contrast, standard corners $\{(x, y), (x+a, y), (x, y+a)\}$ are not full rank.) As a special case, this resolves a conjecture of Ackelsberg, Bergelson, and Best [ABB21, Question 1.21], which concerns the case of rotated corners specifically. Kovač [Kov21] has independently proved this rotated corners conjecture with similar methods.

**Theorem 3.1.1.** *Let $M_1, M_2$ be $k \times k$ invertible integer matrices so that $M_1 - M_2$[1] is invertible.*

[1]The condition $M_1 - M_2$ is invertible is necessary for the theorem as stated to be nontrivial for a single common difference as otherwise one can simply take $\vec{d}$ to be in the kernel of $M_1 - M_2$. Restricting to differences

*For any $\alpha, \epsilon > 0$ there exists $N_0(\alpha, \epsilon, M_1, M_2)$ so that the following holds. If $N \geq N_0$, then for any $A \subseteq [N]^k$, $|A| \geq \alpha N^k$, there is a popular difference $\vec{d} \neq 0$ so that*

$$\#\{\vec{x} \in [N]^k : \vec{x}, \vec{x} + M_1\vec{d}, \vec{x} + M_2\vec{d} \in A\} \geq (\alpha^3 - \epsilon)N^k.$$

We additionally prove an analogous version of the result where the interval $[N]$ is replaced by an arbitrary compact abelian group $G$. See Section 3.7 for the precise statement and proof of this result.

We turn next to four-point patterns of matrices. There are a few natural restrictions on generic patterns $\vec{x}, \vec{x} + M_1\vec{d}, \vec{x} + M_2\vec{d}, \vec{x} + M_3\vec{d}$ that arise when trying to prove a popular differences result. First, we impose $M_3 = M_1 + M_2$, which is a generalization of the "parallelogram" condition in the popular differences result of Green and Tao [GT10a]. Second, we require that $M_1, M_2, M_1 - M_2, M_1 + M_2$ are all invertible; in this case we call the pattern full rank.[2] The combination of these two conditions is analogous to the "admissibility" condition of [ABB21], and essentially appears in [Pre15]. One might guess that they are sufficient to guarantee popular differences. In this chapter we show that this guess is incorrect by demonstrating the necessity of an additional spectral condition on the pattern. In the spirit of the finite field philosophy advocated by Green [Gre05a], we restrict attention to the finite field model $G = \mathbb{F}_p^n$ with $p$ an odd prime. We suspect our methods can be extended to handle more general abelian groups, but choose to avoid the complexity of the inverse theorems for the $U^3$-norm over general finite abelian groups.

**Theorem 3.1.2.** *Fix $k \geq 1$ and $p$ an odd prime. Let $M_1, M_2$ be $k \times k$ matrices with coefficients in $\mathbb{F}_p$ such that $M_1$, $M_2$, $M_1 - M_2$, and $M_1 + M_2$ are invertible and no pair of eigenvalues of $M_1 M_2^{-1}$ (viewed over the algebraic closure $\overline{\mathbb{F}}_p$) are negatives of each other. For $\alpha, \epsilon > 0$, there exists $n_0(\alpha, \epsilon, p)$ such that the following holds. If $n \geq n_0$, then for any $A \subseteq (\mathbb{F}_p^n)^k$, $|A| \geq \alpha p^{nk}$, there is a popular difference $\vec{d} \neq 0$ so that*

$$\#\{\vec{x} \in (\mathbb{F}_p^n)^k : \vec{x}, \vec{x} + M_1\vec{d}, \vec{x} + M_2\vec{d}, \vec{x} + (M_1 + M_2)\vec{d} \in A\} \geq (\alpha^4 - \epsilon)p^{nk}.$$

outside this kernel can likely be done (see e.g. [Pre15, Theorem 1.1]) but requires technical modifications which take us too far afield.

[2] Axis-aligned squares are an example of a four-point pattern that is not full rank, and for which the version of popular differences we would like to prove is known to be false; see [SSZ21, Theorem 3.1].

*In fact, there are $\Omega_{\alpha,\epsilon,p}(p^{nk})$ values of $\vec{d}$ that work.*

Furthermore, we show that one cannot completely remove the spectral condition.

**Theorem 3.1.3.** *There is an absolute constant $c > 0$ such that the following holds. If $\alpha \in (0, c)$, then for all sufficiently large $n$ (depending on $\alpha$) there is a set $A \subseteq (\mathbb{F}_5^n)^2$ satisfying $|A| \geq \alpha 5^{2n}$ and*

$$\max_{(a,b)\neq 0} \#\{(x,y) : (x,y), (x+a, y+b), (x+b, y-a), (x+a+b, y+b-a) \in A\} \leq (1-c)\alpha^4 5^{2n}.$$

Here the associated matrices are

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Note that the eigenvalues of $M_1 M_2^{-1}$ indeed are negatives of each other. We believe it is likely that one can construct a counterexample for all $k \times k$ matrices with some pair of negated eigenvalues by lifting the ideas involved in this construction.

Although there is no direct implication, this can be seen as a combinatorial finite field analogue of [ABB21, Question 1.11] and we expect our counterexample can be extended to the ergodic setting. In particular, we answer the combinatorial analogue in the negative but point to a potential new condition under which their question might be resolved positively.

**Notation:** We use $O, o, \Omega$ as standard asymptotic notation. Subscripts in said notation denote dependence of the implicit constants on those subscripts.

**Structure:** The majority of this chapter, Sections 3.2, 3.3, 3.4 and 3.5, is devoted to the proof of Theorem 3.1.2, the popular difference result for four-point patterns. See Section 3.2 for an outline of that argument. In Section 3.6, we construct the counterexample that proves Theorem 3.1.3. Finally, in Section 3.7, we show Theorem 3.1.1, the three-point pattern result.

## 3.2 Gowers norms and arithmetic regularity

The proof of Theorem 3.1.2 proceeds in three steps, following the now-standard framework of the arithmetic regularity method.

First, we show that the matrix patterns we are interested in are controlled by an appropriate Gowers $U^s$-norm. Results of this nature are sometimes referred to as "generalized von Neumann theorems". The definition of the Gowers norms and the proof of this result are given in this section.

Second, we prove an arithmetic regularity lemma, which gives a decomposition of an arbitrary function $f\colon G \to \mathbb{C}$ as $f = f_{\text{str}} + f_{\text{sml}} + f_{\text{psr}}$ into a "structured", "small", and "pseudorandom" piece. For our application in groups $G^k$, it will be necessary to carefully define "structured" in a way that is adapted to the product structure of $G^k$. This definition and the proof of this result is given in Section 3.3.

Third, we prove novel equidistribution results in order to understand the counts of matrix patterns inside the structured piece, $f_{\text{str}}$. These results occur in Section 3.4. Combining these three steps, we prove Theorem 3.1.2 in Section 3.5.

**Definition 3.2.1.** Fix an integer $s \geq 1$ and a finite abelian group $G$. For a function $f\colon G \to \mathbb{C}$, the *Gowers $U^s$-norm* is defined by

$$\|f\|_{U^s(G)} = \left( \mathbb{E}_{x,h_1,\dots,h_s \in G} \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f(x + \omega_1 h_1 + \cdots + \omega_s h_s) \right)^{1/2^s},$$

where $\mathcal{C}$ denotes the complex conjugation operator and $|\omega| = \omega_1 + \cdots + \omega_s$.

It is well-known that the above is indeed a norm when $s \geq 2$. (For $s = 1$ it is the seminorm $f \mapsto |\mathbb{E}_{x \in G} f(x)|$, so the term "Gowers norm" is a slight misnomer.) A useful equivalent definition is that

$$\|f\|_{U^s(G)}^{2^s} = \mathbb{E}_{h \in G} \|\partial_h f\|_{U^{s-1}(G)}^{2^{s-1}}$$

where the *multiplicative derivative* $\partial_h f$ is defined by $(\partial_h f)(x) = f(x)\overline{f(x+h)}$.

We now prove that full-rank matrix patterns are controlled by an appropriate $U^s$-norm. The typical setup in this chapter is to consider a pattern of the form $\vec{x}+M_1\vec{d}, \vec{x}+M_2\vec{d}, \dots, \vec{x}+M_s\vec{d}$ in $G^k$ where $M_1, \dots, M_s$ are $k \times k$ matrices with certain non-degeneracy conditions. In particular we assume that $M_i$ and $M_i - M_j$ are invertible for each $i \neq j$.

This lemma is true even in the general setting where we replace the matrix $M_i$ acting on $G^k$ by an arbitrary autmorphism $A_i$ acting on $G^k$. In this general setting, the product structure on $G^k$ is no longer important. As the proof of the more general version is no more difficult than the original result, we include it here. The proof follows by an application of the Cauchy–Schwarz inequality; similar results for specific patterns are implicit in the literature (e.g., [Pre15]).

**Lemma 3.2.2.** *Let $s \geq 2$, and $G$ be a finite abelian group. Let $A_1, \dots, A_s$ be automorphisms of $G$ such that $A_i - A_j$ is an automorphism for each $i \neq j$. Then for functions $f_i \colon G \to \mathbb{C}$ satisfying $\|f_i\|_\infty \leq 1$ we have*

$$|\mathbb{E}_{x,d\in G} f_1(x + A_1 d) \cdots f_s(x + A_s d)| \leq \min_{i\in[s]} \|f_i\|_{U^{s-1}(G)}.$$

*Proof.* We induct on $s$. For $s = 2$, note that

$$\begin{aligned} |\mathbb{E}_{x,d\in G} f_1(x + A_1 d) f_2(x + A_2 d)| &= |\mathbb{E}_{x,d\in G} f_1(x) f_2(x + (A_2 - A_1)d)| \\ &= |\mathbb{E}_{x,y\in G} f_1(x) f_2(y)| \\ &= \|f_1\|_{U^1(G)} \|f_2\|_{U^1(G)}. \end{aligned}$$

Since $\|f_i\|_\infty \leq 1$, the result follows in this case. Now suppose $s \geq 3$. We have

$$\begin{aligned} &|\mathbb{E}_{x,d\in G} f_1(x + A_1 d) \cdots f_s(x + A_s d)| \\ &\quad = |\mathbb{E}_{x,d\in G} f_1(x + (A_1 - A_s)d) \cdots f_{s-1}(x + (A_{s-1} - A_s)d) f_s(x)| \\ &\quad \leq \mathbb{E}_x |\mathbb{E}_d f_1(x + (A_1 - A_s)d) \cdots f_{s-1}(x + (A_{s-1} - A_s)d)| \\ &\quad \leq \big(\mathbb{E}_x |\mathbb{E}_d f_1(x + (A_1 - A_s)d) \cdots f_{s-1}(x + (A_{s-1} - A_s)d)|^2\big)^{1/2} \\ &\quad = \big(\mathbb{E}_x \mathbb{E}_{d,h} f_1(x + (A_1 - A_s)d) \cdots f_{s-1}(x + (A_{s-1} - A_s)d) \\ &\qquad \cdot \overline{f_1}(x + (A_1 - A_s)d + (A_1 - A_s)h) \cdots \overline{f_{s-1}}(x + (A_{s-1} - A_s)d + (A_{s-1} - A_s)h)\big)^{1/2}. \end{aligned}$$

To bound the last expression, we apply the induction hypothesis with the maps $A_1 - A_s, \dots, A_{s-1} - A_s$ and the functions $\partial_{(A_i - A_s)h} f_i$. Note that by hypothesis, the maps $A_i - A_s$ are automorphisms as are $(A_i - A_s) - (A_j - A_s)$ for $i \neq j$. Therefore we obtain

$$\begin{aligned}
|\mathbb{E}_{x,d \in G} f_1(x + A_1 d) \cdots f_s(x + A_s d)| &\leq \left(\mathbb{E}_h \|\partial_{(A_1 - A_s)h} f_1\|_{U^{s-2}(G)}\right)^{1/2} \\
&\leq \left(\mathbb{E}_h \|\partial_{(A_1 - A_s)h} f_1\|_{U^{s-2}(G)}^{2^{s-2}}\right)^{1/2^{s-1}} \\
&= \|f_1\|_{U^{s-1}(G)}.
\end{aligned}$$

The last equality comes from the recursive definition of the Gowers norms as well as the fact that $A_1 - A_s$ is an automorphism on $G$. By symmetry, the same holds for $f_2, \dots, f_s$, completing the proof. □

## 3.3 The $U^3$-arithmetic regularity lemma

From now until Section 3.7, we restrict our attention to the case where $G = \mathbb{F}_p^n$ and $p$ is an odd prime. The goal of this section is to prove a $U^3$-arithmetic regularity lemma for functions $f \colon G^k \to \mathbb{C}$. Since $G^k \cong \mathbb{F}_p^{nk}$, we could apply a standard result (say [Gre07, Proposition 3.12]) to deduce some $U^3$-regularity statement. However such regularity statement would ignore the product structure on $G^k$ which will become very important in our application.

The main novelty of this section is our definition of a $k$-symmetrized quadratic factor which gives an appropriate notion of structured function adapted to the product structure of $G^k$. We then prove Theorem 3.3.2, our $k$-symmetrized $U^3$-arithmetic regularity lemma. The structure of the proof closely follows [Gre07].

An element $\vec{x} \in G^k$ is a tuple $\vec{x} = (x_1, \dots, x_k)$ with $x_1, \dots, x_k \in \mathbb{F}_p^n$. It will simplify the following arguments to introduce the following slightly awkward notation: we view the elements of $G^k$ as $k \times n$ matrices $X$ where the rows of $X$ correspond to the elements of the

$k$-tuple. In particular, the element $\vec{x} \in G^k$ is alternatively represented as

$$X = \begin{pmatrix} \text{———} x_1^\intercal \text{———} \\ \vdots \\ \text{———} x_k^\intercal \text{———} \end{pmatrix},$$

where we are viewing tuples as column vectors by default.

Finally, define $\mathcal{S}_k$ (respectively, $\mathcal{S}'_k$) to be the set of symmetric (respectively, skew-symmetric) matrices in $\mathbb{F}_p^{k \times k}$.

**Definition 3.3.1.** A ($k$-)*symmetrized quadratic factor* $\mathfrak{B} = (\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ is given by a list $\mathfrak{B}_1 = (r_1, \dots, r_{d_1})$ of column vectors in $\mathbb{F}_p^n$, a list $\mathfrak{B}_2 = (M_1, \dots, M_{d_2})$ of symmetric matrices in $\mathbb{F}_p^{n \times n}$, and a list $\mathfrak{B}_3 = (N_1, \dots, N_{d_3})$ of skew-symmetric matrices in $\mathbb{F}_p^{n \times n}$. The *complexity* of $\mathfrak{B}$ is $(d_1, d_2, d_3)$. We say that $\mathfrak{B}$ has *rank* at least $r$ if $r_1, \dots, r_{d_1}$ are linearly independent and all nontrivial linear combinations

$$\sum_{i=1}^{d_2} a_i M_i + \sum_{j=1}^{d_3} b_j N_j$$

have $\mathbb{F}_p$-rank at least $r$. (This is equivalent to the same condition on $M_1, \dots, M_{d_2}$ and $N_1, \dots, N_{d_3}$ separately up to an absolute multiplicative constant in the rank, since rank is sub-additive.)

A $k$-symmetrized quadratic factor $\mathfrak{B}$ defines maps $\mathsf{B}_{1,i} \colon G^k \to \mathbb{F}_p^k$, $\mathsf{B}_{2,i} \colon G^k \to \mathcal{S}_k$, and $\mathsf{B}_{3,i} \colon G^k \to \mathcal{S}'_k$ given by

$$\mathsf{B}_{1,i}(X) = Xr_i, \quad \mathsf{B}_{2,i}(X) = XM_iX^\intercal, \quad \mathsf{B}_{3,i}(X) = XN_iX^\intercal.$$

We additionally define

$$\mathsf{B}_1(X) = (\mathsf{B}_{1,i}(X))_{i \in [d_1]}, \quad \mathsf{B}_2(X) = (\mathsf{B}_{2,i}(X))_{i \in [d_2]}, \quad \mathsf{B}_3(X) = (\mathsf{B}_{3,i}(X))_{i \in [d_3]},$$

$$\mathsf{B}(X) = (\mathsf{B}_1(X), \mathsf{B}_2(X), \mathsf{B}_3(X)).$$

For a function $f \colon G^k \to \mathbb{C}$, we use the notation $\mathbb{E}[f|\mathfrak{B}]$ to represent the condition

expectation of $f$ with respect to $\mathfrak{B}$, or equivalently the projection of $f$ onto $\mathfrak{B}$. Here we abuse notation and use $\mathfrak{B}$ to denote the $\sigma$-algebra generated by the fibers of $\mathsf{B}$ in $G^k$. Explicitly, $\mathbb{E}[f|\mathfrak{B}]\colon G^k \to \mathbb{C}$ is defined by $\mathbb{E}[f|\mathfrak{B}](X) = \mathbb{E}_{Y \in \mathsf{B}^{-1}(\mathsf{B}(X))}[f(Y)]$.

Finally, we say that a factor $\mathfrak{B}'$ *refines* a factor $\mathfrak{B}$ if the $\sigma$-algebra corresponding to $\mathfrak{B}'$ refines the $\sigma$-algebra corresponding to $\mathfrak{B}$.

The main result of this section is the following arithmetic regularity statement which guarantees that the desired factor is $k$-symmetrized.

**Theorem 3.3.2** (Arithmetic regularity lemma)**.** *Fix $k \geq 1$. Let $\delta > 0$ and let $\omega_1, \omega_2\colon \mathbb{R}^+ \to \mathbb{R}^+$ be arbitrary growth functions (which may depend on $\delta$). Let $G = \mathbb{F}_p^n$, let $f\colon G^k \to [0,1]$ be a function, and let $(\mathfrak{B}_1^{(0)}, \mathfrak{B}_2^{(0)}, \mathfrak{B}_3^{(0)})$ be a $k$-symmetrized quadratic factor of complexity $(d_1^{(0)}, d_2^{(0)}, d_3^{(0)})$. Then there is a refinement $(\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ of complexity $(d_1, d_2, d_3)$ and a decomposition $f = f_{\mathrm{str}} + f_{\mathrm{sml}} + f_{\mathrm{psr}}$ such that:*

1. *$f_{\mathrm{str}} = \mathbb{E}[f|\mathfrak{B}]$;*
2. *$\|f_{\mathrm{sml}}\|_2 \leq \delta$;*
3. *$\|f_{\mathrm{psr}}\|_{U^3(G^k)} \leq 1/\omega_2(d_1 + d_2 + d_3)$;*
4. *$f_{\mathrm{str}}$ and $f_{\mathrm{str}} + f_{\mathrm{sml}}$ take values in $[0,1]$ and $f_{\mathrm{psr}}, f_{\mathrm{sml}}$ take values in $[-1,1]$;*
5. *the complexity of $\mathfrak{B}$ is $(d_1, d_2, d_3)$ where*

$$d_1, d_2, d_3 \leq C(k, \delta, \omega_1, \omega_2, d_1^{(0)}, d_2^{(0)}, d_3^{(0)})$$

   *for a fixed function $C$;*

6. *the rank of $\mathfrak{B}$ is at least $\omega_1(d_1 + d_2 + d_3)$.*

The proof closely follows the proof of arithmetic regularity given in [Gre05b]; the only additional ingredient is guaranteeing at each stage that the factor introduced is $k$-symmetrized.

**Lemma 3.3.3.** *Fix $k \geq 1$. Let $\delta > 0$. There exists $\epsilon > 0$ such that the following holds. Let $\mathfrak{B}^{(0)} = (\mathfrak{B}_1^{(0)}, \mathfrak{B}_2^{(0)}, \mathfrak{B}_3^{(0)})$ be a $k$-symmetrized quadratic factor with complexity $(d_1, d_2, d_3)$ and*

*let* $f\colon G^k \to [-1,1]$ *be a function such that*

$$\|f - \mathbb{E}[f|\mathfrak{B}^{(0)}]\|_{U^3(G^k)} \geq \delta.$$

*Then there exists a refinement* $\mathfrak{B} = (\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ *with complexity at most* $(d_1 + k, d_2 + \binom{k+1}{2}, d_3 + \binom{k}{2})$ *such that*

$$\|\mathbb{E}[f|\mathfrak{B}]\|_2^2 \geq \|\mathbb{E}[f|\mathfrak{B}^{(0)}]\|_2^2 + \epsilon^2.$$

*Proof.* By the inverse theorem for the Gowers $U^3$-norm applied to $G^k \cong \mathbb{F}_p^{nk}$ (see [Gre07, GT08a]), there exist $\epsilon > 0$ (only depending on $\delta$), a vector $r \in \mathbb{F}_p^{nk}$, and a symmetric matrix $M \in \mathbb{F}_p^{nk \times nk}$ such that

$$\left|\mathbb{E}_{x \in \mathbb{F}_p^{nk}} \left(f(x) - \mathbb{E}[f|\mathfrak{B}^{(0)}](x)\right) e_p(r^\intercal x + x^\intercal M x)\right| \geq \epsilon.$$

Say $r = (r_1, \ldots, r_k) \in \mathbb{F}_p^{nk}$ where $r_1, \ldots, r_k \in \mathbb{F}_p^n$ and $M = (M_{ij})_{i,j \in [k]} \in \mathbb{F}_p^{nk \times nk}$ where $M_{ij} \in \mathbb{F}_p^{n \times n}$. Note that the matrices $M_{ii}$ are symmetric, while $M_{ij} = M_{ji}^\intercal$. For $i < j$, write $M_{ij} = M'_{ij} + M''_{ij}$ where $M'_{ij}$ is symmetric and $M''_{ij}$ is skew-symmetric. (Here we use that $p > 2$; explicitly, to decompose a matrix $A$ in such a form, we let $A_{ij}^{\mathrm{sym}} = (A_{ij} + A_{ji})/2$ and $A_{ij}^{\mathrm{skew}} = (A_{ij} - A_{ji})/2$ so that $A = A^{\mathrm{sym}} + A^{\mathrm{skew}}$.)

We define the factor $\mathfrak{B}$ by appending the vectors $r_1, \ldots, r_k$ to the list $\mathfrak{B}_1^{(0)}$, appending the symmetric matrices $(M_{ii})_{i \in [k]}$ and $(M'_{ij})_{i<j}$ to the list $\mathfrak{B}_2^{(0)}$, and appending the skew-symmetric matrices $(M_{ij})_{i<j}$ to the list $\mathfrak{B}_3^{(0)}$. To conclude, all that remains to show is that

$$\|\mathbb{E}[f|\mathfrak{B}]\|_2^2 \geq \|\mathbb{E}[f|\mathfrak{B}^{(0)}]\|_2^2 + \epsilon^2.$$

Define $g\colon \left(\mathbb{F}_p^n\right)^k \to \mathbb{C}$ by $g(x) = e_p(r^\intercal x + x^\intercal M x)$. Note that $\|g\|_2 = 1$ and crucially that $g$ is $\mathfrak{B}$-measurable by the simple equality

$$g(x_1, \ldots, x_k) = e_p\left(\sum_{i=1}^k r_i^\intercal x_i + \sum_{i=1}^k x_i^\intercal M_{ii} x_i + 2\sum_{i<j} x_i^\intercal M'_{ij} x_j + 2\sum_{i<j} x_i^\intercal M''_{ij} x_j\right).$$

Now the desired inequality follows from the Pythagorean theorem and Cauchy–Schwarz

inequality since

$$\begin{aligned}\|\mathbb{E}[f|\mathfrak{B}]\|_2^2 - \|\mathbb{E}[f|\mathfrak{B}^{(0)}]\|_2^2 &= \|\mathbb{E}[f|\mathfrak{B}] - \mathbb{E}[f|\mathfrak{B}^{(0)}]\|^2 \\ &\geq \left|\left\langle \mathbb{E}[f|\mathfrak{B}] - \mathbb{E}[f|\mathfrak{B}^{(0)}], g\right\rangle\right|^2 \\ &= \left|\left\langle f - \mathbb{E}[f|\mathfrak{B}^{(0)}], \mathbb{E}[g|\mathfrak{B}]\right\rangle\right|^2 \\ &\geq \epsilon^2. \end{aligned}$$

□

**Lemma 3.3.4.** *Fix $k \geq 1$. Let $\omega\colon \mathbb{R}^+ \to \mathbb{R}^+$ be an arbitrary growth function. There exists a growth function $\tau\colon \mathbb{R}^+ \to \mathbb{R}^+$ such that the following holds. Let $\mathfrak{B} = (\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ be a $k$-symmetrized quadratic factor with complexity $(d_1, d_2, d_3)$. There exists a refinement $\mathfrak{B}' = (\mathfrak{B}'_1, \mathfrak{B}'_2, \mathfrak{B}'_3)$ with complexity $(d'_1, d'_2, d'_3)$ that satisfies the following:*

1. *the rank of $\mathfrak{B}'$ is at least $\omega(d'_1 + d'_2 + d'_3)$;*
2. *$d'_2 \leq d_2$ and $d'_3 \leq d_3$ and $d'_1 \leq \tau(d_1 + d_2 + d_3)$.*

*Proof.* Consider a $k$-symmetrized quadratic factor $\mathfrak{B} = (\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ defined by $\mathfrak{B}_1 = (r_1, \ldots, r_{d_1})$ and $\mathfrak{B}_2 = (M_1, \ldots, M_{d_2})$ and $\mathfrak{B}_3 = (N_1, \ldots, N_{d_3})$. (Recall that the $r_i$ are vectors of length $n$ while the $M_i$ are symmetric $n \times n$ matrices and the $N_i$ are skew-symmetric $n \times n$ matrices.)

If the rank of $\mathfrak{B}$ is less than $r$, then either the $r_1, \ldots, r_{d_1}$ are linearly dependent or there exists a non-trivial linear combination

$$\sum_{i=1}^{d_2} a_i M_i + \sum_{j=1}^{d_3} b_j N_j$$

that has $\mathbb{F}_p$-rank less than $r$.

We do the following. First if there is some linear combination with rank less than $r$, then choose column vectors $s_1, \ldots, s_{r-1}, t_1, \ldots, t_{r-1}$ such that

$$\sum_{i=1}^{r-1} s_i t_i^\intercal = \sum_{i=1}^{d_2} a_i M_i + \sum_{j=1}^{d_3} b_j N_j.$$

Add $s_1, \ldots, s_{r-1}, t_1, \ldots, t_{r-1}$ to $\mathfrak{B}_1$ and remove the first $M_i$ or $N_j$ with nonzero coefficient

(i.e., if $a_1 = \cdots = a_{i-1} = 0$ while $a_i \neq 0$, then remove $M_i$ from $\mathfrak{B}_2$; if $a_1 = \cdots = a_{d_2} = b_1 = \cdots = b_{j-1} = 0$ and $b_j \neq 0$, then remove $N_j$ from $\mathfrak{B}_3$). Then remove any element of the modified $\mathfrak{B}_1$ that is linearly dependent on the previous vectors in $\mathfrak{B}_1$. Note that the factor produced refines the original factor.

We iterate the above process, producing a sequence of $k$-symmetrized quadratic factors $\mathfrak{B} = \mathfrak{B}^{(0)}, \mathfrak{B}^{(1)}, \ldots, \mathfrak{B}^{(M)}$ as follows. Suppose that $\mathfrak{B}^{(m)}$ has complexity $(d_1^{(m)}, d_2^{(m)}, d_3^{(m)})$. If $\mathfrak{B}^{(m)}$ has rank at least $\omega(d_1^{(m)} + d_2^{(m)} + d_3^{(m)})$, then halt and set $m = M$. Otherwise refine $\mathfrak{B}^{(m)}$ to $\mathfrak{B}^{(m+1)}$ as described above. Note that $M \leq d_2 + d_3 + 1$ since every step (except possibly the first) reduces $d_2^{(m)} + d_3^{(m)}$ by 1. Furthermore, one can easily see that $d_1^{(M)}$ is bounded by some function of $d_1, d_2, d_3$ and $\omega$, as desired. □

**Lemma 3.3.5.** *Fix $k \geq 1$. Let $\delta > 0$ and let $\omega\colon \mathbb{R}^+ \to \mathbb{R}^+$ be an arbitrary growth function. Let $\mathfrak{B}^{(0)} = (\mathfrak{B}_1^{(0)}, \mathfrak{B}_2^{(0)}, \mathfrak{B}_3^{(0)})$ be a $k$-symmetrized quadratic factor with complexity $(d_1^{(0)}, d_2^{(0)}, d_3^{(0)})$ and let $f\colon G^k \to [0,1]$ be a function. Then there exists a refinement $\mathfrak{B} = (\mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{B}_3)$ and a decomposition $f = f_{\mathrm{str}} + f_{\mathrm{psr}}$ such that:*

1. *$f_{\mathrm{str}} = \mathbb{E}[f|\mathfrak{B}]$;*
2. *$\|f_{\mathrm{psr}}\|_{U^3(G^k)} \leq \delta$;*
3. *$f_{\mathrm{str}}$ takes values in $[0,1]$ and $f_{\mathrm{psr}}$ takes values in $[-1,1]$;*
4. *the complexity of $\mathfrak{B}$ is $(d_1, d_2, d_3)$ where*

$$d_1, d_2, d_3 \leq C_\omega(k, \delta, d_1^{(0)}, d_2^{(0)}, d_3^{(0)})$$

   *for a fixed function $C_\omega$ depending on the function $\omega$;*

5. *the rank of $\mathfrak{B}$ is at least $\omega(d_1 + d_2 + d_3)$.*

*Proof.* This follows immediately by iterating Lemma 3.3.3 and Lemma 3.3.4 at most $\epsilon(\delta)^{-2}$ times.

In detail, we construct a sequence of $k$-symmetrized quadratic factors $\mathfrak{B}^{(0)}, \mathfrak{B}^{(1)}, \ldots, \mathfrak{B}^{(M)}$ each refining the last as follows. If $\|f - \mathbb{E}[f|\mathfrak{B}^{(m)}]\|_{U^3(G^k)} < \delta$, halt the process and set

$M = m$. Otherwise, let $\tilde{\mathfrak{B}}^{(m+1)}$ be the factor produced by applying Lemma 3.3.3 to $\mathfrak{B}^{(m)}$ and $f$ with parameter $\delta$. Then let $\mathfrak{B}^{(m+1)}$ be the factor produced by applying Lemma 3.3.4 to $\tilde{\mathfrak{B}}^{(m+1)}$ with parameter $\omega$. By definition, at every step of this process, the rank of $\mathfrak{B}^{(m)}$ is at least $\omega_1(d_1^{(m)} + d_2^{(m)} + d_3^{(m)})$ where $(d_1^{(m)}, d_2^{(m)}, d_3^{(m)})$ is the complexity of $\mathfrak{B}$.

Since

$$\|\mathbb{E}[f|\mathfrak{B}^{(m+1)}\|_2^2 \geq \|\mathbb{E}[f|\tilde{\mathfrak{B}}^{(m+1)}\|_2^2 \geq \|\mathbb{E}[f|\mathfrak{B}^{(m)}\|_2^2 + \epsilon(\delta)^2$$

and this quantity is bounded between 0 and 1, we conclude that the process must stop after $M \leq \epsilon(\delta)^{-2}$ steps.

At the conclusion of this process, we have produced a $k$-symmetrized quadratic factor $\mathfrak{B}^{(M)}$ that refines $\mathfrak{B}^{(0)}$ such that $\|f - \mathbb{E}[f|\mathfrak{B}^{(M)}]\|_{U^3(G^k)} < \delta$. Defining $f_{\mathrm{str}} = \mathbb{E}[f|\mathfrak{B}^{(M)}]$ and $f_{\mathrm{psr}} = f - \mathbb{E}[f|\mathfrak{B}^{(M)}]$ gives the desired result. □

*Proof of Theorem 3.3.2.* The desired result follows by iterating Lemma 3.3.5 at most $\delta^{-2}$ times.

In detail, we construct a sequence of $k$-symmetrized quadratic factors $\mathfrak{B}^{(0)}, \mathfrak{B}^{(1)}, \ldots, \mathfrak{B}^{(M)}$ each refining the last as follows. If $\|\mathbb{E}[f|\mathfrak{B}^{(m)}] - \mathbb{E}[f|\mathfrak{B}^{(m-1)}]\|_2^2 < \delta^2$, halt the process and set $M = m$. Otherwise, let $\mathfrak{B}^{(m+1)}$ be the factor produced by applying Lemma 3.3.5 to $\mathfrak{B}^{(m)}$ and $f$ with parameter $1/\omega_2(d_1^{(m)} + d_2^{(m)} + d_3^{(m)})$ and growth function $\omega_1$.

Note that by the Pythagorean theorem,

$$\|\mathbb{E}[f|\mathfrak{B}^{(m)}] - \mathbb{E}[f|\mathfrak{B}^{(m-1)}]\|_2^2 = \|\mathbb{E}[f|\mathfrak{B}^{(m)}]\|_2^2 - \|\mathbb{E}[f|\mathfrak{B}^{(m-1)}]\|_2^2.$$

Since these $L^2$-norms are bounded between 0 and 1, we see that the process must stop after $M \leq \delta^{-2}$ steps.

At the conclusion of this process, we have produced a $k$-symmetrized quadratic factor $\mathfrak{B}^{(M-1)}$ that refines $\mathfrak{B}^{(0)}$ with complexity $(d_1^{(M-1)}, d_2^{(M-1)}, d_3^{(M-1)})$ and rank at least $\omega_1(d_1^{(M-1)} + d_2^{(M-1)} + d_3^{(M-1)})$. Defining $f_{\mathrm{str}} = \mathbb{E}[f|\mathfrak{B}^{(M-1)}]$ and $f_{\mathrm{sml}} = \mathbb{E}[f|\mathfrak{B}^{(M)}] - \mathbb{E}[f|\mathfrak{B}^{(M-1)}]$ and $f_{\mathrm{psr}} = f - \mathbb{E}[f|\mathfrak{B}^{(M)}]$ gives the desired result. □

## 3.4 Equidistribution and counting lemma

The goal of this section is to study the counts of matrix patterns of the form $\{0, M_1, M_2, M_1 + M_2\}$ in the "structured term" $f_{\text{str}}$. Recall that a $k$-symmetrized quadratic factor $\mathfrak{B}$ defines a map $\mathsf{B}\colon G^k \to \left(\mathbb{F}_p^k\right)^{d_1} \times \mathcal{S}_k^{d_2} \times \mathcal{S}_k'^{d_3}$ where $\mathcal{S}_k$ and $\mathcal{S}_k'$ are the spaces of $k \times k$ symmetric (resp. skew-symmetric) matrices. We call the fibers of this map *atoms* of $\mathfrak{B}$.

Understanding the counts of patterns in $f_{\text{str}}$ is equivalent to understanding how occurrences of these patterns are distributed among tuples of atoms. The first result of this section is simply that the atoms of $\mathfrak{B}$ are approximately the same size; in other words, as $X \in G^k$ varies, $\mathsf{B}(X)$ is equidistributed in $\left(\mathbb{F}_p^k\right)^{d_1} \times \mathcal{S}_k^{d_2} \times \mathcal{S}_k'^{d_3}$.

The main result of this section describes how the 4-tuple $(\mathsf{B}(X), \mathsf{B}(X + M_1 D), \mathsf{B}(X + M_2 D), \mathsf{B}(X + (M_1 + M_2)D))$ is distributed as $X, D \in G^k$ vary. This 4-tuple is not equidistributed across all possible 4-tuples of atoms; instead it is equidistributed on a certain linear subspace. We need a somewhat unfortunate amount of notation in this section to describe this linear subspace.

Note that this is also the place where the "mysterious" spectral condition that $M_1 M_2^{-1}$ has no pair of eigenvalues that are negatives of each other appears. It turns out that the dimension of the space that the relevant 4-tuples are equidistributed over changes depending on whether or not this spectral condition is satisfied.

Finally in this section we restrict our attention to matrix patterns of the form $\{0, I, J, I+J\}$ where $I = I_{k\times k}$ is the identity, $J, I - J, I + J$ are invertible, and that $J$ satisfies the spectral condition (that no pair of eigenvalues of $J$ over the algebraic closure $\overline{\mathbb{F}}_p$ are negatives of each other). By a change of variables, all cases can be reduced to this one.

### 3.4.1 Equidistribution results

We first quote the following result on equidistribution in $\mathbb{F}_p^n$.

**Proposition 3.4.1** ([Gre07, Lemma 4.2])**.** *Define* $\Gamma(x) = (r_1^\intercal x, \dots, r_{d_1}^\intercal x)$ *and* $\Phi(x) = (x^\intercal M_1 x, \dots, x^\intercal M_{d_2} x)$ *where the* $M_i$ *are symmetric. Furthermore suppose that* $\{r_i\}_{i\in[d_1]}$ *are linearly independent and for any nonzero vector* $(\lambda_1, \dots, \lambda_{d_2})$ *in* $\mathbb{F}_p^{d_2}$ *we have* $\operatorname{rank}(\sum_{i=1}^{d_2} \lambda_i M_i) \geq$

*$r$. Then for any $a \in \mathbb{F}_p^{d_1}$ and $b \in \mathbb{F}_p^{d_2}$ we have*

$$\mathbb{P}_{x \in \mathbb{F}_p^n}[\Gamma(x) = a, \Phi(x) = b] = p^{-d_1 - d_2} + O(p^{-r/2}).$$

Note [Gre07] only states the above for $\mathbb{F}_5^n$ but the proof in general is completely analogous. Given this we can immediately derive the necessary equidistribution result on factors for the specialized factors constructed in the previous section.

**Proposition 3.4.2.** *Let $\mathfrak{B}$ be a $k$-symmetrized quadratic factor with rank at least $r$. Then*

$$\mathbb{P}_{X \in \mathbb{F}_p^{k \times n}}[\mathsf{B}(X) = ((v_i)_{i \in [d_1]}, (U_i)_{i \in [d_2]}, (V_i)_{i \in [d_3]})] = p^{-kd_1 - \binom{k+1}{2} d_2 - \binom{k}{2} d_3} + O(p^{-r/2})$$

*for all $v_i \in \mathbb{F}_p^k$, $U_i \in \mathcal{S}_k$, and $V_i \in \mathcal{S}'_k$.*

*Proof.* We view $X \in \mathbb{F}_p^{k \times n}$ instead as a vector $\tilde{X} \in \mathbb{F}_p^{kn}$ formed by concatenating the rows of $X$ into a single column vector. We deduce the desired result from Proposition 3.4.1 by building appropriate linear and quadratic forms out of the $r_i, M_i, N_i$ that define $\mathfrak{B}$.

For example for $N_i$, one of the $n \times n$ skew-symmetric matrices that define $\mathfrak{B}$, and $V_i \in \mathcal{S}'_k$ we have $XN_iX^\intercal = V_i$ if and only if $\tilde{X}^\intercal N_i^{(s,t)} \tilde{X} = (V_i)_{s,t}$ for all $1 \leq s < t \leq k$. Here $N_i^{(s,t)}$ is the $kn \times kn$ symmetric matrix formed out of $k \times k$ blocks of size $n \times n$ as follows: each block is the zero matrix except for the $(s,t)$ and $(t,s)$ blocks which are the matrices $N_i/2$ and $-N_i/2$, respectively. A similar construction can be used to handle the linear and symmetric quadratic parts of $\mathfrak{B}$. Note that the high rank assumption on $\mathfrak{B}$ implies the same rank bound on the auxiliary matrices. Thus the desired result now follows immediately from Proposition 3.4.1. □

Say that a random variable is *$\epsilon$-equidistributed* if it takes each value in its range with equal probability within a *multiplicative* error of $\epsilon$. A convenient property of this definition is that it is preserved under linear maps.

**Lemma 3.4.3.** *Suppose* $\mathbf{x}$ *is a random variable taking values in $\mathbb{F}_p^r$ satisfying*

$$\sup_{a \in \mathbb{F}_p^r} |p^r \mathbb{P}[\mathbf{x} = a] - 1| \leq \epsilon$$

*and* $L\colon \mathbb{F}_p^r \to \mathbb{F}_p^s$ *is a linear map with image of dimension* $t$*. Then for any* $a \in L\mathbb{F}_p^n$ *we have* $|p^t\mathbb{P}[L\mathbf{x} = a] - 1| \le \epsilon$*, whereas if* $a \notin L\mathbb{F}_p^n$ *then* $\mathbb{P}[L\mathbf{x} = a] = 0$*.*

*Proof.* This follows immediately from the fact that the preimage of every point in $L\mathbb{F}_p^r$ has size $p^{r-t}$. □

We now explicitly define the relevant lattice that the image of our pattern under $\mathsf{B}$ will equidistribute over, in order to state the main result of this section. Recall we have a given $J \in \mathbb{F}_p^{k\times k}$. Let

$$\Xi_J = \{A \in \mathbb{F}_p^{k\times k} : (JA)^\intercal = JA\},$$

and let

$$\Lambda_J = \{(-A, -A(I+J)(I-J)^{-1}, A(I+J)(I-J)^{-1}, A)\colon A^\intercal = +A, A \in \Xi_J\},$$
$$\Lambda'_J = \{(-A, -A(I+J)(I-J)^{-1}, A(I+J)(I-J)^{-1}, A)\colon A^\intercal = -A, A \in \Xi_J\}.$$

Also, let

$$\Psi_J = \{(x_1, x_2, x_3, x_4) \in (\mathbb{F}_p^k)^4\colon x_1 - x_2 - x_3 + x_4 = 0, x_4 - x_2 = J(x_2 - x_1)\}.$$

We make $\mathbb{F}_p^{k\times k}$ an inner product space with the standard inner product

$$\langle A, B\rangle = \langle A, B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^\intercal B)$$

on $\mathbb{F}_p^{k\times k}$. We extend this inner product to $\left(\mathbb{F}_p^{k\times k}\right)^4$ in the natural way, that is,

$$\langle (X_1, X_2, X_3, X_3), (Y_1, Y_2, Y_3, Y_4)\rangle = \langle X_1, Y_1\rangle + \langle X_2, Y_2\rangle + \langle X_3, Y_3\rangle + \langle X_4, Y_4\rangle.$$

We wish to study the equidistribution of the tuple

$$(\mathsf{B}(X), \mathsf{B}(X+D), \mathsf{B}(X+JD), \mathsf{B}(X+(I+J)D))$$

as $X, D$ range over $\mathbb{F}_p^{k\times n}$, for a $k$-symmetrized quadratic factor $\mathfrak{B}$. Ultimately, we will find

that the components corresponding to each $\mathsf{B}_{1,i}, \mathsf{B}_{2,i}, \mathsf{B}_{3,i}$ are all "independent", and that each equidistributes in the following way:

$$\begin{aligned}
&(\mathsf{B}_{1,i}(X), \mathsf{B}_{1,i}(X+D), \mathsf{B}_{1,i}(X+JD), \mathsf{B}_{1,i}(X+(I+J)D)) \quad \text{equidistributes on} \quad \Psi_J, \\
&(\mathsf{B}_{2,i}(X), \mathsf{B}_{2,i}(X+D), \mathsf{B}_{2,i}(X+JD), \mathsf{B}_{2,i}(X+(I+J)D)) \quad \text{equidistributes on} \quad \Lambda_J^{\perp} \cap (\mathcal{S}_k)^4, \\
&(\mathsf{B}_{3,i}(X), \mathsf{B}_{3,i}(X+D), \mathsf{B}_{3,i}(X+JD), \mathsf{B}_{3,i}(X+(I+J)D)) \quad \text{equidistributes on} \quad \Lambda_J'^{\perp} \cap (\mathcal{S}_k')^4.
\end{aligned}$$

Here the $\perp$ means the orthogonal subspace with respect to the inner product defined above. For ease of notation, we will write $\Lambda_J^{\perp}$ for $\Lambda_J^{\perp} \cap (\mathcal{S}_k)^4$ and $\Lambda_J'^{\perp}$ for $\Lambda_J'^{\perp} \cap (\mathcal{S}_k')^4$.

**Theorem 3.4.4.** *Suppose $J \in \mathbb{F}_p^{k\times k}$ is such that $J, I-J, I+J$ are invertible and $J$ has no pair of eigenvalues that are negatives of each other (over the algebraic closure $\overline{\mathbb{F}}_p$), let $G = \mathbb{F}_p^n$, and suppose $\mathfrak{B}$ is a $k$-symmetrized quadratic factor of rank $r$. Then for any $a \in \Psi_J^{d_1} \times (\Lambda_J^{\perp})^{d_2} \times (\Lambda_J'^{\perp})^{d_3}$*

$$\begin{aligned}
&\mathbb{P}_{X,D\in G^k}[(\mathsf{B}(X), \mathsf{B}(X+D), \mathsf{B}(X+JD), \mathsf{B}(X+(I+J)D)) = a] \\
&= p^{-d_1 \dim(\Psi_J) - d_2 \dim(\Lambda_J^{\perp}) - d_3 \dim(\Lambda_J'^{\perp})}(1 + O(p^{-r/2 + 2kd_1 + (2\binom{k+1}{2} + k^2)d_2 + (2\binom{k}{2} + k^2)d_3}))
\end{aligned}$$

The approach is similar to the proof of Proposition 3.4.2. We want to consider $(X, D)$ as a $2kn$-dimensional vector and apply Proposition 3.4.1, but now some linear dependencies will appear.[3] We will instead apply equidistribution on a set of "abstractly independent" forms to which we can indeed apply Proposition 3.4.1. Then we realize $(\mathsf{B}(X), \mathsf{B}(X+D), \mathsf{B}(X+JD), \mathsf{B}(X+(I+J)D))$ as the image of those elements under a linear map, and apply Lemma 3.4.3.

As a first step, we state the necessary equidistribution over these "abstract atoms". For convenience, given a $k$-symmetrized quadratic factor $\mathfrak{B}$, define an attached map

$$\mathsf{B}'(X, D) = ((XM_iD^{\intercal})_{i\in[d_2]}, (XN_iD^{\intercal})_{i\in[d_3]}).$$

[3]These linear dependencies appear for the same reason that $x^2, (x+d)^2, (x+2d)^2, (x+3d)^2$ satisfies a linear equation.

**Proposition 3.4.5.** *Suppose $\mathfrak{B}$ is a $k$-symmetrized quadratic factor of rank $r$. Then for any $a \in ((\mathbb{F}_p^k)^{d_1} \times (\mathcal{S}_k)^{d_2} \times (\mathcal{S}'_k)^{d_3})^2 \times (\mathbb{F}_p^{k\times k})^{d_2+d_3}$ we have*

$$\mathbb{P}_{X,D\in G^k}[(\mathsf{B}(X), \mathsf{B}(D), \mathsf{B}'(X,D)) = a] = p^{-2kd_1-(2\binom{k+1}{2}+k^2)d_2-(2\binom{k}{2}+k^2)d_3} + O(p^{-r/2}).$$

This follows immediately by applying Proposition 3.4.1 to $(X, D)$ viewed as an element of $\mathbb{F}_p^{2kn}$. The proof is exactly analogous to the proof of Proposition 3.4.2 from Proposition 3.4.1.

To complete the proof of Theorem 3.4.4, we note that the desired map, $(X, D) \mapsto (\mathsf{B}(X), \mathsf{B}(X + D), \mathsf{B}(X + JD), \mathsf{B}(X + (I + J)D))$, can be written as the map $(X, D) \mapsto (\mathsf{B}(X), \mathsf{B}(D), \mathsf{B}'(X, D))$ composed with a linear transformation. For example,

$$(X + JD)M_i(X + JD)^\intercal = XM_iX^\intercal + J(DM_iX^\intercal) + XM_iD^\intercal J^\intercal + J(DM_iD^\intercal)J^\intercal$$

and $XM_iD^\intercal = (DM_iX^\intercal)^\intercal$. Therefore it suffices to understand the linear constraints induced by this last linear transformation.

### 3.4.2 Deriving the linear constraints

To this end we prove the following abstract linear algebra statement, which essentially encodes the eigenvalue condition in Theorem 3.1.2.

**Lemma 3.4.6.** *If $A \in \mathbb{F}_p^{k\times k}$ is invertible and has no pair of eigenvalues (over the algebraic closure $\overline{\mathbb{F}}_p$) which are negatives of each other, then $A \in \mathbb{F}_p[A^2]$.*

*Proof.* Given any matrix $M$, let $Q_M \in \mathbb{F}_p[t]$ be the monic polynomial of minimum degree satisfying $Q_M(M) = 0$ (this exists by the Cayley–Hamilton theorem and the fact that $\mathbb{F}_p[t]$ is a principal ideal domain). Then $\mathbb{F}_p[M] \cong \mathbb{F}_p[t]/(Q_M(t))$ as $\mathbb{F}_p$-algebras, and the dimension as an $\mathbb{F}_p$-vector space is $\deg Q_M$.

We clearly have $A \in \mathbb{F}_p[A^2]$ if and only if $\mathbb{F}_p[A] = \mathbb{F}_p[A^2]$, which will certainly follow from

$$\dim_{\mathbb{F}_p} \mathbb{F}_p[A] \le \dim_{\mathbb{F}_p} \mathbb{F}_p[A^2]$$

due to the obvious containment. Now note that if $g(A^2) = 0$ then $Q_A(t)|g(t^2)$ in $\mathbb{F}_p[t]$. By

hypothesis, we have $\gcd(Q_A(t), Q_A(-t)) = 1$, hence $Q_A(t)Q_A(-t)|g(t^2)$. Thus

$$2\dim_{\mathbb{F}_p} \mathbb{F}_p[A] = \deg(Q_A(t)Q_A(-t)) \leq 2\deg g.$$

Since this holds for all such $g$, it in particular holds for $g = Q_{A^2}$, which implies the result. $\square$

Next we need the following abstract matrix equation which is used to derive the desired equidistribution statement.

**Lemma 3.4.7.** *Suppose $J \in \mathbb{F}_p^{k\times k}$ is such that $J, I-J, I+J$ are invertible and $J$ has no pair of eigenvalues that are negatives of each other.*

- *Let $M = M^\intercal$ be nonzero. Then $(A_1, A_2, A_3, A_4) \in (\mathcal{S}_k)^4$ and*

$$\begin{aligned}\operatorname{tr}(A_1^\intercal XMX^\intercal + A_2^\intercal(X+D)M(X+D)^\intercal + A_3^\intercal(X+JD)M(X+JD)^\intercal \\ + A_4^\intercal(X+(I+J)D)M(X+(I+J)D)^\intercal) = 0\end{aligned}$$

  *for all $X, D \in (\mathbb{F}_p^n)^k$ if and only if $(A_1, A_2, A_3, A_4) \in \Lambda_J$.*

- *Let $M = -M^\intercal$ be nonzero. Then $(A_1, A_2, A_3, A_4) \in (\mathcal{S}_k')^4$ and*

$$\begin{aligned}\operatorname{tr}(A_1^\intercal XMX^\intercal + A_2^\intercal(X+D)M(X+D)^\intercal + A_3^\intercal(X+JD)M(X+JD)^\intercal \\ + A_4^\intercal(X+(I+J)D)M(X+(I+J)D)^\intercal) = 0\end{aligned}$$

  *for all $X, D \in (\mathbb{F}_p^n)^k$ if and only if $(A_1, A_2, A_3, A_4) \in \Lambda_J'$.*

*Proof.* We prove the first claim as the second is analogous. Note that

$$\begin{aligned}\operatorname{tr}(A_1^\intercal XMX^\intercal + A_2^\intercal(X+D)M(X+D)^\intercal + A_3^\intercal(X+JD)M(X+JD)^\intercal \\ + A_4^\intercal(X+(I+J)D)M(X+(I+J)D)^\intercal) = 0\end{aligned}$$

implies that

$$\operatorname{tr}((A_1^\intercal + A_2^\intercal + A_3^\intercal + A_4^\intercal)XMX^\intercal) = 0,$$

$$\operatorname{tr}(A_2^\intercal(D)M(D)^\intercal + A_3^\intercal(JD)M(JD)^\intercal + A_4^\intercal((I+J)D)M((I+J)D)^\intercal) = 0,$$

taking $D = 0$ and $X = 0$ respectively. Using that $\mathrm{tr}(\cdot)$ is additive along with the initial condition, we derive

$$\begin{aligned}
&\mathrm{tr}((A_1^\intercal + A_2^\intercal + A_3^\intercal + A_4^\intercal)XMX^\intercal) = 0,\\
&\mathrm{tr}(A_2^\intercal(D)M(D)^\intercal + A_3^\intercal(JD)M(JD)^\intercal + A_4^\intercal((I+J)D)M((I+J)D)^\intercal) = 0,\\
&\mathrm{tr}(A_2^\intercal(DMX^\intercal + XMD^\intercal) + A_3^\intercal(JDMX^\intercal + XM(JD)^\intercal)\\
&\qquad + A_4^\intercal((J+I)DMX^\intercal + XM((J+I)D)^\intercal) = 0.
\end{aligned}$$

Using that $\mathrm{tr}(A^\intercal B^\intercal) = \mathrm{tr}(BA) = \mathrm{tr}(AB)$ the above conditions are equivalent to

$$\begin{aligned}
&\mathrm{tr}((A_1^\intercal + A_2^\intercal + A_3^\intercal + A_4^\intercal)XMX^\intercal) = 0\\
&\mathrm{tr}((A_2^\intercal + J^\intercal A_3^\intercal J + (I+J)^\intercal A_4^\intercal(I+J))(DMD)^\intercal) = 0\\
&\mathrm{tr}((2A_2^\intercal + 2A_3^\intercal J + 2A_4^\intercal(I+J))(DMX^\intercal)) = 0.
\end{aligned}$$

To derive the last, we used both $A_i^\intercal = A_i$ and $M^\intercal = M$. (In fact, one obtains identical equations in the skew-symmetric case.)

Since $M$ is nonzero and symmetric we have that $\{XMX^\intercal\}$, $\{DMD^\intercal\}$ each span the space of all $k \times k$ symmetric matrices while $\{DMX^\intercal\}$ spans the space of all $k \times k$ matrices. This implies that

$$\begin{aligned}
A_1^\intercal + A_2^\intercal + A_3^\intercal + A_4^\intercal &= 0\\
A_2^\intercal + J^\intercal A_3^\intercal J + (I+J)^\intercal A_4^\intercal(I+J) &= 0\\
A_2^\intercal + A_3^\intercal J + A_4^\intercal(I+J) &= 0
\end{aligned}$$

since the $A_i$ are symmetric.

Note that the second and third equations imply that

$$(J^\intercal - I)A_2^\intercal = A_4^\intercal(I+J).$$

Noting that $A_2^\intercal, A_4^\intercal$ are symmetric we find that

$$A_2^\intercal = (J^\intercal - I)^{-1} A_4^\intercal (I + J) = (I + J^\intercal) A_4^\intercal (J - I)^{-1}.$$

This is equivalent to

$$(J^\intercal)^2 A_4^\intercal = A_4^\intercal J^2.$$

It follows that for every polynomial $Q(t) \in \mathbb{F}_p[t]$ we have

$$Q((J^\intercal)^2) A_4^\intercal = A_4^\intercal Q(J^2).$$

Now $J \in \mathbb{F}_p[J^2]$ by Lemma 3.4.6, so it follows that

$$J^\intercal A_4^\intercal = A_4^\intercal J.$$

A similar combination of the second and third equations implies that

$$J^\intercal A_2^\intercal = -A_3^\intercal J$$

and therefore

$$(J^\intercal)^{-1} A_3^\intercal J = (J^\intercal) A_3^\intercal J^{-1}.$$

This similarly implies that $J^\intercal A_3^\intercal = A_3^\intercal J$. Therefore we deduce

$$\begin{aligned} A_1^\intercal + A_2^\intercal + A_3^\intercal + A_4^\intercal &= 0 \\ A_2^\intercal + A_3^\intercal J^2 + A_4^\intercal (I + J)^2 &= 0 \\ A_2^\intercal + A_3^\intercal J + A_4^\intercal (I + J) &= 0. \end{aligned}$$

Subtracting the last two equations gives that

$$A_3^\intercal = A_4^\intercal (I + J)(I - J)^{-1}.$$

Substituting into the last equation gives

$$A_2^\intercal = A_4^\intercal(I+J)(-I-J(I-J)^{-1}) = -A_3^\intercal.$$

Finally using the first equation this implies that $A_1^\intercal = -A_4^\intercal$. Therefore we have proven that $(A_1, A_2, A_3, A_4) \in \Lambda_J$. The reverse implication is a straightforward calculation which we omit. □

We are now ready to prove Theorem 3.4.4.

*Proof of Theorem 3.4.4.* This is almost immediate from Proposition 3.4.5 and Lemmas 3.4.3 and 3.4.7. The noted fact that the desired function is the image of the function in Proposition 3.4.5 under a linear mapping along with Lemma 3.4.3 demonstrates that there is equidistribution over some subspace (with the multiplicative error term preserved).

This subspace is precisely the span of all possible vectors $(\mathsf{B}(X), \mathsf{B}(X+D), \mathsf{B}(X+JD), \mathsf{B}(X+(I+J)D))$. Due to the independence demonstrated in Proposition 3.4.5, we see that the subspace factors as a direct sum.

Lemma 3.4.7 characterizes the resulting vector spaces for $\mathsf{B}_{2,i}, \mathsf{B}_{3,i}$, since it demonstrates the form of all orthogonal vectors in corresponding host spaces (either tuples of symmetric or skew-symmetric matrices). For $\mathsf{B}_1$, it is easy to check that vectors of the form $(Xr, (X+D)r, (X+JD)r, (X+(I+J)D)r)$ span the space $\Psi_J$ when $r \in \mathbb{F}_p^n \setminus 0$. Indeed, the orthogonal vectors of the form $(\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_3)$ are precisely those that satisfy

$$\vec{a}_1 + \vec{a}_2 + \vec{a}_3 + \vec{a}_4 = \vec{a}_2 + \vec{a}_3 J + \vec{a}_4(I+J) = 0.$$

All such vectors are spans of $(\vec{t}, -\vec{t}, -\vec{t}, \vec{t})$ and $(J\vec{t}, -(I+J)\vec{t}, 0, \vec{t})$, which matches the orthogonal space of $\Psi_J$. This completes the proof. □

## 3.5 Popular differences for 4-point patterns

We now have the tools to prove the following popular difference result for four-point patterns. This theorem immediately implies the result stated in the introduction, Theorem 3.1.2.

**Theorem 3.5.1.** *Fix $k \geq 1$ and an odd prime $p$. Let $M_1, M_2$ be $k \times k$ matrices with coefficients in $\mathbb{F}_p$ such that $M_1$, $M_2$, $M_1 - M_2$, and $M_1 + M_2$ are invertible and no pair of eigenvalues of $M_1 M_2^{-1}$ (viewed over the algebraic closure $\overline{\mathbb{F}}_p$) are negatives of each other. For any $\alpha, \epsilon > 0$, letting $G = \mathbb{F}_p^n$ for $n > n_0(\alpha, \epsilon, p)$, if $f \colon G^k \to [0,1]$ satisfies $\mathbb{E}_{X \in G^k} f(X) \geq \alpha$ then there are $\Omega_{\alpha,\epsilon,p}(p^{kn})$ values $D \in G^k$ such that*

$$\mathbb{E}_{X \in G^k} f(X) f(X + M_1 D) f(X + M_2 D) f(X + (M_1 + M_2)D) \geq \alpha^4 - \epsilon.$$

Given the previous developments the proof is similar to that of [Gre07, Theorem 4.1] modulo deriving the necessary positivity.

*Proof of Theorem 3.5.1.* By replacing $M_1 D$ by $D$ (note that $M_1$ is invertible) we can reduce the case of $(M_1, M_2)$ to the case of $(I, M_2 M_1^{-1})$. Hence from now on we assume that $M_1 = I$ and $M_2 = J$. Applying Theorem 3.3.2, we decompose

$$f = f_{\mathrm{str}} + f_{\mathrm{sml}} + f_{\mathrm{psr}}$$

where $f_{\mathrm{str}} = \mathbb{E}[f|\mathfrak{B}]$ for a $k$-symmetrized quadratic factor $\mathfrak{B}$ of complexity $(d_1, d_2, d_3)$ and rank at least $\omega_1(d_1 + d_2 + d_3)$, where $\|f_{\mathrm{sml}}\|_2 \leq \epsilon/250$, and where $\|f_{\mathrm{psr}}\|_{U^3(G^k)} \leq 1/\omega_2(d_1 + d_2 + d_3)$. The complexity $(d_1, d_2, d_3)$ is bounded in terms of the given parameters and growth functions. We define $H$ to be the subspace of $(\mathbb{F}_p^n)^k$ such that the linear factors of $\mathfrak{B}$ are zero, i.e., $H = \{X \in \mathbb{F}_p^{k \times n} : Xr_1 = \cdots = Xr_{d_1} = 0\}$. We will prove that

$$\mathbb{E}_{X, D \in G^k}[f(X) f(X + D) f(X + JD) f(X + (J + I)D) \mathbb{1}_H(D)] \geq p^{-kd_1}(\alpha^4 - \epsilon). \tag{3.1}$$

This suffices to prove the result as we are summing over a density $p^{-kd_1}$ subset of the differences $D$ and hence at least one difference achieves the necessary bound. Furthermore, by Markov's inequality, a fraction of at least $\Omega_{\alpha,\epsilon}(1)$ of the differences in $H$ satisfy the weaker bound of $\alpha^4 - 2\epsilon$. Adjusting $\epsilon$ appropriately will prove the desired upon noting that a positive fraction of $G^k$ lies in $H$ due to the bounded complexity of $\mathfrak{B}$.

Now we focus attention on Eq. (3.1). Expanding $f = f_{\mathrm{str}} + f_{\mathrm{sml}} + f_{\mathrm{psr}}$ allows us to turn

the left side into 81 terms of the form

$$\mathbb{E}_{X,D\in G^k}[f_1(X)f_2(X+D)f_3(X+JD)f_4(X+(J+I)D)\mathbb{1}_H(D)]$$

where $f_1, f_2, f_3, f_4 \in \{f_{\text{str}}, f_{\text{sml}}, f_{\text{psr}}\}$. If $f_{\text{sml}}$ appears in the expression, we have that

$$\begin{aligned}|\mathbb{E}_{X,D}[f_1(X)f_2(X+D)f_3(X+JD)f_4(X+(J+I)D)\mathbb{1}_H(D)]| \\ \leq \mathbb{E}_X[|f_{\text{sml}}(X)|\mathbb{1}_H(D)] \leq p^{-kd_1}(\mathbb{E}_X[f_{\text{sml}}(X)^2])^{1/2} \leq p^{-kd_1}\epsilon/250.\end{aligned}$$

This bounds the 65 terms that include $f_{\text{sml}}$.

Next we bound the terms that include $f_{\text{psr}}$. Say $f_3 = f_{\text{psr}}$ (the other cases are analogous). Note that we have

$$\mathbb{1}_H(D) = \sum_{T\in H^\perp} \mathbb{1}_{T+H}(X+D)\mathbb{1}_{T+H}(X)$$

for all $X$. Therefore

$$\begin{aligned}&|\mathbb{E}[f_1(X)f_2(X+D)f_{\text{psr}}(X+JD)f_4(X+(I+J)D)\mathbb{1}_H(D)]| \\ &\quad= \left|\sum_{T\in H^\perp} \mathbb{E}[f_1(X)f_2(X+D)f_{\text{psr}}(X+JD)f_4(X+(I+J)D)\mathbb{1}_{T+H}(X+D)\mathbb{1}_{T+H}(X)]\right| \\ &\quad\leq |H^\perp|\|f_{\text{psr}}\|_{U^3(G^k)}\end{aligned}$$

where in the final line we have used Lemma 3.2.2 (since $I, J, I+J, I-J$ are invertible). Taking the growth function $\omega_2$ to be a sufficiently fast growing exponential (depending on $p, k$), we can ensure that $\|f_{\text{psr}}\|_{U^3(G^k)} \leq p^{-2kd_1}\epsilon/250$.

Thus it suffices to prove

$$\mathbb{E}[f_{\text{str}}(X)f_{\text{str}}(X+D)f_{\text{str}}(X+JD)f_{\text{str}}(X+(J+I)D)\mathbb{1}_H(D)] \geq p^{-kd_1}(\alpha^4 - \epsilon/2). \quad (3.2)$$

Since $f_{\text{str}} = \mathbb{E}[f|\mathfrak{B}]$, we see that $f_{\text{str}}$ is $\mathfrak{B}$-measurable and $[0,1]$-valued. Recall that a factor $\mathfrak{B}$ defines a $\mathsf{B}\colon G^k = (\mathbb{F}_p^n)^k \to (\mathbb{F}_p^k)^{d_1} \times (\mathcal{S}_k)^{d_2} \times (\mathcal{S}_k')^{d_3}$. Hence there is an associated function $\mathbf{f}\colon (\mathbb{F}_p^k)^{d_1} \times (\mathcal{S}_k)^{d_2} \times (\mathcal{S}_k')^{d_3} \to [0,1]$ such that $f_{\text{str}}(X) = \mathbf{f}(\mathsf{B}(X))$.

**Claim 3.5.2.**

$$\mathbb{E}[f_{\mathrm{str}}(X)f_{\mathrm{str}}(X+D)f_{\mathrm{str}}(X+JD)f_{\mathrm{str}}(X+(J+I)D)\mathbb{1}_H(D)]$$

*and*

$$p^{-kd_1}\mathbb{E}_{\vec{v}\in\mathbb{F}_p^k,\vec{M}_2\in(\Lambda_J^\perp)^{d_2},\vec{M}_3\in(\Lambda_J'^\perp)^{d_3}}[\mathbf{f}(\vec{v},M_2^{(1)},M_3^{(1)})\mathbf{f}(\vec{v},M_2^{(2)},M_3^{(2)})\mathbf{f}(\vec{v},M_2^{(3)},M_3^{(3)})\mathbf{f}(\vec{v},M_2^{(4)},M_3^{(4)})]$$

*are equal to up to a multiplicative factor of* $1+O(p^{-r/2+2kd_1+(2\binom{k+1}{2}+k^2)d_2+(2\binom{k}{2}+k^2)d_3})$*, where the rank* $r$ *is the rank of* $\mathfrak{B}$*.*

*Proof.* Restricting $D$ to lie in $H$ we need to understand the equidistribution of $(\mathsf{B}(X),\mathsf{B}(X+D),\mathsf{B}(X+JD),\mathsf{B}(X+(I+J)D))$. We apply Theorem 3.4.4. We have that $(\mathsf{B}_1(X),\mathsf{B}_1(X+D),\mathsf{B}_1(X+JD),\mathsf{B}_1(X+(I+J)D))$ equidistributes over the diagonal $\{(\vec{v},\vec{v},\vec{v},\vec{v}):\vec{v}\in(\mathbb{F}_p^k)^{d_1}\}$, since looking at $D\in H$ corresponds precisely to looking at atoms for which linear factors in $\mathfrak{B}$ for $\mathsf{B}(X),\mathsf{B}(X+D)$ are equal. (We are also implicitly using that $D\in H$ implies each row of $D$ is in the orthogonal space of the defining vectors $\{r_1,\ldots,r_{d_1}\}$ of the linear factors of $\mathfrak{B}$, which implies $JD\in H$ as well.)

Furthermore

$$(M_2^{(1)},M_2^{(2)},M_2^{(3)},M_2^{(4)})=(\mathsf{B}_2(X),\mathsf{B}_2(X+D),\mathsf{B}_2(X+JD),\mathsf{B}_2(X+(I+J)D))$$

equidistributes over $(\Lambda_J^\perp)^{d_2}$ and finally

$$(M_3^{(1)},M_3^{(2)},M_3^{(3)},M_3^{(4)})=(\mathsf{B}_3(X),\mathsf{B}_3(X+D),\mathsf{B}_3(X+JD),\mathsf{B}_3(X+(I+J)D))$$

equidistributes over $(\Lambda_J'^\perp)^{d_3}$ with each of the components distributing independently. Therefore, even upon restricting the image of $\mathsf{B}_1$ to be zero, we have equidistribution over the remaining space. □

It remains to study the expectation in Claim 3.5.2. Recalling the definitions of $\Xi_J,\Lambda_J,\Lambda_J'$

from Section 3.4, we define the following subspaces:

$$\Omega_J = \{(-A, -A(I+J)(I-J)^{-1})\colon A^\intercal = +A, A \in \Xi_J\},$$
$$\Omega'_J = \{(-A, -A(I+J)(I-J)^{-1})\colon A^\intercal = -A, A \in \Xi_J\}.$$

**Claim 3.5.3.** *Given $(M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}) \in \mathcal{S}_k^4$, we have $(M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}) \in \Lambda_J^\perp$ if and only if we have the equality of cosets*

$$(M^{(1)}, M^{(2)}) + \Omega_J^\perp = (M^{(4)}, M^{(3)}) + \Omega_J^\perp.$$

*Similarly given $(M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}) \in \mathcal{S}_k'^4$, we have $(M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}) \in \Lambda_J'^\perp$ if and only if we have the equality of cosets*

$$(M^{(1)}, M^{(2)}) + \Omega_J'^\perp = (M^{(4)}, M^{(3)}) + \Omega_J'^\perp.$$

This claim follows by inspection of the definitions of $\Lambda_J$ and $\Omega_J$. From this we deduce

$$\begin{aligned}
&\mathbb{E}[\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})\mathbf{f}(\vec{v}, M_2^{(2)}, M_3^{(2)})\mathbf{f}(\vec{v}, M_2^{(3)}, M_3^{(3)})\mathbf{f}(\vec{v}, M_2^{(4)}, M_3^{(4)})] \\
&\qquad = \mathbb{E}_{\vec{v}}\mathbb{E}_\tau\left(\mathbb{E}_{\left((M_2^{(1)}, M_2^{(2)})+\Omega_J^\perp, (M_3^{(1)}, M_3^{(2)})+\Omega_J'^\perp\right)=\tau}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})\mathbf{f}(\vec{v}, M_2^{(2)}, M_3^{(2)})\right)^2 \\
&\qquad \geq \left(\mathbb{E}_{\vec{v}}\mathbb{E}_\tau\mathbb{E}_{\left((M_2^{(1)}, M_2^{(2)})+\Omega_J^\perp, (M_3^{(1)}, M_3^{(2)})+\Omega_J'^\perp\right)=\tau}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})\mathbf{f}(\vec{v}, M_2^{(2)}, M_3^{(2)})\right)^2 \\
&\qquad = \left(\mathbb{E}_{\vec{v}}\mathbb{E}_{(M_2^{(1)}, M_2^{(2)}, M_3^{(1)}, M_3^{(2)})}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})\mathbf{f}(\vec{v}, M_2^{(2)}, M_3^{(2)})\right)^2 \\
&\qquad = \mathbb{E}_{\vec{v}}\left(\mathbb{E}_{(M_2^{(1)}, M_3^{(1)})}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})^2\right)^2 \\
&\qquad \geq \left(\mathbb{E}_{\vec{v}, (M_2^{(1)}, M_3^{(1)})}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})\right)^4
\end{aligned}$$

where we have used the Cauchy–Schwarz inequality twice. To finish note that by the equidistibution derived in Proposition 3.4.2 we have that $\mathbb{E}_{\vec{v}, (M_2^{(1)}, M_2^{(2)})}\mathbf{f}(\vec{v}, M_2^{(1)}, M_3^{(1)})$ is $\mathbb{E}[f_{\text{str}}]$ up to a multiplicative factor of $1 + O(p^{-r/2+kd_1+\binom{k+1}{2}d_2+\binom{k}{2}d_3})$. Finally since $\mathbb{E}[f_{\text{str}}] = \mathbb{E}[\mathbb{E}[f|\mathfrak{B}]] = \alpha$ the desired result follows upon taking the growth function $\omega_2$ (and thus $r$, the rank of $\mathfrak{B}$) to be large enough. □

*Remark* 3.5.4. The key reason that this argument works is a "positivity" result in the final expectation. This "positivity" occurs from the symmetry of $\Lambda_J^\perp$ described in Claim 3.5.3 which ultimately is derived from the spectral condition on $J$. In the next section we consider what occurs when the spectral condition on $J$ is no longer satisfied. In essence, $\Lambda_J^\perp$ will be one dimension larger than otherwise in such a way that it no longer has this symmetry property.

## 3.6 Counterexample to popular differences for rotated squares in $\mathbb{F}_5^n$

Given the results of the last section it is natural to ask whether the popular difference result holds for all matrix patterns which are controlled by the $U^3$-norm. We now prove that this is false and demonstrate Theorem 3.1.3.

As is standard, it suffices to give a construction for a set of fixed positive density, as smaller sets will follow from subsampling and simple concentration facts (see for instance [SSZ21, Lemma 3.3]). In fact, we will merely give a function in $[0, 1]$ (which one can scale down appropriately and sample from). The counterexample proceeds in stages. In the first stage we give a construction which rules out all sufficiently generic differences (i.e., $(a, b)$ such that $\mathrm{span}_{\mathbb{F}_5}\{a, b\}$ is 2-dimensional), using the failure of the key positivity in Section 3.4. We then modify the function to rule out the non-generic directions using techniques from earlier work of Sah, Sawhney, and Zhao in [SSZ21, Sections 2, 3] (which deals with, for instance, axis-aligned squares) with [SSZ21, Section 3] itself building on a construction of Mandache [Man21a]. Finally, we expect that the construction here can be used to disprove the ergodic analogue [ABB21, Question 1.11] but we do not pursue this direction here.

First, if $\Gamma$ is the linear automorphism of $(\mathbb{F}_5^n)^2$ defined by $(x, y) \mapsto (x - 2y, x + 2y)$ then by replacing $A$ by $\Gamma^{-1}A$ and reparametrizing the pattern we are counting (making use of the fact we are in $\mathbb{F}_5$), it suffices to consider the "diagonalized" pattern

$$(x, y), (x + a, y + b), (x + 2a, y - 2b), (x + 3a, y - b).$$

### 3.6.1 Initial construction

Let $f_1 \colon (\mathbb{F}_5^n)^2 \to [0,1]$ be defined by

$$f_1(x,y) = g_1(x \cdot x, x \cdot y),$$

where $g_1 \colon \mathbb{F}_5^2 \to [0,1]$ is to be chosen later. With this choice, the pattern count for $f$ depends only on $g_1$ and the distribution of 8-tuples of the form

$$(x \cdot x, x \cdot y, \ldots, (x+3a) \cdot (x+3a), (x+3a) \cdot (y-b)).$$

Let $\Lambda_2' \leq \mathbb{F}_5^8$ be the space orthogonal to the vectors

$$(1,0,-1,0,-1,0,1,0), \quad (0,1,0,-1,0,-1,0,1), \quad (1,0,-3,0,3,0,-1,0)$$

We claim that these 8-tuples equidistribute over $\Lambda_2'$.

**Proposition 3.6.1.** *Fix nonzero $a, b \in \mathbb{F}_5^n$ that are not multiples of each other. Then*

$$(x \cdot x, x \cdot y, (x+a) \cdot (x+a), (x+a) \cdot (y+b), (x+2a) \cdot (x+2a), \\ (x+2a) \cdot (y-2b), (x+3a) \cdot (x+3a), (x+3a) \cdot (y-b))$$

*obtains every value in $(0,0,a \cdot a, a \cdot b, 4a \cdot a, -4a \cdot b, 9a \cdot a, -3a \cdot b) + \Lambda_2'$ with probability*

$$5^{-5} + O(5^{-n/2})$$

*Remark* 3.6.2. Note that $(0,0,0,1,0,-4,0,-3) \in \Lambda_2'$, so it actually equidistributes in $(0,0,a \cdot a, 0, 4a \cdot a, 0, 9a \cdot a, 0) + \Lambda_2'$.

*Proof.* This is a hands-on computation of equidistribution. Apply Proposition 3.4.1 to the concatenation $x' = (x,y)$ with linear forms $x \cdot a$, $x \cdot b$, $a \cdot y$ and quadratic forms $x \cdot x$, $x \cdot y$ (the last of which can be written as a symmetric matrix when $p \neq 2$). We conclude that the image of $(x,y)$ under this 5-tuple of forms obtains each point in $\mathbb{F}_5^5$ with probability $5^{-5} + O(5^{-n/2})$.

Given these five values one can solve for all values in the 8-tuple, and solving we obtain equidistribution of the 8-tuple over the appropriate 5-dimensional subspace $\Lambda_2'$ as desired. $\square$

Consequently, if $a, b \in \mathbb{F}_5^n$ are nonzero and not multiples of each other,

$$\begin{aligned}\beta_1(a,b) &:= \mathbb{E} f_1(x,y) f_1(x+a, y+b) f_1(x+2a, y-2b) f_1(x+3a, y-b) \\ &= \mathbb{E} g_1(x \cdot x, x \cdot y) g_1((x+a)\cdot(x+a), (x+a)\cdot(y+b)) \\ &\qquad g_1((x+2a)\cdot(x+2a), (x+2a)\cdot(y-2b)) \\ &\qquad g_1((x+3a)\cdot(x+3a), (x+3a)\cdot(y-b)) \\ &= \mathbb{E}_{v \in \Lambda_2'} g_1(v_1, v_2) g_1(v_3 + a \cdot a, v_4) g_1(v_5 + 4a \cdot a, v_6) g_1(v_7 + 9a \cdot a, v_8) + O(5^{-n/2}),\end{aligned}$$

using the remark after Proposition 3.6.1 in the last line.

Let $g_1(x,y) = \mathbb{1}_S(x,y)$ be the indicator of the set

$$S = \{(0,2), (0,3), (0,4), (1,0), (1,3), (1,4), (2,1), (2,2), (3,0), (3,1)\}.$$

Then a direct verification proves that

$$\sup_{a \cdot a \in \mathbb{F}_5} \mathbb{E}_{v \in \Lambda_2'} g_1(v_1, v_2) g_1(v_3 + a \cdot a, v_4) g_1(v_5 + 4a \cdot a, v_6) g_1(v_7 + 9a \cdot a, v_8) = \frac{73}{5^5}$$

while clearly

$$\alpha_1 = \mathbb{E} f_1(x,y) = \mathbb{E}_{v_1, v_2 \in \mathbb{F}_5} g_1(v_1, v_2) + O(5^{-n/2}) = \frac{2}{5} + O(5^{-n/2})$$

by Proposition 3.6.1. (Code verifying this explicit finite computation is attached on the arXiv version of the paper [BSST22].)

If we merely wish to establish Theorem 3.1.3 for all but $O(\sqrt{|G|})$ directions, where $G = (\mathbb{F}_5^n)^2$, then we are done due to $73/5^5 < (2/5)^4$. Now we introduce further constructions which allow us to fix the directions where $a, b$ are nontrivially related.

### 3.6.2 Fixing most special directions

We next proceed with a modification of a construction which appears in [Man21a] and was extended in [SSZ21]. Let $X_x, Y_y, Z_z, X'_x, Y'_y, Z'_z$ for $x, y, z \in \mathbb{F}_5^n$ be independent random variables uniform on $[0,1]$. Let $F_2, F_3 \colon (\mathbb{F}_5^n)^2 \to [0,1]$ be defined by

$$F_2(x,y) = g_2(X_{-x-y}, Y_{-2x+2y}, Z_{2x+y}), \quad F_3(x,y) = g_2(X'_{-x-2y}, Y'_{-2x-y}, Z'_{2x+2y})$$

where $g_2 \colon [0,1]^3 \to [0,1]$ is a function to be chosen later. This is a random function. Let $h(x,y) = f_1(x,y)F_2(x,y)F_3(x,y)$, which is also a random function. Let

$$\alpha_h = \mathbb{E}_{x,y \in \mathbb{F}_5^n} h(x,y)$$

and let

$$\beta_h(a,b) = \mathbb{E}_{x,y \in \mathbb{F}_5^n} h(x,y)h(x+a, y+b)h(x+2a, y-2b)h(x+3a, y-b),$$

which are both random in the $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}'$ variables. We have

$$\begin{aligned}\alpha_2 &= \mathbb{E}_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \alpha_h = \mathbb{E}_{x,y \in \mathbb{F}_5^n} \mathbb{E}_{\substack{\mathbf{X},\mathbf{Y},\mathbf{Z} \\ \mathbf{X}',\mathbf{Y}',\mathbf{Z}'}} f_1(x,y) g_2(X_{-x-y}, Y_{-2x+2y}, Z_{2x+y}) g_2(X'_{-x-y}, Y'_{-2x+2y}, Z'_{2x+y}) \\ &= \mathbb{E}_{x,y \in \mathbb{F}_5^n} f_1(x,y) \cdot (\mathbb{E}_{u,v,w \in [0,1]} g_2(u,v,w))^2\end{aligned}$$

since $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are independent, etc. We also have

$$\begin{aligned}\beta_2(a,b) &= \mathbb{E}_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \beta_h(a,b) \\ &= \mathbb{E}_{x,y \in \mathbb{F}_5^n} \mathbb{E}_{\substack{\mathbf{X},\mathbf{Y},\mathbf{Z} \\ \mathbf{X}',\mathbf{Y}',\mathbf{Z}'}} f_1(x,y) f_1(x+a, y+b) f_1(x+2a, y-2b) f_1(x+3a, y-b) \\ &\quad g_2(X_{-x-y}, Y_{-2x+2y}, Z_{2x+y}) g_2(X_{-x-y-a-b}, Y_{-2x+2y-2a+2b}, Z_{2x+y+2a+b}) \\ &\quad g_2(X_{-x-y-2a+2b}, Y_{-2x+2y+a+b}, Z_{2x+y-a-2b}) g_2(X_{-x-y+2a+b}, Y_{-2x+2y-a-2b}, Z_{2x+y+a-b}) \\ &\quad g_2(X'_{-x-2y}, Y'_{-2x-y}, Z'_{2x+2y}) g_2(X'_{-x-2y-a-2b}, Y'_{-2x-y-2a-b}, Z'_{2x+2y+2a+2b}) \\ &\quad g_2(X'_{-x-2y-2a-b}, Y'_{-2x-y+a+2b}, Z'_{2x+2y-a+b}) g_2(X'_{-x-2y+2a+2b}, Y'_{-2x-y-a+b}, Z'_{2x+2y+a-2b}).\end{aligned}$$

If $a, b \in \mathbb{F}_5^n$ are not linearly dependent, then we easily see that all the terms in the product are independent, and linearity of expectation demonstrates

$$\begin{aligned}\beta(a,b) &= \mathbb{E}_{x,y\in\mathbb{F}_5^n} f_1(x,y)f_1(x+a,y+b)f_1(x+2a,y-2b)f_1(x+3a,y-b)(\mathbb{E}g_2(u,v,w))^8 \\ &= \beta_1(a,b)(\mathbb{E}g_2(u,v,w))^8.\end{aligned}$$

Otherwise we have cases depending on the 6 possible linear dependencies. The key point is that, for example,

$$\mathbb{E}_{\mathbf{X},\mathbf{Y},\mathbf{Z}} g_2(X_0,Y_0,Z_0)g_2(X_0,Y_1,Z_1) = \mathbb{E}_{u_0,v_0,w_0,v_1,w_1\in[0,1]} g_2(u_0,v_0,w_0)g_2(u_0,v_1,w_1),$$

so in each of these 6 cases we can reduce to some sort of expectation of $g_2$, or rather, the product of two such terms coming from the independent sets of variables $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ and $(\mathbf{X}',\mathbf{Y}',\mathbf{Z}')$.

Explicit computation yields

$$\begin{aligned}\beta_2(a,0) &= \beta_1(a,0)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_1)g_2(u_2,v_2,w_2)g_2(u_3,v_3,w_3))^2 \\ &= \beta_1(a,0)(\mathbb{E}g_2(u,v,w))^8, \\ \beta_2(a,a) &= \beta_1(a,a)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_0,w_1)g_2(u_0,v_2,w_2)g_2(u_1,v_2,w_0)) \\ &\qquad (\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_1)g_2(u_1,v_2,w_0)g_2(u_3,v_0,w_1)), \\ \beta_2(a,-a) &= \beta_1(a,-a)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_0,v_1,w_1)g_2(u_2,v_0,w_1)g_2(u_2,v_1,w_3)) \\ &\qquad (\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_0)g_2(u_2,v_1,w_2)g_2(u_0,v_3,w_2)), \\ \beta_2(a,2a) &= \beta_1(a,2a)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_1)g_2(u_1,v_2,w_0)g_2(u_3,v_0,w_1)) \\ &\qquad (\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_0,v_1,w_1)g_2(u_2,v_0,w_1)g_2(u_2,v_1,w_3)), \\ \beta_2(a,-2a) &= \beta_1(a,-2a)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_0)g_2(u_2,v_1,w_2)g_2(u_0,v_3,w_2)) \\ &\qquad (\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_0,w_1)g_2(u_0,v_2,w_2)g_2(u_1,v_2,w_0)), \\ \beta_2(0,b) &= \beta_1(0,b)(\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_1,w_1)g_2(u_2,v_2,w_2)g_2(u_3,v_3,w_3))^2 \\ &= \beta_1(0,b)(\mathbb{E}g_2(u,v,w))^8.\end{aligned}$$

Here all variables other than $a, b \in \mathbb{F}_5^n$ are uniform on $[0,1]$. It is worth noting every complicated product of expectations for the middle four terms are very similar. In particular, they are all equal to the product of the densities of two tripartite 3-uniform hypergraphs in the symmetric tripartite hypergraphon $g_2$ (with embeddings respecting the tripartition). Furthermore, the two hypergraphs attained for each term is the same pair of hypergraphs if we disregard the data of the tripartition. For the function $g_2$ we will choose, these considerations will not affect the bounds we prove, so we will focus on a single term.

We first define $g_2$. Let $L \geq 1$ and let $\Lambda$ be a subset of $\mathbb{Z}/L\mathbb{Z}$ avoiding arithmetic progressions of length 3 of size $L\exp(-C\sqrt{\log L})$ for some absolute constant $C > 0$. Let $U = V = W = \mathbb{Z}/L\mathbb{Z}$ and $H$ be a tripartite graph on $U \times V \times W$ with $(s, s+t) \in U \times V$ an edge when $t \in \Lambda$, $(s, s+t) \in V \times W$ an edge when $t \in \Lambda$, and $(s, s+2t) \in U \times W$ an edge when $t \in \Lambda$. Note that the triangles in $H$ are of the form $(s, s+t, s+2t)$ for $t \in \Lambda$ precisely since $\Lambda$ has no nontrivial arithmetic progressions of length 3. This has the special property that every edge is in a unique triangle of $H$ (this is the Ruzsa-Szemerédi graph).

Now let $g_2(u,v,w) = 1$ if $(\lfloor Lu \rfloor, \lfloor Lv \rfloor, \lfloor Lw \rfloor) \pmod L$ encodes one of these triangles. We have

$$\mathbb{E}g_2(u,v,w) = \frac{L|\Lambda|}{L^3} = L^{-1}\exp(-C\sqrt{\log L}).$$

Now we find

$$\mathbb{E}g_2(u_0,v_0,w_0)g_2(u_1,v_0,w_1)g_2(u_0,v_2,w_2)g_2(u_1,v_2,w_0).$$

When we refer to $u \in [0,1]$ in the following discussion, we mean $\lfloor Lu \rfloor \pmod L$. In order for a term to be 1, we must have $u_0v_0w_0$, $u_1v_0w_1$, $u_0v_2w_2$, $u_1v_2w_0$ be triangles. But then $u_0w_0, u_0v_2, v_2w_0$ are all edges in $H$, so $u_0v_2w_0$ is also a triangle. This forces $v_2 = v_0$ (as edges are in unique triangles). Then $u_1v_0w_0$ is a triangle, so $u_1 = u_0$. Finally, this means $u_0v_0w_1$ is a triangle so $w_0 = w_1$. Thus the expectation equals

$$L^{-4}\mathbb{E}g_2(u,v,w) = \frac{L|\Lambda|}{L^7} = L^{-5}\exp(-C\sqrt{\log L}).$$

Next, we find

$$\mathbb{E}g_2(u_0, v_0, w_0)g_2(u_1, v_1, w_1)g_2(u_1, v_2, w_0)g_2(u_3, v_0, w_1) \leq \frac{L^4}{L^8} = L^{-4}$$

since there are at most $L^4$ ways to choose $w_0, u_1, w_1, v_0$. After that, the variables $u_0, v_1, v_2, u_3$ are forced in order to guarantee a nonzero term, since edges are in unique triangles. This demonstrates that

$$\beta_2(a, \lambda a) \leq \beta_1(a, \lambda a)L^{-9} \exp(-C\sqrt{\log L}) \leq L^{-9+o(1)}$$

for $\lambda \in \mathbb{F}_5^*$. Now we take $L$ sufficiently large. First, if $a, b \in \mathbb{F}_5^n$ are linearly independent, then by Section 3.6.1 and the above we have

$$\beta_2(a, b) = \beta_1(a, b)(\mathbb{E}g_2(u, v, w))^8 \leq \left(\frac{73}{80} + O(5^{-n/2})\right)\alpha_2^4.$$

On the other hand, if $b = \lambda a$ for $\lambda \in \mathbb{F}_5^*$ then

$$\beta_2(a, b) \leq L^{-9+o(1)} \leq \frac{1}{2}\alpha_2^4$$

since $\alpha_2$ grows as $L^{-2+o(1)}$ times $\mathbb{E}f_1(x, y) = 2/5 + O(5^{-n/2})$.

Therefore, in expectation our random function satisfies the conclusion of Theorem 3.1.3 for $(a, b) \in (\mathbb{F}_5^n \setminus 0)^2$. To obtain a single function which satisfies all these inequalities (of which there are $O(|G|)$), we use a concentration argument. A straightforward extension of [SSZ21, Lemma 3.2] suffices. Roughly note that each of the above quantities can be controlled using Azuma–Hoeffding inequality on the Doob-martingale where one successively conditions on each of the random variables $X_x, Y_y, Z_z, X'_x, Y'_y, Z'_z$ for $x, y, z \in \mathbb{F}_5^n$ as each random variables only participates in a small number of terms. The quality of concentration is at least $\exp(-|G|^c)$ for deviations on the scale of $1/|G|^{1/4}$, which is more than sufficient. Let $h(x, y)$ now denote a specific instantiation of the above defined random function which

satisfies

$$\sup_{(a,b)\in(\mathbb{F}_5^n\setminus 0)^2} \mathbb{E}[h(x,y)h(x+a,y+b)h(x+2a,y-2b)h(x+3a,y-b)] \leq (1-\delta)\mathbb{E}[h(x,y)]^4$$

for an absolute constant $\delta > 0$.

### 3.6.3 Finishing the construction

Finally we are in position to fix the directions $(a,b)$ where $a = 0$ or $b = 0$. We now use a randomized version of the construction from [SSZ21, Section 2] in order to eliminate these final special differences. Define $T = \{0,1,2\}^\gamma \times (\mathbb{F}_5)^{n-\gamma}$, where $\gamma \geq 1$ is an integer to be chosen later. Note that this set has density $\beta = (3/5)^\gamma$ but a four term arithmetic progression density of $(3/25)^\gamma \leq \beta^{4.15}$ noting that the set $\{0,1,2\}$ has no nontrivial four term arithmetic progressions. Now for each $g \in \mathbb{F}_5^n$ choose a two uniformly random bijective affine transformations $\phi(g), \phi'(g)$ of $\mathbb{F}_5^n$. Then let

$$f(x,y) = h(x,y)\mathbb{1}_{x\in\phi(y)S}\mathbb{1}_{y\in\phi'(x)S}.$$

First, we have

$$\mathbb{E}_{\phi,\phi'}[\mathbb{E}_{x,y}f(x,y)] = \mathbb{E}_{x,y}[h(x,y)\mathbb{E}_{\phi,\phi'}[\mathbb{1}_{x\in\phi(y)S}\mathbb{1}_{y\in\phi'(x)S}]] = \beta^2\mathbb{E}[h(x,y)].$$

Similarly for $(a,b)$ with nonzero coordinates we find that

$$\begin{aligned}
&\mathbb{E}_{\phi,\phi'}[\mathbb{E}_{x,y}f(x,y)f(x+a,y+b)f(x+2a,y-2b)f(x+3a,y-b)] \\
&\quad = \beta^8\mathbb{E}_{x,y}[h(x,y)h(x+a,y+b)h(x+2a,y-2b)h(x+3a,y-b)] \\
&\quad \leq (1-\delta)\beta^8(\mathbb{E}h(x,y))^4.
\end{aligned}$$

Finally assume that exactly one of $a$ or $b$ is zero. We handle the case when $a$ is zero as the other is analogous. We find

$$\begin{aligned}
&\mathbb{E}_{\phi,\phi'}[\mathbb{E}_{x,y} f(x,y)f(x,y+b)f(x,y-2b)f(x,y-b)] \\
&\quad \leq \mathbb{E}_{\phi,\phi',x,y}[\mathbb{1}_{x\in\phi(y)S}\mathbb{1}_{y\in\phi'(x)S}\mathbb{1}_{x\in\phi(y+b)S}\mathbb{1}_{y+b\in\phi'(x)S}\mathbb{1}_{x\in\phi(y-2b)S}\mathbb{1}_{y-2b\in\phi'(x)S}\mathbb{1}_{x\in\phi(y-b)S}\mathbb{1}_{y-b\in\phi'(x)S}] \\
&\quad = \beta^4 \mathbb{E}_{\phi',x,y}[\mathbb{1}_{y\in\phi'(x)S}\mathbb{1}_{y+b\in\phi'(x)S}\mathbb{1}_{y-2b\in\phi'(x)S}\mathbb{1}_{y-b\in\phi'(x)S}] \\
&\quad \leq \beta^{8.15}.
\end{aligned}$$

The last line follows since for every random map $\phi'(x)$, there is at most a density of $\beta^{4.15}$ of four term arithmetic progressions in $\phi'(x)S$. Therefore taking $\gamma$ to be a sufficiently large multiple of $\log(1/\mathbb{E}[h(x,y)])$ (which is of constant order due to Section 3.6.2) guarantees that each of these expectations is at most $(1-\delta)\beta^8(\mathbb{E}h(x,y))^4$. Furthermore, each of the above densities concentrates (with respect to the randomness of $\phi, \phi'$) with high probability by Azuma–Hoeffding, so Theorem 3.1.3 follows.

## 3.7 Popular differences for 3-point patterns

We end by proving popularity for all admissible three-point patterns, namely, Theorem 3.1.1. As is standard, we actually prove an analogous result over all compact abelian groups. We deduce the version over $\mathbb{Z}$ from results (mod $N$) using a trick of Green [Gre05b].

These patterns can be handled in a direct Fourier-analytic manner. The proof here is closely modeled after the proof for 3-term arithmetic progressions. For a compact abelian group $G$ with Haar measure $\mu$, finite $S \subseteq \widehat{G}$, and $\delta > 0$, define the Bohr set

$$B(S,\delta) = \{x \in G : \max_{\xi\in S}(\|\xi x\|_{\mathbb{R}/\mathbb{Z}}) < \delta\}.$$

We recall the standard fact that $\mu(B(S,\delta)) = \Omega_{|S|,\delta}(1)$ (see [TV06, Lemma 4.20]). We will further write for a Bohr set $B$ that $\mu_B$ is the uniform measure on it obtained by restricting the Haar measure appropriately.

**Theorem 3.7.1.** *For any $\alpha, \epsilon > 0$ there exists $C(\epsilon) > 0$ so that the following holds. Let*

*$G$ be a compact abelian group with Haar probability measure $\mu$. Let $M_1, M_2$ be continuous automorphisms of $G$ such that $M_1 - M_2$ is an automorphism. Then for any function $f : G \to [0,1]$ with $\int_{g\in G} f(g)\mu(g) \geq \alpha$, there is a measure $\nu_D$ with $\|\nu_D\|_{L^\infty} \leq C$ such that*

$$\int_{x\in G, d\in G} f(x)f(x+M_1d)f(x+M_2d)\nu_D(d)\mu(x) \geq \alpha^3 - \epsilon. \tag{3.3}$$

*Moreover, given a Bohr set $B_0 = B(S_0, \rho_0)$, one may take $\nu_D = \mu_B * \mu_B$, where $B = B(S, \rho)$ with $S \supseteq S_0$ and $\rho \leq \rho_0$. In this case, the absolute constant $C$ will also depend on $|S_0|, \rho_0$.*

From this general statement one can easily obtain popular difference results in their standard forms for finite abelian groups (Theorem 3.7.3) and for $\mathbb{Z}$ (Theorem 3.1.1). The latter answers (as a special case) [ABB21, Question 1.16] and is the combinatorial analog of [ABB21, Theorem 1.10]. As much of the proof is identical to that of the three-term arithmetic progression case in [Tao14], we collect the necessary results in the following proposition. Essentially we are extracting a statement of the strong regularity lemma from [Tao14]. Closely related statements appear in [Bou86, Gre07].

**Proposition 3.7.2** ([Tao14])**.** *Fix a compact abelian group $G$, parameters $\delta, \epsilon > 0$, and a set $S_0 \subseteq \widehat{G}$, as well as growth functions $\omega_1, \omega_2 : \mathbb{R}^+ \to \mathbb{R}^+$. Given a function $f : G \to [0,1]$, there are $\gamma_1, \gamma_2 > 0$ and a finite set $T$ satisfying $S_0 \subseteq T \subseteq \widehat{G}$ with the following properties.*

1. *$|T| = O_{\delta,\epsilon,|S_0|,\omega_1,\omega_2}(1)$.*
2. *$\gamma_1 \leq 1/\omega_1(|T| + \delta^{-1} + \epsilon^{-1})$ and $\gamma_2 \leq 1/\omega_2(\gamma_1^{-1})$, but $\gamma_2$ is bounded away from 0 independent of $f$.*
3. *We have the decomposition $f = f_{\mathrm{str}} + f_{\mathrm{sml}} + f_{\mathrm{psr}}$, where*
   (a) *$f_{\mathrm{str}}$, $f_{\mathrm{sml}}$, $f_{\mathrm{psr}}$ are 1-bounded.*
   (b) *$f_{\mathrm{str}}$ is nonnegative, has mean $\int f_{\mathrm{str}}\, d\mu = \int f\, d\mu$, and obeys the bound*

$$f_{\mathrm{str}}(x+r) = f_{\mathrm{str}}(x) + O(\epsilon)$$

*whenever $x \in G$ and $r \in B(T, \gamma_1)$.*

*(c)* $\|f_{\text{sml}}\|_{L^2(G)} \leq \epsilon$.

*(d)* $\|\widehat{f}_{\text{psr}}\|_{\ell^\infty(\widehat{G})} \leq \gamma_2$.

At this point, Tao [Tao14] studies progressions with common difference in $B(T, \gamma_1)$. We require a small modification to handle more general matrix patterns. Define

$$B' = \{r \in G : M_1 r, M_2 r \in B(T, \gamma_1)\}$$

Observe that $B'$ is the Bohr set $B(T', \gamma_1)$, where

$$T' = \{\xi \circ M_1 : \xi \in T\} \cup \{\xi \circ M_2 : \xi \in T\}$$

The fact that $\xi \circ M_1$ and $\xi \circ M_2$ are elements of $\widehat{G}$ is immediate by the identification $\widehat{G} = \text{Hom}(G, \mathbb{R}/\mathbb{Z})$ and the fact that $M_1$ and $M_2$ are automorphisms. As an immediate consequence, we have for any $x \in G, d \in B'$ that

$$f_{\text{str}}(x + M_1 d), f_{\text{str}}(x + M_2 d) = f_{\text{str}}(x) + O(\epsilon). \tag{3.4}$$

With this additional Lipschitz condition in the directions of $M_1$ and $M_2$ within $B'$, the remainder of the proof follows in the standard manner. We will now evaluate

$$\int f(x) f(x + M_1 d) f(x + M_2 d) \mu_{B'} * \mu_{B'}(d) \mu(x).$$

Decompose each occurrence of $f$ into $f_{\text{str}} + f_{\text{sml}} + f_{\text{psr}}$, so that the integral has 27 terms of the form $f_1(x) f_2(x + M_1 d) f_3(x + M_2 d)$. We will show that all terms other than the $f_{\text{str}}, f_{\text{str}}, f_{\text{str}}$ term are bounded in magnitude by $O(\epsilon)$.

We first bound terms containing $f_{\text{sml}}$; we consider the representative case where the second term is $f_{\text{sml}}$. In this case, take absolute values, and bound $f_1$, $f_3$ in magnitude by 1. We have

$$\int f_1(x) f_{\text{sml}}(x + M_1 d) f_3(x + M_2 d) \mu_{B'} * \mu_{B'}(d) \mu(x) \leq \int \mu_{B'} * \mu_{B'}(d) \int |f_{\text{sml}}(x + M_1 d)| \mu(x).$$

The inner integral is bounded by $\|f_{\text{sml}}\|_{L^1} \leq \|f_{\text{sml}}\|_{L^2} \leq \epsilon$. Thus we are left with a bound of

$O(\epsilon)$ as desired.

We next bound the terms containing $f_{\text{psr}}$. For the sake of simplicity we consider the case where the second term is $f_{\text{psr}}$. Standard Fourier analysis allows us to obtain that

$$\begin{aligned}&\int f_1(x)f_{\text{psr}}(x+M_1d)f_3(x+M_2d)\mu_{B'}*\mu_{B'}(d)\mu(x)\\&\quad=\sum_{\xi_4}|\widehat{\mu}_{B'}|^2(\xi_4)\sum_{\xi_1M_1+\xi_3(M_1-M_2)=\xi_4}\widehat{f_1}(\xi_1)\widehat{f}_{\text{psr}}(-\xi_1-\xi_3)\widehat{f_3}(\xi_3)\end{aligned}$$

We take absolute values and use $|\widehat{f}_{\text{psr}}|\leq\gamma_2$. As $M_1$ and $M_1-M_2$ are invertible, Cauchy–Schwarz lets us bound the inner sum by

$$\gamma_2\sum_{\xi_1}|\widehat{f_1}(M_1^{-1}\xi_1)|^2\sum_{\xi_3}|\widehat{f_3}((M_1-M_2)^{-1}\xi_3)|^2\leq\gamma_2$$

where the final inequality follows by Plancherel. Plancherel also implies $\sum_{\xi_4}|\widehat{\mu}_{B'}|^2=\|\mu_{B'}\|_{L^2}^2\leq\|\mu_{B'}\|_{L^\infty}=O_{|T'|,\gamma_1^{-1}}(1)$. Recalling $|T'|\leq2|T|$ and that $\gamma_1$ is small with respect to $|T|$, we find that

$$\sum_{\xi_4}|\widehat{\mu}_{B'}|^2=O_{\gamma_1^{-1}}(1).$$

The total contribution is therefore bounded by $\gamma_2\cdot O_{\gamma_1^{-1}}(1)\leq\epsilon$ as long as $\omega_1,\omega_2$ grow fast enough.

Finally we are left with the term

$$\int f_{\text{str}}(x)f_{\text{str}}(x+M_1d)f_{\text{str}}(x+M_2d)\mu_{B'}*\mu_{B'}(d)\mu(x).$$

Since $\mu_{B'}*\mu_{B'}(d)$ is supported on $B'+B'$, for any $d$ in this support we have

$$f_{\text{str}}(x+M_id)=f_{\text{str}}(x)+O(2\epsilon)$$

for $i\in\{1,2\}$ by triangle inequality and two applications of (3.4) each. Rewriting, and recalling $f_{\text{str}}$ is $[0,1]$-valued, we have

$$\begin{aligned}\int f_{\mathrm{str}}(x) f_{\mathrm{str}}(x+M_1 d) f_{\mathrm{str}}(x+M_2 d)\mu_{B'} * \mu_{B'}(d)\mu(x) & \\ &= \int f_{\mathrm{str}}^3(x)\mu_{B'} * \mu_{B'}(d)\mu(x) + O(\epsilon) \\ &\geq \left(\int f_{\mathrm{str}}(x)\mu(x)\right)^3 + O(\epsilon) \qquad \text{(Hölder's inequality)} \\ &\geq \alpha^3 + O(\epsilon).\end{aligned}$$

This completes the proof. □

We now deduce a pair of corollaries of this result.

**Theorem 3.7.3.** *For any $\alpha, \epsilon > 0$ there exists $N_0$ so that the following holds. Let $G$ be a finite abelian group of order $N \geq N_0$. Let $M_1, M_2$ be automorphisms of $G$ so that $M_1 - M_2$ is an automorphism. Then for any $A \subseteq G$ with $|A| \geq \alpha N$, there is a popular difference $d \neq 0$ so that*

$$\#\{x \in G : x, x + M_1 d, x + M_2 d \in A\} \geq (\alpha^3 - \epsilon)N.$$

*Proof.* Choose $N_0 = C/\epsilon$, where $C = C(\epsilon)$ is the constant from Theorem 3.7.1. Given this choice, the contribution from $d = 0$ in (3.3) is at most $\epsilon N$, and therefore by the pigeonhole principle, some nonzero $d$ in the support of $\mu_D$ must have at least $\alpha^3 - O(\epsilon)$ density of this three-point pattern. This gives the desired result with an adjusted value of $\epsilon$. □

Finally we deduce the statement over the interval $[N]$ which was stated in the introduction.

*Proof of Theorem 3.1.1.* Choose $N_0 = C/\epsilon$, where $C = C(\epsilon, |S_0| = 2k, \delta_0 = \epsilon/(2k))$ is the constant from Theorem 3.7.1. Embed $S \subseteq [p]^k \hookrightarrow (\mathbb{Z}/p\mathbb{Z})^k$ naturally, where $p$ is a prime with $N < p < (1 + \epsilon/k)N$. Initialize $\delta_0 = \epsilon/(2k)$, and

$$S_0 = \left\{x \mapsto \frac{(M_1 x)_i}{p}\right\}_{i \in [k]} \cup \left\{x \mapsto \frac{(M_2 x)_i}{p}\right\}_{i \in [k]}.$$

Given this setup, we apply Theorem 3.7.1. Our assumptions ensure that our matrices have nonzero determinant (and are therefore invertible) modulo $p$. The contribution from $d = 0$ is at most $\epsilon$ as before, and we can find at least $(\alpha^3 - O(\epsilon))N^k$ patterns with common

difference $d \neq 0$ for some $d \in B(S,\delta) + B(S,\delta) \subseteq B(S,2\delta)$, where $S \supseteq S_0$ and $\delta \leq \delta_0$. (Note that the density may decrease by $O(\epsilon)$ under this embedding.) Plugging in the elements of $S_0$ and using $\delta \leq \epsilon/(2k)$ gives us the constraints

$$(M_1 x)_i, (M_2 x)_i \in [-\epsilon p/k, \epsilon p/k], \text{ for all } i \in [k], x = (x_i)_{i=1}^k \in B(S, 2\delta).$$

Now we attempt to lift these $(\alpha^3 - \epsilon)N^k$ patterns into $\mathbb{Z}$ by viewing $d$ and each choice of $x$ as integer vectors. Throw out at most $\epsilon N^k$ choices of $x$ for which $x_i \notin [(\epsilon/k)p, (1-\epsilon/k)p]$ for some $i$. For the remainder, we see by triangle inequality in each coordinate that $x, x+M_1 d, x+M_2 d \in [p]^k$. By assumption, this triple must map to a three-point pattern under our embedding $[p]^k \hookrightarrow (\mathbb{Z}/p\mathbb{Z})^k$, so each of these points must have been an element of $A$ originally, and the triple forms a three-point pattern over $\mathbb{Z}$. We have therefore found $(\alpha^3 - O(\epsilon))N^k$ of the necessary 3-point patterns in $[N]^k$ as desired, which completes the proof upon adjusting $\epsilon$. □

*Remark* 3.7.4. As usual, one can actually find $\Omega_{\alpha,\epsilon}(N^k)$ differences by using Markov's inequality and adjusting $\epsilon$.

# Chapter 4

# Testing linear-invariant properties

In property testing, the aim is to find randomized algorithms that distinguish objects that have some property from those that are far from satisfying the property by querying the given large object at a small number of locations. Property testing emerged from the linearity test of Blum, Luby, and Rubinfeld [BLR93], and was formally defined and systematically studied by Rubinfeld and Sudan [RS07] and Goldreich, Goldwasser, and Ron [GGR98]. There have been important developments especially in the following two settings: graph property testing and arithmetic property testing.

Two representative problems are: (1) given a large graph, test whether the graph is triangle-free or $\epsilon$-far from triangle-free (an $n$-vertex graph is *$\epsilon$-far* from a graph property if one needs to add and/or remove more than $\epsilon n^2$ edges in order to satisfy the property), and (2) given a function $f\colon \mathbb{F}_p^n \to \mathbb{F}_p$, test whether $f$ is linear or $\epsilon$-far from linear (a function is *$\epsilon$-far* from an arithmetic property if one needs to change the value of the function on more than an $\epsilon$-fraction of the domain in order to satisfy the property). In both cases, it is known that one can achieve the desired goal by sampling a fixed number of entries repeatedly $C(\epsilon)$ times. For testing whether a graph is triangle-free [RS78], one samples a uniform random triple of vertices and checks whether they form a triangle, and for testing linearity [BLR93], one samples $x, y \in \mathbb{F}_p^n$ uniformly and checks if $f(x) + f(y) = f(x + y)$.

In this chapter we give a property testing algorithm for a very general class of arithmetic properties. The goal is to determine whether a function $f\colon \mathbb{F}_p^n \to [R] := \{1, \ldots, R\}$ (with fixed prime $p$ and positive integer $R$) satisfies some given property or is $\epsilon$-far from satisfying

the property. All the properties we consider are *linear-invariant* in the sense that they are invariant under automorphisms of the domain vector space.[1] Linear-invariant properties form an important general class of arithmetic properties, e.g., the work of Kaufman and Sudan [KS08] "highlights linear-invariance as a central theme in algebraic property testing."

An oblivious tester is one which works as follows: the tester produces a positive integer $d = d(\epsilon)$ and an oracle provides the tester with the restriction $f|_U$ where $U$ is a uniform random $d$-dimensional linear subspace of the domain (if the domain is large enough that such a subspace exists; if the domain has dimension strictly less than $d$, the oracle provides the tester with all of $f$). We say that a property $\mathcal{P}$ is *testable* if there exists an oblivious tester with one-sided error (and constant query complexity) for the property. That is, we require our tester to accept functions $f$ satisfying $\mathcal{P}$ with probability 1 and reject functions that are $\epsilon$-far from satisfying $\mathcal{P}$ with probability at least $\delta = \delta(\epsilon)$ for some function $\delta\colon (0,1) \to (0,1)$. Furthermore, we say that $\mathcal{P}$ is *proximity oblivious-testable* (PO-testable) if $d = d(\epsilon)$ is a constant independent of $\epsilon$. The idea of PO-testability was introduced by Goldreich and Ron [GR11] who, among other results, classified the PO-testable graph properties.

One surprising feature of property testing is that many natural properties, such as linearity, are testable and even PO-testable. A key feature of linearity is that it is subspace-hereditary meaning that if $f\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is linear, then the same is true for $f|_U$ for every linear subspace $U \leq \mathbb{F}_p^n$. To be precise, we say that a linear-invariant property $\mathcal{P}$ is *subspace-hereditary* if for every $f\colon \mathbb{F}_p^n \to [R]$ satisfying $\mathcal{P}$ and every linear subspace $U \leq \mathbb{F}_p^n$, the restriction $f|_U$ also satisfies $\mathcal{P}$.

A central conjecture in this field, first proposed by Bhattacharyya, Grigorescu, and Shapira, is that all linear-invariant, subspace-hereditary properties are testable [BGS15, Conjecture 4]. In fact, they conjecture that the slightly larger class of *semi subspace-hereditary* properties are testable and prove that no other properties can be tested.

**Definition 4.0.1.** A linear-invariant property $\mathcal{P}$ is *semi subspace-hereditary* if there exists a subspace-hereditary property $\mathcal{Q}$ such that

[1] It would also be possible to consider this problem with base field $\mathbb{F}_q$ for a prime power $q$. While many of our techniques extend to this setting, there are also many increased technical difficulties, so we do not attempt to consider this problem here.

(i) every function satisfying $\mathcal{P}$ also satisfies $\mathcal{Q}$;

(ii) for all $\epsilon > 0$, there exists $N(\epsilon)$ such that if $f\colon \mathbb{F}_p^n \to [R]$ satisfies $\mathcal{Q}$ and is $\epsilon$-far from satisfying $\mathcal{P}$, then $n < N(\epsilon)$.

It is known that there are subspace-hereditary properties where the dimension $d$ sampled must grow as the proximity parameter $\epsilon$ approaches 0. To be PO-testable, a property must satisfy the following more restrictive condition.

**Definition 4.0.2.** A linear-invariant property $\mathcal{P}$ is *locally characterized* if there exists some $d$ such that the following holds. For every $f\colon \mathbb{F}_p^n \to [R]$ with $n \geq d$, the function $f$ satisfies $\mathcal{P}$ if and only if $f|_U$ satisfies $\mathcal{P}$ for every $U \leq \mathbb{F}_p^n$ of dimension $d$.

Our first result is a resolution of the conjecture of Bhattacharyya, Grigorescu, and Shapira, classifying the testable linear-invariant properties. Our second result is a classification of the PO-testable linear-invariant properties.

**Theorem 4.0.3.** *A linear-invariant property is testable if and only if it is semi subspace-hereditary.*

**Theorem 4.0.4.** *A linear-invariant property is PO-testable if and only if it is locally characterized.*

*Remark* 4.0.5. Note that under our definition of an oblivious tester, the tester does not know the dimension of the domain. This rules out some "unnatural" properties such as those properties that behave differently depending on whether the dimension of the domain is even or odd.

Previous work in arithmetic property testing has focused on a number of special cases including monotone properties [KSV12, Sha10], "complexity 1" properties over $\mathbb{F}_2$ [BGS15], and bounded complexity translation-invariant properties [BFH$^+$13].

We note that very little was previously known about general linear-invariant properties. One simple way to define a class of linear-invariant properties can be done by, for example, choosing an arbitrary subset of "allowable" maps $\mathbb{F}_p^2 \to [R]$ and defining a property of functions $\mathbb{F}_p^n \to [R]$ to consist of those whose restriction to every 2-dimensional linear subspace is

allowable. Even this class of 2-dimensionally-defined patterns was not known to be testable in general prior to this work.

Our innovations are two-fold. We prove a strong arithmetic regularity lemma which, unlike previous arithmetic regularity lemmas, mixes different levels of Gowers uniformity. This allows us to give a more powerful version of the *compactness argument* first introduced by Alon and Shapira [AS08]. With this tool we can remove the *bounded complexity* restriction that was present in all previous work.

Our second innovation is a novel recoloring technique we call *patching*. This technique is critical for working in the linear-invariant setting and allows us to handle an important obstacle encountered by previous works.

This chapter is based on the paper [TZ20].

## 4.1 Proof techniques and previous work

### 4.1.1 Graph removal lemmas and property testing

We begin with an overview of graph removal lemmas and their proof techniques (see the survey [CF13] for a more detailed discussion).

The foundational result in this field is the triangle removal lemma of Ruzsa and Szemerédi [RS78]. This result states that for all $\epsilon > 0$ there exists $\delta > 0$ such that any $n$-vertex graph with at most $\delta n^3$ triangles can be made triangle-free by removing $\epsilon n^2$ edges. It is immediate from the definitions that this result implies that triangle-freeness is testable (and in fact PO-testable). In general we will work with such *removal lemmas* and deduce corresponding property testing results from them. The triangle removal lemma was generalized to the graph removal lemma, first stated explicitly by Alon, Duke, Lefmann, Rödl, and Yuster [ADL$^+$94] and by Füredi [Für95].

A key tool for proving the graph removal lemma is a regularity lemma, namely Szemerédi's graph regularity lemma. Roughly speaking, the proof proceeds by using this regularity lemma to partition the input graph $G$ into a small number of structured components. Then we "clean up" $G$ by removing at most $\epsilon n^2$ edges. This is done in such a way that either the

resulting graph is $H$-free or the original graph $G$ contains many copies of $H$.

An important extension of the graph removal lemma is the induced graph removal lemma, proved by Alon, Fischer, Krivelevich, and Szegedy [AFKS00]. The induced graph removal lemma states that for every graph $H$, for all $\epsilon > 0$ there exists $\delta > 0$ such that every $n$-vertex graph with at most $\delta n^{v(H)}$ induced copies of $H$ can be made induced $H$-free by adding and/or removing at most $\epsilon n^2$ edges (here *induced $H$-free* means not containing any induced subgraph isomorphic to $H$).

The original proof of the induced graph removal lemma relies on an extension of Szemerédi's graph regularity lemma known as the "strong regularity lemma." Using such a regularity lemma combined with a random sampling argument, one can produce a "regular model", that is, a large induced subgraph $X := G[U]$ (on a constant fraction of the vertices of $G$) that is very regular and approximates the original graph well in a certain sense. Then we "clean up" $G$ by adding and/or removing at most $\epsilon n^2$ edges in such a way that if the resulting graph is not induced $H$-free then $X$ (in the original graph) must contain many induced copies of $H$.

Both the graph removal lemma and induced graph removal lemma can easily be extended to remove any finite collection $\mathcal{H}$ of graphs. Alon and Shapira [AS08] extended the induced graph removal lemma to an infinite collection of graphs. Namely they prove that for a (possibly infinite) set $\mathcal{H}$ of graphs and for $\epsilon > 0$ there exist $\delta > 0$ and $k$ such that the following holds: if $G$ is an $n$-vertex graph with at most $\delta n^{v(H)}$ copies of $H$ for all $H \in \mathcal{H}$ with $k$ or fewer vertices, then $G$ can be made induced $\mathcal{H}$-free by adding and/or removing at most $\epsilon n^2$ edges (meaning the modified graph has no induced subgraph isomorphic to $H$ for every $H \in \mathcal{H}$). Despite the strange statement of this result (the hypothesis that $G$ has low $H$-density only for the small graphs $H \in \mathcal{H}$) this theorem immediately implies (and is equivalent to) the fact that every hereditary graph property is testable with constant query-complexity and one-sided error.

This series of works, in addition to being important results in their own right, gives a framework for proving constant query-complexity property testing algorithms in other settings, given an appropriate regularity lemma. For example a hypergraph regularity lemma is known, proven by Gowers [Gow07] and independently by to Rödl et al. [RNS+05]. Using the above framework, one can use the hypergraph regularity lemma to deduce an infinite induced

hypergraph removal lemma [RS07]. Consequently, every hereditary hypergraph property is testable with constant query-complexity and one-sided error. We will be interested in seeing if this framework, combined with an appropriate arithmetic regularity lemma, can prove an infinite induced arithmetic removal lemma.

### 4.1.2 Arithmetic analogs

The problem of property testing for functions $f\colon \mathbb{F}_p^n \to [R]$ has been intensively studied, starting with the the classic work of Blum, Luby, and Rubinfeld [BLR93] on linearity testing. Much of the work focuses on testing whether some function $f\colon \mathbb{F}_p^n \to \mathbb{F}_p$ has certain algebraic properties (e.g., a polynomial of some given type) [AKK$^+$05, KS08]. There is also much interest in testing properties that do not arrive from algebraic characterizations. Below we give an overview of the developments related to property testing in $\mathbb{F}_p^n$ from a perspective that is parallel to the graph regularity method developments discussed earlier.

The first arithmetic regularity lemma was proved by Green [Gre05b] using Fourier-analytic techniques, and it laid the groundwork for further developments of the regularity method in the arithmetic setting. These regularity lemmas have since found many applications in additive combinatorics and related fields. In particular, combined with the graph removal framework described above, Green's regularity lemma is suitable for proving an arithmetic removal lemma for "complexity 1" systems of linear forms (see Section 4.3 for the definition of complexity; roughly speaking, a system of linear forms is complexity 1 if it can be controlled by Fourier-analytic means); e.g., see [BCSX11].

Král', Sera, and Vena [KSV12] and independently Shapira [Sha10] bypass the need for an arithmetic regularity lemma and prove the full arithmetic removal lemma by a direct reduction from the hypergraph removal lemma. Their results imply that all linear-invariant, subspace-hereditary *monotone* properties are testable with constant query-complexity and one-sided error. (A property of functions $\mathbb{F}_p^n \to \{0,1\}$ is *monotone* if changing 1's to 0's preserves the property.)

Note that the above result is an arithmetic removal lemma and not an *induced* arithmetic removal lemma (hence the restriction to monotone properties). Due to the nature of the reduction, the techniques do not seem to be capable of deducing the induced arithmetic

removal lemma from the induced hypergraph removal lemma.

An alternative approach is to apply the strong graph regularity approach [AFKS00] of proving the induced graph removal lemma to Green's arithmetic regularity lemma. However there is also a major obstacle to the approach, related to the fact that the origin plays a special role in a vector space while there is no corresponding feature of graphs. It turns out that it is not always possible to regularize the space in a neighborhood of the origin [GS15].

Bhattacharyya, Grigorescu, and Shapira [BGS15] managed to overcome this obstacle in the special case of vector spaces over $\mathbb{F}_2$. They follow the above strategy, implementing the strong regularity idea [AFKS00] in the style of Green's arithmetic regularity [Gre05b] along with one additional tool, namely a Ramsey-theoretic result, to prove an infinite induced arithmetic removal lemma for complexity 1 patterns over $\mathbb{F}_2$. Unfortunately, it is known [GS15] that this Ramsey-theoretic result fails over all finite fields other than $\mathbb{F}_2$.

Bhattacharyya, Fischer, and Lovett [BFL12] managed to overcome this obstacle in a different special case, namely for translation-invariant patterns. When all patterns considered are translation-invariant, the origin no longer plays a special role and one can essentially ignore it while carrying out the strong regularity framework. In addition, [BFL12] allows one to handle higher complexity patterns, which requires developing and applying tools from higher-order Fourier analysis.

Higher-order Fourier analysis plays a central role in modern additive combinatorics. These techniques were initiated by Gowers [Gow01] in his celebrated new proof of Szemerédi's theorem, and further developed in a sequence of works by Green, Tao, and Ziegler [GT08b, GT10c, GTZ12] settling classical conjectures on the asymptotics of prime numbers patterns. A parallel theory of higher-order Fourier analysis was developed in finite field vector spaces by Bergelson, Tao, and Ziegler [BTZ10, TZ10, TZ12], leading to an inverse Gowers theorem over finite fields vector spaces.

For applications to property testing, this line of work culminated in the work of Bhattacharyya, Fischer, Hatami, Hatami, and Lovett [BFH$^+$13] (extending [BFL12]), who applied the inverse Gowers theorem over finite fields and developed further equidistribution tools to prove an infinite induced arithmetic removal lemma for all linear-invariant, subspace-hereditary properties that are also translation-invariant and bounded-complexity. Their work

follows the strong regularity framework of [AFKS00, AS08]. Our results improve upon this work by removing the translation-invariant and bounded-complexity restrictions.

In addition to their property testing algorithm, Bhattacharyya, Fischer, Hatami, Hatami, and Lovett [BFH+13] proved that a large class of somewhat algebraically structured properties are indeed affine-invariant, subspace-hereditary, and locally characterized. These are the so-called "degree-structural properties". A simple extension of their result [HHL19, Theorem 16.3] implies that the larger class of "homogeneous degree-structural properties" are linear-invariant, linear subspace-hereditary (but not affine-invariant and not subspace-hereditary), and locally characterized, and thus these properties are testable by our main theorem. As an example, one can test whether a function $\mathbb{F}_p^n \to \mathbb{F}_p$ can be written as $A^2 + B^2$ where both $A$ and $B$ are homogeneous polynomials of some given degree $d$.

### 4.1.3 Our contributions

#### Patching

In this chapter, building on the work in Chapter 2 for complexity 1 patterns, we develop a new technique called "patching" that allows us to overcome the obstacle faced by earlier approaches, namely that a neighborhood of the origin cannot be regularized and fails certain Ramsey properties (unless working over $\mathbb{F}_2$). In essence, the patching result states that if there exists some map $f\colon \mathbb{F}_p^n \to [R]$ that has low density of some colored patterns $\mathcal{H}$ for $n$ large enough, then for all $m$ there must exist some map $g\colon \mathbb{F}_p^m \to [R]$ that has no $\mathcal{H}$-instances. (Informally, a colored pattern is a linear configuration with specified colors of the points.)

**Theorem 4.1.1** (Informal patching result)**.** *For every set of colored patterns $\mathcal{H}$, there exist $\epsilon_0 > 0$ and $n_0$ such that the following holds. Either:*

- *for every $n$, there exists a function $f\colon \mathbb{F}_p^n \to [R]$ that is $\mathcal{H}$-free; or*
- *for every function $f\colon \mathbb{F}_p^n \to [R]$ with $n \geq n_0$, the $H$-density in $f$ is at least $\epsilon_0$ for some $H \in \mathcal{H}$.*

Our proof proceeds in two steps. First, as in [BFH+13], following the strong regularity framework of [AFKS00] for proving induced graph removal lemmas, we apply a strong

arithmetic regularity lemma, which produces a partition $\mathfrak{B}$ of $\mathbb{F}_p^n$ and a "regular model" $X \subseteq \mathbb{F}_p^n$ made up of a randomly sampled set of atoms from $\mathfrak{B}$. Unlike in the graph setting, we cannot ensure that the map $f\colon \mathbb{F}_p^n \to [R]$ is very regular on every atom of $\mathfrak{B}|_X$. In particular, it may be impossible to guarantee that $f$ is regular on the atom containing the origin. Instead we only ensure that almost every atom of $X$ is very regular. Unlike earlier proofs of removal lemmas, our "recoloring algorithm" has two components: for the regular atoms we "clean up" $f$ as usual, while for the irregular atoms we apply our patching result. Our patching result implies that there is some new global coloring $g\colon \mathbb{F}_p^n \to [R]$ that avoids some appropriate set of colored patterns. To complete the proof we "patch" $f$ by replacing it by $g$ on all of the irregular atoms. If $f$ has low density of some set of colored pattern, then our argument shows that these patterns cannot appear in the recoloring, thereby completing the proof of the induced arithmetic removal lemma. This patching result is stronger than the corresponding result in [FTZ22] since for this application the patching must also respect the structure of the partition given by $\mathfrak{B}$.

Our proof does not give effective bounds on the rejection probability function $\delta(\epsilon)$ guaranteed by Theorem 4.0.3. The ineffectiveness is due to the fact that the current best-known bounds on the inverse theorem for non-classical polynomials are ineffective (the same occurs in [BFH$^+$13]).

**Unbounded complexity**

The technique used to handle infinite removal lemmas is a compactness argument first introduced by Alon and Shapira [AS08] in the graph setting. A key ingredient of their proof is a strong regularity lemma.

Bhattacharyya, Fischer, Hatami, Hatami, and Lovett [BFH$^+$13] prove that all linear-invariant subspace-hereditary properties that are also translation-invariant and bounded-complexity are testable. Their result follows from an infinite removal lemma for arithmetic patterns of bounded complexity. The proof of this result involves a strong arithmetic regularity lemma and a compactness argument in the spirit of Alon and Shapira.

To remove the bounded complexity assumption from [BFH$^+$13], we prove a new strong arithmetic regularity lemma obtained by iterating a weaker arithmetic regularity lemma. The

key innovation here is the level of Gowers uniformity used in each iteration is allowed to increase at each step of the process.

## 4.2 Colored patterns and removal lemmas

Theorem 4.0.3 and Theorem 4.0.4 both follow from an arithmetic removal lemma for colored linear patterns. In this section we define these objects and state the main removal lemma.

**Definition 4.2.1.** A **linear form** over $\mathbb{F}_p$ in $\ell$ variables is an expression $L$ of the form

$$L(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} c_i x_i$$

with $c_i \in \mathbb{F}_p$. For any $\mathbb{F}_p$-vector space $V$, the linear form $L$ gives rise to a function $L\colon V^\ell \to V$ that is linear in each variable.

**Definition 4.2.2.** For a prime $p$ and a finite set $\mathcal{S}$, an **$\mathcal{S}$-colored pattern** over $\mathbb{F}_p$ consisting of $m$ linear forms in $\ell$ variables is a pair $(\boldsymbol{L}, \psi)$ given by a system $\boldsymbol{L} = (L_1, \dots, L_m)$ of $m$ linear forms in $\ell$ variables and a coloring $\psi\colon [m] \to \mathcal{S}$. Given a finite-dimensional $\mathbb{F}_p$-vector space $V$ and a function $f\colon V \to \mathcal{S}$, an **$(\boldsymbol{L}, \psi)$-instance** in $f$ is some $\boldsymbol{x} \in V^\ell$ such that $f(L_i(\boldsymbol{x})) = \psi(i)$ for all $i \in [m]$. An instance is called **generic** if $x_1, \dots, x_\ell$ are linearly independent. We say that $(\boldsymbol{L}, \psi)$ is **translation-invariant** if the coefficient of $x_1$ is 1 in each of $L_1, \dots, L_m$.

Given a finite-dimensional $\mathbb{F}_p$-vector space $V$ and functions $f_1, \dots, f_m\colon V \to [-1, 1]$, we write

$$\Lambda_{\boldsymbol{L}}(f_1, \dots, f_m) := \mathbb{E}_{\boldsymbol{x} \in V^k}[f_1(L_1(\boldsymbol{x})) \cdots f_m(L_m(\boldsymbol{x}))].$$

**Definition 4.2.3.** For an $\mathcal{S}$-colored pattern over $\mathbb{F}_p$ consisting of $m$ linear forms in $k$ variables $(\boldsymbol{L}, \psi)$, a finite dimensional $\mathbb{F}_p$-vector space $V$, and a function $f\colon V \to \mathcal{S}$, define the **$(\boldsymbol{L}, \psi)$-density** in $f$ to be $\Lambda_{\boldsymbol{L}}(f_1, \dots, f_m)$ where $f_i := 1_{f^{-1}(\psi(i))}$ for each $i \in [m]$.

Our main removal lemma is the following result.

**Theorem 4.2.4** (Main removal lemma)**.** *Fix a prime $p$ and a finite set $\mathcal{S}$. Let $\mathcal{H}$ be a (possibly infinite) set of $\mathcal{S}$-colored patterns over $\mathbb{F}_p$. For every $\epsilon > 0$, there exists a finite set*

*$\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and $\delta = \delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. If $f\colon V \to \mathcal{S}$ has $H$-density at most $\delta$ for every $H \in \mathcal{H}_\epsilon$, then there exists a recoloring $g\colon V \to \mathcal{S}$ that agrees with $f$ on all but an at most $\epsilon$-fraction of $V$ such that $g$ has no generic $H$-instances for every $H \in \mathcal{H}$.*

There are several difficulties in the proof of the main removal lemma. The first is that individual patterns $H \in \mathcal{H}$ may have "infinite complexity". Second, the set of patterns $\mathcal{H}$ may be infinite. Complicating this, even if all patterns in $\mathcal{H}$ have finite complexity, these complexities can be unbounded. Finally, there are major difficulties related to the fact that the patterns in $\mathcal{H}$ are not necessarily translation-invariant.

We use a trick we call "projectivization" to reduce to the case where all patterns have finite complexity.

A "compactness argument" due to Alon and Shapira [AS08] reduces the problem of an infinite collection of patterns to a finite one at the expense of requiring a stronger arithmetic regularity lemma. If the collection of patterns is all complexity at most $d$, we only require a strong $U^{d+1}$-regularity lemma with rapidly decreasing error parameter. In the most general case when the collection of patterns has unbounded complexity we require an even stronger regularity lemma where the error parameter rapidly decreases and the degree of the uniformity norm rapidly increases.

Unless we restrict to the special case where all patterns in $\mathcal{H}$ are translation-invariant, the origin of the vector space plays a special role. This is unfortunate because it is impossible to regularize a function in the neighborhood of the origin. Since regularity methods are useless here, we turn to a new technique called patching, introduced in Chapter 2, to deal with the portions of the vector space that cannot be regularized.

We now explain the "projectivization" trick. This will allow us to reduce to the case where all patterns have finite complexity at the expense of requiring that the functions considered are "projective". This restriction can be intuitively thought of as requiring that the functions live on projective space instead of affine space. We first define what it means for a function to be projective and then state the projective removal lemma, Theorem 4.2.8, from which we will later deduce Theorem 4.2.4.

**Definition 4.2.5.** Let $\mathcal{S}$ be a finite set equipped with a group action of $\mathbb{F}_p^\times$ that we denote $c \cdot s$ for $c \in \mathbb{F}_p^\times$ and $s \in \mathcal{S}$. Given a finite-dimensional $\mathbb{F}_p$-vector space $V$, a function $f\colon V \to \mathcal{S}$ is **projective** if it preserves the action of $\mathbb{F}_p^\times$, i.e., $f(cx) = c \cdot f(x)$ for all $c \in \mathbb{F}_p^\times$ and all $x \in V$.

We give one example of a class of projective functions which will be the one we need at the end of the argument. For a finite set $\mathcal{S}$, define $\overline{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$ with $\mathbb{F}_p^\times$-action defined by

$$b' \cdot (c_b)_{b \in \mathbb{F}_p^\times} := (c_{b'b})_{b \in \mathbb{F}_p^\times}.$$

Then for any function $f\colon V \to \mathcal{S}$, one can create a projective function $\overline{f}\colon V \to \overline{\mathcal{S}}$ as

$$\overline{f}(x) := (f(bx))_{b \in \mathbb{F}_p^\times}.$$

**Definition 4.2.6.** A list of linear forms $\boldsymbol{L} = (L_1, \dots, L_m)$ is **finite complexity** if no form is identically equal to zero, i.e., $L_i \not\equiv 0$ for all $i \in [m]$, and no two forms are linearly dependent, i.e., $L_i \not\equiv cL_j$ for all $i \neq j$ and $c \in \mathbb{F}_p$. (Later we will define the complexity of a system of linear forms to be a non-negative integer or infinity.)

**Definition 4.2.7.** Fix a prime $p$ and a positive integer $\ell$. We consider two particular systems of linear forms. For $\boldsymbol{i} = (i_1, \dots, i_\ell) \in \mathbb{F}_p^\ell$, define

$$L^\ell_{\boldsymbol{i}}(x_1, \dots, x_\ell) := i_1 x_1 + \cdots + i_\ell x_\ell.$$

Then define $\boldsymbol{L}^\ell := (L^\ell_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{F}_p^\ell}$, the system of $p^\ell$ linear forms in $\ell$ variables that defines an $\ell$-dimensional subspace.

Let $E_\ell \subset \mathbb{F}_p^\ell$ be the set of non-zero vectors whose first non-zero coordinate is 1. Then define $\overline{\boldsymbol{L}}^\ell := (L^\ell_{\boldsymbol{i}})_{\boldsymbol{i} \in E_\ell}$, a system of $(p^\ell - 1)/(p-1)$ linear forms in $\ell$ variables.

Note that unlike $\boldsymbol{L}^\ell$, the system $\overline{\boldsymbol{L}}^\ell$ has finite complexity. For technical reasons, it will be convenient to reduce the removal lemma for general patterns to the case where all patterns are defined by a system of the form $\boldsymbol{L}^\ell$. Then we reduce this to the following projective removal lemma where all patterns are defined by a system of the form $\overline{\boldsymbol{L}}^\ell$.

**Theorem 4.2.8** (Projective removal lemma)**.** *Fix a prime $p$ and a finite set $\mathcal{S}$ equipped with an $\mathbb{F}_p^\times$-action. Let $\mathcal{H}$ be a (possibly infinite) set consisting of $\mathcal{S}$-colored patterns over $\mathbb{F}_p$ of the form $(\overline{\boldsymbol{L}}^\ell, \psi)$ where $\ell$ is some positive integer and $\psi\colon E_\ell \to \mathcal{S}$ is some map (see Definition 4.2.7 for the definition of $\overline{\boldsymbol{L}}^\ell$ and $E_\ell$). For every $\epsilon > 0$, there exists a finite subset $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and $\delta = \delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. If $f\colon V \to \mathcal{S}$ is a projective function with $H$-density at most $\delta$ for every $H \in \mathcal{H}_\epsilon$, then there exists a projective recoloring $g\colon V \to \mathcal{S}$ that agrees with $f$ on all but an at most $\epsilon$-fraction of $V$ such that $g$ has no generic $H$-instances for every $H \in \mathcal{H}$.*

## 4.3 Preliminaries on higher-order Fourier analysis

### 4.3.1 Gowers norms and complexity

Our main tool for studying arithmetic patterns is the Gowers uniformity norms. Informally, if a colored pattern $H$ has complexity $d$, then any coloring which is "quasirandom with respect to the Gowers $U^{d+1}$-norm" will have close to the expected number of copies of $H$.

**Definition 4.3.1.** Fix a prime $p$, a finite-dimensional $\mathbb{F}_p$-vector space $V$, and an abelian group $G$. Given a function $f\colon V \to G$ and a shift $h \in V$, define the **additive derivative** $\Delta_h f\colon V \to G$ by

$$(\Delta_h f)(x) := f(x+h) - f(x).$$

Given a function $f\colon V \to \mathbb{C}$ and a shift $h \in V$, define the **multiplicative derivative** $\partial_h f\colon V \to \mathbb{C}$ by

$$(\partial_h f)(x) := f(x+h)\overline{f(x)}.$$

**Definition 4.3.2.** Fix a prime $p$ and a finite-dimensional $\mathbb{F}_p$-vector space $V$. Given a function $f\colon V \to \mathbb{C}$ and $d \geq 1$, the **Gowers uniformity norm** $\|f\|_{U^d}$ is defined by

$$\|f\|_{U^d} := \left|\mathbb{E}_{x,h_1,\ldots,h_d \in V} (\partial_{h_1} \cdots \partial_{h_d} f)(x)\right|^{1/2^d}.$$

**Definition 4.3.3.** A system $\boldsymbol{L} = (L_1, \ldots, L_m)$ of $m$ linear forms in $\ell$ variables is **complexity at most** $d$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $f_1, \ldots, f_\ell\colon V \to [-1,1]$ it holds

that

$$|\Lambda_{\boldsymbol{L}}(f_1, \dots, f_\ell)| \leq \epsilon \qquad \text{whenever} \qquad \min_{1 \leq i \leq \ell} \|f_i\|_{U^{d+1}} \leq \delta. \tag{4.1}$$

The **complexity** of $\boldsymbol{L}$ is the smallest $d$ such that the above holds, and infinite otherwise.

Eq. (4.1) is known as the counting lemma and will be used frequently throughout the chapter.

*Remark* 4.3.4. The above definition is sometimes known as true complexity. It is known that a pattern $(L_1, \dots, L_m)$ is complexity at most $d$ if and only if $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent as $(d+1)$th order tensors [GW11a, HHL16].

Let $(L_1, \dots, L_m)$ be any pattern such that no form is identically zero and no two forms are linearly dependent. It is known (for example, because Cauchy-Schwarz complexity is an upper bound for true complexity [GT10c]) that $(L_1, \dots, L_m)$ has complexity at most $m-2$.

It follows from the above discussion that the definition of complexity given in Definition 4.3.3 agrees with the definition of finite complexity given in Definition 4.2.6.

### 4.3.2 Non-classical polynomials and homogeneity

The main technical result which is the basis for all tools in higher-order Fourier analysis is the inverse theorem for the Gowers uniformity norms (stated as Theorem 4.4.1 in this chapter). This result states that a function which is not quasirandom with respect to the Gowers $U^{d+1}$-norm has some sort of structure. The specific structure involves the notion of non-classical polynomials which we will need to work with heavily in this chapter.

For ease of notation we write

$$\mathbb{U}_k := \tfrac{1}{p^k}\mathbb{Z}/\mathbb{Z} \subset \mathbb{R}/\mathbb{Z} \tag{4.2}$$

through the chapter.

**Definition 4.3.5.** Fix a prime $p$, and a non-negative integer $d \geq 0$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. A **non-classical polynomial** of degree at most $d$ is a map $P \colon V \to \mathbb{R}/\mathbb{Z}$ that satisfies

$$(\Delta_{h_1} \cdots \Delta_{h_{d+1}} P)(x) = 0$$

for all $h_1, \dots, h_{d+1}, x \in V$. The **degree** of $P$ is the smallest $d > 0$ such that the above holds. The **depth** of $P$ is the smallest $k \geq 0$ such that $P$ takes values in a coset of $\mathbb{U}_{k+1}$.

See [TZ12, Lemma 1.7] for some basic facts about non-classical polynomials. We record one such fact here.

**Lemma 4.3.6** ([TZ12, Lemma 1.7(iii)])**.** *Fix a prime $p$, and a finite-dimensional $\mathbb{F}_p$-vector space $V \simeq \mathbb{F}_p^n$. Then $P\colon V \to \mathbb{R}/\mathbb{Z}$ is a non-classical polynomial of degree at most $d$ if and only if it can be expressed in the form*

$$P(x_1, \dots, x_n) = \alpha + \sum_{\substack{0 \leq i_1, \dots, i_n < p,\, j \geq 0: \\ 0 < i_1 + \cdots + i_n \leq d - k(p-1)}} \frac{c_{i_1,\dots,i_n,k} |x_1|^{i_1} \cdots |x_n|^{i_n}}{p^{k+1}} \pmod 1,$$

*for some $\alpha \in \mathbb{R}/\mathbb{Z}$ and coefficients $c_{i_1,\dots,i_n,k} \in \{0, \dots, p-1\}$ where $|\cdot|$ is the standard map $\mathbb{F}_p \to \{0, \dots, p-1\}$. Furthermore, this representation is unique.*

As a corollary we see that in characteristic $p$, every non-classical polynomial of degree at most $d$ has depth at most $\lfloor (d-1)/(p-1) \rfloor$.

For technical reasons we will often need to work with homogeneous non-classical polynomials. These are defined analogously to usual homogeneous polynomials and satisfy many of the expected properties.

**Definition 4.3.7.** A **homogeneous non-classical polynomial** is a non-classical polynomial $P\colon V \to \mathbb{R}/\mathbb{Z}$ that also satisfies the following. For all $b \in \mathbb{F}_p$ there exists $\sigma_b^{(P)} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ such that $P(bx) = \sigma_b^{(P)} P(x)$ for all $x \in V$.

As an example, consider the non-classical cubic polynomial $P\colon \mathbb{F}_3 \to \mathbb{R}/\mathbb{Z}$ defined by $P(x) = |x|/9$. This is not homogeneous since $2P(1) = P(2)$, but $2P(2) \neq P(1)$. However we can write $P = Q_3 + Q_2$ where $Q_3 = |x|/9 + |x|^2/3$ and $Q_2(x) = -|x|^2/3$. One can check that these are homogeneous non-classical polynomials of degree 3 and 2 respectively.

**Lemma 4.3.8** ([HHL16, Lemma 3.3])**.** *Fix a prime $p$ and integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$. For each $b \in \mathbb{F}_p$ there exists $\sigma_b^{(d,k)} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ such that $\sigma_b^{(P)} = \sigma_b^{(d,k)}$ for all homogeneous non-classical polynomials $P$ of degree $d$ and depth $k$. Furthermore, for $b \neq 0$, the number $\sigma_b^{(d,k)}$ is uniquely determined by the following two properties:*

*(i)* $\sigma_b^{(d,k)} \equiv b^d \pmod{p}$

*(ii)* $\left(\sigma_b^{(d,k)}\right)^{p-1} = 1$ *(in* $\mathbb{Z}/p^{k+1}\mathbb{Z}$*).*

**Theorem 4.3.9** ([HHL16, Theorem 3.4])**.** *Let $P$ be a non-classical polynomial of degree $d$ and depth $k$. Then $P$ can be written as the sum of homogeneous non-classical polynomials of degree at most $d$ and depth at most $k$.*

### 4.3.3 Polynomial factors

The main objects we work with are polynomial factors, which are lists of polynomials which partition $\mathbb{F}_p^n$ into pieces called atoms. There is a lot of data associated with a polynomial factor, so we introduce some notation which we use to keep track of this data.

**Definition 4.3.10.** Fix a prime $p$. Define

$$D_p := \{(d,k) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{\geqslant 0} : k \leq \lfloor (d-1)/(p-1) \rfloor\},$$

and

$$\mathcal{I}_p := \left\{ I \in \mathbb{Z}_{\geqslant 0}^{D_p} : \sum_{(d,k)\in D_p} I_{d,k} < \infty \right\}.$$

We call $I \in \mathcal{I}_p$ a **parameter list**. For $I \in \mathcal{I}_p$, we write $\|I\| := p^{\sum_{d,k}(k+1)I_{d,k}}$ and $\deg I$ for the largest $d$ such that $I_{d,k} \neq 0$ for some $k$. We add and subtract parameter lists coordinatewise. For $I, I' \in \mathcal{I}_p$, we write $I \leq I'$ if $I_{d,k} \leq I'_{d,k}$ for all $(d,k) \in D_p$.

**Definition 4.3.11.** For $p$ a prime and $I \in \mathcal{I}_p$, define the **atom-indexing set** of $I$ to be

$$A_I := \prod_{(d,k)\in D_p} \left(\tfrac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}\right)^{I_{d,k}}. \tag{4.3}$$

Note that $|A_I| = \|I\|$.

For $I, I' \in \mathcal{I}_p$ with $I \leq I'$, write $\pi\colon A_{I'} \to A_I$ for the standard projection map defined by

$$\pi\left(\left(a_{d,k}^i\right)_{\substack{(d,k)\in D_p \\ i\in[I'_{d,k}]}}\right) \mapsto \left(a_{d,k}^i\right)_{\substack{(d,k)\in D_p \\ i\in[I_{d,k}]}}. \tag{4.4}$$

$A_I$ is equipped with the following $\mathbb{F}_p^\times$-action:

$$c \cdot \left(a^i_{d,k}\right)_{\substack{(d,k)\in D_p \\ i\in\left[I_{d,k}\right]}} := \left(\sigma_c^{(d,k)} a^i_{d,k}\right)_{\substack{(d,k)\in D_p \\ i\in\left[I_{d,k}\right]}}, \tag{4.5}$$

where $\sigma_c^{(d,k)}$ is defined in Lemma 4.3.8.

**Definition 4.3.12.** Fix a prime $p$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $I \in \mathcal{I}_p$ be a parameter list. A **polynomial factor** on $V$ with parameters $I$, denoted $\mathfrak{B}$, is a collection

$$\left(P^i_{d,k}\right)_{\substack{(d,k)\in D_p \\ i\in\left[I_{d,k}\right]}}$$

where $P^i_{d,k}$ is a homogeneous non-classical polynomial of degree $d$ and depth $k$. We also use $\mathfrak{B}$ to denote the map $\mathfrak{B}\colon V \to A_I$ defined by evaluation of the polynomials. We also associate to $\mathfrak{B}$ the partition of $V$ given by the fibers of this map. The atoms of this partition are called the atoms of $\mathfrak{B}$. We write $\|\mathfrak{B}\| := \|I\|$ and $\deg \mathfrak{B} := \deg I$.

Note that if $\mathfrak{B}$ is a polynomial factor on $V$ with parameters $I$, then $\mathfrak{B}(cx) = c \cdot \mathfrak{B}(x)$ for all $c \in \mathbb{F}_p^\times$ and $x \in V$ where the $\mathbb{F}_p^\times$-action on $A_I$ is defined in Eq. (4.5).

**Definition 4.3.13.** Fix a prime $p$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $I, I' \in \mathcal{I}_p$ be two parameter lists. Let $\mathfrak{B}$ and $\mathfrak{B}'$ be two polynomial factors on $V$ with parameters $I$ and $I'$. We say that $\mathfrak{B}'$ is a **refinement** of $\mathfrak{B}$ if $I \leq I'$ and the lists of polynomials defining $\mathfrak{B}'$ are extensions of the lists of polynomials defining $\mathfrak{B}$. We say that $\mathfrak{B}'$ is a **weak refinement** of $\mathfrak{B}$ if the partition of $V$ associated to $\mathfrak{B}'$ is a refinement of the partition associated to $\mathfrak{B}$.

Note that if $\mathfrak{B}'$ is a refinement of $\mathfrak{B}$, then $\mathfrak{B} = \pi \circ \mathfrak{B}'$ where $\pi\colon A_{I'} \to A_I$ is the projection defined in Eq. (4.4).

We would like our polynomial factors to satisfy certain regularity conditions. For example, it will be convenient to know that all atoms are approximately the same size. All the regularity condition we need will be guaranteed if the factor is "high rank".

**Definition 4.3.14.** Fix a prime $p$ and integer $d \geq 0$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. For a non-classical polynomial $P\colon V \to \mathbb{R}/\mathbb{Z}$, define the **$d$-rank of** $P$,

denoted $\operatorname{rank}_d P$, to be the smallest integer $r$ such that there exists non-classical polynomials $Q_1, \dots, Q_r \colon V \to \mathbb{R}/\mathbb{Z}$ of degree at most $d-1$ and a function $\Gamma \colon (\mathbb{R}/\mathbb{Z})^r \to \mathbb{R}/\mathbb{Z}$ such that $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$ for all $x \in V$.

For a polynomial factor $\mathfrak{B}$ on $V$ with parameters $I \in I_p$, defined by a collection $(P^i_{d,k})_{(d,k)\in D_p, i\in[I_{d,k}]}$ where $P^i_{d,k}$ is a homogeneous non-classical polynomial of degree $d$ and depth $k$, we define the **rank of $\mathfrak{B}$**, denoted $\operatorname{rank}\mathfrak{B}$, to be

$$\min_{\boldsymbol{\lambda}\in\prod_{(d,k)\in D_p}\left(\mathbb{Z}/p^{k+1}\mathbb{Z}\right)^{I_{d,k}}} \operatorname{rank}_{d'}\left(\sum_{(d,k)\in D_p}\sum_{i=1}^{I_{d,k}} \lambda^i_{d,k} P^i_{d,k}\right)$$

where

$$d' := \min_{(d,k)\in D_p, i\in[I_{d,k}]} \deg\left(\lambda^i_{d,k} P^i_{d,k}\right).$$

We need a couple of easy properties of rank, the first simply being that some high rank factor exists.

**Lemma 4.3.15.** *Fix a prime $p$, a parameter list $I \in \mathcal{I}_p$, and a positive integer $r$. There exists a constant $n_{\mathsf{high-rank}}(p, I, r)$ such that for every $n \geq n_{\mathsf{high-rank}}(p, I, r)$, there exists a polynomial factor $\mathfrak{B}$ on $\mathbb{F}_p^n$ with parameters $I$ and satisfying $\operatorname{rank}\mathfrak{B} \geq r$.*

*Proof.* For $(d,k) \in D_p$, write $d = (k+a)(p-1)+b$ where $a \geq 0$ and $b \in \{1, \dots, p-1\}$. By Lemma 4.3.6 there exists a non-classical polynomial of degree $d$ and depth $k$. For example, consider

$$\frac{|x_1|^{p-1}|x_2|^{p-1}\cdots|x_a|^b}{p^{k+1}} \pmod 1.$$

By Theorem 4.3.9, we can decompose this non-classical polynomial as the sum of homogeneous non-classical polynomials of degree at most $d$ and depth at most $k$. Let $P \colon \mathbb{F}_p^a \to \mathbb{U}_{k+1}$ be the homogeneous part of degree $d$ and depth $k$.

Next define $Q \colon \mathbb{F}_p^{a\oplus N} \to \mathbb{U}_{k+1}$ by

$$Q(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) := P(\boldsymbol{x}_1) + \cdots + P(\boldsymbol{x}_N).$$

This is clearly a homogeneous non-classical polynomial of degree $d$ and depth $k$. We claim that for $N$ large enough, we have $\operatorname{rank}_d Q \geq r$. The proof of this fact uses several basic

results from [TZ12] that are not used elsewhere in this chapter.

First, since $P$ has degree exactly $d$, we have that $(\Delta_{h_1} \cdots \Delta_{h_d} P)(x)$ is a constant, independent of $x$, but is not identically zero. This implies that

$$c := \mathbb{E}_{h_1,\dots,h_d \in \mathbb{F}_p^a} e^{2\pi i (\Delta_{h_1} \cdots \Delta_{h_d} P)} < 1.$$

The quantity $-\log_p c$ is known as the *analytic rank* of $P$. Now a simple calculation shows that

$$\mathbb{E}_{\boldsymbol{h}_1,\dots,\boldsymbol{h}_d \in (\mathbb{F}_p^a)^{\oplus N}} e^{2\pi i (\Delta_{\boldsymbol{h}_1} \cdots \Delta_{\boldsymbol{h}_d} Q)} = c^N.$$

To conclude we use [TZ12, Lemma 1.15(iii)] which implies that $\operatorname{rank}_d Q \geq -c_d \log_p(c^N)$ for some constant $c_d > 0$ only depending on $d$. Since $c < 1$, taking $N$ large enough gives $\operatorname{rank}_d Q \geq r$, as desired.

Thus there exist homogeneous non-classical polynomials $Q_{d,k} \colon V_{d,k} \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ that satisfy $\operatorname{rank}_d Q_{d,k} \geq r$ for each $(d,k) \in D_p$. Define the vector space

$$V := \bigoplus_{(d,k) \in D_p} V_{d,k}^{\oplus I_{d,k}}.$$

Define the homogeneous non-classical polynomials $\overline{Q}_{d,k}^i \colon V \to \mathbb{R}/\mathbb{Z}$ for each $(d,k) \in D_p$ and $i \in [I_{d,k}]$ such that $\overline{Q}_{d,k}^i$ is equal to $Q_{d,k}$ evaluated on the $i$th copy of $V_{d,k}$ and does not depend on the other coordinates. In particular, we define

$$\overline{Q}_{d,k}^i \left( (x_{d',k'}^{i'})_{\substack{(d',k') \in D_p \\ i' \in [I_{d',k'}]}} \right) := Q_{d,k}(x_{d,k}^i).$$

These polynomials define a polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$ such that each of the homogeneous non-classical polynomials defining $\mathfrak{B}$ has rank at least $r$. Furthermore, since the polynomials defining $\mathfrak{B}$ depend on disjoint sets of variables, it follows that all non-trivial linear combinations of the polynomials defining $\mathfrak{B}$ also have high rank. Setting $n_{\mathsf{high-rank}}(p, I, R) := \dim V$, we have constructed the desired polynomial factor on $\mathbb{F}_p^{n_{\mathsf{high-rank}}(p,I,r)}$. To extend this construction to $\mathbb{F}_p^n$ with $n \geq n_{\mathsf{high-rank}}(p, I, r)$ one can simply add on extra variables that none of the polynomials depend on. $\square$

**Lemma 4.3.16** ([BFH+13, Lemma 2.13])**.** *Fix a prime $p$ and positive integers $d, r$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $P\colon V \to \mathbb{R}/\mathbb{Z}$ be a non-classical polynomial of degree $d$ such that $\operatorname{rank}_d(P) \geq r + p$. Let $U \leq V$ be a codimension-1 hyperplane. Then $\operatorname{rank}_d(P|_U) \geq r$ unless $d = 1$ and $P|_U$ is identically zero.*

*As a consequence, let $\mathfrak{B}$ be a polynomial factor on $V$ and let $P\colon V \to \mathbb{R}/\mathbb{Z}$ be a linear polynomial. Write $\mathfrak{B}'$ for the common refinement of $\mathfrak{B}$ and $\{P\}$. If $\operatorname{rank} \mathfrak{B}' \geq r + p$, then $\operatorname{rank} \mathfrak{B}|_U \geq r$ where $U$ is the codimension-1 hyperplane where $P$ vanishes.*

### 4.3.4 Equidistribution and consistency sets

In this section we state a very useful result known as equidistribution and introduce several related notations. One consequence of equidistribution is the previously-mentioned fact that high rank polynomial factors have atoms of approximately equal sizes.

**Definition 4.3.17.** Fix a prime $p$, integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$, and a system $\boldsymbol{L} = (L_1, \ldots, L_m)$ of $m$ linear forms in $\ell$ variables. Define the $(d,k)$**-consistency set of** $\boldsymbol{L}$, denoted $\Phi_{d,k}(\boldsymbol{L})$, to be the subset of $\mathbb{U}_{k+1}^m$ consisting of the tuples $\boldsymbol{a} = (a_1, \ldots, a_m)$ such that there exists a finite-dimensional $\mathbb{F}_p$-vector space $V$, a homogeneous non-classical polynomial $P\colon V \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$, and a tuple $\boldsymbol{x} \in V^\ell$ such that $a_i = P(L_i(\boldsymbol{x}))$ for all $i \in [m]$.

For a parameter list $I \in \mathcal{I}_p$, define the $I$**-consistency set of** $\boldsymbol{L}$ to be the set of tuples $\boldsymbol{a} = (a_1, \ldots, a_m) \in A_I^m$ such that for each $(d,k) \in D_p$ and $j \in [I_{d,k}]$ the tuple $\left((a_1)_{d,k}^j, \ldots, (a_m)_{d,k}^j\right)$ lies in $\Phi_{d,k}(\boldsymbol{L})$.

**Lemma 4.3.18.** *Fix a prime $p$, integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$, and a system $\boldsymbol{L} = (L_1, \ldots, L_m)$ of $m$ linear forms. The $(d,k)$-consistency set of $\boldsymbol{L}$ is a subgroup of $\mathbb{U}_{k+1}^m$.*

*Proof.* Suppose $\boldsymbol{a}, \boldsymbol{b} \in \Phi_{d,k}(\boldsymbol{L})$. We wish to show that $-\boldsymbol{a}$ and $\boldsymbol{a} + \boldsymbol{b}$ both lie in this set. By definition, there exist finite-dimensional $\mathbb{F}_p$-vector spaces $V, W$, homogeneous non-classical polynomials $P\colon V \to \mathbb{U}_{k+1}$ and $Q\colon W \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$, and tuples $\boldsymbol{x} \in V^k$ and $\boldsymbol{y} \in W^k$ such that $a_i = P(L_i(\boldsymbol{x}))$ and $b_i = Q(L_i(\boldsymbol{y}))$ for all $i \in [m]$.

Note that $-\boldsymbol{a} \in \Phi_{d,k}(\boldsymbol{L})$ since $-P\colon V \to \mathbb{U}_{k+1}$ is a homogeneous non-classical polynomial of degree $d$ and depth $k$ that satisfies $(-P)(L_i(\boldsymbol{x})) = -a_i$.

Now define $P \oplus Q\colon V \oplus W \to \mathbb{U}_{k+1}$ by $(P \oplus Q)(v \oplus w) := P(v) + Q(w)$. One can easily check that $P \oplus Q$ is a homogeneous non-classical polynomial of degree $d$ and depth $k$. Finally note that $(P \oplus Q)(L_i(\boldsymbol{x} \oplus \boldsymbol{y})) = a_i + b_i$, as desired.[2] □

**Theorem 4.3.19** (Equidistribution [HHL16, Theorem 3.10])**.** *Fix a prime $p$, a positive integer $d > 0$, and a parameter $\epsilon > 0$. There exists $r_{\mathsf{equi}}(p, d, \epsilon)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $\mathfrak{B}$ be a polynomial factor on $V$ with parameters $I$ such that $\deg \mathfrak{B} \le d$ and $\operatorname{rank}(\mathfrak{B}) \ge r_{\mathsf{equi}}(p, d, \epsilon)$. Then for a system of linear forms $\boldsymbol{L} = (L_1, \dots, L_m)$ consisting of $m$ forms in $\ell$ variables, and a tuple of atoms $\boldsymbol{a} = (a_1, \dots, a_m) \in \Phi_I(\boldsymbol{L})$,*

$$\left| \mathbb{P}_{\boldsymbol{x} \in V^\ell} \left( \mathfrak{B}(L_i(\boldsymbol{x})) = a_i \text{ for all } i \in [m] \right) - \frac{1}{|\Phi_I(\boldsymbol{L})|} \right| \le \epsilon.$$

*Remark* 4.3.20. To be completely correct, the statement given above follows by combining [HHL16, Theorem 3.10] and [HHL16, Corollary 2.13].

Note that the probability above is 0 if $\boldsymbol{a} \notin \Phi_I(\boldsymbol{L})$. We typically apply the above theorem with $\epsilon$ that decreases rapidly with $\|I\|$, for example, taking $\epsilon = 1/(2\|I\|^m)$ and using the fact that $|\Phi_I(\boldsymbol{L})| \le \|\mathfrak{B}\|^m$, we see that in this case the probability above is at least $1/(2|\Phi_I(\boldsymbol{L})|)$.

Consistency sets are often hard to compute exactly. The next two lemmas give exact relations on the sizes of consistency sets in two special cases that occur in this chapter.

**Definition 4.3.21.** Fix a prime $p$. A system $\boldsymbol{L}$ of $m$ linear forms in $\ell$ variables over $\mathbb{F}_p$ is **full dimensional** if $|\Phi_{d,k}(\boldsymbol{L})| = |\Phi_{d,k}(\boldsymbol{L}^\ell)|$ for all $(d, k) \in D_p$ (recall the system $\boldsymbol{L}^\ell$ defined in Definition 4.2.7 defines an $\ell$-dimensional subspace).

**Lemma 4.3.22.** *Fix a prime $p$ and a positive integer $\ell$. Let $J \subseteq \mathbb{F}_p^\ell$ be a set that contains at least one vector in each direction (i.e., for each $\boldsymbol{i} \in \mathbb{F}_p^\ell$, there exists $\boldsymbol{j} \in J$ and $b \in \mathbb{F}_p^\times$ such*

[2] To be completely correct, we also need to show that $\Phi_{d,k}(\boldsymbol{L})$ is non-empty, which follows from, for example Lemma 4.3.6 and Theorem 4.3.9, which together show the existence of homogeneous non-classical polynomial of degree $d$ and depth $k$ for every $(d, k) \in D_p$.

*that $\boldsymbol{i} = b\boldsymbol{j}$). Consider the system $\boldsymbol{L}_J := (L^\ell_{\boldsymbol{i}})_{\boldsymbol{i}\in J}$ of $|J|$ linear forms in $\ell$ variables (recall the linear form $L^\ell_{\boldsymbol{i}}$ defined in Definition 4.2.7)). Then $\boldsymbol{L}_J$ is full dimensional.*

As a special case of this result we see that the system $\overline{\boldsymbol{L}}^\ell$, defined in Definition 4.2.7, is full dimensional.

*Proof.* Note that the system $\boldsymbol{L}_J$ is a subsystem of $\boldsymbol{L}^\ell$. This immediately implies that $|\Phi_{d,k}(\boldsymbol{L}_J)| \leq |\Phi_{d,k}(\boldsymbol{L}^\ell)|$, since if $(a_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathbb{F}_p^\ell} \in \Phi_{d,k}(\boldsymbol{L}^\ell)$, then $(a_{\boldsymbol{i}})_{\boldsymbol{i}\in J} \in \Phi_{d,k}(\boldsymbol{L}_J)$.

To go the other direction, we use the homogeneity of our polynomials. Suppose $(a_{\boldsymbol{j}})_{\boldsymbol{j}\in J} \in \Phi_{d,k}(\boldsymbol{L}_J)$. Thus there exists a finite-dimensional $\mathbb{F}_p$-vector space $V$ and a homogeneous non-classical polynomial $P\colon V \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$, and a vector $\boldsymbol{x} \in V^\ell$ such that $P(L^\ell_{\boldsymbol{j}}(\boldsymbol{x})) = a_{\boldsymbol{j}}$ for all $\boldsymbol{j} \in J$. Now by assumption, for $\boldsymbol{i} \in \mathbb{F}_p^\ell$, there exists $\boldsymbol{j} \in J$ and $b \in \mathbb{F}_p$ such that $\boldsymbol{i} = b\boldsymbol{j}$. Define $a_{\boldsymbol{i}} := \sigma_b^{(d,k)} a_{\boldsymbol{j}}$ (recall the definition of $\sigma_b$ from Lemma 4.3.8). We claim that $(a_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathbb{F}_p^\ell} \in \Phi_{d,k}(\boldsymbol{L}^\ell)$. This is true simply because

$$P(L^\ell_{\boldsymbol{i}}(x)) = P(bL^\ell_{\boldsymbol{j}}(\boldsymbol{x})) = \sigma_b^{(d,k)} P(L^\ell_{\boldsymbol{j}}(x)) = \sigma_b^{(d,k)} a_{\boldsymbol{j}} = a_{\boldsymbol{i}}$$

where $\boldsymbol{j} \in J$ and $b \in \mathbb{F}_p$ are defined as above. Thus $|\Phi_{d,k}(\boldsymbol{L}_J)| \geq |\Phi_{d,k}(\boldsymbol{L}^\ell)|$, as desired. □

Several times in the chapter, an applications of the Cauchy-Schwarz inequality we lead us to consider counts of triples of patterns $\boldsymbol{L}, \boldsymbol{L}', \boldsymbol{L}''$ of the following form.

**Lemma 4.3.23.** *Fix a prime $p$. Let $\boldsymbol{L}$ be a system of $m$ linear forms in $\ell$ variables over $\mathbb{F}_p$. Say $\boldsymbol{L}$ is defined by $M$, an $m\times\ell$ matrix. (By this we mean that $L_i(x_1,\dots,x_\ell) = M_{i,1}x_1+\cdots+M_{i,\ell}x_\ell$ for all $i \in [m]$.) Let $\boldsymbol{L}'$ be a system of $m(n+1)$ linear forms in $\ell+\ell'$ variables, defined by a matrix of the form*

$$\left(\begin{array}{c|c} M & 0 \\ \hline c_1 M & \\ \vdots & N \\ c_n M & \end{array}\right)$$

*where $c_1,\dots,c_n \in \mathbb{F}_p$ and $N$ is an $mn \times \ell'$ matrix. Let $\boldsymbol{L}''$ be the system of $m(2n+1)$ linear*

*forms in $\ell + 2\ell'$ variables, defined by the matrix*

$$\left(\begin{array}{c|c|c} M & 0 & 0 \\ \hline c_1 M & & \\ \vdots & N & 0 \\ c_n M & & \\ \hline c_1 M & & \\ \vdots & 0 & N \\ c_n M & & \end{array}\right).$$

*Then for all $(d,k) \in D_p$, we have*

$$|\Phi_{d,k}(\boldsymbol{L})| \cdot |\Phi_{d,k}(\boldsymbol{L}'')| = |\Phi_{d,k}(\boldsymbol{L}')|^2.$$

*Proof.* We construct injections between $\Phi_{d,k}(\boldsymbol{L}) \times \Phi_{d,k}(\boldsymbol{L}'')$ and $\Phi_{d,k}(\boldsymbol{L}') \times \Phi_{d,k}(\boldsymbol{L}')$ in both directions. Write $\sigma_i := \sigma_{c_i}^{d,k}$ for $i \in [n]$ for the rest of the proof (see Lemma 4.3.8 for the definition).

Consider $\boldsymbol{a} = (a_1, \ldots, a_m) \in \Phi_{d,k}(\boldsymbol{L})$ and $\boldsymbol{b} = (b_1, \ldots, b_{m(2n+1)}) \in \Phi_{d,k}(\boldsymbol{L}'')$. By definition, there exists a finite-dimensional $\mathbb{F}_p$-vector space $V$, a homogeneous non-classical polynomial $P\colon V \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ and a vector $\boldsymbol{x} \in V^\ell$ such that $P(L_i(\boldsymbol{x})) = a_i$ for all $i \in [m]$. Also by definition, there exists a finite-dimensional $\mathbb{F}_p$-vector space $W$, a homogeneous non-classical polynomial $Q\colon W \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ and a vector $(\boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}') \in W^\ell \times W^{\ell'} \times W^{\ell'}$ such that $Q(L_i''(\boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}')) = b_i$ for all $i \in [m(2n+1)]$.

Now we map $(\boldsymbol{a}, \boldsymbol{b})$ to the pair $(\boldsymbol{a}', \boldsymbol{b}')$ where $b_i' = b_i$ for $i \in [m(n+1)]$ and $a_i' = a_i + b_i$ for $i \in [m]$ and $a_{tm+i}' = \sigma_t a_i + b_{m(n+t)+i}$ for $i \in [m]$ and $t \in [n]$. We can easily check that no two pairs $(\boldsymbol{a}, \boldsymbol{b})$ map to the same pair $(\boldsymbol{a}', \boldsymbol{b}')$. All that remains is to check that $\boldsymbol{a}', \boldsymbol{b}' \in \Phi_{d,k}(\boldsymbol{L}')$.

Define $P \oplus Q\colon V \oplus W \to \mathbb{U}_{k+1}$ by $(P \oplus Q)(x \oplus y) := P(x) + Q(y)$. This is clearly a homogeneous non-classical polynomial of degree $d$ and depth $k$. Note that $\boldsymbol{z} := (\boldsymbol{x} \oplus \boldsymbol{x}', \boldsymbol{0} \oplus \boldsymbol{y}') \in (V \oplus W)^{\ell+\ell'}$ satisfies $(P \oplus Q)(L_i'(\boldsymbol{z})) = a_i'$ for all $i \in [m(n+1)]$. Similarly note that $\boldsymbol{z}' := (\boldsymbol{0} \oplus \boldsymbol{x}', \boldsymbol{0} \oplus \boldsymbol{y}) \in (V \oplus W)^{\ell+\ell'}$ satisfies $(P \oplus Q)(L_i'(\boldsymbol{z}')) = b_i'$ for all $i \in [m(n+1)]$. This demonstrates the first injection.

Now consider $\boldsymbol{a} = (a_1, \ldots, a_{m(n+1)}) \in \Phi_{d,k}(\boldsymbol{L}')$ and $\boldsymbol{b} = (b_1, \ldots, b_{m(n+1)}) \in \Phi_{d,k}(\boldsymbol{L}')$. By definition, there exists a finite-dimensional $\mathbb{F}_p$-vector space $V$, a homogeneous non-classical polynomial $P\colon V \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ and a vector $(\boldsymbol{x}, \boldsymbol{y}) \in V^\ell \times V^{\ell'}$ such that $P(L'_i(\boldsymbol{x}, \boldsymbol{y})) = a_i$ for $i \in [m(n+1)]$. Also by definition, there exists a finite-dimensional $\mathbb{F}_p$-vector space $W$, a homogeneous non-classical polynomial $Q\colon W \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ and a vector $(\boldsymbol{x}', \boldsymbol{y}') \in W^\ell \times W^{\ell'}$ such that $Q(L'_i(\boldsymbol{x}', \boldsymbol{y}')) = b_i$ for $i \in [m(n+1)]$.

We map $(\boldsymbol{a}, \boldsymbol{b})$ to the pair $(\boldsymbol{a}', \boldsymbol{b}')$ where $a'_i = a_i$ for $i \in [m]$ and $b'_i = a_i + b_i$ for $i \in [m]$ and $b'_{m+i} = a_{m+i}$ for $i \in [mn]$ and $b'_{m(n+1)+i} = b_{m+i}$ for $i \in [mn]$. We can easily check that no two pairs $(\boldsymbol{a}, \boldsymbol{b})$ map to the same pair $(\boldsymbol{a}', \boldsymbol{b}')$. All that remains is to check that $\boldsymbol{a}' \in \Phi_{d,k}$ and $\boldsymbol{b}' \in \Phi_{d,k}(\boldsymbol{L}'')$.

As above, define the homogeneous non-classical polynomial $P \oplus Q\colon V \oplus W \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ by $(P \oplus Q)(x \oplus y) := P(x) + Q(y)$. Note that $\boldsymbol{z} := \boldsymbol{x} \oplus \boldsymbol{0} \in (V \oplus W)^\ell$ satisfies $(P \oplus Q)(\boldsymbol{L}_i(\boldsymbol{z})) = a'_i$ for $i \in [m]$ and $\boldsymbol{z}' := (\boldsymbol{x} \oplus \boldsymbol{x}', \boldsymbol{y} \oplus \boldsymbol{0}, \boldsymbol{0} \oplus \boldsymbol{y}') \in (V \oplus W)^\ell \times (V \oplus W)^{\ell'} \times (V \oplus W)^{\ell'}$ satisfies $(P \oplus Q)(L''_i(\boldsymbol{z}')) = b'_i$ for $i \in [m(2n+1)]$, as desired. □

### 4.3.5 Subatom selection functions

A situation that often occurs is the following. We have a polynomial factor $\mathfrak{B}$ with parameters $I$ and a refinement $\mathfrak{B}'$ with parameters $I'$. We use the word atom to refer to the atoms of the partition induced by $\mathfrak{B}$; these atoms are indexed by $A_I$. We use the word subatom to refer to the atoms of the partition induced by $\mathfrak{B}'$; these atoms are indexed by $A_{I'}$. The projection map $\pi\colon A_{I'} \to A_I$, defined in Eq. (4.4), maps a subatom to the atom that it is contained in.

We wish to designate one subatom inside each atom as special. This choice is given by a map $s\colon A_I \to A_{I'}$ that is a right inverse for $\pi$. In this chapter we define a certain class of these maps that we call subatom selection functions that have several desirable properties.

First we define certain polynomials $P_{d,k}\colon \mathbb{F}_p \to \mathbb{U}_{k+1}$ for each $(d,k) \in D_p$. Recall from Definition 4.3.10 that $(d,k) \in D_p$ means that $d \geq k(p-1)+1$. First, for $k(p-1)+1 \leq d \leq (k+1)(p-1)$, let $P_{d,k}$ be an arbitrary homogeneous non-classical polynomial $\mathbb{F}_p \to \mathbb{U}_{k+1}$ of degree $d$ and depth $k$ in one variable. (The existence of such a polynomial follows from Lemma 4.3.6 and Theorem 4.3.9.) Then for the remaining cases, i.e., $d > (k+1)(p-1)$, simply define $P_{d,k} := P_{d',k}$ where $k(p-1)+1 \leq d' \leq (k+1)(p-1)$ satisfies $d' \equiv d \pmod{p-1}$.

**Definition 4.3.24.** Fix a prime $p$ and parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \leq I'$. Let $c_{d,k}^{i,j} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ be arbitrary elements for $(d,k) \in D_p$ and $i \in [I_{1,0}]$ and $I_{d,k} < j \leq I'_{d,k}$. A **subatom selection function** is a map of the form $s_{\boldsymbol{c}} \colon A_I \to A_{I'}$, defined by

$$[s_{\boldsymbol{c}}(a)]_{d,k}^i = \begin{cases} a_{d,k}^i & \text{if } i \leq I_{d,k} \\ \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(|a_{1,0}^j|) & \text{otherwise,} \end{cases}$$

where the maps $P_{d,k} \colon \mathbb{F}_p \to \mathbb{U}_{k+1}$ were defined in the preceding paragraph and $|\cdot|$ is the standard map $\mathbb{U}_1 \to \mathbb{F}_p$.

**Lemma 4.3.25.** *Fix a prime $p$, parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \leq I'$, and a subatom selection function $s_{\boldsymbol{c}} \colon A_I \to A_{I'}$. The following hold:*

*(i)* $\pi \circ s_{\boldsymbol{c}} = \mathrm{Id}$ *(where $\pi \colon A_{I'} \to A_I$ is defined in Eq. (4.4));*

*(ii) for $a \in A_I$ and $b \in \mathbb{F}_p^\times$, we have $b \cdot s_{\boldsymbol{c}}(a) = s_{\boldsymbol{c}}(b \cdot a)$ (where the action of $\mathbb{F}_p^\times$ on $A_I$ and $A_{I'}$ is defined in Eq. (4.5));*

*(iii) for every system $\boldsymbol{L}$ of $m$ linear forms and every consistent tuple of atoms $(a_1, \ldots, a_m) \in \Phi_I(\boldsymbol{L})$, we have*

$$(s_{\boldsymbol{c}}(a_1), \ldots, s_{\boldsymbol{c}}(a_m)) \in \Phi_{I'}(\boldsymbol{L})$$

*(see Definition 4.3.17 for the definition of the consistency sets $\Phi_I(\boldsymbol{L})$ and $\Phi_{I'}(\boldsymbol{L})$).*

*Proof.* Property (i) is immediate.

For the property (ii), by definition, we have

$$[b \cdot s_{\boldsymbol{c}}(a)]_{d,k}^i = \sigma_b^{(d,k)} [s_{\boldsymbol{c}}(a)]_{d,k}^i,$$

where $\sigma_b^{(d,k)}$ is defined in Lemma 4.3.8. Now $s_{\boldsymbol{c}}(b \cdot a)_{d,k}^i = \sigma_b^{(d,k)} a_{d,k}^i$ if $i \leq I_{d,k}$, so we are done in this case.

Assume otherwise. Define $d'$ such that $d' \equiv d \pmod{p-1}$ and $k(p-1)+1 \leq d' \leq (k+1)(p-1)$. Remember that $P_{d,k} = P_{d',k}$ is a homogeneous non-classical polynomial of

degree $d'$ and depth $k$. Also note that $\sigma_b^{(1,0)} = b \in \mathbb{Z}/p\mathbb{Z}$. Then we have

$$[s_{\boldsymbol{c}}(b \cdot a)]_{d,k}^i = \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(b|a_{1,0}^j|) = \sigma_b^{(d',k)} \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(|a_{1,0}^j|).$$

To complete the proof of (ii) we need to show that $\sigma_b^{d,k} = \sigma_b^{d',k}$ whenever $d \equiv d' \pmod{p-1}$. This follows from Lemma 4.3.8 which implies that $\sigma_b^{d,k}$ is uniquely determined by the facts that $\sigma_b^{d,k} \equiv b^d \pmod{p}$ and $\left(\sigma_b^{(d,k)}\right)^{p-1} = 1$ in $\mathbb{Z}/p^{k+1}\mathbb{Z}$. The first property does not change when $d$ changes by a multiple of $p-1$ (by Fermat's little theorem) and the second property does not depend on $d$ at all. Thus we conclude the desired result.

Now we prove (iii). We know, by Eq. (4.1), that the consistency set $\Phi_{d,k}(\boldsymbol{L})$ is a subgroup of $\mathbb{U}_{k+1}^m$. Thus it suffices to prove that for $(a_1, \ldots, a_m) \in \Phi_{1,0}(\boldsymbol{L}) \subseteq \mathbb{U}_1^m$ we have $(P_{d,k}(|a_1|), \ldots, P_{d,k}(|a_m|)) \in \Phi_{d,k}(\boldsymbol{L})$. Given that $(a_1, \ldots, a_m) \in \Phi_{1,0}$ we know that there exists a finite-dimensional $\mathbb{F}_p$-vector space $V$, a linear function $P\colon V \to \mathbb{U}_1$, and vectors $\boldsymbol{x} = (x_1, \ldots, x_\ell) \in V^\ell$ such that $P(L_i(\boldsymbol{x})) = a_i$ for $i \in [m]$. Since $P$ and $L_i$ are linear, they commute, and thus $L_i(\boldsymbol{y}) = |a_i|$ for all $i \in [m]$ where $\boldsymbol{y} \in \mathbb{F}_p^\ell$ is defined by $y_i = |P(x_i)|$. Finally, since $P_{d,k}$ is a homogeneous non-classical polynomial of degree $d'$ and depth $k$, we have $(P_{d,k}(|a_1|), \ldots, P_{d,k}(|a_m|)) = (P_{d,k}(L_1(\boldsymbol{y})), \ldots, P_{d,k}(L_m(\boldsymbol{y})) \in \Phi_{d',k}(\boldsymbol{L})$.

To complete the proof, we show that let $\Phi_{d',k}(\boldsymbol{L}) \subseteq \Phi_{d,k}(\boldsymbol{L})$. Let $Q\colon \mathbb{F}_p^n \to \mathbb{U}_{k+1}$ be a homogeneous non-classical polynomial of degree $d$ and depth $k$. (This exists as long as $n \geq \lceil (d-1)/(p-1) \rceil - k$.) Then consider the map $P_{d,k} \oplus Q\colon \mathbb{F}_p \oplus \mathbb{F}_p^n \to \mathbb{U}_{k+1}$ defined as usual by $(P_{d,k} \oplus Q)(x \oplus y) := P_{d,k}(x) + Q(y)$. This is clearly a non-classical polynomial of degree $d$ and depth $k$. Furthermore,

$$(P_{d,k} \oplus Q)(b(x \oplus y)) = \sigma_b^{(d',k)} P_{d,k}(x) + \sigma_b^{(d,k)} Q(y) = \sigma_b^{(d,k)} (P_{d,k} \oplus Q)(x \oplus y),$$

since $d' \equiv d \pmod{p-1}$. Considering $\boldsymbol{y} \oplus \boldsymbol{0} \in (\mathbb{F}_p \oplus \mathbb{F}_p^n)^\ell$ shows that $(P_{d,k}(|a_1|), \ldots, P_{d,k}(|a_m|)) \in \Phi_{d,k}(\boldsymbol{L})$, as desired. □

## 4.4 Arithmetic regularity and subatom selection

In this section we prove the subatom selection result, Theorem 4.4.6, that will be one of the main technical tools used in the proof of our main theorem. To prove this result we first prove a strong arithmetic regularity lemma, Lemma 4.4.5, via a chain of regularity lemmas of varying strengths. From the strong regularity lemma we deduce the subatom selection result via the probabilistic method.

The first two regularity lemmas, Lemma 4.4.3 and Lemma 4.4.4, are very similar to results in [BFL12, BFH$^+$13], differing only in some technical details. The main innovation in this section is that Lemma 4.4.5 is much stronger than the corresponding result in [BFL12, BFH$^+$13]. To accomplish this, we iterate Lemma 4.4.4 with the complexity parameter (i.e., degree of the non-classical polynomials) increasing at each step of the iteration. To our knowledge, this idea has not appeared previously in the literature and it is crucial to handle patterns of unbounded complexity.

**Notation and conventions:** Recall that a polynomial factor $\mathfrak{B}$ on a vector space $V$ with parameters $I$ gives rise to a partition (or $\sigma$-algebra) on $V$ whose atoms are the fibers of the map $\mathfrak{B}\colon V \to A_I$. For a function $f\colon V \to \mathbb{C}$, we write $\mathbb{E}[f|\mathfrak{B}]\colon V \to \mathbb{C}$ for the projection of $f$ onto the $\sigma$-algebra generated by $\mathfrak{B}$. Concretely, $\mathbb{E}[f|\mathfrak{B}](x)$ is defined to be the average of $f$ over the atom of $\mathfrak{B}$ which contains $x$.

In this section we have to deal with many growth functions. Without loss of generality we always assume that these growth functions are monotone in all their parameters.

### 4.4.1 Arithmetic regularity lemmas

Given functions $f^{(1)}, \dots, f^{(R)}\colon V \to [0,1]$ and a polynomial factor $\mathfrak{B}$, we call the quantity

$$\sum_{\ell=1}^{R} \left|\mathbb{E}[f^{(\ell)}|\mathfrak{B}]\right|_2^2$$

the energy of $\mathfrak{B}$. The main technique we use in the proofs of regularity lemmas is an energy increment technique; using the fact that the energy is bounded between 0 and $R$, we show that some iteration must halt after a bounded number of steps.

By applying this iteration to the inverse theorem, we deduce a weak regularity lemma, Lemma 4.4.3, which produces a decomposition $f = f_{\mathsf{str}} + f_{\mathsf{psr}}$ where $f_{\mathsf{str}}$ is structured in the sense that $f_{\mathsf{str}} = \mathbb{E}[f|\mathfrak{B}]$ for some polynomial factor $\mathfrak{B}$ and $f_{\mathsf{psr}}$ is $U^{d+1}$-pseudorandom. Furthermore, it will be important that we can make the rank of $\mathfrak{B}$ arbitrarily large in terms of its size.

Iterating again gives a regularity lemma, Lemma 4.4.4, which produces a decomposition $f = f_{\mathsf{str}} + f_{\mathsf{psr}} + f_{\mathsf{sml}}$ where $f_{\mathsf{str}}$ is structured, $f_{\mathsf{psr}}$ is $U^{d+1}$-pseudorandom, and $f_{\mathsf{sml}}$ is small in $L^2$-norm. The main benefit of this result is that the amount of pseudorandomness can be arbitrarily small in terms of the size of $\mathfrak{B}$.

Iterating a third time gives a strong regularity lemma, Lemma 4.4.5, (sometimes called a "two-level decomposition theorem" or a "super decomposition theorem") which produces a pair of polynomial factors $\mathfrak{B}$ refined by $\mathfrak{B}'$ and a decomposition $f = f_{\mathsf{str}} + f_{\mathsf{psr}} + f_{\mathsf{sml}}$ where $f_{\mathsf{str}} = \mathbb{E}[f|\mathfrak{B}']$. This result has two main strengths. First, that $\|f_{\mathsf{sml}}\|_2$ can be made arbitrarily small in terms of the size of $\mathfrak{B}$. And second, the main innovation of this section, is that $f_{\mathsf{psr}}$ is $U^{d+1}$-pseudorandom where $d$ can be made arbitrarily larger in terms of the size of $\mathfrak{B}$. To accomplish the second property, in the proof we simply iterate Lemma 4.4.4 with the degree of the non-classical polynomials involved increasing at each step of the iteration.

Other than this innovation, most of the proofs in this section are fairly standard in this field. We include their proofs for completeness and because the exact technical details we require differ slightly from other results that have appeared previously.

**Theorem 4.4.1** (Inverse theorem [TZ12, Theorem 1.10])**.** *Fix a prime $p$, a positive integer $d$, and a parameter $\delta > 0$. There exists $\epsilon_{\mathsf{inv}}(p, d, \delta) > 0$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Given a function $f \colon V \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^{d+1}} > \delta$, there exists a non-classical polynomial $P \colon V \to \mathbb{R}/\mathbb{Z}$ of degree at most $d$ such that*

$$\left|\mathbb{E}_{x \in V} f(x) e^{-2\pi i P(x)}\right| \geq \epsilon_{\mathsf{inv}}(p, d, \delta).$$

The next lemma is important for making factors high rank and its second claim is critical in proving the stronger regularity lemma where we need to produce a refinement (instead of a weak refinement).

**Lemma 4.4.2** (Making factors high rank [HHL16], c.f. [BFH+13, Theorem 2.19])**.** *Fix a prime $p$, positive integers $d, C_0$, and a non-decreasing function $r\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. There exist constants $C_{\mathsf{rank}}(p, d, C_0, r)$ and $r_{\mathsf{rank}}(p, d, C_0, r)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Suppose that $\mathfrak{B}$ and $\mathfrak{B}'$ are polynomial factors on $V$ with degree at most $d$ such that $\mathfrak{B}'$ refines $\mathfrak{B}$ and $\|\mathfrak{B}'\| \leq C_0$ and*

$$\operatorname{rank} \mathfrak{B} \geq r_{\mathsf{rank}}(p, d, C_0, r).$$

*Then there is a polynomial factor $\mathfrak{B}''$ on $V$ that weakly refines $\mathfrak{B}'$, refines $\mathfrak{B}$, and satisfies $\|\mathfrak{B}''\| \leq C_{\mathsf{rank}}(p, d, C_0, r)$ and $\deg \mathfrak{B}'' \leq d$ and $\operatorname{rank} \mathfrak{B}'' \geq r(\|\mathfrak{B}''\|)$.*

**Lemma 4.4.3** (Weak arithmetic regularity)**.** *Fix a prime $p$, positive integers $d, R, C_0$, a parameter $\eta > 0$, and a non-decreasing function $r\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. There exist constants $C_{\mathsf{reg}'}(p, d, R, C_0, \eta, r)$ and $r_{\mathsf{reg}'}(p, d, R, C_0, \eta, r)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $\mathfrak{B}_0$ be a polynomial factor on $V$ satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg \mathfrak{B}_0 \leq d$ and $\operatorname{rank} \mathfrak{B}_0 \geq r_{\mathsf{reg}'}(p, d, R, C_0, \eta, r)$. Given functions $f^{(1)}, \dots, f^{(R)}\colon V \to [0, 1]$, there exists a polynomial factor $\mathfrak{B}$ on $V$ that refines $\mathfrak{B}_0$ with the following properties. There exists a decomposition*

$$f^{(\ell)} = f^{(\ell)}_{\mathsf{str}} + f^{(\ell)}_{\mathsf{psr}}$$

*for each $\ell \in [R]$ such that:*

*(i)* $f^{(\ell)}_{\mathsf{str}} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}]$ *for each $\ell \in [R]$;*

*(ii)* $\|f^{(\ell)}_{\mathsf{psr}}\|_{U^{d+1}} < \eta$ *for each $\ell \in [R]$;*

*(iii)* $f^{(\ell)}_{\mathsf{str}}$ *has range $[0, 1]$ and $f^{(\ell)}_{\mathsf{psr}}$ has range $[-1, 1]$ for each $\ell \in [R]$;*

*(iv)* $\operatorname{rank} \mathfrak{B} \geq r(\|\mathfrak{B}\|)$;

*(v)* $\|\mathfrak{B}\| \leq C_{\mathsf{reg}'}(p, d, R, C_0, \eta, r)$ *and* $\deg \mathfrak{B} \leq d$.

*Proof.* Set $M := \lceil R\epsilon_{\mathsf{inv}}(p, d, \eta)^{-2} \rceil$. Define non-decreasing functions $r_i\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ for each $i = 0, \dots, M$ such that $r_0 = r$ and $r_{i+1}(N) \geq r_{\mathsf{rank}}(p, d, p^{d^3}N, r_i)$ and such that $r_{i+1}(N) \geq r_i(N)$ for all $i = 0, \dots, M-1$ and all $N \in \mathbb{Z}_{>0}$. Define $r_{\mathsf{reg}'}(p, d, R, C_0, \eta, r) := r_M(C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0, \mathfrak{B}_1, \dots, \mathfrak{B}_m$ on $V$ such that

- $\mathfrak{B}_i$ refines $\mathfrak{B}_{i-1}$ for $i = 1, \dots, m$;
- $\operatorname{rank} \mathfrak{B}_i \geq r_{M-i}(\|\mathfrak{B}_i\|)$ for $i = 0, \dots, m$;
- $\|\mathfrak{B}_i\| \leq C_{\mathsf{rank}}(p, d, p^{d^3}\|\mathfrak{B}_{i-1}\|, r_{M-i})$ and $\deg \mathfrak{B}_i \leq d$ for $i = 1, \dots, m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0, \dots, \mathfrak{B}_i$ with the above properties. If $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_{U^{d+1}} < \eta$ for each $\ell \in [R]$ we halt the iteration. Otherwise there is some $\ell \in [R]$ such that, writing $g := f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]$, we have

$$\|g\|_{U^{d+1}} \geq \eta.$$

By Theorem 4.4.1, there exists a non-classical polynomial $P\colon V \to \mathbb{R}/\mathbb{Z}$ of degree at most $d$ such that

$$\left|\mathbb{E}_{x \in V} g(x) e^{-2\pi i P(x)}\right| \geq \epsilon_{\mathsf{inv}}(p, d, \eta).$$

By Theorem 4.3.9, we can write $P = P_1 + \cdots + P_C$ as the sum of homogeneous non-classical polynomials. There are at most $\sum_{i=1}^{d} 1 + \lfloor (i-1)/(p-1) \rfloor \leq d^2$ terms in this sum. Let $\mathfrak{B}_i'$ to be the polynomial factor defined by the polynomials defining $\mathfrak{B}$ as well as the polynomials $P_1, \dots, P_C$. Note that $\|\mathfrak{B}_i'\| \leq p^{d^3}\|\mathfrak{B}_i\|$. Finally let $\mathfrak{B}_{i+1}$ be the polynomial factor produced by applying Lemma 4.4.2 to $\mathfrak{B}_i$ and $\mathfrak{B}_i'$ with parameters $p, d, r_{M-i-1}$. In particular $\mathfrak{B}_i'$ refines $\mathfrak{B}_i$ and

$$\begin{aligned}
\operatorname{rank} \mathfrak{B}_i &\geq r_{M-i}(\|\mathfrak{B}_i\|) \\
&\geq r_{\mathsf{rank}}(p, d, p^{d^3}\|\mathfrak{B}_i\|, r_{M-i-1}) \\
&\geq r_{\mathsf{rank}}(p, d, \|\mathfrak{B}_i'\|, r_{M-i-1}),
\end{aligned}$$

so the hypotheses of Lemma 4.4.2 are satisfied. Thus we have defined $\mathfrak{B}_{i+1}$ with all the desired properties.

Next we claim that this iteration must stop after at most $M$ steps. We claim that

$$\sum_{\ell=1}^{R} \left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\right\|_2^2$$

increases by at least $\epsilon_{\mathsf{inv}}(p,d,\eta)^2$ each time $i$ increases. Since this sum is clearly bounded between 0 and $R$, it suffices to prove this claim.

First note that by the Cauchy-Schwarz inequality, $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i+1}]\|_2^2 \geq \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2$ holds for all $\ell$. Now pick $\ell \in [R]$ as in the $i$th iteration, define $g := f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]$, and let $P$ be the non-classical polynomial defined in the $i$th iteration. Note in particular that $e^{-2\pi i P(x)}$ is in the $\sigma$-algebra defined by $\mathfrak{B}_i'$. Then we compute

$$\begin{aligned}
\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i+1}]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 &\geq \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i']\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 \\
&= \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i'] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 \\
&= \|\mathbb{E}[g|\mathfrak{B}_i']\|_2^2 \\
&\geq \left\langle \mathbb{E}[g|\mathfrak{B}_i'], e^{2\pi i P}\right\rangle^2 \\
&= \left\langle g, e^{2\pi i P}\right\rangle^2 \\
&\geq \epsilon_{\mathsf{inv}}(p,d,\eta)^2.
\end{aligned}$$

Thus we have produced $\mathfrak{B}_m$ with $m \leq M$ such that $\mathfrak{B}_m$ refines $\mathfrak{B}_0$ and $\operatorname{rank}\mathfrak{B}_m \geq r_{M-m}(\|\mathfrak{B}_m\|) \geq r(\|\mathfrak{B}_m\|)$ and $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m]\|_{U^{d+1}} < \eta$ for each $\ell \in [R]$. Defining $f_{\mathsf{str}}^{(\ell)} := \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m]$ and $f_{\mathsf{psr}}^{(\ell)} := f^{(\ell)} - f_{\mathsf{str}}^{(\ell)}$, we immediately see that conclusions (i), (ii), (iii), and (iv) hold. Conclusion (v) holds by defining $C_{\mathsf{reg}'}(p,d,R,C_0,\eta,r)$ to be the $M$-fold iteration of the function $N \mapsto C_{\mathsf{rank}}(p,d,p^{d^3}N,r_M)$ applied to $C_0$. □

**Lemma 4.4.4** (Arithmetic regularity)**.** *Fix a prime $p$, positive integers $d, R, C_0$, a parameter $\theta > 0$, a non-increasing function $\eta\colon \mathbb{Z}_{>0} \to (0,1)$, and a non-decreasing function $r\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. There exist constants $C_{\mathsf{reg}''}(p,d,R,C_0,\theta,\eta,r)$ and $r_{\mathsf{reg}''}(p,d,R,C_0,\theta,\eta,r)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $\mathfrak{B}_0$ be a polynomial factor on $V$ satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg\mathfrak{B}_0 \leq d$ and $\operatorname{rank}\mathfrak{B}_0 \geq r_{\mathsf{reg}''}(p,d,r,I_0,\theta,\eta,r)$. Given functions $f^{(1)},\dots,f^{(R)}\colon V \to [0,1]$, there exists a polynomial factor $\mathfrak{B}$ on $V$ that refines*

$\mathfrak{B}_0$ *with the following properties. There exists a decomposition*

$$f^{(\ell)} = f^{(\ell)}_{\mathsf{str}} + f^{(\ell)}_{\mathsf{psr}} + f^{(\ell)}_{\mathsf{sml}}$$

*for each* $\ell \in [R]$ *such that:*

*(i)* $f^{(\ell)}_{\mathsf{str}} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}]$ *for each* $\ell \in [R]$*;*

*(ii)* $\|f^{(\ell)}_{\mathsf{psr}}\|_{U^{d+1}} < \eta(\|\mathfrak{B}\|)$ *for each* $\ell \in [R]$*;*

*(iii)* $f^{(\ell)}_{\mathsf{str}}$ *and* $f^{(\ell)}_{\mathsf{str}} + f^{(\ell)}_{\mathsf{sml}}$ *have range* $[0,1]$ *and* $f^{(\ell)}_{\mathsf{psr}}$ *and* $f^{(\ell)}_{\mathsf{sml}}$ *have range* $[-1,1]$ *for each* $\ell \in [R]$*;*

*(iv)* $\operatorname{rank}\mathfrak{B} \geq r(\|\mathfrak{B}\|)$*;*

*(v)* $\|f^{(\ell)}_{\mathsf{sml}}\|_2 < \theta$ *for each* $\ell \in [R]$*;*

*(vi)* $\|\mathfrak{B}\| \leq C_{\mathsf{reg}''}(p,d,R,C_0,\theta,\eta,r)$ *and* $\deg\mathfrak{B} \leq d$*.*

*Proof.* Set $M := \lceil R\theta^{-2}\rceil$. Define non-decreasing functions $r_i\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ for each $i = 0,\dots,M$ such that $r_0 = r$ and $r_{i+1}(N) \geq r_{\mathsf{reg}'}(p,d,R,N,\eta(N),r_i)$ and such that $r_{i+1}(N) \geq r_i(N)$ for all $i = 0,\dots,M-1$ and all $N \in \mathbb{Z}_{>0}$. Define $r_{\mathsf{reg}''}(p,d,R,C_0,\theta,\eta,r) := r_M(C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0,\mathfrak{B}_1,\dots,\mathfrak{B}_m$ on $V$ such that

- $\mathfrak{B}_i$ refines $\mathfrak{B}_{i-1}$ for $i = 1,\dots,m$;
- $\operatorname{rank}\mathfrak{B}_i \geq r_{M-i}(\|\mathfrak{B}_i\|)$ for $i = 0,\dots,m$;
- $\|\mathfrak{B}_i\| \leq C_{\mathsf{reg}'}(p,d,R,\|\mathfrak{B}_{i-1}\|,\eta(\|\mathfrak{B}_{i-1}\|),r_{M-i})$ and $\deg\mathfrak{B}_i \leq d$ for $i = 1,\dots,m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0,\dots,\mathfrak{B}_i$ with the above properties. If $i \geq 1$ and $\|E[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\|_2^2 < \theta^2$ for each $\ell \in [R]$ we halt the iteration. Otherwise let $\mathfrak{B}_{i+1}$ be the polynomial factor produced by applying Lemma 4.4.3 to $\mathfrak{B}_i$ with parameters $p,d,R,\|\mathfrak{B}_i\|,\eta(\|\mathfrak{B}_i\|),r_{M-i-1}$. Note that

$$\begin{aligned}\operatorname{rank}\mathfrak{B}_i &\geq r_{M-i}(\|\mathfrak{B}_i\|)\\ &\geq r_{\mathsf{reg}'}(p,d,R,\|\mathfrak{B}_i\|,\eta(\|\mathfrak{B}_i\|),r_{M-i-1}),\end{aligned}$$

so the hypotheses of Lemma 4.4.3 are satisfied.

Next we claim that this iteration must stop after at most $M$ steps. This is obvious since

$$\sum_{\ell=1}^{R} \left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\right\|_2^2 - \sum_{\ell=1}^{R} \left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\right\|_2^2 \geq \theta^2$$

for $i = 1, \dots, m$ and the sum is bounded between 0 and $R$.

Thus we have produced $\mathfrak{B}_{m-1}$ with $m \leq M$ such that $\mathfrak{B}_{m-1}$ refines $\mathfrak{B}_0$ and $\operatorname{rank} \mathfrak{B}_{m-1} \geq r_{M-m+1}(\|\mathfrak{B}_{m-1}\|) \geq r(\|\mathfrak{B}_{m-1}\|)$ and $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m]\|_{U^{d+1}} < \eta(\|\mathfrak{B}_{m-1}\|)$ for each $\ell \in [R]$ and $\left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]\right\|_2 < \theta$ for each $\ell \in [R]$. Defining $f_{\mathsf{str}}^{(\ell)} := \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]$ and $f_{\mathsf{psr}}^{(\ell)} := f^{(\ell)} - E[f^{(\ell)}|\mathfrak{B}_m]$ and $f_{\mathsf{sml}}^{(\ell)} := \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]$, we immediately see that conclusions (i), (ii), (iii), (iv), and (v) hold. Conclusion (vi) holds by defining $C_{\mathsf{reg}''}(p, d, R, C_0, \theta, \eta, r)$ to be the $M$-fold iteration of the function $N \mapsto C_{\mathsf{reg}'}(p, d, R, N, \eta(N), r_M)$ applied to $C_0$. $\square$

Now we are ready to prove the final arithmetic regularity lemma. In this result we produce a pair of polynomial factors, $\mathfrak{B}$ refined by $\mathfrak{B}'$. This result is stronger in the sense that the $\theta$ that appeared in Lemma 4.4.4(v) and the $d$ in Lemma 4.4.4(ii) are allowed to be functions of the size of $\mathfrak{B}$. Furthermore we have an additional property, Lemma 4.4.5(vi), that states that a typical subatom $a$ of $\mathfrak{B}'$ "looks like" the atom $\pi(a)$ of $\mathfrak{B}$ that it is contained in.

**Lemma 4.4.5** (Strong arithmetic regularity)**.** *Fix a prime $p$, positive integers $R, C_0, d_0$, a parameter $\zeta > 0$, non-increasing functions $\eta, \theta \colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to (0,1)$, and non-decreasing functions $d, r \colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. There exist constants $C_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ and $D_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ and $r_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and let $\mathfrak{B}_0$ be a polynomial factor on $V$ satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg \mathfrak{B}_0 \leq d_0$ and $\operatorname{rank} \mathfrak{B}_0 \geq r_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$. Given functions $f^{(1)}, \dots, f^{(R)} \colon V \to [0,1]$, there exist a polynomial factor $\mathfrak{B}$ and a refinement $\mathfrak{B}'$ both on $V$ with parameters $I$ and $I'$ with the following properties. There exists a decomposition*

$$f^{(\ell)} = f_{\mathsf{str}}^{(\ell)} + f_{\mathsf{psr}}^{(\ell)} + f_{\mathsf{sml}}^{(\ell)}$$

*for each $\ell \in [R]$ such that:*

*(i) $f_{\mathsf{str}}^{(\ell)} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}']$ for each $\ell \in [R]$;*

*(ii)* $\|f_{\mathsf{psr}}^{(\ell)}\|_{U^{d(\deg \mathfrak{B}, \|\mathfrak{B}\|)+1}} < \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$ *for each* $\ell \in [R]$*;*

*(iii)* $f_{\mathsf{str}}^{(\ell)}$ *and* $f_{\mathsf{str}}^{(\ell)} + f_{\mathsf{sml}}^{(\ell)}$ *have range* $[0,1]$ *and* $f_{\mathsf{psr}}^{(\ell)}$ *and* $f_{\mathsf{sml}}^{(\ell)}$ *have range* $[-1,1]$ *for each* $\ell \in [R]$*;*

*(iv)* $\operatorname{rank} \mathfrak{B} \geq r(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ *and* $\operatorname{rank} \mathfrak{B}' \geq r(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$*;*

*(v)* $\|f_{\mathsf{sml}}^{(\ell)}\|_2 < \theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ *for each* $\ell \in [R]$*;*

*(vi) for all but at most a* $\zeta$*-fraction of* $a \in A_{I'}$ *it holds that*

$$\left|\mathbb{E}_{x \in \mathfrak{B}'^{-1}(a)}[f^{(\ell)}(x)] - \mathbb{E}_{x \in \mathfrak{B}^{-1}(\pi(a))}[f^{(\ell)}(x)]\right| < \zeta$$

*for each* $\ell \in [R]$ *(recall the definition of the atom indexing sets from Eq.* (4.3) *and the projection map* $\pi \colon A_{I'} \to A_I$ *from Eq.* (4.4)*);*

*(vii)* $\|\mathfrak{B}'\| \leq C_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ *and* $\deg \mathfrak{B}' \leq D_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$*.*

*Proof.* Set $M := \lceil R\zeta^{-3} \rceil$. Define non-decreasing functions $r_i \colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ for each $i = 0, \ldots, M$ such that $r_0 = r$,

$$r_{i+1}(D, N) \geq r_{\mathsf{reg}''}(p, d(D,N), R, N, \theta(D,N), \eta(d(D,N), \cdot), r_i(d(D,N), \cdot)),$$

and such that $r_{i+1}(D,N) \geq r_i(D,N)$ for all $i = 0, \ldots, M-1$ and all $D, N \in \mathbb{Z}_{>0}$. Define $r_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r) := r_M(d_0, C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0, \mathfrak{B}_1, \ldots, \mathfrak{B}_m$ on $V$ such that

- $\mathfrak{B}_i$ refines $\mathfrak{B}_{i-1}$ for $i = 1, \ldots, m$;
- $\operatorname{rank} \mathfrak{B}_i \geq r_{M-i}(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|)$ for $i = 0, \ldots, m$;
- $\|\mathfrak{B}_i\| \leq C_{\mathsf{reg}''}(p, d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|), R, \|\mathfrak{B}_{i-1}\|, \theta, \eta, r_{M-i}(d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|), \cdot))$ and $\deg \mathfrak{B}_i \leq d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|)$ for $i = 1, \ldots, m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0, \ldots, \mathfrak{B}_i$ with the above properties. If $i \geq 1$ and $\|E[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\|_2^2 < \zeta^3$ for each $\ell \in [R]$ we halt the iteration. Otherwise let $\mathfrak{B}_{i+1}$ be the polynomial factor produced by applying Lemma 4.4.4 to $\mathfrak{B}_i$ with parameters

$p, d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), R, \|\mathfrak{B}_i\|, \theta(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \eta(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot), r_{M-i-1}(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot)$. Note that

$$\begin{aligned}\operatorname{rank} \mathfrak{B}_i &\geq r_{M-i}(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|) \\ &\geq r_{\mathsf{reg}''}(p, d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), R, \|\mathfrak{B}_i\|, \theta(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \\ &\qquad \eta(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot), r_{M-i-1}(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot)),\end{aligned}$$

so the hypotheses of Lemma 4.4.4 are satisfied.

Next we claim that this iteration must stop after at most $M$ steps. This is obvious since

$$\sum_{\ell=1}^{R} \left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\right\|_2^2 - \sum_{\ell=1}^{R} \left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\right\|_2^2 \geq \zeta^3$$

for $i = 1, \ldots, m$ and the sum is bounded between 0 and $R$.

Thus we have produced $\mathfrak{B}_{m-1}$ and $\mathfrak{B}_m$ with $m \leq M$ such that $\mathfrak{B}_{m-1}$ refines $\mathfrak{B}_0$ and $\mathfrak{B}_m$ refines $\mathfrak{B}_{m-1}$. Also $\operatorname{rank} \mathfrak{B}_{m-1} \geq r_{M-m+1}(\deg \mathfrak{B}_{m-1}, \|\mathfrak{B}_{m-1}\|) \geq r(\deg \mathfrak{B}_{m-1}, \|\mathfrak{B}_{m-1}\|)$ and $\operatorname{rank} \mathfrak{B}_m \geq r_{M-m}(\deg \mathfrak{B}_m, \|\mathfrak{B}_m\|) \geq r(\deg \mathfrak{B}_m, \|\mathfrak{B}_m\|)$. In addition, we know that $\left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]\right\|_2^2 < \zeta^3$ for each $\ell \in [R]$. Let $f^{(\ell)} = f_{\mathsf{str}}^{(\ell)} + f_{\mathsf{psr}}^{(\ell)} + f_{\mathsf{sml}}^{(\ell)}$ be the decomposition produced by the last application of Lemma 4.4.4. This decomposition satisfies conclusions (i), (ii), (iii), and (v). Conclusion (iv) we already verified, and conclusion (vi) follows from Markov's inequality applied to the bound $\left\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]\right\|_2^2 < \zeta^3$. Finally conclusion (vii) holds where we define the pair

$$(D_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r), C_{\mathsf{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r))$$

to be the $M$-fold iteration of the function

$$(D, N) \mapsto (C_{\mathsf{reg}''}(p, d(D, N), R, N, \theta, \eta, r_M(d(D, N), \cdot)), d(D, N))$$

applied to $(d_0, C_0)$. □

### 4.4.2 Subatom selection

Let us pause to recall our setup. We have a polynomial factor $\mathfrak{B}$ and a refinement $\mathfrak{B}'$. The atoms of $\mathfrak{B}$ are indexed by $A_I$ and the atoms of $\mathfrak{B}'$, which we call "subatoms", are indexed by $\mathfrak{B}'$. The map $\pi\colon A_{I'} \to A_I$, defined by Eq. (4.4) maps a subatom $a$ to the atom $\pi(a)$ which contains it.

The goal of this section is to produce a function $s\colon A_I \to A_{I'}$, called a subatom selection function, which for each atom $a \in A_I$ selects a single subatom $s(a) \in A_{I'}$. In Definition 4.3.24 we defined subatom selection functions to have a very specific form. In this section we will prove that selecting a uniform random subatom selection function will have many desirable properties with positive probability.

The concept of subatom selection has appeared in past works [BFL12, BFH$^+$13]. However, as those works were restricted to translation-invariant patterns, their subatom selection functions are allowed to be affine in some sense. In fact, in those papers the subatom selection function was essentially chosen to be a uniform random constant function. In our case our subatom selection function must be linear in an appropriate sense. This constraint is why the form of our subatom selection functions is more complicated.

For an atom $a \in A_I$, we write $a_{d,k} \in \mathbb{U}_{k+1}^{I_{d,k}}$ to be the degree $d$, depth $k$ part of $a$. Say that $a$ is "good" if $a_{1,0} \neq 0$. Our subatom selection function will be chosen so that $s(a)$ is a regular subatom for every good $a$: see Theorem 4.4.6(v) below. For bad $a$, we cannot ensure that $s(a)$ is regular; this is the set of atoms which we will later apply patching to. For an intuitive explanation of the good/bad distinction, note that from Definition 4.3.24 that if $a$ is bad, then $s(a)$ is constant as $s$ varies. We will prove that if $a$ is good, when $s$ is chosen uniformly random, then $s(a)$ is uniformly distributed over $\pi^{-1}(a)$.

**Theorem 4.4.6** (Subatom selection)**.** *Fix a prime $p$, positive integers $R, c_0$, a parameter $\zeta > 0$, non-increasing functions $\eta, \theta\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to (0,1)$, and non-decreasing functions $d, r\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. There exist constants $C_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ and $D_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ and $n_{\mathsf{reg}}(p, c_0, \zeta)$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space satisfying $\dim V \geq n_{\mathsf{reg}}(p, c_0, \zeta)$. Given functions $f^{(1)}, \dots, f^{(R)}\colon V \to [0,1]$, there exist a polynomial factor $\mathfrak{B}$ and a refinement $\mathfrak{B}'$ both on $V$ with parameters $I$ and $I'$ with the following*

*properties. There exists a subatom selection function* $s\colon A_I \to A_{I'}$ *and a decomposition*

$$f^{(\ell)} = f^{(\ell)}_{\mathsf{str}} + f^{(\ell)}_{\mathsf{psr}} + f^{(\ell)}_{\mathsf{sml}}$$

*for each* $\ell \in [R]$ *such that:*

*(i)* $f^{(\ell)}_{\mathsf{str}} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}']$ *for each* $\ell \in [R]$*;*

*(ii)* $\|f^{(\ell)}_{\mathsf{psr}}\|_{U^{d(\deg \mathfrak{B}, \|\mathfrak{B}\|)+1}} < \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$ *for each* $\ell \in [R]$*;*

*(iii)* $f^{(\ell)}_{\mathsf{str}}$ *and* $f^{(\ell)}_{\mathsf{str}} + f^{(\ell)}_{\mathsf{sml}}$ *have range* $[0,1]$ *and* $f^{(\ell)}_{\mathsf{psr}}$ *and* $f^{(\ell)}_{\mathsf{sml}}$ *have range* $[-1,1]$ *for each* $\ell \in [R]$*;*

*(iv)* $\operatorname{rank} \mathfrak{B} \geq r(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ *and* $\operatorname{rank} \mathfrak{B}' \geq r(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$*;*

*(v) for each* $a \in A_I$ *with* $a_{1,0} \neq 0$*, it holds that*

$$\|f^{(\ell)}_{\mathsf{sml}} 1_{\mathfrak{B}'^{-1}(s(a))}\|_2 < \theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)\|1_{\mathfrak{B}'^{-1}(s(a))}\|_2$$

*for each* $\ell \in [R]$*;*

*(vi) for all but at most a* $\zeta$*-fraction of* $a \in A_I$ *it holds that*

$$\left|\mathbb{E}_{x \in \mathfrak{B}^{-1}(a)}[f^{(\ell)}(x)] - \mathbb{E}_{x \in \mathfrak{B}'^{-1}(s(a))}[f^{(\ell)}(x)]\right| < \zeta$$

*for each* $\ell \in [R]$*;*

*(vii)* $I_{1,0} \geq c_0$*;*

*(viii)* $\|\mathfrak{B}'\| \leq C_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ *and* $\deg \mathfrak{B}' \leq D_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$*.*

*Proof.* Define $n_{\mathsf{reg}}(p, c_0, \zeta) := \max\{c_0, \lceil \log_p(2/\zeta) \rceil\}$. Let $\mathfrak{B}_0$ be a polynomial factor on $V$ defined by $n_{\mathsf{reg}}(p, c_0, \zeta)$ linearly independent linear functions. Define $\theta'\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ by $\theta'(D, N) := \theta(D, N)/(2\sqrt{R}N)$. We apply Lemma 4.4.5 to $\mathfrak{B}_0$ with parameters $p, R, \|\mathfrak{B}_0\|, 1, \zeta/4, \eta, \theta', d, r$. Let $\mathfrak{B}$ and $\mathfrak{B}'$ be the polynomial factors produced. We immediately have conclusions (i), (ii), (iii), and (iv). Conclusion (vii) follows since $\mathfrak{B}$ refines

$\mathfrak{B}_0$ which is defined by at least $c_0$ linear forms. Conclusion (viii) also follows by defining the constants $C_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r) = C_{\mathsf{reg}'''}(p, R, n_{\mathsf{reg}}(p, c_0, \zeta), 1, \zeta/4, \eta, \theta', d, r)$ and $D_{\mathsf{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r) = D_{\mathsf{reg}'''}(p, R, p^{n_{\mathsf{reg}}(p,c_0,\zeta)}, 1, \zeta/4, \eta, \theta', d, r)$.

Let $c_{d,k}^{i,j} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ be elements chosen independently and uniformly at random for each $(d,k) \in D_p$ and $i \in [I_{1,0}]$ and $I_{d,k} < j \leq I'_{d,k}$. Consider the subatom selection function $s_{\boldsymbol{c}} \colon A_I \to A_{I'}$ defined in Definition 4.3.24. We claim that with positive probability, this $s_{\boldsymbol{c}}$ satisfies conclusions (v) and (vi).

Fix $a \in A_I$ with $a_{1,0} \neq 0$. We first claim that for this fixed $a$, as $\boldsymbol{c}$ varies, the subatom $s(a)$ is uniformly distributed over $\pi^{-1}(a) \subset A_{I'}$. To see this, first note that the univariate homogeneous non-classical polynomials $P_{d,k} \colon \mathbb{F}_p \to \mathbb{U}_{k+1}$ used in the definition of $s_{\boldsymbol{c}}$ satisfy $P_{d,k}(x) \notin \mathbb{U}_k$ for all $x \neq 0$. This follows by homogeneity: if $P_{d,k}(x) \in U_k$ for some $x \neq 0$, then $P_{d,k}$ always takes values in $\mathbb{U}_k$, contradicting the assumption that $P_{d,k}$ has depth exactly $k$. Thus for $a \in A_I$ with $a_{1,0} \neq 0$, we find that the vector $(P_{d,k}(|a_{1,0}^1|), \dots, P_{d,k}(|a_{1,0}^{I_{1,0}}|)) \in \mathbb{U}_{k+1}^{I_{1,0}}$ does not lie in $\mathbb{U}_k^{I_{1,0}}$.

Considering the definition of $s_{\boldsymbol{c}}$, we see that the vector $([s_{\boldsymbol{c}}]_{d,k}^i)_{I_{d,k} < i \leq I'_{d,k}}$ is produced by the following matrix multiplication.

$$\begin{pmatrix} [s_{\boldsymbol{c}}]_{d,k}^{I_{d,k}+1} \\ [s_{\boldsymbol{c}}]_{d,k}^{I_{d,k}+2} \\ \vdots \\ [s_{\boldsymbol{c}}]_{d,k}^{I'_{d,k}} \end{pmatrix} = \begin{pmatrix} c_{d,k}^{I_{d,k}+1,1} & c_{d,k}^{I_{d,k}+1,2} & \cdots & c_{d,k}^{I_{d,k}+1,I_{d,k}} \\ c_{d,k}^{I_{d,k}+2,1} & c_{d,k}^{I_{d,k}+2,2} & \cdots & c_{d,k}^{I_{d,k}+2,I_{d,k}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{d,k}^{I'_{d,k},1} & c_{d,k}^{I'_{d,k},2} & \cdots & c_{d,k}^{I'_{d,k},I_{d,k}} \end{pmatrix} \begin{pmatrix} P_{d,k}(|a_{1,0}^1|) \\ P_{d,k}(|a_{1,0}^2|) \\ \vdots \\ P_{d,k}(|a_{1,0}^{I_{d,k}}|) \end{pmatrix}$$

As stated above, the vector on the right lies in $\mathbb{U}_{k+1}^{I_{1,0}}$ but not in $\mathbb{U}_k^{I_{1,0}}$ while the matrix is uniform random with entries in $\mathbb{Z}/p^{k+1}\mathbb{Z}$. This implies that the vector on the left is uniformly distributed over $\mathbb{U}_{k+1}^{I'_{d,k}-I_{d,k}}$, as desired. This holds for each $(d,k) \in D_p$ and furthermore, since for each $(d,k)$ the $c$ matrices are chosen independently, one can see that the resulting vectors are also independent, proving the desired result.

Now note that by Lemma 4.4.5(vi), for each $\ell \in [R]$,

$$\sum_{a \in A_{I'}} \|f_{\mathsf{sml}}^{(\ell)} 1_{\mathfrak{B}'^{-1}(a)}\|_2^2 \leq \frac{\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2}{4R\|\mathfrak{B}\|^2} \sum_{a \in A_{I'}} \|1_{\mathfrak{B}'^{-1}(a)}\|_2^2.$$

Thus by Markov's inequality, for $a' \in A_{I'}$ chosen uniformly at random, with probability at least $1 - 1/(4R\|\mathfrak{B}\|^2)$ the following holds

$$\|f^{(\ell)}_{\mathsf{sml}} 1_{\mathfrak{B}'^{-1}(a')}\|_2^2 \leq \theta(\|\mathfrak{B}\|)^2 \|1_{\mathfrak{B}'^{-1}(a')}\|_2^2. \tag{4.6}$$

Thus for a fixed $a \in A_I$ with $a_{1,0} \neq 0$, we have that for $\boldsymbol{c}$ chosen at random, we have $s_{\boldsymbol{c}}(a)$ satisfies Eq. (4.6) with probability at least $1 - 1/(4R\|\mathfrak{B}\|)$. Taking a union bound over all choices of $a \in A_I$ and $\ell \in [R]$, we see that with probability at least $3/4$, conclusion (v) holds.

Now we deduce conclusion (vi). First note that the fraction of $a \in A_I$ which satisfy $a_{1,0} = 0$ is $p^{-I_{1,0}} \leq \zeta/2$. For the other $a \in A_I$, the expected fraction of $a \in A_I$ that fail to satisfy the desired inequality is at most $\zeta/4$. Thus by Markov's inequality, with probability at least $1/2$, at most $\zeta/2$ fraction of $a \in A_I$ satisfy $a_{1,0} \neq 0$ and fail to satisfy the desired inequality. Thus with probability at least $1/2$, conclusion (vi) holds. □

## 4.5 Patching

In this section we prove the main patching theorem we need, Theorem 4.5.9. We call this result a supersaturation dichotomy theorem for reasons we will explain shortly. This is proved from a Ramsey dichotomy result, Theorem 4.5.6. This Ramsey dichotomy result is fairly easy to deduce from an appropriate Ramsey-type theorem which we also prove in this section.

### 4.5.1 A motivating example

To motivate the kind of results proved in this section, we give a series of dichotomy results in the much simpler setting of edge-colored graphs. These results are analogous to the dichotomy results we will prove later in this section.

We start with a tautological dichotomy which is essentially trivial.

**Fact 0** (Tautological dichotomy)**.** *Let $\mathcal{H}$ be a finite set of red/blue edge-colored graphs. There exists an integer $n_0 = n_0(\mathcal{H})$ such that the following holds. Either:*

*(a) for every $n$, there exists a red/blue edge-coloring of $K_n$ that contains no subgraph from $\mathcal{H}$; or*

*(b) every red/blue edge-coloring of $K_n$ with $n \geq n_0$ contains a subgraph from $\mathcal{H}$.*

In case (a), we find $\mathcal{H}$-free colorings of arbitrarily-large cliques. Using Ramsey's theorem, we can find arbitrarily-large monochromatic cliques in these $\mathcal{H}$-free colorings. This lets us boost the previous result into the following stronger Ramsey dichotomy result. (One could also prove this result directly from Ramsey's theorem.)

**Fact 1** (Ramsey dichotomy)**.** *Let $\mathcal{H}$ be a finite set of red/blue edge-colored graphs. There exists an integer $n_0 = n_0(\mathcal{H})$ such that the following holds. Either:*

*(a) either the all-red coloring of $K_n$ or the all-blue coloring of $K_n$ contains no subgraph from $\mathcal{H}$ for every $n$; or*

*(b) every red/blue edge-coloring of $K_n$ with $n \geq n_0$ contains a subgraph from $\mathcal{H}$.*

A sampling argument allows one to boost this dichotomy result into a supersaturation dichotomy result. This result allows one to find many copies of the desired graph instead of just one.

**Fact 2** (Supersaturation dichotomy)**.** *Let $\mathcal{H}$ be a finite set of red/blue edge-colored graphs. There exist $n_0 = n_0(\mathcal{H})$ and $\epsilon = \epsilon(\mathcal{H}) > 0$ such that the following holds. Either:*

*(a) either the all-red coloring of $K_n$ or the all-blue coloring of $K_n$ contains no subgraph from $\mathcal{H}$ for every $n$; or*

*(b) for every red/blue edge-coloring of $K_n$ with $n \geq n_0$ there exists a subgraph $H \in \mathcal{H}$ which has density at least $\epsilon$ in the coloring.*

The main dichotomy result which we will prove is analogous to Fact 1 and the main patching result is analogous to Fact 2.

### 4.5.2 Colored labeled patterns and canonical colorings

In this section we define the setting in which we prove our dichotomy results. Unfortunately for our application it is not sufficient to study colored patterns in colorings $f\colon V \to \mathcal{S}$. In our setting we also have a polynomial factor $\mathfrak{B}$ on $V$ and it is important to know how the pattern interacts with the factor. To do so we introduce the notion of a colored labeled pattern.

**Definition 4.5.1.** For a prime $p$, a finite set $\mathcal{S}$, and a parameter list $I \in \mathcal{I}_p$, an **$\mathcal{S}$-colored $I$-labeled pattern** over $\mathbb{F}_p$ consisting of $m$ linear forms in $\ell$ variables is a triple $(\boldsymbol{L}, \psi, \phi)$ given by:

- a system $\boldsymbol{L} = (L_1, \ldots, L_m)$ of $m$ linear forms in $\ell$ variables,
- a coloring $\psi\colon [m] \to \mathcal{S}$, and
- a labeling $\phi\colon [m] \to A_I$ (recall the definition of the atom-indexing set $A_I$ from Eq. (4.3)).

Given a finite-dimensional $\mathbb{F}_p$-vector space $V$, a function $f\colon V \to \mathcal{S}$, and a polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$, an **$(\boldsymbol{L}, \psi, \phi)$-instance** in $(f, \mathfrak{B})$ is some $\boldsymbol{x} \in V^\ell$ such that $f(L_i(\boldsymbol{x})) = \psi(i)$ for all $i \in [m]$ and $\mathfrak{B}(L_i(\boldsymbol{x})) = \phi(i)$ for all $i \in [m]$. An instance is called **generic** if $x_1, \ldots, x_\ell$ are linearly independent.

**Definition 4.5.2.** For an $\mathcal{S}$-colored $I$-labeled pattern $(\boldsymbol{L}, \psi, \phi)$ consisting of $m$ linear forms, a finite dimensional $\mathbb{F}_p$-vector space $V$, a function $f\colon V \to \mathcal{S}$, and a polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$, define the **$(\boldsymbol{L}, \psi, \phi)$-density** in $(f, \mathfrak{B})$ to be

$$\Lambda_{\boldsymbol{L}}(1_{f^{-1}(\psi(1)) \cap \mathfrak{B}^{-1}(\phi(1))}, \ldots, 1_{f^{-1}(\psi(m)) \cap \mathfrak{B}^{-1}(\phi(m))}).$$

Given a set $X \subseteq V$, define the **relative density of $(\boldsymbol{L}, \psi, \phi)$ in** $X$ to be

$$\frac{\Lambda_{\boldsymbol{L}}(f_1, \ldots, f_m)}{\Lambda_{\boldsymbol{L}}(1_X, \ldots, 1_X)}$$

where $f_i := 1_{X \cap f^{-1}(\psi(i)) \cap \mathfrak{B}^{-1}(\phi(i))}$.

In the graph example we gave in Section 4.5.1, the all-red and all-blue colorings played a special role. In this setting the analogous role is played by so-called canonical colorings which we define below. A similar concept appears in the simpler application of patching given in Chapter 2. In fact, similar ideas can be extracted from classical results in Ramsey theorem such as Rado's theorem.

**Definition 4.5.3.** Define the **first non-zero coordinate** function $\mathrm{fnz}\colon \mathbb{F}_p^n \to \mathbb{F}_p$ by $\mathrm{fnz}(0, \ldots, 0) := 0$ and $\mathrm{fnz}(x_1, \ldots, x_n) := x_k$ where $x_1 = \cdots = x_{k-1} = 0$ and $x_k \neq 0$.

Given a finite-dimensional $\mathbb{F}_p$-vector space $V$ equipped with an isomorphism $\iota\colon V \xrightarrow{\sim} \mathbb{F}_p^n$, define the function $\mathrm{fnz}_\iota\colon V \to \mathbb{F}_p$ by $\mathrm{fnz}_\iota(x) := \mathrm{fnz}(\iota(x))$.

**Definition 4.5.4.** Fix a prime $p$, a finite set $\mathcal{S}$, a parameter list $I \in \mathcal{I}_p$, and a function $\xi\colon \mathbb{F}_p \times A_I \to \mathcal{S}$. For a finite-dimensional $\mathbb{F}_p$-vector space $V$ equipped with an isomorphism $\iota\colon V \xrightarrow{\sim} \mathbb{F}_p^n$ and a polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$, define the **$\xi$-canonical coloring** $\Xi_{\xi,\iota,\mathfrak{B}}\colon V \to \mathcal{S}$ by $\Xi_{\xi,\iota,\mathfrak{B}}(x) := \xi(\mathrm{fnz}_\iota(x), \mathfrak{B}(x))$. Furthermore, if $\mathcal{S}$ is equipped with an $\mathbb{F}_p^\times$-action, say that $\xi$ is **projective** if the same is true for every function $\Xi_{\xi,\iota,\mathfrak{B}}$. (Note that this property is equivalent to the condition that $\xi$ preserves the action of $\mathbb{F}_p^\times$, i.e., $\xi(cx, c\cdot a) = c\cdot\xi(x,a)$ for all $c \in \mathbb{F}_p^\times$, all $x \in \mathbb{F}_p$, and all $a \in A_I$. Recall the action of $\mathbb{F}_p^\times$ on $A_I$ defined in Eq. (4.5).)

We will soon prove that if there exists a generic $H$-instance in some $\xi$-canonical coloring $\Xi_{\xi,\iota,\mathfrak{B}}$ of some $\mathbb{F}_p$-vector space, then there exists an $H$-instance in every $\xi$-canonical coloring of every sufficiently-large vector space equipped with a polynomial factor of sufficiently-high rank. We say that $\xi$ canonically induces $H$ if the former condition holds.

**Definition 4.5.5.** Given a prime $p$, a finite set $\mathcal{S}$, a parameter list $I \in \mathcal{I}_p$, a function $\xi\colon \mathbb{F}_p \times A_I \to \mathcal{S}$, and a $\mathcal{S}$-colored $I$-labeled pattern $H = (\boldsymbol{L}, \psi, \phi)$, say that $\xi$ **canonically induces** $H$ if the following holds. There exists some $n \geq 0$ and a polynomial factor $\mathfrak{B}$ on $\mathbb{F}_p^n$ with parameters $I$ such that there exists a generic $H$-instance in $(\Xi_{\xi,\mathrm{Id},\mathfrak{B}}, \mathfrak{B})$. For a finite set of $\mathcal{S}$-colored $I$-labeled patterns $\mathcal{H}$, say that $\xi$ **canonically induces** $\mathcal{H}$ if $\xi$ canonically induces some $H \in \mathcal{H}$.

### 4.5.3 Ramsey dichotomy result

In this section we prove our Ramsey dichotomy result, Theorem 4.5.6. To do so we will need two Ramsey-type theorems for this setting, Lemma 4.5.7 and Lemma 4.5.8.

Ramsey-type results, specifically the affine Ramsey theorem of Graham and Rothschild, allow one to find a large monochromatic *affine* subspace in a coloring $f\colon V \to \mathcal{S}$. In our setting we have a polynomial factor $\mathfrak{B}$ on $V$. The main technical result of this section, Lemma 4.5.8, is analogous to the affine Ramsey theorem with the additional restriction

that the affine subspace $U$ produced respects the structure of $\mathfrak{B}$ (essentially that $\mathfrak{B}|_U$ has sufficiently-high rank).

From this result, a fairly elementary argument lets us deduce Lemma 4.5.7, which says that one to find a large *linear* subspace which is canonically colored. Using this Ramsey result we can easily deduce the main Ramsey dichotomy result, Theorem 4.5.6. One has to be careful with some technical details related to high rank. For example, we will need to check that there exists at least one $\boldsymbol{L}$-instance lying in the correct atoms of $\mathfrak{B}$.

**Theorem 4.5.6** (Ramsey dichotomy)**.** *Fix a prime $p$, a finite set $\mathcal{S}$ with an $\mathbb{F}_p^\times$-action, a parameter list $I \in \mathcal{I}_p$, and a positive integer $\ell_0$. There exist constants $n_{\mathsf{dich}} = n_{\mathsf{dich}}(p, |\mathcal{S}|, I, \ell_0)$ and $r_{\mathsf{dich}} = r_{\mathsf{dich}}(p, |\mathcal{S}|, I, \ell_0)$ such that the following holds. Let $\mathcal{H}$ be a finite set of $\mathcal{S}$-colored, $I$-labeled patterns each defined by a system of linear forms in at most $\ell_0$ variables. Either:*

*(a) there exists a projective $\xi \colon \mathbb{F}_p \times A_I \to \mathcal{S}$ that does not canonically induce $\mathcal{H}$; or*

*(b) for every finite-dimensional $\mathbb{F}_p$-vector space $V$ satisfying $\dim V \geq n_{\mathsf{dich}}$, every projective function $f \colon V \to \mathcal{S}$, and every polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$ which has rank at least $r_{\mathsf{dich}}$, there is a generic $H$-instance in $(f, \mathfrak{B})$ for some $H \in \mathcal{H}$.*

*Proof.* Define $m$ to be the smallest positive integer such that the following holds. Let $H$ be an $\mathcal{S}$-colored, $I$-labeled pattern defined by a system of linear forms in at most $\ell_0$ variables and let $\xi \colon \mathbb{F}_p \times A_I \to \mathcal{S}$ be a projective function. If $\xi$ canonically induces $H$, then there exists some $n_H \leq m$ and a polynomial factor $\mathfrak{B}_H$ on $\mathbb{F}_p^{n_H}$ with parameters $I$ such that there exists a generic $H$-instance in $(\Xi_{\xi,\mathrm{Id},\mathfrak{B}_H}, \mathfrak{B}_H)$. This is well defined since there are only a finite number of $\mathcal{S}$-colored, $I$-labeled patterns defined by a system of linear forms in at most $\ell_0$ variables.

**Lemma 4.5.7.** *Fix a prime $p$, a finite set $\mathcal{S}$, a parameter list $I \in \mathcal{I}_p$, and positive integers $m, r_0, n_0$. There exist constants $n_{\mathsf{ramsey}} = n_{\mathsf{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0)$ and $r_{\mathsf{ramsey}} = r_{\mathsf{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0)$ such that the following holds. Let $V$ be a finite dimensional $\mathbb{F}_p$-vector space satisfying $\dim V \geq n_{\mathsf{ramsey}}$, let $\mathfrak{B}$ be a polynomial factor on $V$ with parameters $I$ such that $\operatorname{rank} \mathfrak{B} \geq r_{\mathsf{ramsey}}$, and let $f \colon V \to \mathcal{S}$ be a function. Then there exists a subspace $U \leq V$, and linear functions $P_1, \ldots, P_m \colon V \to \mathbb{F}_p$, and a function $\xi \colon \mathbb{F}_p \times A_I \to \mathcal{S}$ such that the following holds:*

*(i)* $\xi(\mathrm{fnz}(P_1(x),\dots,P_m(x)),\mathfrak{B}(x)) = f(x)$ *for all* $x \in U$ *that also satisfy* $(P_1(x),\dots,P_m(x)) \neq (0,\dots,0)$*;*

*(ii)* $\xi(0,0) = f(0)$

*(iii) the polynomial factor* $\mathfrak{B}'$ *on* $U$ *defined by the homogeneous non-classical polynomials that define* $\mathfrak{B}|_U$ *in addition to the polynomials* $P_1,\dots,P_m$ *satisfies* $\operatorname{rank}\mathfrak{B}' \geq r_0$*;*

*(iv)* $\dim U \geq n_0$.

Let us show how this lemma completes the proof. Define

$$r_0 := r_{\mathsf{equi}}\left(p, \deg I, \frac{1}{2(p^m\|I\|)^{p^m}}\right) \qquad \text{and} \qquad n_0 := 2p^m(m + \lceil \log_p \|I\| \rceil).$$

Then define

$$n_{\mathsf{dich}}(p,|\mathcal{S}|,I,\ell_0) := n_{\mathsf{ramsey}}(p,|\mathcal{S}|,I,m,r_0,n_0),$$

and

$$r_{\mathsf{dich}}(p,|\mathcal{S}|,I,\ell_0) := r_{\mathsf{ramsey}}(p,|\mathcal{S}|,I,m,r_0,n_0).$$

Let $\mathcal{H}$ be a finite set of $\mathcal{S}$-colored, $I$-labeled patterns each defined by a system of linear forms in at most $\ell_0$ variables. Suppose (a) does not hold. Thus for every projective $\xi\colon \mathbb{F}_p \times A_I \to \mathcal{S}$, there exists an $H \in \mathcal{H}$ such that $\xi$ canonically induces $H$.

Now we apply Lemma 4.5.7 to $f\colon V \to \mathcal{S}$. This produces a subspace $U \leq V$, linear functions $P_1,\dots,P_m$, and a function $\xi\colon \mathbb{F}_p \times A_I \to \mathcal{S}$ with several desirable properties.

First note that since $f\colon V \to \mathcal{S}$ is projective, the same is true of $\xi\colon \mathbb{F}_p \times A_I \to \mathcal{S}$. Thus by assumption there exists an $H \in \mathcal{H}$ such that $\xi$ canonically induces $H$. By the choices in the first paragraph, there exists a $n_H \leq m$ and a polynomial factor $\mathfrak{B}_H$ on $\mathbb{F}_p^{n_H}$ such that $(\Xi_{\xi,\mathrm{Id},\mathfrak{B}_H}, \mathfrak{B}_H)$ contains a generic $H$-instance.

To complete the proof, all we need to show is that there exists an injective linear map $\kappa\colon \mathbb{F}_p^{n_H} \to U$ with the following two properties. First $\mathfrak{B}_H(x) = \mathfrak{B}(\kappa(x))$ for all $x \in \mathbb{F}_p^{n_H}$ and second that $\mathrm{fnz}(x) = \mathrm{fnz}(P_1(\kappa(x)),\dots,P_m(\kappa(x)))$ for all $x \in \mathbb{F}_p^{n_H}$. This follows by an application of equidistribution (Theorem 4.3.19).

Recall the definition of $\boldsymbol{L}^{n_H}$ (Definition 4.2.7), the system of $p^{n_H}$ linear forms in $n_H$ variables that define an $n_H$-dimensional subspace.

Say that $\mathfrak{B}'$ has parameters $I'$. Thus the atom-indexing set of $\mathfrak{B}'$ can be written as $A_{I'} \simeq \mathbb{F}_p^m \times A_I$. We define the following tuple of atoms $\boldsymbol{a} = (a_{\boldsymbol{i}})_{i \in \mathbb{F}_p^{n_H}}$ by $a_{\boldsymbol{i}} = ((i_1, \dots, i_{n_H}, 0, \dots, 0), \mathfrak{B}_H(\boldsymbol{i}))$ where there are $m - n_H$ 0's. We claim that $\boldsymbol{a} \in \Phi_{I'}(\boldsymbol{L}^{n_H})$. We can check this separately for the first and second coordinate; each is trivial.

Thus by Theorem 4.3.19 and the rank bound on $\mathfrak{B}'$, we have

$$\begin{aligned}\mathbb{P}_{x_1,\dots,x_{n_H} \in U}\left(\mathfrak{B}'(L_{\boldsymbol{i}}^{n_H}(x_1, \dots, x_{n_H})) = a_{\boldsymbol{i}} \text{ for all } \boldsymbol{i} \in \mathbb{F}_p^{n_H}\right) &\geq \frac{1}{|\Phi_{I'}(\boldsymbol{L}^{n_H})|} - \frac{1}{2(p^m \|I\|)^{p^m}} \\ &\geq \frac{1}{2(p^m \|I\|)^{p^m}}.\end{aligned}$$

We wish to find a single tuple $(x_1, \dots, x_{n_H}) \in V^{n_H}$ that satisfies the above condition and also has $x_1, \dots, x_{n_H}$ linearly independent. The number of linearly dependent tuples is small, so we calculate that the number of good tuples is at least

$$\frac{|U|^{n_H}}{2(p^m \|I\|)^{p^m}} - |U|^{n_H - 1} p^{n_H}.$$

This is positive by our assumption that $\dim U \geq n_0$. Thus there exists some good tuple $(x_1, \dots, x_{n_H}) \in V^{n_H}$. Defining $\kappa \colon \mathbb{F}_p^{n_H} \to U$ by $\kappa(\boldsymbol{i}) := L_{\boldsymbol{i}}^{n_H}(x_1, \dots, x_{n_H})$ has all the desired properties. Thus we have shown (b) assuming that (a) does not hold. □

*Proof of Lemma 4.5.7.* Define $M := m|\mathcal{S}|^{p\|I\|}$. Our strategy is to find a large subspace $U_M$ and linear functions $P_1, \dots, P_M \colon V \to \mathbb{F}_p$ such that for $x \in U_M$, the value of $f(x)$ only depends on $\mathfrak{B}(x)$, $\operatorname{fnz}(P_1(x), \dots, P_M(x))$, and the index $k$ such that $P_1(x) = \cdots = P_{k-1}(x) = 0$ and $P_k(x) \neq 0$. Once we have found such a configuration, we can complete the proof by a simple pigeonhole argument.

Define

$$\mathcal{L} := \{p^s P^i_{(p-1)s+1,s} : s \geq 0, i \in [I_{(p-1)s+1,s}]\},$$

where $P^i_{d,k}$ is the $i$th homogeneous non-classical polynomial of degree $d$ and depth $k$ defining $\mathfrak{B}$. Note that $\mathcal{L}$ is a finite set of linear functions (in particular, it is the set of all $p^s P$

that are classical linear polynomials where $s \geq 0$ is a non-negative integer and $P$ is one of the homogeneous non-classical polynomials that define $\mathfrak{B}$.) It is immediate from the definition of rank that if $P_1, \ldots, P_m$ are linear functions such that $\{P_1, \ldots, P_m\} \cup \mathcal{L}$ are linearly independent, then $\operatorname{rank} \mathfrak{B}' = \operatorname{rank} \mathfrak{B}$. We will use this fact to guarantee conclusion (iii).

Our main tool is the following Ramsey-type lemma which is somewhat analogous to the affine Ramsey theorem of Graham and Rothschild.

**Lemma 4.5.8.** *Fix a prime $p$, a finite set $\mathcal{S}$, a parameter list $I \in \mathcal{I}_p$, and positive integers $n_0, r_0$. There exist constants $n_{\mathsf{ramsey}'} = n_{\mathsf{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0)$, $r_{\mathsf{ramsey}'} = r_{\mathsf{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0)$ such that the following holds. Let $V$ be a finite dimensional $\mathbb{F}_p$-vector space satisfying $\dim V \geq n_{\mathsf{ramsey}'}$ and let $P \colon V \to \mathbb{F}_p$ be a non-trivial linear function. Let $\mathfrak{B}$ be a polynomial factor on $V$ with parameters $I$ and let $\mathfrak{B}'$ be the common refinement of $\mathfrak{B}$ and $\{P\}$. Suppose that $\operatorname{rank} \mathfrak{B}' \geq r_{\mathsf{ramsey}'}$. Let $f \colon V \to \mathcal{S}$ be a function. Then there exists a subspace $U \leq V$ contained in the zero set of $P$, a vector $z \in V$ such that $P(z) = 1$, and a function $\chi \colon A_I \to \mathcal{S}$ such that the following holds:*

*(i)* $\chi(\mathfrak{B}(x)) = f(x)$ *for all* $x \in z + U$;

*(ii)* $\operatorname{rank} \mathfrak{B}|_U \geq r_0$;

*(iii)* $\dim U \geq n_0$.

Define $r_1, \ldots, r_M$ and $n_1, \ldots, n_M$ by

$$n_i := \max\{n_{\mathsf{ramsey}'}(p, |\mathcal{S}|^{p-1}, I, n_{i-1}, r_{i-1}), |\mathcal{L}| + 1\}$$

and

$$r_i := r_{\mathsf{ramsey}'}(p, |\mathcal{S}|^{p-1}, I, n_{i-1}, r_{i-1}).$$

Then define

$$n_{\mathsf{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0) := n_M \qquad \text{and} \qquad r_{\mathsf{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0) := r_M.$$

We will find nested subspaces $V = U_0 \geq U_1 \geq \cdots \geq U_M$, linear functions $P_i \colon V \to \mathbb{F}_p$, and functions $\xi_i \colon (\mathbb{F}_p \setminus \{0\}) \times A_I \to \mathcal{S}$ such that the following holds for each $i \in [M]$:

- $\xi_i(P_i(x), \mathfrak{B}(x)) = f(x)$ for all $x \in U_i$ that satisfy $P_1(x) = \cdots = P_{i-1}(x) = 0$ and $P_i(x) \neq 0$;
- $\{P_1, \ldots, P_i\} \cup \mathcal{L}$ are linearly independent;
- $\operatorname{rank} \mathfrak{B}|_{W_i} \geq r_{M-i}$ where $W_i := \{x \in U_i : P_1(x) = \cdots = P_i(x)\}$;
- $\dim W_i \geq n_{M-i}$ where $W_i := \{x \in U_i : P_1(x) = \cdots = P_i(x)\}$.

Suppose we have defined $V = U_0 \geq U_1 \geq \cdots \geq U_i$, linear functions $P_1, \ldots, P_i \colon V \to \mathbb{F}_p$, and functions $\xi_1, \ldots, \xi_i \colon (\mathbb{F}_p \setminus \{0\}) \times A_I \to \mathcal{S}$ with the above properties.

Define $W := \{x \in U_i : P_1(x) = \cdots = P_i(x) = 0\}$. We have $\dim W \geq m_i > |\mathcal{L}|$. Pick an arbitrary $y \in W$ such that $y \neq 0$ but all of the linear functions in $\mathcal{L}$ vanish on $y$. Let $P_{i+1} \colon V \to \mathbb{F}_p$ be an arbitrary linear function such that $P_{i+1}(y) = 1$. Note that automatically we have $\{P_1, \ldots, P_i, P_{i+1}\} \cup \mathcal{L}$ are linearly independent.

Define $W' := \{x \in W : P_{i+1}(x) = 0\}$. Note that the subspace $W$ is partitioned into hyperplanes as $W = W' \sqcup (y + W') \sqcup (2y + W') \sqcup \cdots$. Write $\overline{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$. Then define $\overline{f} \colon W \to \overline{\mathcal{S}}$ by

$$\overline{f}(x + ty) := (f(bx + by))_{b \in \mathbb{F}_p^\times} \qquad \text{for } x \in W' \text{ and } t \in \mathbb{F}_p.$$

We apply Lemma 4.5.8 to $\overline{f}, \mathfrak{B}|_W, P_{i+1}$ with parameters $n_{M-i-1}, r_{M-i-1}$ to produce a subspace $U_i' \leq W'$, a vector $z \in y + U_i'$ and a function $\chi \colon A_I \to \overline{S}$ with several desirable properties.

We have $\chi(\mathfrak{B}(x + z)) = \overline{f}(x + z)$ for all $x \in U_i'$. Looking at the $b$th coordinate of this equation for some $b \in \mathbb{F}_p^\times$ gives $\chi(\mathfrak{B}(x + z))_b = f(bx + bz)$. Finally using the homogeneity of $\mathfrak{B}$ (recall the action of $\mathbb{F}_p^\times$ on $A_I$ defined in Eq. (4.5)) gives $\chi(b^{-1} \cdot \mathfrak{B}(bx + bz))_b = f(bx + bz)$ for all $x \in U_i'$ and all $b \in \mathbb{F}_p^\times$.

Define $U_{i+1} \leq U_i$ to be a $(\dim W + i + 1)$-dimensional subspace of $U_i$ that contains $z$ and $W$ and such that none of $P_1, \ldots, P_i$ are identically 0 on $U_{i+1}$. Then define $\xi_{i+1} \colon (\mathbb{F}_p \setminus \{0\}) \times A_I \to \mathcal{S}$ by

$$\xi_{i+1}(b, a) := \chi(b^{-1} \cdot a)_b.$$

Note $\dim U_{i+1} \geq \dim W' \geq m_{i+1}$ and $\operatorname{rank} \mathfrak{B}|_{U_{i+1}} \geq \operatorname{rank} \mathfrak{B}|_{W'} \geq r_{i+1}$. Furthermore, for $x \in U_{i+1}$ such that $P_1(x) = \cdots = P_{i-1}(x) = 0$ and $P_i(x) = b \neq 0$, we can write $x = bx' + bz$ for some $x' \in W'$. Then $f(x) = \overline{f}(x' + z)_b = \chi(\mathfrak{B}(x' + z))_b = \chi(b^{-1} \cdot \mathfrak{B}(bx' + bz))_b = \xi_{i+1}(b, \mathfrak{B}(x))$, as desired.

Thus we have defined a sequence of nested subspaces $V = U_0 \leq \cdots \leq U_M$, linear functions $P_1, \ldots, P_m \colon V \to \mathbb{F}_p$ and functions $\xi_1, \ldots, \xi_M \colon (\mathbb{F}_p \setminus \{0\}) \times A_I \to \mathcal{S}$ with the above properties.

Finally note that the number of possible functions $(\mathbb{F}_p \setminus \{0\}) \times A_I \to \mathcal{S}$ is at most $|\mathcal{S}|^{p\|I\|}$. Thus by the pigeonhole principle, there exists $1 \leq i_1 < \cdots < i_m \leq M$ such that $\xi_1 = \cdots = \xi_m$. Define $\xi \colon \mathbb{F}_p \times A_I \to \mathcal{S}$ by $\xi(0, a) = f(0)$ for all $a \in A$ and $\xi(b, a) = \xi_1(b, a)$ for all $b \in (\mathbb{F}_p \setminus \{0\})$ and all $a \in A$. Define $W_M := \{x \in U_M : P_1(x) = \cdots = P_M(x) = 0\}$ and let $U$ be a $(\dim W_M + m)$-dimensional subspace of $V$ that contains $W_M$ and such that none of $P_{i_1}, \ldots, P_{i_m}$ are identically 0 on $U$. Then $U$, $P_{i_1}, \ldots, P_{i_m}$, and $\xi$ have all the desired properties. □

We now prove our Ramsey-type lemma. We are given a coloring $f \colon V \to \mathcal{S}$ and a polynomial factor $\mathfrak{B}'$. Our goal is to find an affine subspace $z + U$ such that the coloring $f|_{z+U}$ only depends on $\mathfrak{B}$ and additionally $\mathfrak{B}|_U$ has rank at least $r_0$.

We satisfy the second condition in a quite crude way. We start by fixing some polynomial factor $\mathfrak{B}_1$ on $\mathbb{F}_p^{n_1}$ with rank at least $r_0$. Then we require $\mathfrak{B}|_U$ to be isomorphic to $\mathfrak{B}_1$ which ensures that it has rank at least $r_0$.

The proof works by applying the arithmetic regularity lemma, Lemma 4.4.4, to produce a polynomial factor $\mathfrak{B}''$ which regularizes the coloring. The arithmetic regularity lemma is a powerful tool that when combined with the equidistribution (Theorem 4.3.19), allows us to not just find one, but actually count the number of affine subspaces $U$ where each point of $U$ lies in a specified atom of $\mathfrak{B}''$ and is colored a specified color by $f$. While there is a fair amount of technical work that goes into applying arithmetic regularity, the high-level overview is that these counts of $U$ are "close to expected" which allows us to find at least one $U$ with all the desired properties.

*Proof of Lemma 4.5.8.* Define constants

$$n_1 := \max\{n_0, n_{\mathsf{high-rank}}(p, I, r_0 + p)\},$$

and

$$\theta := \frac{1}{8(2|\mathcal{S}|)^{2p^{n_1}}} \qquad \text{and} \qquad \theta' := \frac{\theta}{\|I\|\sqrt{2p|\mathcal{S}|}}$$

Define the non-increasing functions $\alpha\colon \mathbb{Z}_{>0} \to (0,1)$ by

$$\alpha(N) := \frac{1}{2N^{2p^{n_1}}},$$

and $\eta\colon \mathbb{Z}_{>0} \to (0,1)$ by

$$\eta(N) := \frac{1}{8(3|\mathcal{S}|N)^{p^{n_1}}},$$

and define the non-decreasing function $r\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ by

$$r(N) := r_{\mathsf{equi}}(p, p^{n_1}, \alpha(N)).$$

Define

$$n_{\mathsf{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0) := 2p^{n_1} \left\lceil \log_p(16|\mathcal{S}|C_{\mathsf{reg}''}(p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r)) \right\rceil$$

and

$$r_{\mathsf{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0) := r_{\mathsf{reg}''}(p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r)$$

For $c \in \mathcal{S}$, define $f^{(c)}\colon V \to [0,1]$ by $f^{(c)} := 1_{f^{-1}(c)}$. We are now ready to proceed with the proof. We apply the arithmetic regularity lemma, Lemma 4.4.4, to the polynomial factor $\mathfrak{B}'$ on $V$ and the functions $(f^{(c)})_{c\in\mathcal{S}}$ with parameters $p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r$. This produces a polynomial factor $\mathfrak{B}''$ refining $\mathfrak{B}'$ and decompositions $f^{(c)} = f^{(c)}_{\mathsf{str}} + f^{(c)}_{\mathsf{psr}} + f^{(c)}_{\mathsf{sml}}$ with several desirable properties.

Say that $\mathfrak{B}'$ has parameters $I'$ and $\mathfrak{B}''$ has parameters $I''$ (note that $\|I'\| = p\|I\|$). We consider the atom-indexing set of $\mathfrak{B}''$ (see Eq. (4.3) for the definition) as $A_{I''} \simeq \mathbb{F}_p \times A_I \times A_{I''-I'}$.

Say that an atom $a \in A_{I''}$ is *regular* if

$$\|f_{\mathsf{sml}}^{(c)} 1_{\mathfrak{B}''^{-1}(a)}\|_2 \leq \theta \|1_{\mathfrak{B}''^{-1}(a)}\|_2 \qquad \text{for all } c \in \mathcal{S}.$$

Our first goal is to find $s \in A_{I''-I'}$ such that all atoms of the form $(1, a, s) \in A_{I''} \simeq \mathbb{F}_p \times A_I \times A_{I''-I'}$ are regular.

By Lemma 4.4.4(v), for each $c \in \mathcal{S}$,

$$\sum_{a \in A_{I''}} \|f_{\mathsf{sml}}^{(c)} 1_{\mathfrak{B}''^{-1}(a)}\|_2^2 \leq \frac{\theta^2}{2p|\mathcal{S}|\|I\|^2} \sum_{a \in A_{I''}} \|1_{\mathfrak{B}''^{-1}(a)}\|_2^2.$$

Thus at least a $(1 - 1/(2p\|I\|^2))$-fraction of atoms are regular. For each $a \in A_I$, at least a $(1 - 1/(2\|I\|))$-fraction of the atoms of the form $(1, a, s)$ are regular, for $s \in A_{I''-I'}$. Thus by a union bound there exists some $s \in A_{I''-I'}$ such that $(1, a, s)$ is regular for all $a \in A_I$. Fix this value of $s$ for the rest of the proof.

Define $\chi \colon A_I \to \mathcal{S}$ such that $\chi(a)$ is a color that appears in the atom $\mathfrak{B}^{-1}(1, a, s)$ with density at least $1/|\mathcal{S}|$.

By the definition of $n_1$ and Lemma 4.3.15, there exists a polynomial factor $\mathfrak{B}_1$ on $\mathbb{F}_p^{n_1}$ with parameters $I$ and satisfying $\operatorname{rank} \mathfrak{B}_1 \geq r_0 + p$. Our goal is to find vectors $x_0, x_1, \ldots, x_{n_1} \in V$ such that

- $\mathfrak{B}''(x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}) = (1, \mathfrak{B}_1(i_1, \ldots, i_{n_1}), s)$ for all $(i_1, \ldots, i_{n_1}) \in \mathbb{F}_p^{n_1}$;
- $f(x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}) = \chi(\mathfrak{B}_1(i_1, \ldots, i_{n_1}))$ for all $(i_1, \ldots, i_{n_1}) \in \mathbb{F}_p^{n_1}$;
- $x_1, \ldots, x_{n_1}$ are linearly independent.

We choose $x_0, x_1, \ldots, x_{n_1} \in V$ independently and uniformly at random. Let $p_1$ be the probability that this choice of $\boldsymbol{x}$ satisfies all three conditions above. First note that the probability that $x_0, \ldots, x_{n_1}$ are linearly dependent is at most $p^{n_1+1}/|V|$. Let $p_2$ be the probability that this choice of $\boldsymbol{x}$ satisfies the first two conditions above. We have shown that

$$p_1 \geq p_2 - p^{n_1+1}/|V|.$$

Let $\boldsymbol{L} = (L_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{F}_p^{n_1}}$ to be the system of $p^{n_1}$ linear forms in $n_1 + 1$ variables defined by

$$L_{\boldsymbol{i}}(x_0, x_1, \ldots, x_{n_1}) := x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}.$$

This system defines an $n_1$-dimensional affine subspace. One can easily see that $\boldsymbol{L}$ is finite complexity and in fact its complexity is at most $p^{n_1}$ (see Remark 4.3.4).

Define $g^{(\boldsymbol{i})} := f^{(\chi(\mathfrak{B}_1(\boldsymbol{i})))}$ and define $g^{(\boldsymbol{i})}_{\mathsf{str}}, g^{(\boldsymbol{i})}_{\mathsf{sml}}, g^{(\boldsymbol{i})}_{\mathsf{psr}}$ similarly. Define $h^{(\boldsymbol{i})} \colon V \to [0,1]$ by $h^{(\boldsymbol{i})} := 1_{\mathfrak{B}''^{-1}(1, \mathfrak{B}_1(\boldsymbol{i}), s)}$.

We compute $p_2$ as

$$\begin{aligned}
p_2 &= \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{\boldsymbol{i} \in \mathbb{F}_p^{n_1}} g^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right] \\
&= \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{\boldsymbol{i}} \left( g^{(\boldsymbol{i})}_{\mathsf{str}}(L_{\boldsymbol{i}}(\boldsymbol{x})) + g^{(\boldsymbol{i})}_{\mathsf{sml}}(L_{\boldsymbol{i}}(\boldsymbol{x})) + g^{(\boldsymbol{i})}_{\mathsf{psr}}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right) h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right] \\
&\geq \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{\boldsymbol{i}} \left( g^{(\boldsymbol{i})}_{\mathsf{str}}(L_{\boldsymbol{i}}(\boldsymbol{x})) + g^{(\boldsymbol{i})}_{\mathsf{sml}}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right) h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right] - 3^{p^{n_1}} \eta(\|\mathfrak{B}''\|).
\end{aligned}$$

The inequality follows from Lemma 4.4.4(ii) and the counting lemma, Eq. (4.1).

Write $p_3$ for the expectation in the last line above. We have $p_2 \geq p_3 - 3^{p^{n_1}} \eta(\|\mathfrak{B}''\|)$. Expanding the product, there are at most $2^{p^{n_1}}$ terms involving $g^{(\boldsymbol{j})}_{\mathsf{sml}}$ for some $\boldsymbol{j} \in \mathbb{F}_p^{n_1}$. Each of these is bounded in magnitude by

$$\mathbb{E}_{\boldsymbol{x}} \left[ \left| g^{(\boldsymbol{j})}_{\mathsf{sml}}(L_{\boldsymbol{j}}(\boldsymbol{x})) \right| \prod_{\boldsymbol{i}} h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right].$$

By applying a change of coordinates, we can transform to the case that $\boldsymbol{j} = 0$ (and $L_{\boldsymbol{0}}(\boldsymbol{x}) = x_0$). Then by the Cauchy-Schwarz inequality, the square of the above expression is bounded by

$$\mathbb{E}_{x_0} \left[ \left| g^{(0)}_{\mathsf{sml}}(x_0) \right|^2 h^{(0)}(x_0) \right] \mathbb{E}_{x_0} \left[ h^{(0)}(x_0) \mathbb{E}_{x_1, \ldots, x_{n_1}} \left[ \prod_{\boldsymbol{i} \neq 0} h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right]^2 \right].$$

The first term is at most $\theta(\|\mathfrak{B}\|)^2 \|h^{(0)}\|_2^2$ by the fact that $(1, \mathfrak{B}_1(\boldsymbol{i}), s)$ is a regular atom for all $\boldsymbol{i}$. The second term can be counted by equidistribution applied to the system $\boldsymbol{L}'$ of $2p^{n_1} - 1$

linear forms in $2n_1 - 1$ variables defined as follows. Set

$$L'_0(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0,$$

and for $\boldsymbol{i} \in \mathbb{F}_p^{n_1} \setminus \{0\}$, define

$$L'_{\boldsymbol{i},1}(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1},$$

$$L'_{\boldsymbol{i},2}(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0 + i_1 x'_1 + \cdots + i_{n_1} x'_{n_1}.$$

By Lemma 4.3.23, we know that $\|\mathfrak{B}''\| \cdot |\Phi_{I''}(\boldsymbol{L}')| = |\Phi_{I''}(\boldsymbol{L})|^2$ (see also [BFH+13, Lemma 5.13].)

Thus by equidistribution, Theorem 4.3.19, and the rank bound on $\mathfrak{B}''$, we have the second term is at most

$$\frac{1}{|\Phi_{I''}(\boldsymbol{L}')|} + \alpha(\|\mathfrak{B}''\|) = \frac{\|\mathfrak{B}''\|}{|\Phi_{I''}(\boldsymbol{L})|^2} + \alpha(\|\mathfrak{B}''\|) \leq \frac{2\|\mathfrak{B}''\|}{|\Phi_{I''}(\boldsymbol{L})|^2}.$$

Applying equidistribution again we have that the first term is at most

$$\theta^2 \left( \frac{1}{\|\mathfrak{B}''\|} + \alpha(\|\mathfrak{B}''\|) \right) \leq \frac{2\theta^2}{\|\mathfrak{B}''\|}.$$

Combining these bounds and summing over all terms that contain some $g_{\mathsf{sml}}^{(\boldsymbol{j})}$, we see that

$$p_3 \geq \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{\boldsymbol{i}} g_{\mathsf{str}}^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right] - 2^{p^{n_1}+1} \frac{\theta}{|\Phi_{I''}(\boldsymbol{L})|}.$$

Write $p_4$ for the expectation in the last line. The quantity $g_{\mathsf{str}}^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x}))$ is the density of $\chi(\mathfrak{B}_1(\boldsymbol{i}))$ in the atom of $\mathfrak{B}''$ that $L_{\boldsymbol{i}}(\boldsymbol{x})$ lies in. When $\mathfrak{B}(L_{\boldsymbol{i}}(\boldsymbol{x})) = (1, \mathfrak{B}_1(\boldsymbol{i}), s)$, the choice of $\chi$ implies that this density is at least $1/|\mathcal{S}|$. Thus

$$p_4 \geq \frac{1}{|\mathcal{S}|^{p^{n_1}}} \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{\boldsymbol{i}} h^{(\boldsymbol{i})}(L_{\boldsymbol{i}}(\boldsymbol{x})) \right].$$

Write $p_5$ for the expectation in the last line. Unwrapping the definition of $h^{(\boldsymbol{i})}$, this can be written as

$$p_5 = \mathbb{P}_{\boldsymbol{x}}\left(\mathfrak{B}''(L_{\boldsymbol{i}}(\boldsymbol{x})) = (1, \mathfrak{B}_1(\boldsymbol{i}), s) \text{ for all } \boldsymbol{i} \in \mathbb{F}_p^{n_1}\right).$$

We claim that $((1, \mathfrak{B}_1(\boldsymbol{i}), s))_{\boldsymbol{i}\in\mathbb{F}_p^{n_1}}$ is an $\boldsymbol{L}$-consistent tuple of atoms. We check this coordinate by coordinate. Obviously $(\mathfrak{B}_1(\boldsymbol{i}))_{\boldsymbol{i}\in\mathbb{F}_p^{n_1}}$ is $\boldsymbol{L}$-consistent. Furthermore any constant tuple is also obviously $\boldsymbol{L}$-consistent (this follows from the fact that $\boldsymbol{L}$ is a translation-invariant pattern). Thus by another application of equidistribution and the rank bound on $\mathfrak{B}''$, we have

$$p_5 \geq \frac{1}{|\Phi_{I''}(\boldsymbol{L}')|} - \alpha(\|\mathfrak{B}'\|) \geq \frac{1}{2|\Phi_{I''}(\boldsymbol{L}')|}.$$

Combining all these inequalities, we see that

$$p_1 \geq \frac{1}{|\mathcal{S}|^{p^{n_1}}}\frac{1}{2|\Phi_{I''}(\boldsymbol{L}')|} - 2^{p^{n_1}+1}\frac{\theta}{|\Phi_{I''}(\boldsymbol{L})|} - 3^{p^{n_1}}\eta(\|\mathfrak{B}''\|) - \frac{p^{n_1+1}}{|V|}.$$

This expression is positive by the definition of $\theta, \eta$ and the assumption that $\dim V \geq n_{\mathsf{ramsey}'}$.

Thus we have defined a function $\chi\colon A_I \to \mathcal{S}$ and found linearly independent $x_0, x_1, \ldots, x_{n_1} \in V$ with several desirable properties. Define $z := x_0$ and $U := \operatorname{span}\{x_1, \ldots, x_{n_1}\}$. Since $P\colon V \to \mathbb{F}_p$ is a linear function and $P(x_0 + i_1x_1 + \cdots + i_{n_1}x_{n_1}) = 1$ for all $\boldsymbol{i} \in \mathbb{F}_p^{n_1}$, we conclude that $P(z) = P(x_0) = 1$ and $U$ is contained in the zero set of $P$. Furthermore, $\mathfrak{B}(x_0 + i_1x_1 + \cdots + i_{n_1}x_{n_1}) = \mathfrak{B}_1(i_1, \ldots, i_{n_1})$ for all $\boldsymbol{i} \in \mathbb{F}_p^{n_1}$. Since $\mathfrak{B}_1$ was chosen such that $\operatorname{rank}\mathfrak{B}_1 \geq r_0 + p$ and $n_1 \geq n_0$, we see that $\dim U \geq n_0$ and $\operatorname{rank}\mathfrak{B}|_{\operatorname{span}\{x_0,U\}} \geq r_0 + p$. By Lemma 4.3.16, we conclude that $\operatorname{rank}\mathfrak{B}|_U \geq r_0$. Finally, $f(x_0 + i_1x_1 + \cdots + i_{n_1}x_{n_1}) = \chi(\mathfrak{B}_1(i_1, \ldots, i_{n_1})) = \chi(\mathfrak{B}(x_0 + i_1x_1 + \cdots + i_{n_1}x_{n_1}))$ for all $\boldsymbol{i} \in \mathbb{F}_p^{n_1}$, which proves the desired result. □

### 4.5.4 Main patching result

We boost the Ramsey dichotomy result to a supersaturation dichotomy result which is our main patching theorem. For our application, we will need this result to hold inside a "subvariety" of a vector space, i.e., the zero set of a sufficiently high-rank collection of non-classical polynomials. Also for technical reasons, this supersaturation argument only

works for full dimensional patterns (recall Definition 4.3.21).

Despite the complicated nature of this statement, the high-level overview of the proof is quite simple. If we are in case (b) of Theorem 4.5.6, we know that there exists an $\mathcal{H}$-instance not just in $V$, but in essentially every $n_{\mathsf{dich}}$-dimensional subspace of $V$. Adding up over all subspaces of this dimension gives us a lot of $\mathcal{H}$-instances.

There are two issues with this proof sketch which we have to fix. First, it is not true that for every subspace $U$, the polynomial factor $\mathfrak{B}|_U$ has high rank. Fortunately, we will be able to show that a positive proportion of these subspaces have sufficiently-high rank. Second, we have overcounted some of the $\mathcal{H}$-instances since each instances lives in many $n_{\mathsf{dich}}$-dimensional subspace. To deal with this issue, we will have to show that the $\mathcal{H}$-instances are "distributed evenly" in some sense. The most elegant way which we are aware of carrying out this argument is through a Cauchy-Schwarz trick which is Eq. (4.7) below. It is this last argument which we are only able to carry out for full dimensional patterns.

**Theorem 4.5.9** (Patching)**.** *Fix a prime $p$, a finite set $\mathcal{S}$ with an $\mathbb{F}_p^\times$-action, parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \le I'$, and a positive integer $\ell_0$. There exist constants $n_{\mathsf{patch}} = n_{\mathsf{patch}}(p, |\mathcal{S}|, I', \ell_0)$ and $\beta_{\mathsf{patch}} = \beta_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0) > 0$ and a non-decreasing function $r_{\mathsf{patch}} = r_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0) \colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that the following holds. Let $\mathcal{H}$ be a finite set of $\mathcal{S}$-colored, $I$-labeled patterns such that each pattern is defined by a full dimension system of linear forms in at most $\ell_0$ variables (recall Definition 4.3.21). Either:*

*(a) there exists a projective $\xi \colon \mathbb{F}_p \times A_I \to \mathcal{S}$ that does not canonically induce $\mathcal{H}$; or*

*(b) for every finite-dimensional $\mathbb{F}_p$-vector space $V$ satisfying $\dim V \ge n_{\mathsf{patch}}$, every projective function $f \colon V \to \mathcal{S}$, every polynomial factor $\mathfrak{B}$ on $V$ with parameters $I$ that satisfies $\operatorname{rank} \mathfrak{B} \ge r_{\mathsf{patch}}(\deg \mathfrak{B}, \|\mathfrak{B}\|)$, and every polynomial factor $\mathfrak{B}'$ on $V$ with parameters $I'$ that refines $\mathfrak{B}$ and satisfies $\operatorname{rank} \mathfrak{B}' \ge r_{\mathsf{patch}}(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$, there is a pattern $H \in \mathcal{H}$ such that in $(f, \mathfrak{B})$, the relative density of $H$ in $\mathfrak{B}'^{-1}(A_I \times \{0\})$ is at least $\beta_{\mathsf{patch}}$.*

Note that $n_{\mathsf{patch}}$ may depend on $I'$, but critically $\beta_{\mathsf{patch}}$ does not depend on $I'$.

*Proof.* First we define several parameters.

Write $r := r_{\mathsf{dich}}(p, |\mathcal{S}|, I, \ell_0)$ for brevity. Define

$$n_1 := \max\{n_{\mathsf{dich}}(p, |\mathcal{S}|, I, \ell_0), n_{\mathsf{high-rank}}(p, I, r)\}$$

where $n_0$ is defined in Lemma 4.3.15.

Define the constants

$$n_{\mathsf{patch}}(p, |\mathcal{S}|, I', \ell_0) := 2p^{n_1} \left\lceil \log_p(\|I'\|) \right\rceil,$$

and

$$\beta_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0) := \frac{1}{6400\ell_0^2 p^{2n_1 \cdot \ell} \|I\|^{2p^{n_1}}},$$

and define the non-increasing function $\alpha\colon \mathbb{Z}_{>0} \to (0, 1)$ by

$$\alpha(N) := \frac{1}{2N^{2p^{n_1}}},$$

and the non-decreasing function $r_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0)\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ by

$$r_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0)(D, N) := r_{\mathsf{equi}}(p, D, \alpha(N)).$$

We now proceed to the proof. Let $\mathcal{H}$ be a finite set of $\mathcal{S}$-colored, $I$-labeled patterns each defined by a full dimension system of linear forms in at most $\ell_0$ variables. We apply Theorem 4.5.6 to $\mathcal{H}$. If Theorem 4.5.6(a) holds, then clearly conclusion (a) holds. Now assume that Theorem 4.5.6(b) holds. We wish to show conclusion (b).

Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space satisfying $\dim V \geq n_{\mathsf{patch}}$, let $f\colon V \to \mathcal{S}$ be a projective function, let $\mathfrak{B}$ be a polynomial factor on $V$ with parameters $I$ that satisfies $\operatorname{rank} \mathfrak{B} \geq r_{\mathsf{patch}}(\|\mathfrak{B}\|)$, and let $\mathfrak{B}'$ be a polynomial factor on $V$ with parameters $I'$ that refines $\mathfrak{B}$ and satisfies $\operatorname{rank} \mathfrak{B}' \geq r_{\mathsf{patch}}(\|\mathfrak{B}'\|)$.

Write $X := \mathfrak{B}'^{-1}(A_I \times \{0\})$ and $n := \dim V$. By assumption, $n \geq n_{\mathsf{patch}}$. We wish to count $\mathcal{H}$ instances in $(f, \mathfrak{B})$ that are contained in $X$. Write $\mathcal{H} = \bigsqcup_{\ell=1}^{\ell_0} \mathcal{H}_\ell$ where $\mathcal{H}_\ell$ is defined to be the subset of $\mathcal{H}$ consisting of colored labeled patterns defined by a system of linear forms in exactly $\ell$ variables. We define sets $\mathfrak{U}_1, \ldots, \mathfrak{U}_{\ell_0}$ as follows. $\mathfrak{U}_\ell$ is the set of $\ell$-dimensional

subspaces $U$ of $V$ which satisfy the following:

- $U \subseteq X$;
- there exists a colored labeled pattern $H \in \mathcal{H}_\ell$ in $\ell$ variables and a generic $H$-instance $x_1, \ldots, x_\ell \in U$.

Note that the requirement that $x_1, \ldots, x_\ell \in U$ are generic implies that $U = \operatorname{span}\{x_1, \ldots, x_\ell\}$. Thus $\sum_{\ell=1}^{\ell_0} |\mathfrak{U}_\ell|$ is a lower bound on the number of $\mathcal{H}$-instances in $X$.

Define the following counting function $c\colon \mathfrak{U}_\ell \to \mathbb{Z}_{\geqslant 0}$ for each $\ell \in [\ell_0]$ as follows. For $U \in \mathfrak{U}_\ell$, let $c(U)$ be the number of $n_1$-dimensional subspaces that contain $U$ and are contained in $X$. An application of the Cauchy-Schwarz inequality implies that

$$|\mathfrak{U}_\ell| \geq \frac{\left(\sum_{U \in \mathfrak{U}_\ell} c(U)\right)^2}{\sum_{U \in \mathfrak{U}_\ell} c(U)^2}. \tag{4.7}$$

Define $S_1$ to be the number of $n_1$-dimensional subspaces $W$ of $V$ such that $W \subseteq X$ and $\operatorname{rank}(\mathfrak{B}|_W) \geq r$. By Theorem 4.5.6(b), every such $W$ contains a generic $\mathcal{H}$-instance. Thus

$$\sum_{\ell=1}^{\ell_0} \sum_{U \in \mathfrak{U}_\ell} c(U) \geq S_1.$$

By the pigeonhole principle, there exists some $\ell \in [\ell_0]$ such that

$$\sum_{U \in \mathfrak{U}_\ell} c(U) \geq \frac{S_1}{\ell_0}. \tag{4.8}$$

We fix such a value of $\ell \in [\ell_0]$ for the rest of the proof.

Define $S_2$ to be the number of $n_1$-tuples $(x_1, \ldots, x_{n_1}) \in V^{n_1}$ such that $x_1, \ldots, x_{n_1}$ are linearly independent, $\operatorname{span}\{x_1, \ldots, x_{n_1}\} \subseteq X$, and $\operatorname{rank}(\mathfrak{B}|_{\operatorname{span}\{x_1, \ldots, x_{n_1}\}}) \geq r$. We can compute

$$S_2 = S_1 \prod_{i=0}^{n_1-1} (p^{n_1} - p^i) \leq p^{n_1^2} S_1. \tag{4.9}$$

Define $S_3$ to be the number of $n_1$-tuples $(x_1, \ldots, x_{n_1}) \in V^{n_1}$ such that $\operatorname{span}\{x_1, \ldots, x_{n_1}\} \subseteq$

$X$ and $\operatorname{rank}(\mathfrak{B}|_{\operatorname{span}\{x_1,\ldots,x_{n_1}\}}) \geq r$. We can easily bound

$$S_2 \geq S_3 - p^{n\cdot n_1} p^{n_1-n}. \tag{4.10}$$

By the definition of $n_1$ and Lemma 4.3.15, there exists a polynomial factor $\mathfrak{B}_1$ on $\mathbb{F}_p^{n_1}$ with parameters $I$ and rank at least $r$. Define $S_4$ to be the number of ordered $n_1$-tuples $(x_1,\ldots,x_{n_1}) \in V^{n_1}$ such that $\mathfrak{B}_1(i_1,\ldots,i_{n_1}) = \mathfrak{B}(i_1x_1+\cdots+i_{n_1}x_{n_1})$ and $\operatorname{span}\{x_1,\ldots,x_n\} \subseteq X$. Notice that $S_3 \geq S_4$.

Write $\boldsymbol{L}' := \boldsymbol{L}^{n_1}$, the system of $p^{n_1}$ linear forms in $n_1$ variables that define an $n_1$-dimensional subspace (see Definition 4.2.7). By definition, $(P^i_{d,k}(i_1,\ldots,i_{n_1}))_{\boldsymbol{i}\in\mathbb{F}_p^{n_1}} \in \Phi_{d,k}(\boldsymbol{L}')$ for every $(d,k) \in D_p$ and $i \in [I_{d,k}]$ where $P^i_{d,k}$ is the $i$th non-classical polynomial of degree $d$ and depth $k$ defining $\mathfrak{B}_1$. Also, $(0,\ldots,0) \in \Phi_{d,k}(\boldsymbol{L}')$ for all $(d,k) \in D_p$. Define $\boldsymbol{a} \in A_{I'}^{\mathbb{F}_p^{n_1}}$ by

$$(a_{\boldsymbol{i}})^i_{d,k} = \begin{cases} P^i_{d,k}(i_1,\ldots,i_{n_1}) & \text{if } i \leq I_{d,k}, \\ 0 & \text{if } I_{d,k} < i \leq I'_{d,k}. \end{cases}$$

By the above discussion, $\boldsymbol{a}$ is $\boldsymbol{L}'$-consistent, so by Theorem 4.3.19 and the rank assumption of $\mathfrak{B}'$, we find

$$\begin{aligned} S_3 \geq S_4 &= \left|\{\boldsymbol{x} \in V^{n_1} : \mathfrak{B}'(L'_{\boldsymbol{i}}(\boldsymbol{x})) = a_{\boldsymbol{i}} \text{ for all } \boldsymbol{i} \in \mathbb{F}_p^{n_1}\}\right| \\ &\geq \left(\frac{1}{|\Phi_{I'}(\boldsymbol{L}')|} - \alpha(\|I'\|)\right) p^{n\cdot n_1} \geq \frac{p^{n\cdot n_1}}{2|\Phi_{I'}(\boldsymbol{L}')|}. \end{aligned} \tag{4.11}$$

Combining Eq. (4.8), Eq. (4.9), Eq. (4.10), and Eq. (4.11), we conclude

$$\sum_{U\in\mathfrak{U}_\ell} c(U) \geq \frac{1}{\ell_0 p^{n_1^2}} p^{n\cdot n_1} \left(\frac{1}{2|\Phi_{I'}(\boldsymbol{L}')|} - p^{n_1-n}\right) \geq \frac{1}{4\ell_0 p^{n_1^2}|\Phi_{I'}(\boldsymbol{L}')|} p^{n\cdot n_1}. \tag{4.12}$$

Next we find an upper bound on $\sum_{U\in\mathfrak{U}_\ell} c(U)^2$. Define $T_1$ to be the number of triples $(U,W,W')$ where $U$ is a $\ell$-dimensional subspace of $V$ and $W,W'$ are both $n_1$-dimensional

subspaces of $V$ that contain $U$ and are contained in $X$. First note that

$$\sum_{U \in \mathfrak{U}_\ell} c(U)^2 \leq T_1. \tag{4.13}$$

Now define $T_2$ to be the number of $(2n_1-\ell)$-tuples $(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) \in V^{2n_1-\ell}$ such that $x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}$ are linearly independent, $x_1, \dots, x_\ell, z_1, \dots, z_{n_1-\ell}$ are linearly independent, $\operatorname{span}\{x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}\} \subseteq X$, and $\operatorname{span}\{x_1, \dots, x_\ell, z_1, \dots, z_{n_1-\ell}\} \subseteq X$. We compute

$$T_2 = \left(\prod_{i=0}^{\ell-1}(p^\ell - p^i)\right)\left(\prod_{i=\ell}^{n_1-1}(p^{n_1} - p^i)\right)^2 T_1 \geq \frac{p^{2n_1(n_1-\ell)}}{100} T_1. \tag{4.14}$$

Next define $T_3$ to be the number of $(2n_1-\ell)$-tuples $(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) \in V^{2n_1-\ell}$ such that $\operatorname{span}\{x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}\} \subseteq X$ and $\operatorname{span}\{x_1, \dots, x_\ell, z_1, \dots, z_{n_1-\ell}\} \subseteq X$. Clearly $T_2 \leq T_3$.

Define $\boldsymbol{L}''$ to be the following system of $2p^{n_1} - p^\ell$ linear forms in $2n_1 - n_\ell$ variables. For $\boldsymbol{i} \in \mathbb{F}_p^\ell$, define

$$L''_{\boldsymbol{i}}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) := i_1 x_1 + \cdots + i_\ell x_\ell.$$

For $\boldsymbol{i} \in \mathbb{F}_p^{n_1} \setminus \left(\mathbb{F}_p^\ell \times \{0\}^{n_1-\ell}\right)$, define

$$L''_{\boldsymbol{i},1}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) := i_1 x_1 + \cdots + i_\ell x_\ell + i_{\ell+1} y_1 + \cdots + i_{n_1} y_{n_1-\ell},$$
$$L''_{\boldsymbol{i},2}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) := i_1 x_1 + \cdots + i_\ell x_k + i_{\ell+1} z_1 + \cdots + i_{n_1} z_{n_1-\ell}.$$

Now let $\mathfrak{B}''$ be the polynomial factor on $V$ with parameters $I' - I$ defined by homogeneous non-classical polynomials

$$\left(P^i_{d,k}\right)_{\substack{(d,k)\in D_p \\ I_{d,k} < i \leq I'_{d,k}}}$$

where $P^i_{d,k}$ is the $i$th non-classical polynomial of degree $d$ and depth $k$ that defines $\mathfrak{B}'$. The important property of this polynomial factor is that $X = \mathfrak{B}''^{-1}(\mathbf{0})$. Also note that

$\operatorname{rank}\mathfrak{B}'' \geq \operatorname{rank}\mathfrak{B}' \geq r(\|\mathfrak{B}'\|)$. By Theorem 4.3.19 and the bounds on $\operatorname{rank}\mathfrak{B}''$, we find

$$\begin{aligned} T_2 \leq T_3 &= \left|\{\boldsymbol{x} \in V^{2n_1-\ell} : \mathfrak{B}''(\boldsymbol{L}''(\boldsymbol{x})) = \mathbf{0}\}\right| \\ &\leq \left(\frac{1}{|\Phi_{I'-I}(\boldsymbol{L}'')|} + \alpha(\|I'\|)\right) p^{n(2n_1-\ell)} \leq \frac{2p^{n(2n_1-\ell)}}{|\Phi_{I'-I}(\boldsymbol{L}'')|}. \end{aligned} \tag{4.15}$$

Combining Eq. (4.13), Eq. (4.14), and Eq. (4.15), we conclude

$$\sum_{U \in \mathfrak{U}_\ell} c(U)^2 \leq \frac{200p^{2n_1\cdot\ell}}{p^{2n_1^2}|\Phi_{I'-I}(\boldsymbol{L}'')|} p^{n(2n_1-\ell)}. \tag{4.16}$$

Finally, we combine Eq. (4.7), Eq. (4.12), and Eq. (4.16) to find

$$|\mathfrak{U}_\ell| \geq \frac{|\Phi_{I'-I}(\boldsymbol{L}'')|}{3200\ell_0^2 p^{2n_1\cdot\ell}|\Phi_{I'}(\boldsymbol{L}')|^2} p^{n\cdot\ell}.$$

Consider $\boldsymbol{L}^\ell$, the system of $p^\ell$ linear forms in $\ell$ variables that define an $\ell$-dimensional subspace (see Definition 4.2.7). By Lemma 4.3.23, we have $|\Phi_{I'-I}(\boldsymbol{L}^\ell)| \cdot |\Phi_{I'-I}(\boldsymbol{L}'')| = |\Phi_{I'-I}(\boldsymbol{L}')|^2$. Thus the above expression simplifies to

$$|\mathfrak{U}_\ell| \geq \frac{1}{3200\ell_0^2 p^{2n_1\cdot\ell}|\Phi_I(\boldsymbol{L}')|^2} \cdot \frac{p^{n\cdot\ell}}{|\Phi_{I'-I}(\boldsymbol{L}^\ell)|} \geq \frac{1}{3200\ell_0^2 p^{2n_1\cdot\ell}\|I\|^{2p^{n_1}}} \cdot \frac{p^{n\cdot\ell}}{|\Phi_{I'-I}(\boldsymbol{L}^\ell)|}.$$

Therefore there exists some colored labeled pattern $H = (\boldsymbol{L}, \psi, \phi) \in \mathcal{H}_\ell$ where $\boldsymbol{L}$ is a full dimension system of linear forms in $\ell$ variables and such that the number of generic $H$-instances in $(f, \mathfrak{B})$ which are contained in $X$ is at least $1/|\mathcal{H}_\ell|$ times the right-hand side of the above equation. Note that by equidistribution, Theorem 4.3.19, and the rank bound on $\mathfrak{B}''$,

$$\Lambda_{\boldsymbol{L}}(1_X, \ldots, 1_X) \leq \frac{1}{|\Phi_{I'-I}(\boldsymbol{L})|} + \alpha(\|I'\|) \leq \frac{2}{|\Phi_{I'-I}(\boldsymbol{L})|}. \tag{4.17}$$

Noting that since $\boldsymbol{L}$ is full dimensional, we have $|\Phi_{I'-I}(\boldsymbol{L})| = |\Phi_{I'-I}(\boldsymbol{L}^\ell)|$. Thus dividing the two above quantities, we find that the relative density of the above $H$ in $X$ is at least

$$\frac{1}{6400\ell_0^2 p^{2n_1\cdot\ell}\|I\|^{2p^{n_1}}} = \beta_{\mathsf{patch}}(p, |\mathcal{S}|, I, \ell_0). \qquad \square$$

## 4.6 Proof of removal lemmas

In this section we prove the main technical result, Theorem 4.2.8, the projective removal lemma. Then we give a short deduction of the full removal lemma, Theorem 4.2.4, from it.

**Notation:** As usual, for an atom $a \in A_I$ (defined in Eq. (4.3)), we use $a_{d,k} \in \mathbb{U}_{k+1}^{I_{d,k}}$ to denote the degree $d$, depth $k$ part of $a$. We use the notation $\tilde{A}_I \subset A_I$ to denote the set $\tilde{A}_I := \{a \in A_I : a_{1,0} = 0\}$. This is the set of atoms that are not regularized by Theorem 4.4.6. Also define $\tilde{I} \in \mathcal{I}_p$ by $\tilde{I}_{1,0} = 0$ and $\tilde{I}_{d,k} = I_{d,k}$ otherwise.

**Overview of proof:** There are four main steps. First, we apply a "compactness argument" inspired by a similar argument of Alon and Shapira. We will talk about this more later, but for now let us say that this essentially allows us to assume that $\mathcal{H}$ is a finite set of colored patterns.

The rest of the proof involves "cleaning up" the coloring $f\colon V \to \mathcal{S}$ so that every pattern $H \in \mathcal{H}$ appears in $f$ either zero or many times. The next step is to apply our subatom selection theorem. This produces a polynomial factor $\mathfrak{B}$, a refinement $\mathfrak{B}'$, and a subatom selection function $s\colon A_I \to A_{I'}$. Define $X$ to be the "subvariety" of $V$ that is the union of the subatoms selected by $s$. We are guaranteed that $X$ is a "regular model" for $V$ in the sense that most subatoms of $X$ are regular and are colored similarly to the corresponding atom they are contained in.

The third step is to "clean up" the regular atoms. For each $a \in A_I \setminus \tilde{A}_I$, we recolor any point colored with a color that is low-density in the subatom $s(a)$. After this cleaning procedure, arithmetic regularity implies that any pattern which appears in the regular atoms of the cleaned coloring must appear a lot (namely, with positive density) in the regular model $X$ in the original coloring.

To complete the proof we apply patching to handle the irregular atoms. By our patching result, there exists some $\xi$ such that every $H$ that $\xi$ canonically induces appears a lot in the irregular atoms of $X$. By replacing our coloring on the irregular atoms with the $\xi$-canonical coloring, we produce a new coloring. The property of the new coloring is that any $H$ which appears in the new coloring must also have appeared with positive density in the regular model $X$ in the original coloring, which is sufficient to prove the removal lemma.

As in the proof of Lemma 4.5.8, there is quite a bit of technical work involved with the application of the arithmetic regularity lemma, but the above contains the high-level ideas of the proof.

We now give the definitions necessary to run the compactness argument. The recoloring produced in the above argument can be summarized by two pieces of data. First, we record which colors appear in each regular atom and second, we record which $\xi$-canonical coloring was used on the irregular atoms. With these two pieces of data we can determine which patterns we could possibly find in the recoloring.

**Definition 4.6.1.** Fix a prime $p$, a finite set $\mathcal{S}$ equipped with an $\mathbb{F}_p^\times$-action, and a parameter list $I \in \mathcal{I}_p$. A **summary function** with parameters $I$ is a pair $(F, \xi)$ consisting of a function $F\colon (A_I \setminus \tilde{A}_I) \to 2^{\mathcal{S}} \setminus \{\emptyset\}$ and a projective function $\xi\colon \mathbb{F}_p \times A_{\tilde{I}} \to \mathcal{S}$.

**Definition 4.6.2.** For an $\mathcal{S}$-colored pattern $H = (\boldsymbol{L}, \psi)$ consisting of $m$ linear forms and a summary function $(F, \xi)$ with parameters $I$, say that $(F, \xi)$ **partially induces** $H$ if there exists a tuple of atoms $\boldsymbol{a} \in A_I^m$ such that the following holds:

(i) $\boldsymbol{a}$ is $\boldsymbol{L}$-consistent, i.e., $\boldsymbol{a} \in \Phi_I(\boldsymbol{L})$;

(ii) for each $i \in [m]$ such that $a_i \notin \tilde{A}_I$, we have $\psi(i) \in F(a_i)$;

(iii) defining $J := \{i \in [m] : a_i \in \tilde{A}_i\}$ and $H_J := ((L_i)_{i\in J}, \psi|_J, (a_i)_{i\in J}\}$, an $\mathcal{S}$-colored $I$-labeled pattern, we have $\xi$ canonically induces $H_J$.

*Proof of Theorem 4.2.8.* We are given a parameter $\epsilon > 0$ and a possibly infinite set $\mathcal{H}$ of $\mathcal{S}$-colored patterns over $\mathbb{F}_p$ of the form $(\overline{\boldsymbol{L}}^\ell, \psi)$ where $\ell$ is some positive integer and $\psi\colon E_\ell \to \mathcal{S}$ is some map. (See Definition 4.2.7 for the definition of $\overline{\boldsymbol{L}}^\ell$ and $E_\ell$.)

We begin with a "compactness argument" based on ideas of Alon and Shapira that allows us to reduce to the case when $\mathcal{H}$ is finite size.

For each parameter list $I \in \mathcal{I}_p$, we define a finite subset $\mathcal{H}_I \subseteq \mathcal{H}$ as follows. Consider the set of all summary functions $(F, \xi)$ with parameters $I$. If there exists any $H \in \mathcal{H}$ such $(F, \xi)$ partially induces $H$, include one such $H$ in $\mathcal{H}_I$. Note that $|\mathcal{H}_I|$ is at most the number of summary functions with parameters $I$, which is finite.

Define the compactness functions $\Psi_{\mathcal{H}}\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ as follows. Let $\Psi_{\mathcal{H}}(D,N)$ be the largest positive integer $\ell$ such that there exists a parameter list $I\in\mathcal{I}_p$ satisfying $\deg I\le D$ and $\|I\|\le N$ such that a pattern of the form $(\overline{\boldsymbol{L}}^{\ell},\psi)$ exists in $\mathcal{H}_I$.

Now we set several parameters. Define non-increasing functions $\eta\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ by

$$\eta(D,N) := \frac{1}{40}\left(\frac{\epsilon}{12N|\mathcal{S}|}\right)^{p^{\Psi_{\mathcal{H}}(D,N)}},$$

$\beta\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ by

$$\beta(D,N) := \min_{I\in\mathcal{I}_p:\deg I\le D,\|I\|\le N}\beta_{\mathsf{patch}}(p,|\mathcal{S}|,I,\Psi_{\mathcal{H}}(D,N)),$$

$\theta\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ by

$$\theta(D,N) := \frac{\beta(D,N)}{40}\left(\frac{\epsilon}{8|\mathcal{S}|}\right)^{p^{\Psi_{\mathcal{H}}(D,N)}},$$

and $\alpha\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to(0,1)$ by

$$\alpha(D,N) := \frac{\beta(D,N)}{10N^{2p^{\Psi_{\mathcal{H}}(D,N)}}}$$

Then define non-decreasing functions $r\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ by

$$r(D,N) := \max\Bigg\{r_{\mathsf{equi}}(p,D,\alpha(D,N)),\\ \max_{I\in\mathcal{I}_p:\deg I\le D,\|I\|\le N} r_{\mathsf{patch}}(p,|\mathcal{S}|,I,\Psi_{\mathcal{H}}(D,N))(D,N)+p\left\lceil\log_p N\right\rceil\Bigg\}$$

and $d\colon \mathbb{Z}_{>0}\times\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ by

$$d(D,N) := p^{\Psi_{\mathcal{H}}(D,N)}.$$

Define parameters

$$\zeta := \frac{\epsilon}{16|\mathcal{S}|} \qquad \text{and} \qquad c_0 := \left\lceil\log_p(2/\epsilon)\right\rceil.$$

Then define

$$C_{\mathsf{max}} := C_{\mathsf{reg}}(p,|\mathcal{S}|,c_0,\zeta,\eta,\theta,d,r),$$

$$D_{\mathsf{max}} := D_{\mathsf{reg}}(p, |\mathcal{S}|, c_0, \zeta, \eta, \theta, d, r),$$

$$n_{\mathsf{min}} := \max\Bigg\{ n_{\mathsf{reg}}(p, c_0, \zeta), \max_{I \in \mathcal{I}_p:\ \deg I \le D_{\mathsf{max}}, \|I\| \le C_{\mathsf{max}}} n_{\mathsf{patch}}(p, |\mathcal{S}|, I, \Psi_{\mathcal{H}}(D_{\mathsf{max}}, C_{\mathsf{max}})) + \left\lceil \log_p C_{\mathsf{max}} \right\rceil \Bigg\}.$$

Finally, define

$$\delta(\epsilon, \mathcal{H}) := \min\left\{ \frac{\beta(D_{\mathsf{max}}, C_{\mathsf{max}})}{40} \left( \frac{\epsilon}{4C_{\mathsf{max}}|\mathcal{S}|} \right)^{p^{\Psi_{\mathcal{H}}(D_{\mathsf{max}}, C_{\mathsf{max}})}}, p^{-n_{\mathsf{min}} \cdot \Psi_{\mathcal{H}}(D_{\mathsf{max}}, C_{\mathsf{max}})} \right\}$$

and

$$\mathcal{H}_\epsilon := \bigcup_{I \in \mathcal{I}_p : \|I\| \le C_{\mathsf{max}},\, \deg I \le D_{\mathsf{max}}} \mathcal{H}_I.$$

Since the union is over a finite set of $I$, we have $\mathcal{H}_\epsilon$ is finite. We will show that this choice of $\delta$ and $\mathcal{H}_\epsilon$ satisfies the desired conclusion.

Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space and $f\colon V \to \mathcal{S}$ be a projective function with $H$-density at most $\delta(\epsilon, \mathcal{H})$ for every $H \in \mathcal{H}_\epsilon$. Our goal is to produce a projective recoloring $g\colon V \to \mathcal{S}$ that agrees with $f$ on all but an at most $\epsilon$-fraction of $V$ that has no generic $H$-instances for every $H \in \mathcal{H}$.

First note that if $\dim V < n_{\mathsf{min}}$, the theorem easily follows. This is because for the pattern $H = (\overline{\boldsymbol{L}}^\ell, \psi)$, if there exists an $H$-instance in $f$, then the $H$-density in $f$ is at least $1/|V|^\ell$. Thus taking $g = f$ and noticing that we chose $\delta(\epsilon, \mathcal{H}) \le p^{-n_{\mathsf{min}} \cdot \Psi_{\mathcal{H}}(C_{\mathsf{max}}, D_{\mathsf{max}})}$, the theorem holds in this case.

Now assume that $\dim V \ge n_{\mathsf{min}}$. We apply Theorem 4.4.6 to the functions $\{1_{f^{-1}(c)}\}_{c \in \mathcal{S}}$ with parameters $p, |\mathcal{S}|, c_0, \zeta, \eta, \theta, d, r$. This produces a polynomial factor $\mathfrak{B}$ and a refinement $\mathfrak{B}'$ both on $V$ with parameters $I$ and $I'$ and a subatom selection function $s\colon A_I \to A_{I'}$ satisfying several other desirable properties.

As above, define $\tilde{A}_I \subset A_I$ to be the set $\tilde{A}_I := \{a \in A_I : a_{1,0} = 0\}$. We call the atoms $a \in \tilde{A}_I$ *irregular* and the remaining atoms $a \in A_I \setminus \tilde{A}_I$ *regular*. Define $\tilde{V}$ to be the codimension-$I_{1,0}$ subspace of $V$ that is the common zero set of all $I_{1,0}$ linear polynomials defining $\mathfrak{B}$. The irregular atoms of $\mathfrak{B}$ exactly consist of $\tilde{V}$, i.e., $\mathfrak{B}^{-1}(\tilde{A}_I) = \tilde{V}$.

Define $\tilde{\mathfrak{B}}$ to be the polynomial factor on $\tilde{V}$ defined by the restrictions of the homogeneous non-classical polynomials that define $\mathfrak{B}$ to $\tilde{V}$, except for the linear polynomials (which restrict to the zero function). Let $\tilde{I} \in \mathcal{I}_p$ be the parameter list of $\tilde{\mathfrak{B}}$ ($\tilde{I}_{1,0} = 0$ and $\tilde{I}_{d,k} = I_{d,k}$ otherwise). Also define $\tilde{\mathfrak{B}}'$ to be the polynomial factor on $\tilde{V}$ defined by the restrictions of the homogeneous non-classical polynomial that define $\mathfrak{B}'$ to $\tilde{V}$, except for the linear polynomials that also define $\mathfrak{B}$. Let $\tilde{I}' \in \mathcal{I}_p$ be the parameter list of $\tilde{\mathfrak{B}}'$ ($\tilde{I}'_{1,0} = I'_{1,0} - I_{1,0}$ and $\tilde{I}'_{d,k} = I'_{d,k}$ otherwise). Note that by Lemma 4.3.16 and our definition of $r$, we have

$$\operatorname{rank} \tilde{\mathfrak{B}} \geq r_{\mathsf{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg \mathfrak{B}, \|\mathfrak{B}\|))(\deg \mathfrak{B}, \|\mathfrak{B}\|),$$

$$\operatorname{rank} \tilde{\mathfrak{B}}' \geq r_{\mathsf{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg \mathfrak{B}, \|\mathfrak{B}\|))(\deg \mathfrak{B}', \|\mathfrak{B}'\|).$$

We will "clean up" the regular atoms by removing low-density colors in a projective manner. We will "patch" the irregular atoms by replacing the coloring by a new coloring $\Xi_{\xi,\iota,\tilde{\mathfrak{B}}}$ for some projective $\xi \colon \mathbb{F}_p \times A_{\tilde{I}} \to \mathcal{S}$ and $\iota \colon \tilde{V} \xrightarrow{\sim} \mathbb{F}_p^{\dim \tilde{V}}$.

Note that to check that the recoloring $g \colon V \to \mathcal{S}$ is projective, it suffices to check this fact separately on $\tilde{V}$ and on $V \setminus \tilde{V}$.

**Clean up regular atoms:** For each $a \in (A_I \setminus \tilde{A}_I)$, say that a color $c \in \mathcal{S}$ is *high-density in $a$* if it appears in $\mathfrak{B}'^{-1}(s(a))$ with density at least $\epsilon/(4|\mathcal{S}|)$. Say that a color is *low-density in $a$* otherwise.

First note that a basic property of subatom selection functions, Lemma 4.3.25(ii), is the following. For $a \in A_I$ and $b \in \mathbb{F}_p^\times$, we have $b \cdot s(a) = s(b \cdot a)$. Combined with the projectiveness of $f$, this implies that for a color $c \in \mathcal{S}$ and $b \in \mathbb{F}_p^\times$, the $c$-density in $\mathfrak{B}'^{-1}(s(a))$ is the same as the $(b \cdot c)$-density in $\mathfrak{B}^{-1}(s(b \cdot a))$. Thus a color $c$ is high-density in $a$ if and only if $b \cdot c$ is high density in $b \cdot a$.

We pick a single high-density color $c_a \in \mathcal{S}$ for each regular atom $a \in (A_I \setminus \tilde{A}_I)$ in a projective way, namely such that $b \cdot c_a = c_{b \cdot a}$ for all $a \in (A_I \setminus \tilde{A}_I)$ and $b \in \mathbb{F}_p^\times$. By the argument in the above paragraph, this is possible.

Now we define our recoloring of the regular atoms, $g \colon (V \setminus \tilde{V}) \to \mathcal{S}$ as follows. For each $a \in (A_I \setminus \tilde{A}_I)$ and $x \in \mathfrak{B}^{-1}(a)$, we define $g(x) := f(x)$ unless $f(x)$ is low-density in $a$, in which case we define $g(x) := c_a$. Note that $g$ is a projective function. Furthermore we claim

that $g$ differs from $f$ on at most an $(\epsilon/2)$-fraction of $V$.

By Theorem 4.4.6(vi), for all but at most a $\zeta$-fraction of $a \in A_I$, the $c$-density in $\mathfrak{B}^{-1}(a)$ and the $c$-density in $\mathfrak{B}'^{-1}(s(a))$ differ by at most $\zeta$ for all $s \in \mathcal{S}$. Thus for most atoms $a$, each low-density color appears in $\mathfrak{B}^{-1}(a)$ with density at most $\epsilon/(4|\mathcal{S}|) + \zeta$, so $f$ and $g$ differ on at most an $(\epsilon/4 + \zeta|\mathcal{S}|)$-fraction of these atoms. The functions $f$ and $g$ may differ completely on the other atoms, but there are at most $\zeta\|\mathfrak{B}\|$ of these. It follows by equidistribution, Theorem 4.3.19, and the rank bound on $\mathfrak{B}$ that each atom of $\mathfrak{B}$ is at most an $(\|\mathfrak{B}\|^{-1} + \alpha(\deg \mathfrak{B}, \|\mathfrak{B}\|))$-fraction of $V$. Putting this all together, we see that $g$ differs from $f$ on at most the following fraction of $V$

$$\zeta\|\mathfrak{B}\| \left( \frac{1}{\|\mathfrak{B}\|} + \alpha(\deg \mathfrak{B}, \|\mathfrak{B}\|) \right) + \left( \frac{\epsilon}{4|\mathcal{S}|} + \zeta \right) |\mathcal{S}| < \frac{\epsilon}{2}.$$

**Patch irregular atoms:** We define $\tilde{\mathcal{H}}$ to be the set of all $\mathcal{S}$-colored $I$-labeled patterns that are defined by a full dimension system of linear forms in at most $\Psi_{\mathcal{H}}(\deg I, \|I\|)$ variables and whose relative density in $\tilde{\mathfrak{B}}'^{-1}(A_{\tilde{I}} \times \{0\})$ is less than $\beta_{\mathsf{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg I, \|I\|))$.

We apply our patching result, Theorem 4.5.9, to the set $\tilde{\mathcal{H}}$. Our definitions are exactly such that $f|_{\tilde{V}}$ with $\tilde{\mathfrak{B}}, \tilde{\mathfrak{B}}'$ demonstrate that Theorem 4.5.9(b) does not hold. In particular, we checked the rank assumptions on $\tilde{\mathfrak{B}}$ and $\tilde{\mathfrak{B}}'$ above when they were defined. Furthermore, we assumed that $\dim V \geq n_{\mathsf{min}}$, which implies that $\dim \tilde{V} \geq n_{\mathsf{patch}}(p, |\mathcal{S}|, \tilde{I}', \Psi_{\mathcal{H}}(\deg I, \|I\|))$. Finally we defined $\tilde{\mathcal{H}}$ to be the set of patterns which appear with very low density in $\tilde{\mathfrak{B}}'^{-1}(A_{\tilde{I}} \times \{0\})$. Thus we conclude that Theorem 4.5.9(a) holds. This means that there exists a projective $\xi \colon \mathbb{F}_p \times A_{\tilde{I}} \to \mathcal{S}$ that does not canonically induce $\tilde{\mathcal{H}}$. In particular, this means that for any fixed isomorphism $\iota \colon \tilde{V} \xrightarrow{\sim} \mathbb{F}_p^{\dim \tilde{V}}$, there are no generic $H$-instances in $(\Xi_{\xi,\iota,\tilde{\mathfrak{B}}})$ for any $H \in \tilde{\mathcal{H}}$.

We complete our definition of $g \colon V \to \mathcal{S}$ by defining $g(x) := \Xi_{\xi,\iota,\tilde{\mathfrak{B}}}(x)$ for all $x \in \tilde{V}$. To conclude this portion of the proof, we make sure that $|\tilde{V}| \leq (\epsilon/2)|V|$. By assumption,

$$|\tilde{V}|/|V| = p^{-I_{1,0}} \leq p^{-c_0} = \epsilon/2,$$

as desired.

**Proof of correctness:** We claim that $g$ has no generic $H$-instances for each $H \in \mathcal{H}$. Define $F\colon \tilde{(}A_I \setminus \tilde{A}_I) \to 2^{\mathcal{S}} \setminus \{0\}$ to map $a$ to the set of high-density colors in $a$ and recall the projective function $\xi\colon \mathbb{F}_p \times A_{\tilde{I}} \to \mathcal{S}$ defined in the "patch irregular atoms" section. Now suppose that the desired conclusion does not hold, i.e., that there is a generic $H'$-instance in $g$ for some $H' \in \mathcal{H}$. By the construction of $g$, this means that $(F, \xi)$ partially induces $H'$. By the definition of the $\mathcal{H}_I$, this means that there is some $H \in \mathcal{H}_I \subseteq \mathcal{H}_\epsilon$ so that $(F, \xi)$ also partially induces $H$. We will reach a contradiction by showing that this implies that the $H$-density in $f$ is larger than $\delta(\epsilon, \mathcal{H})$.

Say that $H = (\overline{\boldsymbol{L}}^\ell, \psi)$ (note that $\ell \leq \Psi_{\mathcal{H}}(\deg I, \|I\|)$). Since $(F, \xi)$ partially induces $H$, this implies that there exists a tuple of atoms $\boldsymbol{a} \in \Phi_I(\overline{\boldsymbol{L}}^\ell) \subseteq A_I^{E_\ell}$ with several desirable properties. Define $J := \{i \in E_\ell : a_i \in \tilde{A}_I\}$. Recalling that $\tilde{A}_I$ is just the set of atoms whose linear part is zero, we can conclude that $J \subseteq E_\ell \subset \mathbb{F}_p^\ell$ is the intersection of $E_\ell$ with some linear subspace $U \leq \mathbb{F}_p^\ell$ of dimension $\ell' \leq \ell$. This means that the system $(L_i^\ell)_{i \in J}$ is equivalent to the system $(L_i^{\ell'})_{i \in J}$ where now we view $J \subset U \simeq \mathbb{F}_p^{\ell'}$. Define, $H_J := ((L_i^{\ell'})_{i \in E_\ell}, \psi|_J, (a_i)_{i \in J})$, an $\mathcal{S}$-colored $I$-labeled pattern. By Lemma 4.3.22, we see that $H_j$ is a full dimension pattern.

The first property, that $\boldsymbol{a}$ is $\overline{\boldsymbol{L}}^\ell$-consistent, implies that $s(\boldsymbol{a})$ is also $\overline{\boldsymbol{L}}^\ell$-consistent, by Lemma 4.3.25(iii). The second property implies that for each $i \in (E_\ell \setminus J)$, the color $\psi(i)$ is high-density in $a_i$. The third property, together with our definition of $\xi$ implies that in $(f, \mathfrak{B})$, the relative density of $H_J$ in $\tilde{\mathfrak{B}}'^{-1}(A_{\tilde{I}} \times \{0\})$ is at least $\beta_{\mathsf{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg I, \|I\|) \geq \beta(\deg I, \|I\|))$.

Now we put everything together as follows. Write $f^{(i)}$ for $1_{f^{-1}(\psi(i))}$. There is a decomposition $f^{(i)} = f^{(i)}_{\mathsf{str}} + f^{(i)}_{\mathsf{sml}} + f^{(i)}_{\mathsf{psr}}$ given by Theorem 4.4.6 for each $i \in E_\ell$. Let $p_1$ be the $H$-density in $f$. We lower bound $p_1$ as follows.

$$\begin{aligned}
p_1 &= \mathbb{E}_{\boldsymbol{x}}\left[\prod_{i \in E_\ell} f^{(i)}(L_i^\ell(\boldsymbol{x}))\right] \\
&= \mathbb{E}_{\boldsymbol{x}}\left[\prod_{i \in J} f^{(i)}(L_i^\ell(\boldsymbol{x})) \prod_{i \in E_\ell \setminus J} \left(f^{(i)}_{\mathsf{str}}(L_i^\ell(\boldsymbol{x})) + f^{(i)}_{\mathsf{sml}}(L_i^\ell(\boldsymbol{x})) + f^{(i)}_{\mathsf{psr}}(L_i^\ell(\boldsymbol{x}))\right)\right] \\
&\geq \mathbb{E}_{\boldsymbol{x}}\left[\prod_{i \in J} f^{(i)}(L_i^\ell(\boldsymbol{x})) \prod_{i \in E_\ell \setminus J} \left(f^{(i)}_{\mathsf{str}}(L_i^\ell(\boldsymbol{x})) + f^{(i)}_{\mathsf{sml}}(L_i^\ell(\boldsymbol{x}))\right)\right] - 3^{|E_\ell|}\eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|).
\end{aligned}$$

The inequality follows from Theorem 4.4.6(iii), the counting lemma (Eq. (4.1)), and the fact that the complexity of $H$ is at most $d(\deg I, \|I\|) = p^{\Psi_{\mathcal{H}}(\deg I, \|I\|)}$.

Write $p_2$ for the expectation in the last line above. By Theorem 4.4.6(iii), the expression inside the expectation is non-negative so we can restrict the expectation to $\boldsymbol{x}$ such that $\overline{\boldsymbol{L}}^{\ell}(\boldsymbol{x}) = s(\boldsymbol{a})$. Thus $p_2$ is at least

$$\mathbb{E}_{\boldsymbol{x}}\left[\prod_{i\in J} f^{(i)}(L_i^{\ell}(\boldsymbol{x}))1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^{\ell}(\boldsymbol{x}))\prod_{i\in E_\ell\setminus J}\left(f_{\mathsf{str}}^{(i)}(L_i^{\ell}(\boldsymbol{x}))+f_{\mathsf{sml}}^{(i)}(L_i^{\ell}(\boldsymbol{x}))\right)1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^{\ell}(\boldsymbol{x}))\right].$$

Write $p_3$ for the expectation in the last line above. Expanding the product, there are at most $2^{|E_\ell|}$ terms involving $f_{\mathsf{sml}}^{(j)}$ for some $j \in E_\ell \setminus J$. Each of these is bounded in magnitude by

$$\mathbb{E}_{\boldsymbol{x}}\left[\left|f_{\mathsf{sml}}^{(j)}(L_j(\boldsymbol{x}))\right|\prod_{i\in E_\ell} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^{\ell}(\boldsymbol{x}))\right].$$

By applying a change of coordinates, we can assume that $L_j^{\ell}(\boldsymbol{x}) = x_1$. Then by the Cauchy-Schwarz inequality, the square of the above expression is bounded by

$$\mathbb{E}_{x_1}\left[\left|f_{\mathsf{sml}}^{(j)}(x_1)\right|^2 1_{\mathfrak{B}'^{-1}(s(a_j))}(x_1)\right]\mathbb{E}_{x_1}\left[1_{\mathfrak{B}'^{-1}(s(a_j))}(x_1)\mathbb{E}_{x_2,\ldots,x_\ell}\left[\prod_{i\neq j} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^{\ell}(\boldsymbol{x}))\right]^2\right].$$

The first term is at most $\theta(\deg\mathfrak{B}, \|\mathfrak{B}\|)^2\|1_{\mathfrak{B}'^{-1}(s(a_j))}\|_2^2$ by Theorem 4.4.6(vi) and the fact that $a_j$ is a regular atom for $j \in E_\ell \setminus J$. The second term can be counted by equidistribution applied to the system $\boldsymbol{L}'$ of $2|E_k| - 1$ linear forms in $2\ell - 1$ variables defined as follows. Set

$$L'_j(x_1, x_2, \ldots, x_\ell, x'_2, \ldots, x'_\ell) := x_1,$$

and for $i \in E_\ell \setminus \{j\}$, define

$$L'_{i,1}(x_1, x_2, \ldots, x_\ell, x'_2, \ldots, x'_\ell) := L_i^{\ell}(x_1, x_2 \ldots, x_\ell),$$

$$L'_{i,2}(x_1, x_2, \ldots, x_\ell, x'_2, \ldots, x'_\ell) := L_i^{\ell}(x_1, x'_2, \ldots, x'_\ell).$$

By Lemma 4.3.23, we know that $\|\mathfrak{B}'\| \cdot |\Phi_{I'}(\boldsymbol{L}')| = |\Phi_{I'}(\overline{\boldsymbol{L}}^{\ell})|^2$ (see also [BFH+13, Lemma 5.13].)

Thus by equidistribution (Theorem 4.3.19) and the rank bound on $\mathfrak{B}'$, we have the second term is at most

$$\frac{1}{|\Phi_{I'}(\boldsymbol{L}')|} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) = \frac{\|\mathfrak{B}'\|}{|\Phi_{I'}(\overline{\boldsymbol{L}}^{\ell})|^2} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) \leq \frac{2\|\mathfrak{B}'\|}{|\Phi_{I'}(\overline{\boldsymbol{L}}^{\ell})|^2}.$$

Applying equidistribution again we have that the first term is at most

$$\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2 \left( \frac{1}{\|\mathfrak{B}'\|} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) \right) \leq \frac{2\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2}{\|\mathfrak{B}'\|}.$$

Combining these bounds and summing over all terms that contain some $f^j_{\mathsf{sml}}$, we see that

$$p_3 \geq p_4 - 2^{|E_\ell|+1} \frac{\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)}{|\Phi_{I'}(\overline{\boldsymbol{L}}^{\ell})|}$$

where

$$p_4 = \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{i \in J} f^{(i)}(L_i^\ell(\boldsymbol{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\boldsymbol{x})) \prod_{i \in E_\ell \setminus J} f^{(i)}_{\mathsf{str}}(L_i^\ell(\boldsymbol{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\boldsymbol{x})) \right].$$

The quantity $f^{(i)}_{\mathsf{str}}(L_i^\ell(\boldsymbol{x}))$ is the density of $\psi(i)$ in the atom of $\mathfrak{B}'$ that $L_i^\ell(\boldsymbol{x})$ lies in. When $\mathfrak{B}(L_i^\ell(\boldsymbol{x})) = s(a_i)$, the fact that $\psi(i)$ is high density in $a_i$ for all $i \in E_\ell \setminus J$ implies that

$$p_4 \geq \left( \frac{\epsilon}{4|\mathcal{S}|} \right)^{|E_\ell|} \mathbb{E}_{\boldsymbol{x}} \left[ \prod_{i \in J} f^{(i)}(L_i^\ell(\boldsymbol{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\boldsymbol{x})) \prod_{i \in E_\ell \setminus J} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\boldsymbol{x})) \right].$$

Write $p_5$ for the expectation above. We write $\boldsymbol{L}_J := (L_i^\ell)_{i \in J}$. By assumption, we know that

$$\mathbb{E}_{\boldsymbol{x}} \left[ \prod_{i \in J} f^{(i)}(L_i^\ell(\boldsymbol{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\boldsymbol{x})) \right] \geq \frac{\beta(\deg I, \|I\|)}{|\Phi_{I'}(\boldsymbol{L}_J)|}. \tag{4.18}$$

We want to use this inequality to show that $p_5$ is at least on the order of $\beta(\deg I, \|I\|)/|\Phi_{I'}(\overline{\boldsymbol{L}}^{\ell})|$. For simplicity, write $\beta := \beta(\deg I, \|I\|)$ in the rest of this argument.

By applying a change of coordinates, we can assume that $\boldsymbol{L}_J$ only depends on $x_1, \dots, x_{\ell'}$ and is independent of $x_{\ell'+1}, \dots, x_\ell$. To lower bound $p_5$, we want to show that each tuple $(x_1, \dots, x_{\ell'})$ that lies in certain atoms extends to a tuple $(x_1, \dots, x_\ell)$ that still lies in certain atoms in approximately the same number of ways. We do this by a Cauchy-Schwarz argument. Define $\boldsymbol{L}''$ to be the following system of $2|E_\ell| - |J|$ linear forms in $2\ell - \ell'$ variables. For $i \in J$, set

$$L_i''(x_1, \dots, x_\ell, x'_{\ell'+1}, \dots, x'_\ell) := L_i^{\ell'}(x_1, \dots, x_{\ell'}),$$

and for $i \in E_\ell \setminus J$, define

$$L_{i,1}''(x_1, \dots, x_\ell, x'_{\ell'+1}, \dots, x'_\ell) := L_i^{\ell}(x_1, \dots, x_{\ell}),$$

$$L_{i,2}''(x_1, \dots, x_\ell, x'_{\ell'+1}, \dots, x'_\ell) := L_i^{\ell}(x_1, \dots, x_{\ell'}, x_{\ell'+1}, \dots, x_{\ell}).$$

By Lemma 4.3.23, we know that $|\Phi_{I'}(\boldsymbol{L}_J)| \cdot |\Phi_{I'}(\boldsymbol{L}'')| = |\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|^2$.

Define $S \subseteq V^{\ell'}$ to be the set of tuples $\boldsymbol{x} = (x_1, \dots, x_{\ell'})$ such that $\mathfrak{B}'(L_i^\ell(\boldsymbol{x})) = s(a_i)$ for each $i \in J$. For $\boldsymbol{x} \in S$, let $c_{\boldsymbol{x}}$ be the number of tuples $\boldsymbol{x}' = (x_1, \dots, x_\ell)$ such that $\mathfrak{B}'(L_i^\ell(\boldsymbol{x})) = s(a_i)$ for each $i \in E_\ell$. By applying equidistribution (Theorem 4.3.19) to $\boldsymbol{L}_J$ and $\overline{\boldsymbol{L}}^\ell$ and $\boldsymbol{L}''$, we find that

$$|S| = \sum_{\boldsymbol{x} \in S} 1 \leq (1 + \beta/10) \frac{|V|^{\ell'}}{|\Phi_{I'}(\boldsymbol{L}_J)|}, \tag{4.19}$$

$$\sum_{\boldsymbol{x} \in S} c_{\boldsymbol{x}} \geq (1 - \beta/10) \frac{|V|^{\ell}}{|\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|}, \tag{4.20}$$

$$\sum_{\boldsymbol{x} \in S} c_{\boldsymbol{x}}^2 \leq (1 + \beta/10) \frac{|V|^{2\ell - \ell'}}{|\Phi_{I'}(\boldsymbol{L}'')|} = (1 + \beta/10) \frac{|V|^{2\ell - \ell'} |\Phi_{I'}(\boldsymbol{L}_J)|}{|\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|^2}. \tag{4.21}$$

Define $T \subseteq S \subseteq V^{\ell'}$ to be the set of tuples $\boldsymbol{x} = (x_1, \dots, x_{\ell'})$ such that $\mathfrak{B}'(L_i^\ell(\boldsymbol{x})) = s(a_i)$ and $f^{(i)}(L_i^\ell(\boldsymbol{x})) = 1$ for each $i \in J$. Eq. (4.18) implies that

$$|T| \geq \beta \frac{|V|^{\ell'}}{|\Phi_{I'}(\boldsymbol{L}_J)|}. \tag{4.22}$$

We express $p_5$ as follows.

$$p_5 = \frac{1}{|V|^\ell}\sum_{\boldsymbol{x}\in T} c_{\boldsymbol{x}} = \frac{1}{|V|^\ell}\left(\sum_{\boldsymbol{x}\in S} c_{\boldsymbol{x}} - \sum_{\boldsymbol{x}\in S\setminus T} c_{\boldsymbol{x}}\right) \geq (1-\beta/10)\frac{1}{|\Phi_{I'}(\boldsymbol{L}_J)|} - \frac{1}{|V|^\ell}\sum_{\boldsymbol{x}\in S\setminus t} c_{\boldsymbol{x}}.$$

Then combining Eq. (4.19), Eq. (4.21), Eq. (4.22) with the Cauchy-Schwarz inequality gives

$$\left(\sum_{\boldsymbol{x}\in S\setminus T} c_{\boldsymbol{x}}\right)^2 \leq |S\setminus T|\cdot\sum_{\boldsymbol{x}\in S\setminus T} c_{\boldsymbol{x}}^2 \leq (1-4\beta/5)\frac{|V|^{2\ell}}{|\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|^2}.$$

Taking the square root and combining the above two inequalities gives

$$p_5 \geq \frac{\beta}{10}\frac{1}{|\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|}.$$

Combining all the above inequalities we see that $p_1$, the $H$-density in $f$, is bounded by

$$p_1 \geq \left(\left(\frac{\epsilon}{4|\mathcal{S}|}\right)^{|E_\ell|}\frac{\beta(\deg\mathfrak{B},\|\mathfrak{B}\|)}{10} - 2^{|E_\ell|+1}\theta(\deg\mathfrak{B},\|\mathfrak{B}\|)\right)\frac{1}{|\Phi_{I'}(\overline{\boldsymbol{L}}^\ell)|} - 3^{|E_\ell|}\eta(\deg\mathfrak{B}',\|\mathfrak{B}'\|).$$

By our choice of parameters, this quantity is greater than $\delta(\epsilon,\mathcal{H})$ which provides the desired contradiction. Therefore we conclude that the recoloring $g\colon V\to\mathcal{S}$ has no generic $H$-instances for every $H\in\mathcal{H}$. □

We are finally in a position to prove the main removal lemma. To do so we take an arbitrary coloring $f\colon V\to\mathcal{S}$ and construct a new projective coloring $\overline{f}$ where the color of $x$ under $\overline{f}$ contains the information of the coloring of the whole line spanned by $x$. Once we have done so, we can deduce the main removal lemma from the projective removal lemma.

*Proof of Theorem 4.2.4.* Define $\overline{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$ with $\mathbb{F}_p^\times$-action defined by

$$b'\cdot(c_b)_{b\in\mathbb{F}_p^\times} := (c_{b'b})_{b\in\mathbb{F}_p^\times}.$$

First we partition

$$\mathcal{H} = \bigsqcup_{c\in\mathcal{S}\sqcup\{0\}}\mathcal{H}_c$$

as follows. If $L_i \equiv 0$ for some $i \in [m]$, place $H$ in the set $\mathcal{H}_{\psi(i)}$. Otherwise, place $H$ in $\mathcal{H}_0$. (Without loss of generality, we can assume that no pattern in $\mathcal{H}$ has multiple linear forms that are identically equal to 0.)

Now we define sets $\overline{\mathcal{H}}_c$ of $\overline{\mathcal{S}}$-colored patterns for each $c \in \mathcal{S}$. Let $H = (\boldsymbol{L}, \psi) \in \mathcal{H}_0 \cup \mathcal{H}_c$ be an $\mathcal{S}$-colored pattern over $\mathbb{F}_p$ consisting of $m$ linear forms in $\ell$ variables. We can write $\boldsymbol{L} = (L_i^\ell)_{i \in J}$ for some set $J \subseteq \mathbb{F}_p^\ell$ of size $m$. We convert $H$ to a set of patterns defined by the system $\overline{\boldsymbol{L}}^\ell = (L_i^\ell)_{i \in E_\ell}$ as follows. (See Definition 4.2.7 for the definitions of $L_i^\ell$ and $E_\ell$.) For each function $\overline{\psi} \colon E_\ell \to \overline{\mathcal{S}}$ that satisfies $\overline{\psi}(i)_b = \psi(bi)$ whenever $i \in E_\ell \subseteq \mathbb{F}_p^\ell$ and $b \in \mathbb{F}_p^\times$ are such that $bi \in J \subseteq \mathbb{F}_p^\ell$, we add $(\overline{\boldsymbol{L}}^\ell, \overline{\psi})$ to $\overline{\mathcal{H}}_c$.

For each $c \in \mathcal{S}$, we apply Theorem 4.2.8 to $\overline{\mathcal{H}}_c$ with parameter $\epsilon$. This produces a finite subset $\overline{\mathcal{H}}_{c,\epsilon} \subseteq \overline{\mathcal{H}}_c$ and $\delta_c = \delta(\epsilon, \overline{\mathcal{H}}_c) > 0$ with several desirable properties.

Let $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ be the finite subset consisting of all patterns $H$ such that some pattern $\overline{H}$ corresponding to $H$ lies in $\overline{\mathcal{H}}_{c,\epsilon}$ for some $c \in \mathcal{S}$. Let $\delta = \min_{c \in \mathcal{S}} \delta_c > 0$. We claim that $\mathcal{H}_\epsilon, \delta$ satisfy the desired conclusion.

Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Let $f \colon V \to \mathcal{S}$ be a function. Suppose that the $H$-density in $f$ is at most $\delta$ for every $H \in \mathcal{H}_\epsilon$. Define $\overline{f} \colon V \to \overline{\mathcal{S}}$ by

$$\overline{f}(x) := (f(bx))_{b \in \mathbb{F}_p^\times}.$$

Note that $\overline{f}$ is a projective function. Furthermore, we claim that the $\overline{H}$-density in $\overline{f}$ is at most $\delta$ for all $\overline{H} \in \overline{\mathcal{H}}_{f(0)}$. This is true simply because if $\boldsymbol{x} = (x_1, \ldots, x_\ell) \in V^\ell$ is an $\overline{H}$-instance in $\overline{f}$, then $\boldsymbol{x}$ is also an $H$-instance in $f$ where $\overline{H} \in \overline{\mathcal{H}}_{f(0)}$ is any pattern corresponding to $H \in \mathcal{H}_0 \sqcup \mathcal{H}_{f(0)}$.

Thus by assumption, there exists a projective recoloring $\overline{g} \colon V \to \overline{\mathcal{S}}$ such that $\overline{g}$ agrees with $\overline{f}$ on all but an at most $\epsilon$-fraction of $V$ and $\overline{g}$ has no generic $\overline{H}$-instances for every $\overline{H} \in \overline{\mathcal{H}}_{f(0)}$.

Define $g \colon V \to \mathcal{S}$ by $g(x) := \overline{g}(x)_1$ for $x \neq 0$ and $g(0) = f(0)$. Note that $g$ agrees with $f$ on all but an at most $\epsilon$-fraction of $V$. Furthermore, note that $g$ has no $H$-instances for $H \in \mathcal{H}_c$ with $c \neq f(0)$ since $f(0) = g(0)$. Finally, $g$ has no generic $H$-instances for $H \in \mathcal{H}_0 \sqcup \mathcal{H}_{f(0)}$ since any such generic $H$-instance in $g$ is a generic $\overline{H}$-instance in $\overline{g}$ for some $\overline{H}$ corresponding

$H$ in $\overline{\mathcal{H}}_{f(0)}$, which we assumed was not the case. □

## 4.7 Proof of property testing results

In this section we prove our two property testing results, Theorem 4.0.3 and Theorem 4.0.4 from the main removal lemma.

*Proof of Theorem 4.0.3.* It follows from [BGS15, Theorem 10] that a linear-invariant property is testable only if it is semi subspace-hereditary.

Now suppose that $\mathcal{P}$ is a linear-invariant semi subspace-hereditary property. By definition, there exists a subspace-hereditary property $\mathcal{Q}$ such that

(i) every function satisfying $\mathcal{P}$ also satisfies $\mathcal{Q}$;

(ii) for all $\epsilon > 0$, there exists $N(\epsilon)$ such that if $f\colon V \to \mathcal{S}$ satisfies $\mathcal{Q}$ and is $\epsilon$-far from satisfying $\mathcal{P}$, then $\dim V < N(\epsilon)$.

We define $\mathcal{H}$ a (possibly infinite) set of $\mathcal{S}$-colored patterns. For each $f\colon \mathbb{F}_p^\ell \to \mathcal{S}$ that does not satisfy $\mathcal{Q}$, include $H = (\boldsymbol{L}^\ell, f)$ in $\mathcal{H}$ ($\boldsymbol{L}^\ell$ is the system of linear forms that defines an $\ell$-dimensional subspace, defined in Definition 4.2.7). Since $\mathcal{Q}$ is subspace-hereditary, it immediately follows that $\mathcal{Q}$ consists exactly of the functions with no generic $H$-instances for any $H \in \mathcal{H}$.

By Theorem 4.2.4, there exist a finite subset $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and some $\delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. If $f\colon V \to \mathcal{S}$ has $H$-density at most $\delta(\epsilon, \mathcal{H})$ for every $H \in \mathcal{H}_\epsilon$, then $f$ is $\epsilon$-close to $\mathcal{Q}$. Define $\ell(\epsilon)$ to be the largest $\ell$ such that some pattern defined by the system $\boldsymbol{L}^\ell$ is present in $\mathcal{H}_\epsilon$.

Now we define the oblivious tester for $\mathcal{P}$. Given $\epsilon > 0$, the tester produces

$$d(\epsilon) := \max\left\{N(\epsilon/2), \lceil \log_p(2/\delta(\epsilon/2, \mathcal{H}))\rceil + \ell(\epsilon/2)\right\}.$$

Given a function $f\colon V \to \mathcal{S}$ our tester receives oracle access to $f|_U$ where

(i) if $\dim V \geq d(\epsilon)$, then $U$ is a uniform random affine subspace of dimension $d(\epsilon)$;

(ii) else, $U = V$.

Our tester works as follows. If $\dim U < d(\epsilon)$ the tester accepts if $f|_U \in \mathcal{P}$. If $\dim U \geq d(\epsilon)$ the tester accepts in $f|_U \in \mathcal{Q}$.

Suppose $f \in \mathcal{P}$. If $\dim U < d(\epsilon)$, then $U = V$, so $f|_U = f \in \mathcal{P}$. Thus the tester always accepts in this case. In the other case, note that since $f \in \mathcal{P}$, it follows that $f \in \mathcal{Q}$, and since $\mathcal{Q}$ is subspace-hereditary, $f|_U \in \mathcal{Q}$. Thus the tester also always accepts in this case.

Now suppose that $f$ is $\epsilon$-far from $\mathcal{P}$. By the definition of $\mathcal{Q}$ we know that either $\dim V < N(\epsilon/2) \leq d(\epsilon)$ or $f$ is $\epsilon/2$-far from $\mathcal{Q}$. Consider the action of the tester. If $\dim U < d(\epsilon)$, then $U = V$ so $f|_U = f \notin \mathcal{P}$. Thus the tester always rejects in this case. In the other case, note that since $f$ is $\epsilon/2$-far from $\mathcal{Q}$, by assumption there is some $H \in \mathcal{H}_{\epsilon/2}$ such that $f$ has $H$-density more than $\delta(\epsilon/2, \mathcal{H})$. Let $H = (\boldsymbol{L}^\ell, \psi)$ for some $\ell \leq \ell(\epsilon/2)$. We claim the fact that the $H$-density in $f$ is large implies that there is at least a $\delta(\epsilon/2, \mathcal{H})/2$-fraction of $\ell$-dimensional subspaces that $f$ colors by $\psi$. (Note that this does not immediately follow since the $H$-density includes the contribution of $H$-instances that are not generic.) We can compute that the probability a uniform random $\boldsymbol{L}^\ell$-instance in $V$ is not generic is at most $p^{\ell - \dim V}$. It follows that the fraction of $\ell$-dimensional subspaces that $f$ colors by $\psi$ is at least

$$\delta(\epsilon/2, \mathcal{H}) - p^{\ell - \dim V} \geq \delta(\epsilon/2, \mathcal{H}) - p^{-\lceil \log_p(2/\delta(\epsilon/2, \mathcal{H})) \rceil} \geq \delta(\epsilon/2, \mathcal{H})/2.$$

We conclude that in this case the tester rejects with probability at least $\delta(\epsilon/2, \mathcal{H})/2$, as desired. □

*Proof of Theorem 4.0.4.* Suppose $\mathcal{P}$ is a linear-invariant property that is PO-testable. By definition, there exists some $d$, independent of $\epsilon$ and $\dim V$, such that to test $f \colon V \to \mathcal{S}$, the tester receives $f|_U$ where $U$ is

(i) if $\dim V \geq d$, then $U$ is a uniform random linear subspace of dimension $d$;

(ii) else, $U = V$.

We define $\mathcal{H}$ to be the set of patterns of the form $(\boldsymbol{L}^d, \psi)$ where $\psi \colon \mathbb{F}_p^d \to \mathcal{S}$ is a restriction that the tester rejects on ($\boldsymbol{L}^d$ is the pattern that defines a $d$-dimensional subspace, defined

in Definition 4.2.7). We claim that for every $f\colon V \to \mathcal{S}$ with $\dim V \geq d$, it holds that $f \in \mathcal{P}$ if and only if $f$ has no generic $\mathcal{H}$-instances. This claim suffices to prove that $\mathcal{H}$ is subspace-hereditary and locally characterized.

Suppose $f\colon V \to \mathcal{S}$ satisfies $\mathcal{P}$ and $\dim V \geq d$. By the definition of PO-testable, the tester must accept $f$ with probability 1. Thus the tester must accept $f|_U$ for every $U \leq V$ of dimension $d$. This implies that $f$ has no generic $\mathcal{H}$-instances. Now suppose that $f\colon V \to \mathcal{S}$ does not satisfy $\mathcal{P}$ and $\dim V \geq d$ holds. By definition, the tester must accept $f$ with positive probability. Thus there must be some $U \leq V$ of dimension $d$ such that $f|_U$ rejects. This is equivalent to the fact that $f$ contains a generic $H$-instance for some $H \in \mathcal{H}$, proving the desired result.

Now we show that every linear-invariant subspace-hereditary locally characterized property is testable. Suppose $\mathcal{P}$ is such a property. It follows that there is some $d$ and (finite) $\mathcal{H}$ consisting of patterns of the form $(\boldsymbol{L}^d, \psi)$ such that for $f\colon V \to \mathcal{S}$ with $\dim V \geq d$, we have $f$ satisfies $\mathcal{P}$ if and only if $f$ has no generic $\mathcal{H}$-instances.

The PO-tester for $\mathcal{P}$ proceeds in the obvious way. The tester is given oracle access to $f|_U$ where $U$ is

(i) if $\dim V \geq d$, then $U$ is a uniform random linear subspace of dimension $d$;

(ii) else, $U = V$.

The tester accepts if and only if $f|_U \in \mathcal{P}$.

Suppose $f\colon V \to \mathcal{S}$ satisfies $\mathcal{P}$. If $\dim V < d$, then $f|_U = f \in \mathcal{P}$, so the tester accepts $f$. If $\dim V \geq d$, then by the fact that $\mathcal{P}$ is subspace-hereditary and locally characterized, it follows that $f|_U \in \mathcal{P}$ for all $d$-dimensional $U \leq V$. Thus the tester accepts $f$ in this case as well.

Now suppose that $f\colon V \to \mathcal{S}$ is $\epsilon$-far from $\mathcal{P}$. If $\dim V < d$, then $f|_U = f \notin \mathcal{P}$, so the tester rejects $f$. If $\dim V \geq d$, by Theorem 4.2.4, there must be some $H = (\boldsymbol{L}^d, \psi) \in \mathcal{H}$ such that the $H$-density in $f$ is more than $\delta(\epsilon, \mathcal{H})$. We claim that this implies that there is a large fraction of $d$-dimensional subspaces that $f$ colors by some $H \in \mathcal{H}$. (Note that this does not immediately follow since the $H$-density includes the contribution of $H$-instances that are not generic.) We can compute that at most a $p^{d-\dim V}$-fraction of $H$-instances are non-generic.

Thus at least the fraction of $d$-dimensional subspaces that $f$ colors by $\psi$ is at least

$$\delta(\epsilon, \mathcal{H}) - p^{d-\dim V}.$$

This is a good bound on the rejection probability when $\dim V$ is large, but it can be quite bad (or even negative) when $\dim V$ is small. In the latter case we use the following easy bound. If $f$ does not satisfy $\mathcal{P}$, there is at least one $d$-dimensional subspace that is colored by $\mathcal{H}$. The probability that the tester finds this one $\mathcal{H}$-instance is lower-bounded by $p^{-d\cdot\dim V}$. Thus the rejection probability of this tester is at least

$$\max\left\{\delta(\epsilon) - p^{d-\dim V}, p^{-d\cdot\dim V}\right\}.$$

Note that for $\dim V \geq d$, this quantity is uniformly bounded away from 0 (independent of $\dim V$), as desired. □

# Chapter 5

# Non-classical polynomials and the inverse theorem

The inverse theorem for the Gowers $U^k$-norm states that a bounded function $f\colon G \to \mathbb{C}$ has large $U^k$-norm if and only if $f$ correlates with a certain structured object. When $G = \mathbb{Z}/N\mathbb{Z}$, these structured objects are quite complicated and need the theory of nilsequences to describe. When $G = \mathbb{F}_p^n$, the situation is somewhat simpler. When $p > k$, a bounded function $f\colon \mathbb{F}_p^n \to \mathbb{C}$ has large $U^k$-norm if and only if $f$ has non-negligible correlation with a polynomial phase function, i.e., $e^{2\pi i P(x)/p}$ where $P\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is a polynomial of degree at most $k-1$.

The situation when $p$ is small compared to $k$ is more delicate. Green and Tao [GT09] and independently Lovett, Meshulam, and Samorodnitsky [LMS11] showed that the corresponding conjecture is false for $k = 4$ and $p = 2$. In other words, there exist bounded functions $f\colon \mathbb{F}_2^n \to \mathbb{C}$ with large $U^4$-norm but with correlation $o_{n\to\infty}(1)$ with every cubic phase function. Tao and Ziegler [TZ12] clarified this situation by proving that for all $k$ and $p$, a bounded function $f\colon \mathbb{F}_p^n \to \mathbb{C}$ has large $U^k$-norm if and only if $f$ has non-negligible correlation with a non-classical polynomial phase function, i.e., $e^{2\pi i P(x)}$ where $P\colon \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ is a non-classical polynomial of degree at most $k-1$. (See Section 5.1 for the relevant definitions.)

A natural question which remains from the above discussion is to determine for which pairs $p, k$ does the $U^k$-inverse theorem over $\mathbb{F}_p^n$ hold with classical polynomials. In the positive direction, it is known due to Samorodnitsky [Sam07] that the $U^3$-inverse theorem over $\mathbb{F}_2^n$ holds with classical polynomials. In the negative direction, Lovett, Meshulam, and Samorodnitsky

[LMS11] proved that the $U^{p^\ell}$-inverse theorem over $\mathbb{F}_p^n$ requires non-classical polynomials for all $p$ and $\ell \geq 2$. (A curious feature of this problem is that it is not monotone in $k$, e.g., the Lovett-Meshulam-Samorodnitsky result does not imply that non-classical polynomials are necessary in the $U^k$-inverse theorem for all $k \geq p^2$.)

In this chapter we completely characterize when classical polynomials suffice in the statement of the inverse theorem. We first prove the inverse theorem for the Gowers $U^{p+1}$-norm with classical polynomials. This result is proved via a short deduction from the usual inverse theorem for the $U^{p+1}$-norm that involves non-classical polynomials.[1]

**Theorem 5.0.1.** *Fix a prime $p$ and $\delta > 0$. There exists $\epsilon > 0$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Given a function $f \colon V \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^{p+1}} > \delta$, there exists a classical polynomial $P \in \mathrm{Poly}_{\leqslant p}(V \to \mathbb{F}_p)$ such that*

$$|\mathbb{E}_{x \in V} f(x) e_p(-P(x))| \geq \epsilon.$$

Second, we give an example showing that non-classical polynomials are necessary in the $U^k$-inverse theorem for all $k \geq p + 2$.

**Theorem 5.0.2.** *Fix a prime $p$ and an integer $k \geq p + 2$. For all $n$, there exists a function $f \colon \mathbb{F}_p^n \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^k} = 1$ such that for all (classical) polynomials $P \in \mathrm{Poly}_{\leqslant k-1}(\mathbb{F}_p^n \to \mathbb{F}_p)$,*

$$|\mathbb{E}_{x \in \mathbb{F}_p^n} f(x) e_p(-P(x))| = o_{p,k;n \to \infty}(1).$$

Our example is fairly simple to write down. For $k \geq p + 2$, we write $k - 1 = r + (p-1)\ell$ where $\ell \geq 1$ and $0 < r < p$. Then our function is

$$f(x) = e^{2\pi i \frac{\sum_{i=1}^n |x_i|^r}{p^{\ell+1}}}$$

(where $|\cdot| \colon \mathbb{F}_p \to \{0, \ldots, p-1\}$ is the standard map). Note that this function $f$ is a non-classical polynomial phase function of degree $k - 1$, so the content of this result is that it does not correlate with any classical polynomial phase functions of the same degree.

[1] See Section 5.1 for the definitions and notation used in the statement of these results.

The $o(1)$ correlation in Theorem 5.0.2 is fairly bad – the inverse of many iterated logarithms. This is due to our use of a Ramsey-theoretic argument inspired by an argument of Alon and Beigel. (A similar argument appeared in the previous works [GT09, LMS11].) We conjecture that this bound on the correlation can be improved.

**Conjecture 5.0.3.** *Fix a prime $p$ and an integer $k \geq p+2$. For all $n$ there exist $f\colon \mathbb{F}_p^n \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^k} \geq c_{p,k} > 0$ such that for all (classical) polynomials $P \in \mathrm{Poly}_{\leqslant k-1}(\mathbb{F}_p^n \to \mathbb{F}_p)$,*

$$|\mathbb{E}_{x\in\mathbb{F}_p^n} f(x)e_p(-P(x))| \leq \exp(-\Omega_{p,k}(n)).$$

In fact, we believe that this conjecture is true with the same functions that we use to prove Theorem 5.0.2.

**Structure:** In Section 5.1 we give the definition of the Gowers $U^k$-norm and of non-classical polynomials. In Section 5.2 we prove Theorem 5.0.1. We prove Theorem 5.0.2 in the remainder of the chapter. Section 5.3 develops the symmetrization tool that we use and Section 5.4 gives the full proof.

**Notation:** We use $|\cdot|$ for the standard map $\mathbb{F}_p \to \{0,\dots,p-1\}$. We often treat $\mathbb{F}_p$ as an additive subgroup of $\mathbb{R}/\mathbb{Z}$ via the map $x \mapsto |x|/p$ and, by some abuse of notation, freely switch between these two viewpoints. We use $e\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ for the function $e(x) = e^{2\pi i x}$ and $e_p\colon \mathbb{F}_p \to \mathbb{C}$ for the function $e_p(x) = e^{2\pi i|x|/p}$.

This chapter is based on the paper [BSST21].

## 5.1 Background on non-classical polynomials

**Definition 5.1.1.** Fix a prime $p$, a finite-dimensional $\mathbb{F}_p$-vector space $V$, and an abelian group $G$. Given a function $f\colon V \to G$ and a shift $h \in V$, define the **additive derivative** $\Delta_h f\colon V \to G$ by

$$(\Delta_h f)(x) = f(x+h) - f(x).$$

Given a function $f\colon V \to \mathbb{C}$ and a shift $h \in V$, define the **multiplicative derivative** $\partial_h f\colon V \to \mathbb{C}$ by

$$(\partial_h f)(x) = f(x+h)\overline{f(x)}.$$

**Definition 5.1.2.** Fix a prime $p$ and a finite-dimensional $\mathbb{F}_p$-vector space $V$. Given a function $f\colon V \to \mathbb{C}$ and $d \geq 1$, the **Gowers uniformity norm** $\|f\|_{U^d}$ is defined by

$$\|f\|_{U^d} = |\mathbb{E}_{x,h_1,\dots,h_d \in V}(\partial_{h_1} \cdots \partial_{h_d} f)(x)|^{1/2^d}.$$

See [TZ12, Lemma B.1] for some basic facts about the Gowers uniformity norms.

**Definition 5.1.3.** Fix a prime $p$, and a non-negative integer $d \geq 0$. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. A **non-classical polynomial** of degree at most $d$ is a map $P\colon V \to \mathbb{R}/\mathbb{Z}$ that satisfies

$$(\Delta_{h_1} \cdots \Delta_{h_{d+1}} P)(x) = 0$$

for all $h_1, \dots, h_{d+1}, x \in V$. We write $\mathrm{Poly}_{\leqslant d}(V \to \mathbb{R}/\mathbb{Z})$ for the set of non-classical polynomials of degree at most $d$.

A classical polynomial is a map $V \to \mathbb{F}_p$ satisfying the same property. By composing with the standard map $x \mapsto |x|/p$ we can view $\mathrm{Poly}_{\leqslant d}(V \to \mathbb{F}_p)$ as a subset of $\mathrm{Poly}_{\leqslant d}(V \to \mathbb{R}/\mathbb{Z})$.

See [TZ12, Lemma 1.7] for some properties of non-classical polynomials. We give one property below which will be used several times in this chapter.

**Lemma 5.1.4** ([TZ12, Lemma 1.7(iii)])**.** *Fix a prime $p$ and a finite-dimensional $\mathbb{F}_p$-vector space $V = \mathbb{F}_p^n$. Then $P\colon V \to \mathbb{R}/\mathbb{Z}$ is a non-classical polynomial of degree at most $d$ if and only if it can be expressed in the form*

$$P(x_1, \dots, x_n) = \alpha + \sum_{\substack{0 \leq i_1, \dots, i_n < p,\, j \geq 0: \\ 0 < i_1 + \cdots + i_n \leq d - j(p-1)}} \frac{c_{i_1,\dots,i_n,j} |x_1|^{i_1} \cdots |x_n|^{i_n}}{p^{j+1}} \pmod 1,$$

*for some $\alpha \in \mathbb{R}/\mathbb{Z}$ and coefficients $c_{i_1,\dots,i_n,j} \in \{0, \dots, p-1\}$. Furthermore, this representation is unique.*

Define $\mathbb{U}_k \subset \mathbb{R}/\mathbb{Z}$ to be $\{0, 1/p^k, \dots, (p^k-1)/p^k\}$. As a corollary we see that in characteristic $p$, every non-classical polynomial of degree at most $d$ takes values in a coset of $\mathbb{U}_{\lfloor (d-1)/(p-1) \rfloor + 1}$.

Finally we state the inverse theorem of Tao and Ziegler.

**Theorem 5.1.5** ([TZ12, Theorem 1.10]). *Fix a prime $p$, a positive integer $k$, and a parameter $\delta > 0$. There exists $\epsilon > 0$ such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Given a function $f : V \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^k} > \delta$, there exists a non-classical polynomial $P \in \mathrm{Poly}_{\leqslant k-1}(V \to \mathbb{R}/\mathbb{Z})$ such that*

$$|\mathbb{E}_{x \in V} f(x) e^{-2\pi i P(x)}| \geq \epsilon.$$

Earlier works of Bergelson, Tao, and Ziegler [BTZ10] and Tao and Ziegler [TZ10] show this result in the high-characteristic regime $p \geq k$, with the additional guarantee that $P$ is a classical polynomial of degree at most $k - 1$.

## 5.2 Classical polynomials for the $U^{p+1}$-inverse theorem

The inverse theorem for the $U^k$-norm does not require non-classical polynomials when $p \geq k$ for the simple reason that every non-classical polynomial of degree at most $p-1$ is a classical polynomial of the same degree (up to a constant shift). To prove Theorem 5.0.1, about the $U^{p+1}$-inverse theorem, we use the following fact. Every non-classical polynomial of degree $p$ agrees with a classical polynomial on a codimension 1 hyperplane (up to a constant shift).

**Proposition 5.2.1.** *Let $P \in \mathrm{Poly}_{\leqslant p}(V \to \mathbb{R}/\mathbb{Z})$ be a non-classical polynomial of degree at most $p$. Then there exists a codimension 1 hyperplane $U \leq V$, a classical polynomial $Q \in \mathrm{Poly}_{\leqslant p}(V \to \mathbb{F}_p)$, and $\alpha \in \mathbb{R}/\mathbb{Z}$ such that $P(x) = \alpha + |Q(x)|/p$ for all $x \in U$.*

*Proof.* Pick an isomorphism $V \simeq \mathbb{F}_p^n$. By Lemma 5.1.4, we have

$$P(x_1, \ldots, x_n) = \alpha + P'(x_1, \ldots, x_n) + \frac{c_1|x_1| + \cdots + c_n|x_n|}{p^2} \pmod 1$$

for $\alpha \in \mathbb{R}/\mathbb{Z}$, a polynomial $P'$ taking values in $\mathbb{U}_1 = \{0, 1/p, \ldots, (p-1)/p\}$, and $c_1, \ldots, c_n \in \{0, \ldots, p-1\}$. Define the codimension 1 hyperplane $U \leq V$ by $c_1 x_1 + \cdots + c_n x_n = 0$. Note that for $(x_1, \ldots, x_n) \in U$, we have $c_1|x_1| + \cdots + c_n|x_n| \equiv 0 \pmod p$. Thus $P|_U$ takes values in $\alpha + \mathbb{U}_1$. Thus by our identification of $\mathbb{U}_1$ with $\mathbb{F}_p$, $P|_U - \alpha$ is a classical polynomial of degree at most $p$. □

*Proof of Theorem 5.0.1.* By the usual inverse theorem, Theorem 5.1.5, there exists $P \in \mathrm{Poly}_{\leqslant p}(V \to \mathbb{R}/\mathbb{Z})$ such that $|\mathbb{E}_{x \in V} f(x) e(-P(x))| \geq \epsilon$. By Proposition 5.2.1, there exists a codimension 1 hyperplane $U$ and $\alpha \in \mathbb{R}/\mathbb{Z}$ such that $P|_U$ takes values in $\alpha + \mathbb{U}_1$, i.e., $P|_U - \alpha$ is classical. Pick an isomorphism $V \simeq \mathbb{F}_p^n$ such that $U$ is the hyperplane defined by $x_1 = 0$. In this basis, there exists a classical polynomial $Q \in \mathrm{Poly}_{\leqslant p}(\mathbb{F}_p^n \to \mathbb{F}_p)$ and $c \in \{0, \ldots, p-1\}$ so that

$$P(x_1, \ldots, x_n) = \alpha + \frac{|Q(x_1, \ldots, x_n)|}{p} + \frac{c|x_1|}{p^2} \pmod 1.$$

Thus we have

$$\begin{aligned}
\epsilon &\leq |\mathbb{E}_{x \in \mathbb{F}_p^n} f(x) e(-P(x))| \\
&\leq \mathbb{E}_{x_1 \in \mathbb{F}_p} \left| \mathbb{E}_{y \in \mathbb{F}_p^{n-1}} f(x_1, y) e(-P(x_1, y)) \right| \\
&= \mathbb{E}_{x_1 \in \mathbb{F}_p} \left| \mathbb{E}_{y \in \mathbb{F}_p^{n-1}} f(x_1, y) e_p(-Q(x_1, y)) \right| \\
&\leq \left( \mathbb{E}_{x_1 \in \mathbb{F}_p} \left| \mathbb{E}_{y \in \mathbb{F}_p^{n-1}} f(x_1, y) e_p(-Q(x_1, y)) \right|^2 \right)^{1/2}.
\end{aligned}$$

By Parseval and the pigeonhole principle, there exists $a \in \mathbb{F}_p$ such that

$$\left| \mathbb{E}_{x_1 \in \mathbb{F}_p} e(-a x_1) \mathbb{E}_{y \in \mathbb{F}_p^{n-1}} f(x_1, y) e_p(-Q(x_1, y)) \right| \geq \epsilon / \sqrt{p}.$$

Therefore $f$ has correlation at least $\epsilon/\sqrt{p}$ with the classical polynomial $Q(x_1, \ldots, x_n) + a x_1$. □

## 5.3 Symmetrization techniques

We now extend a symmetrization technique of Alon and Beigel [AB01] which will be needed to prove the non-correlation property of our example. The original version of this technique uses Ramsey theory to show that if a function correlates with a bounded degree multilinear polynomial, then some restriction of coordinates correlates with a symmetric polynomial. This result was used in the previous works [GT09, LMS11] on this problem.

For our application we need a result which applies to arbitrary polynomials. We show

that if a function correlates with a bounded degree polynomial, then some restriction of coordinates correlates with a so-called quasisymmetric polynomial. These are a generalization of the notion of symmetric polynomials which have found extensive use in enumerative and algebraic combinatorics.

**Definition 5.3.1.** For a prime $p$ and a tuple $(\alpha_1, \ldots, \alpha_s)$ of positive integers satisfying $\alpha_i < p$ for all $i$, the **elementary quasisymmetric polynomial associated to** $(\alpha_1, \ldots, \alpha_s)$ in $n$ variables is the polynomial $Q_\alpha \colon \mathbb{F}_p^n \to \mathbb{F}_p$ defined by

$$Q_\alpha(x_1, \ldots, x_n) = \sum_{1 \le i_1 < \cdots < i_s \le n} \prod_{j=1}^{s} x_{i_j}^{\alpha_j}.$$

We additionally note the total degree is $|\alpha| := \alpha_1 + \cdots + \alpha_s$.

**Theorem 5.3.2.** *Fix a prime $p$ and an integer $d \ge 1$. For any $n$, there exists $m = \omega_{p,d;n\to\infty}(1)$ such that the following holds. Let $P(x_1, \ldots, x_n)$ be a polynomial of degree at most $d$ with coefficients in $\mathbb{F}_p$. There exists $I \subseteq [n]$ of size $|I| = m$ such that for any $y_{[n]\setminus I} \in \mathbb{F}_p^{[n]\setminus I}$, the function $P(x_I, y_{[n]\setminus I})$ (viewed as a polynomial in the $x_I$) can be written as a quasisymmetric polynomial of degree $d$ plus an arbitrary polynomial of degree at most $d-1$.*

*Proof.* We can uniquely express $P$ in the form

$$P(x_1, \ldots, x_n) = \sum_{s=0}^{d} \sum_{1 \le i_1 < \cdots < i_s \le n} \sum_{\substack{\alpha \in \{1,\ldots,p-1\}^s \\ |\alpha| \le d}} c_{i_1,\ldots,i_s,\alpha} \prod_{j=1}^{s} x_{i_j}^{\alpha_j}$$

where the $c_{i_1,\ldots,i_s,\alpha} \in \mathbb{F}_p$ are arbitrary.

Define $\Lambda = \{\lambda \in \{0, \ldots, p-1\}^d : |\lambda| = d\}$. We define a coloring of the complete $d$-uniform hypergraph on $n$ vertices where the set of colors is $\mathbb{F}_p^\Lambda$. For an edge $\{i_1, \ldots, i_d\}$ with $1 \le i_1 < \cdots < i_d \le n$, let the color of this edge be given by $c_{i_1,\ldots,i_d} \colon \Lambda \to \mathbb{F}_p$. We define $c_{i_1,\ldots,i_d}(\lambda) = c_{j_1,\ldots,j_s,\alpha}$ where $\alpha$ is formed by removing the 0's from the tuple $\lambda$ and $(j_1, \ldots, j_s)$ is formed from $(i_1, \ldots, i_d)$ by removing the coordinate $i_k$ if $\lambda_k = 0$.

Applying the hypergraph Ramsey theorem, there exists a subset $I \subseteq [n]$ such that the induced subhypergraph on vertex set $I$ is colored monochromatically with color $c \colon \Lambda \to \mathbb{F}_p$

and $|I| = \omega_{p,d;n\to\infty}(1)$. Unwinding the definitions, we see that

$$P(x_I, y_{[n]\setminus I}) = \sum_{s=0}^{d} \sum_{\substack{\alpha\in\{1,\dots,p-1\}^s \\ |\alpha|=d}} c(\alpha, \underbrace{0,\dots,0}_{d-s \text{ 0's}}) Q_\alpha(x_I) + \text{mixed terms}.$$

Now the mixed terms involve at least one factor of $y_{[n]\setminus I}$, so their total $x_I$-degree is strictly smaller than $d$. □

## 5.4 Non-classical polynomials are necessary

In this section we prove Theorem 5.0.2, namely that non-classical polynomials are necessary in the $U^{k+1}$ inverse theorem when $k > p$. To do this, we use the function $f_n^{(k)} \colon \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ defined by

$$f_n^{(k)}(x) = \frac{1}{p^{\ell+1}} \sum_{i=1}^{n} |x_i|^r \tag{5.1}$$

where $k = r + (p-1)\ell$ with $\ell \geq 1$ and $0 < r < p$. Note that $f_n^{(k)}$ is a non-classical polynomial of degree $k$, so $\|e(f_n^{(k)})\|_{U^{k+1}} = 1$.

In order to motivate our proof, suppose for the sake of contradiction that $f_n^{(k)}$ has correlation at least $\epsilon$ with some classical polynomial of degree at most $k$. By Theorem 5.3.2, we will be able to reduce to the situation

$$\epsilon \leq \left|\mathbb{E}_{x\sim\mathbb{F}_p^n} e(f_n^{(k)}(x) + g(x) + h(x))\right|$$

where $g$ is a homogeneous quasisymmetric polynomial of degree $k$ and $h$ is a classical polynomial of degree at most $k-1$. By the monotonicity of the Gowers norms (alternatively by the Gowers-Cauchy-Schwarz inequality), we deduce

$$\begin{aligned}
\epsilon^{2^k} &\leq \left|\mathbb{E}_x e(f_n^{(k)}(x) + g(x) + h(x))\right|^{2^k} \\
&= \mathbb{E}_{x,h_1,\dots,h_k}(\partial_{h_1}\cdots\partial_{h_k} e(f_n^{(k)} + g + h))(x) \\
&= \mathbb{E}_{h_1,\dots,h_k} e(\Delta_{h_1}\cdots\Delta_{h_k}(f_n^{(k)} + g)).
\end{aligned}$$

Since $f_n^{(k)}$ and $g$ are polynomials of degree $k$, the iterated derivatives $(\Delta_{h_1} \cdots \Delta_{h_k} f_n^{(k)})(x)$ and $(\Delta_{h_1} \cdots \Delta_{h_k} g)(x)$ are constants independent of $x$. Furthermore, they take values in $\mathbb{U}_1$ which (with some abuse of notation) we identify with $\mathbb{F}_p$. Many results on these objects are known, including the fact that in general they are multilinear functions of $h_1, \ldots, h_k$ (see [TZ12, Section 4]). For the purposes of this chapter, it is sufficient to do the following explicit computation.

**Lemma 5.4.1.** *For $k = r + (p-1)\ell$ with $\ell \geq 0$ and $0 < r < p$,*

$$\iota_k(h_1, \ldots, h_k) := \Delta_{h_1} \cdots \Delta_{h_k} f_n^{(k)} = (-1)^\ell r! \sum_{i=1}^{n} (h_1)_i \cdots (h_k)_i,$$

*and for $\alpha = (\alpha_1, \ldots, \alpha_s)$ with $\alpha_1 + \cdots + \alpha_s = k$ and $0 \leq \alpha_i < p$,*

$$\tau_\alpha(h_1, \ldots, h_k) := \Delta_{h_1} \cdots \Delta_{h_k} Q_\alpha = \sum_{\pi \in \mathfrak{S}_k} \sum_{\vec{\imath}} (h_1)_{i_{\pi(1)}} \cdots (h_k)_{i_{\pi(k)}}$$

*where the sum is over sequences $1 \leq i_1 \leq \cdots \leq i_k \leq n$ that satisfy $i_{\alpha_1+\cdots+\alpha_j+1} = \cdots = i_{\alpha_1+\cdots+\alpha_j+\alpha_{j+1}}$ and $i_{\alpha_1+\cdots+\alpha_j} < i_{\alpha_1+\cdots+\alpha_j+1}$ for all $j$.*

*Proof of Lemma 5.4.1.* For classical polynomials $P, Q \colon \mathbb{F}_p^n \to \mathbb{F}_p$, of degrees $d_1, d_2$, the discrete Leibniz rule, $\Delta_h(PQ) = (\Delta_h P)Q + P(\Delta_h Q) + (\Delta_h P)(\Delta_h Q)$, can be easily verified. This implies the more convenient $\Delta_h(PQ) \equiv (\Delta_h P)Q + P(\Delta_h Q) \pmod{\mathrm{Poly}_{\leqslant d_1+d_2-2}(\mathbb{F}_p^n \to \mathbb{F}_p)}$. Note that taking $d$ discrete derivatives kills $\mathrm{Poly}_{\leqslant d}(\mathbb{F}_p^n \to \mathbb{F}_p)$, so if $P \equiv Q \pmod{\mathrm{Poly}_{\leqslant d}(\mathbb{F}_p^n \to \mathbb{F}_p)}$, then $\Delta_{h_1} \cdots \Delta_{h_d} P(x) = \Delta_{h_1} \cdots \Delta_{h_d} Q(x)$.

We compute $\tau_\alpha$ first. Define $f \in \mathrm{Poly}_{\leqslant k}(V \to \mathbb{F}_p)$ by $f(x) = x_{i_1} \cdots x_{i_k}$. By many applications of the discrete Leibniz rule, we see that

$$\Delta_{h_1} \cdots \Delta_{h_k} f(x) = \sum_{\pi \in \mathfrak{S}_k} \Delta_{h_1}(x_{i_{\pi(1)}}) \cdots \Delta_{h_k}(x_{i_{\pi(k)}}) = \sum_{\pi \in \mathfrak{S}_k} (h_1)_{i_{\pi(1)}} \cdots (h_k)_{i_{\pi(k)}}.$$

Extending this result by linearity gives the formula for $\tau_\alpha$.

Now for $\iota_k$. For any $k \geq 1$, write $k = r + (p-1)\ell$ with $0 < r < p$ and $\ell \geq 0$. Define $Q_k \in \mathrm{Poly}_{\leqslant k}(\mathbb{F}_p \to \mathbb{R}/\mathbb{Z})$ by $Q_k(x) = |x|^r / p^{\ell+1}$. By linearity, it suffices to prove that $\Delta_{h_1} \cdots \Delta_{h_k} Q_k(x) = r!(-1)^\ell h_1 \cdots h_k$. We prove that $\Delta_h Q_k(x) \equiv r|h| Q_{k-1}(x)$

(mod $\mathrm{Poly}_{\leqslant k-2}(\mathbb{F}_p \to \mathbb{R}/\mathbb{Z})$) for $k \geq 2$, while obviously $\Delta_h Q_1(x) = |h|/p$. Iterating (and applying the fact that $(p-1)! \equiv -1 \pmod p$) gives the desired result.

We break into two cases. First, if $r = 1$ (and $\ell \geq 1$) then

$$\begin{aligned}\Delta_h Q_k(x) &= \frac{|x+h| - |x|}{p^{\ell+1}} \\ &= \frac{|h|}{p^{\ell+1}} - \frac{\mathbb{1}(|x| + |h| \geq p)}{p^\ell} \\ &= \frac{|h|}{p^{\ell+1}} - \sum_{c=p-|h|}^{p-1} \frac{\mathbb{1}(|x| = c)}{p^\ell} \\ &= \frac{|h|}{p^{\ell+1}} + |h|\frac{|x|^{p-1}}{p^\ell} - \sum_{c=p-|h|}^{p-1} \frac{\mathbb{1}(|x| = c) + |x|^{p-1}}{p^\ell}.\end{aligned}$$

Now $\mathbb{1}(|x| = c) \equiv 1 - (|x| - c)^{p-1} \pmod p$, say $\mathbb{1}(|x| = c) = 1 - (|x| - c)^{p-1} + pE_{c,p}(x)$ for some function $E_{c,p}$. Then we see

$$\Delta_h Q_k(x) - |h|\frac{|x|^{p-1}}{p^\ell} = \frac{|h|}{p^{\ell+1}} - \sum_{c=p-|h|}^{p-1} \frac{1 - (|x| - c)^{p-1} + |x|^{p-1}}{p^\ell} - \sum_{c=p-|h|}^{p-1} \frac{E_{c,p}(x)}{p^{\ell-1}}.$$

We know that every term in this equation is a non-classical polynomial of degree at most $k-1$ except for the last term. Thus we conclude that the last term is also a non-classical polynomial of degree $k-1$. Furthermore, of the three terms on the right-hand side, the first is a constant, the second is a non-classical polynomial of degree at most $(p-1)(\ell-1) + (p-2) = k-2$ (since the $|x|^{p-1}/p^\ell$ terms cancel), and the third has degree at most $(p-1)(\ell-1) = k-p$ (since it takes values in $\mathbb{U}_{\ell-1}$). Thus the right-hand side lies in $\mathrm{Poly}_{\leqslant k-2}(\mathbb{F}_p \to \mathbb{R}/\mathbb{Z})$, proving the desired result in the $r = 1$ case.

Now assume that $k = r + (p-1)\ell$ where $r \geq 2$. We compute

$$\begin{aligned}\Delta_h Q_k(x) &= \frac{|x+h|^r - |x|^r}{p^{\ell+1}} \\ &= \frac{(|x| + |h|)^r - |x|^r}{p^{\ell+1}} - \mathbb{1}(|x| + |h| \geq p)\frac{(|x| + |h|)^r - (|x| + |h| - p)^r}{p^{\ell+1}} \\ &= \frac{\sum_{i=1}^r \binom{r}{i}|h|^i|x|^{r-i}}{p^{\ell+1}} - \mathbb{1}(|x| + |h| \geq p)\left(\sum_{i=1}^r \frac{(-1)^{i-1}\binom{r}{i}(|x| + |h|)^{r-i}}{p^{\ell+1-i}}\right).\end{aligned}$$

We rewrite this as

$$\Delta_h Q_k(x) - r|h|\frac{|x|^{r-1}}{p^{\ell+1}} = \frac{\sum_{i=2}^{r}\binom{r}{i}|h|^i|x|^{r-i}}{p^{\ell+1}} - \mathbb{1}(|x|+|h|\geq p)\left(\sum_{i=1}^{r}\frac{(-1)^{i-1}\binom{r}{i}(|x|+|h|)^{r-i}}{p^{\ell+1-i}}\right).$$

We know that every term in this equation is a non-classical polynomial of degree at most $k-1$ except for the last term, implying that the last term is also a non-classical polynomial of degree $k-1$. Furthermore, of the two terms on the right-hand side, the first is a non-classical polynomial of degree at most $(p-1)\ell + r - 2 = k-2$ and the second has degree at most $(p-1)\ell = k-r$ (since it takes values in $\mathbb{U}_{\ell-1}$). Thus the right-hand side lies in $\mathrm{Poly}_{\leqslant k-2}(\mathbb{F}_p \to \mathbb{R}/\mathbb{Z})$, proving the desired result in the $r \geq 2$ case. □

Define the maps $I_k, T_\alpha \colon (\mathbb{F}_p^n)^{k-1} \to \mathbb{F}_p^n$ by the equations $\iota_k(h_1,\dots,h_k) = I_k(h_1,\dots,h_{k-1})\cdot h_k$ and $\tau_\alpha(h_1,\dots,h_k) = T_\alpha(h_1,\dots,h_{k-1})\cdot h_k$. From Lemma 5.4.1, clearly $I_k(h_1,\dots,h_{k-1})_i = (-1)^\ell r!(h_1)_i\cdots(h_{k-1})_i$. To continue the argument we will need to show that $T_\alpha$ can be expressed in a particularly convenient form.

**Lemma 5.4.2.** *For $\alpha = (\alpha_1,\dots,\alpha_s)$ with $\alpha_1+\cdots+\alpha_s = k$ and $0<\alpha_i<p$ for all $i$,*

$$T_\alpha(h_1,\dots,h_{k-1})_i = \sum_{J\subseteq[k-1]} C_{i,\alpha,J}(h_{[k-1],<i}, (\tau_\beta(h_I))_{\beta,I})\prod_{j\in J}(h_j)_i$$

*for some functions $C_{i,\alpha,J}(\cdot,\cdot)$, evaluated at the tuple of $(h_j)_{i'}$ for all $j\in[k-1]$ and $i'<i$ and the tuple of $\tau_\beta(h_I)$ for all $I\subseteq[k-1]$ and $\beta=(\beta_1,\dots,\beta_t)$ with $\beta_1+\cdots+\beta_t = |I|$ and $0<\beta_i<p$ for all $i$. Furthermore,*

$$C_{i,\alpha,[k-1]} = (-1)^{s-1}\alpha_1(k-1)!.$$

*Proof.* Fix $i\in[n]$ for the rest of the proof. We introduce some notation. We have $h_1,\dots,h_{k-1}\in\mathbb{F}_p^n$. For $I\subseteq[k-1]$, we write $h_I = (h_j)_{j\in I}$. We use

$$h_{I,<} = ((h_j)_1,\dots,(h_j)_{i-1})_{j\in I}\in(\mathbb{F}_p^{i-1})^I \qquad \text{and} \qquad h_{I,>} = ((h_j)_{i+1},\dots,(h_j)_n)_{j\in I}\in(\mathbb{F}_p^{n-i})^I.$$

For $\alpha=(\alpha_1,\dots,\alpha_s)$ we write $\alpha_{<\ell} = (\alpha_1,\dots,\alpha_{\ell-1})$ and $\alpha_{>\ell} = (\alpha_{\ell+1},\dots,\alpha_s)$. We define $\alpha_{\leqslant\ell}$

analogously. Recall that we use $|\alpha| = \alpha_1 + \cdots + \alpha_s$.

By inspection, we can write

$$T_\alpha(h_{[k-1]})_i = \sum_{\ell=1}^{s} \alpha_\ell! \sum_{\substack{I \sqcup J \sqcup K = [k-1]: \\ |I| = |\alpha_{<\ell}|, \\ |J| = \alpha_\ell - 1, \\ |K| = |\alpha_{>\ell}|}} \left( \prod_{j \in J} (h_j)_i \right) \tau_{\alpha_{<\ell}}(h_{I,<}) \tau_{\alpha_{>\ell}}(h_{K,>}). \tag{5.2}$$

We now remove the terms depending on $h_{[k-1],>}$. Take $\beta = (\beta_1, \ldots, \beta_t)$ with $0 < \beta_i < p$ for all $i$ and $L \subseteq [k-1]$ with $|L| = |\beta|$. We have the identity

$$\begin{aligned} \tau_\beta(h_{L,>}) = \tau_\beta(h_L) &- \sum_{\ell=1}^{t} \sum_{\substack{I \sqcup K = L: \\ |I| = |\beta_{\leqslant \ell}|, \\ |K| = |\beta_{>\ell}|}} \tau_{\beta_{\leqslant \ell}}(h_{I,<}) \tau_{\beta_{>\ell}}(h_{K,>}) \\ &- \sum_{\ell=1}^{t} \beta_\ell! \sum_{\substack{I \sqcup J \sqcup K = L: \\ |I| = |\beta_{<\ell}|, \\ |J| = \beta_\ell, \\ |K| = |\beta_{>\ell}|}} \left( \prod_{j \in J} (h_j)_i \right) \tau_{\beta_{<\ell}}(h_{I,<}) \tau_{\beta_{>\ell}}(h_{K,>}). \end{aligned} \tag{5.3}$$

Repeatedly applying this identity eventually puts $T_\alpha(h_{[k-1]})_i$ into the desired form. To see this, note that applying this identity to $\tau_\beta(h_{L,>})$ produces many terms of the form $\tau_{\beta_{>\ell}}(h_{K,>})$ but all of these satisfy $|\beta_{>\ell}| = |K| < |\beta| = |L|$, so we always make progress.

Finally we need to compute $C_{i,\alpha,[k-1]}$. Obviously this coefficient is a constant since the final decomposition that we produce is multilinear in the $h_1, \ldots, h_{k-1}$. Furthermore, the only way to produce a term that is a multiple of $(h_1)_i \cdots (h_{k-1})_i$ is to have no factors of $\tau_{\beta_{<\ell}}(h_{I,<})$ in that term. (However, we have a choice of $I, J, K$ in Eq. (5.2).) This means that in the initial decomposition we need to be in the $\ell = 1$ case of the sum and every time we use the identity Eq. (5.3) we need to be in the $\ell = 1$ case of the second sum. Again, in every subsequent choice although $\ell = 1$ is fixed, we have a choice of $I, J, K$. Tracing through all of these reductions, we see that we produce the coefficient

$$\alpha_1! \binom{k-1}{\alpha_1 - 1} \prod_{j=2}^{s} \left( -(\alpha_j!) \binom{k - \alpha_1 - \cdots - \alpha_{j-1}}{\alpha_j} \right) = (-1)^{s-1} \alpha_1 (k-1)!. \qquad \square$$

So far we have developed the tools to, starting with the assumption of correlation with a

classical polynomial, reduce to a situation in which

$$\epsilon^{2^k} \leq \mathbb{E}\left[e_p\left(\iota_k - \sum_\alpha c_\alpha \tau_\alpha\right)\right] = \mathbb{P}\left(I_k + \sum_\alpha c_\alpha T_\alpha = 0\right)$$
$$= \mathbb{P}_{h_1,\ldots,h_{k-1}\sim\mathbb{F}_p^n}\left(\forall i \in [n],\ I_k(h_1,\ldots,h_{k-1})_i + \sum_\alpha c_\alpha T_\alpha(h_1,\ldots,h_{k-1})_i = 0\right).$$

Therefore, we need a bound on the probability that a multiaffine function equals zero.

**Lemma 5.4.3.** *Let $L\colon \mathbb{F}_p^r \to \mathbb{F}_p$ be a multiaffine function whose leading coefficient (i.e., coefficient of $x_1\cdots x_r$) is non-zero. Then*

$$\mathbb{P}_{x_1,\ldots,x_r\sim\mathbb{F}_p}(L(x_1,\ldots,x_r)=0) \leq 1 - \left(1-\frac{1}{p}\right)^r =: 1 - c_{p,r}.$$

*Proof.* We prove this result by induction on $r$. The bound is trivially true for $r = 0$.

For $r \geq 1$, we can write

$$L(x_1,\ldots,x_r) = x_r M(x_1,\ldots,x_{r-1}) + N(x_1,\ldots,x_{r-1})$$

where $M$ and $N$ are multiaffine and the leading coefficient of $M$ is non-zero. Then for each fixed $x_1,\ldots,x_{r-1}$, there is at most 1 choice of $x_r$ that makes $L$ vanish unless $M(x_1,\ldots,x_r)=0$. Then

$$\mathbb{P}_{x_1,\ldots,x_r\sim\mathbb{F}_p}(L(x_1,\ldots,x_r)\neq 0) \geq \left(1-\frac{1}{p}\right)\mathbb{P}_{x_1,\ldots,x_{r-1}\sim\mathbb{F}_p}(M(x_1,\ldots,x_{r-1})\neq 0).$$

The second term can be handled by the inductive hypothesis. □

We now have the tools to prove the main theorem. The probability we are considering is the probability that $n$ multiaffine functions vanish simultaneously. If these were independent, by the above lemma, we could bound the probability by $(1-c_{p,k})^n = o_{p,k;n\to\infty}(1)$.

In order to introduce such independence, we can take a union bound over all possible $\tau_\beta(h_I)$. Then Lemma 5.4.2 shows that our multiaffine forms have the following property: if we plug in values for $((h_1)_{i'},\ldots,(h_{k-1})_{i'})_{i'<i}$ and $\tau_\beta(h_I)$ for all $\beta, I$, then $T_\alpha(h_1,\ldots,h_{k-1})_i$ is multiaffine in $(h_1)_i,\ldots,(h_{k-1})_i$ with non-zero leading coefficient. Then we may reveal

$((h_1)_i, \dots, (h_{k-1})_i)$ one-by-one for $i \in [n]$, and find that the total probability is bounded by $(1 - c_{p,k})^n$. As the number of possible choices in the union bound is $O_{p,k}(1)$ we will be able to prove the desired bound.

*Proof of Theorem 5.0.2.* Take $k \geq p + 1$. Consider $f_n^{(k)} \colon \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ defined in Eq. (5.1). Since $f_n^{(k)}$ is a non-classical polynomial of degree $k$, we know that $\|e(f_n^{(k)})\|_{U^{k+1}} = 1$. For a classical polynomial $P \in \mathrm{Poly}_{\leqslant k}(\mathbb{F}_p^n \to \mathbb{F}_p)$, set $\epsilon = |\mathbb{E}_x e(f_n^{(k)}(x) + |P(x)|/p)|$. We will prove that $\epsilon = o_{p,k;n\to\infty}(1)$.

By Theorem 5.3.2, there exists $m = \omega_{p,k;n\to\infty}(1)$ and a subset $I \subseteq [n]$ such that for all $y_{[n]\setminus I} \in \mathbb{F}_p^{[n]\setminus I}$,

$$P(x_I, y_{[n]\setminus I}) = Q(x_I) + P_{y_{[n]\setminus I}}(x_I)$$

where $Q$ is a homogeneous quasisymmetric polynomial of degree $k$ and $P_{y_{[n]\setminus I}}$ is a polynomial of degree at most $k - 1$.

Without loss of generality, assume that $I = [m]$. Then

$$\epsilon = |\mathbb{E}_{x\sim\mathbb{F}_p^n} e(f_n^{(k)}(x) + |P(x)|/p)| \leq \mathbb{E}_{y\sim\mathbb{F}_p^{n-m}} |\mathbb{E}_{x\sim\mathbb{F}_p^m} e(f_n^{(k)}(x, y) + |P(x, y)|/p)|.$$

Then by the pigeonhole principle, there exists $y \in \mathbb{F}_p^{n-m}$ such that the inner expectation is at least $\epsilon$. Fix this choice of $y$ for the rest of the proof. Note that $f_n^{(k)}(x, y) = f_m^{(k)}(x) + c_y$ where $c_y = f_{n-m}^{(k)}(y) \in \mathbb{R}/\mathbb{Z}$ is a constant. Thus we have

$$\epsilon \leq |\mathbb{E}_{x\sim\mathbb{F}_p^m} e(f_m^{(k)}(x) + |Q(x)|/p + |P_y(x)|/p + c_y)|.$$

Note that the right-hand side is a $U^1$-norm. Using the monotonicity of the Gowers norms (see, e.g., [TZ12, Lemma B.1.(ii)]) we deduce

$$\begin{aligned}
\epsilon^{2^k} &\leq \|e(f_m^{(k)} + |Q|/p + |P_y|/p + c_y)\|_{U^1}^{2^k} \\
&\leq \|e(f_m^{(k)} + |Q|/p + |P_y|/p + c_y)\|_{U^k}^{2^k} \\
&= \mathbb{E}_{x,h_1,\dots,h_k} \partial_{h_1} \cdots \partial_{h_k} e(f_m^{(k)} + |Q|/p + |P_y|/p + c_y)(x) \\
&= \mathbb{E}_{x,h_1,\dots,h_k} e(\Delta_{h_1} \cdots \Delta_{h_k} (f_m^{(k)} + |Q|/p + |P_y|/p + c_y)(x)).
\end{aligned}$$

Taking $k$ discrete derivatives kills (non-classical) polynomials of degree at most $k-1$ and turns those of degree $k$ into constants. Thus the final expression is equal to

$$\mathbb{E}_{h_1,\dots,h_k} e(\Delta_{h_1} \cdots \Delta_{h_k}(f_m^{(k)} + |Q|/p)).$$

Since $Q$ is a homogeneous quasisymmetric polynomial of degree $k$, it can be written as $\sum_\alpha c_\alpha Q_\alpha$ where $\alpha$ ranges over all tuples $(\alpha_1, \dots, \alpha_s)$ with $|\alpha| = k$ and $0 < \alpha_i < p$ for all $i$ and the $c_\alpha \in \mathbb{F}_p$ are arbitrary coefficients.

We computed $\Delta_{h_1} \cdots \Delta_{h_k} f_m^{(k)}$ and $\Delta_{h_1} \cdots \Delta_{h_k} Q_\alpha$ in Lemma 5.4.1. These are the $k$-linear forms denoted $\iota_k, \tau_\alpha \colon (\mathbb{F}_p^m)^k \to \mathbb{F}_p$ respectively. Thus so far we have shown that

$$\epsilon^{2^k} \le \mathbb{E}_{h_1,\dots,h_k} e_p\left(\iota_k(h_1, \dots, h_k) + \sum_\alpha c_\alpha \tau_\alpha(h_1, \dots, h_k)\right).$$

For an arbitrary $k$-linear form $\sigma \colon (\mathbb{F}_p^m)^k \to \mathbb{F}_p$, there is a unique $(k-1)$-linear function $S \colon (\mathbb{F}_p^m)^{k-1} \to \mathbb{F}_p^m$ that satisfies $\sigma(h_1, \dots, h_k) = S(h_1, \dots, h_{k-1}) \cdot h_k$. Furthermore, we have

$$\mathbb{E}_{h_1,\dots,h_k} e_p(\sigma(h_1, \dots, h_k)) = \mathbb{E}_{h_1,\dots,h_k} e_p(S(h_1, \dots, h_{k-1}) \cdot h_k) = \mathbb{P}_{h_1,\dots,h_{k-1}}(S(h_1, \dots, h_{k-1}) = 0).$$

From this we conclude

$$\epsilon^{2^k} \le \mathbb{P}_{h_1,\dots,h_{k-1}}\left(\forall i \in [m],\ I_k(h_1, \dots, h_{k-1})_i + \sum_\alpha c_\alpha T_\alpha(h_1, \dots, h_{k-1})_i = 0\right).$$

Recall that $I_k(h_1, \dots, h_{k-1})_i = (-1)^\ell r! (h_1)_i \cdots (h_{k-1})_i$ where $k = r + (p-1)\ell$ with $\ell \ge 1$ and $0 < r < p$. Note that $(-1)^\ell r! \neq 0$ in $\mathbb{F}_p$. Furthermore, Lemma 5.4.2 states that

$$T_\alpha(h_1, \dots, h_{k-1})_i = \sum_{J \subseteq [k-1]} C_{i,\alpha,J}(h_{[k-1],<i}, (\tau_\beta(h_I))_{\beta,I}) \prod_{j \in J} (h_j)_i.$$

In other words $T_\alpha(h_1, \dots, h_{k-1})_i$, viewed just as a function of $(h_1)_i, \dots, (h_{k-1})_i$ is multiaffine with coefficients given by $C_{i,\alpha,J}$. Additionally, Lemma 5.4.2 also gives the critical fact that the leading coefficient, $C_{i,\alpha,[k-1]}$, is equal to $(-1)^{s-1}\alpha_1(k-1)!$ for all $i$. Since $k \ge p+1$, we have that $C_{i,\alpha,[k-1]} = 0$ (recall that the coefficients live in $\mathbb{F}_p$).

This implies that $I_k(h_1, \ldots, h_{k-1})_i + \sum_\alpha c_\alpha T_\alpha(h_1, \ldots, h_{k-1})_i$, viewed just as a function of $(h_1)_i, \ldots, (h_{k-1})_i$ is multiaffine with non-zero leading coefficient, say

$$I_k(h_1, \ldots, h_{k-1})_i + \sum_\alpha c_\alpha T_\alpha(h_1, \ldots, h_{k-1})_i = \sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (\tau_\beta(h_I))_{\beta,I}) \prod_{j \in J} (h_j)_i$$

where $C_{i,[k-1]} = (-1)^\ell r! \neq 0$ for all $i$.

By Lemma 5.4.3, if the coefficients are fixed then this function vanishes with probability at most $1 - c_{p,k} < 1$. To complete the proof, we need to show that we can approximately decouple these events. Formally,

$$\begin{aligned}
\epsilon^{2^k} &\leq \mathbb{P}_{h_1,\ldots,h_{k-1}} \left( \forall i \in [m], \sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (\tau_\beta(h_I))_{\beta,I}) \prod_{j \in J} (h_j)_i = 0 \right) \\
&= \sum_{(A_{\beta,I})_{\beta,I}} \mathbb{P}_{h_1,\ldots,h_{k-1}} \left( \forall \beta, I,\ \tau_\beta(h_I) = A_{\beta,I} \cap \sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (\tau_\beta(h_I))_{\beta,I}) \prod_{j \in J} (h_j)_i = 0 \right) \\
&\leq \sum_{(A_{\beta,I})_{\beta,I}} \mathbb{P}_{h_1,\ldots,h_{k-1}} \left( \sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (A_{\beta,I})_{\beta,I}) \prod_{j \in J} (h_j)_i = 0 \right).
\end{aligned}$$

The final replacement simply comes by substituting in the values $A_{\beta,I}$.

Now for each $i \in [m]$, let $E_i$ be the event that $\sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (A_{\beta,I})_{\beta,I}) \prod_{j \in J} (h_j)_i = 0$. We wish to bound

$$\mathbb{P}_{h_1,\ldots,h_{k-1}} \left( E_i \,|\forall i' < i,\ E_{i'} \right).$$

Since the event we are conditioning on only depends on $h_{[k-1],<i}$, the conditional distribution of $(h_1)_i, \ldots, (h_{k-1})_i$ is still uniform. Thus we can upper bound the above probability by

$$\sup_{h_{[k-1],<i}} \mathbb{P}_{(h_1)_i,\ldots,(h_{k-1})_i \sim \mathbb{F}_p} \left( \sum_{J \subseteq [k-1]} C_{i,J}(h_{[k-1],<i}, (A_{\beta,I})_{\beta,I}) \prod_{j \in J} (h_j)_i = 0 \right).$$

By Lemma 5.4.3, and the fact that $C_{i,[k-1]} = (-1)^\ell r! \neq 0$ always, this probability is upper-bounded by $1 - c_{p,k} < 1$. Putting everything together, we have shown that $\epsilon^{2^k} \leq$

$O_{p,k}((1-c_{p,k})^m)$. (The hidden constant is the number of terms in the sum over $(A_{\beta,I})_{\beta,I}$, which depends on $p, k$ but not on $m, n$. It can be bounded by $p^{4^k}$.) We showed that $m = \omega_{p,k;n\to\infty}(1)$, implying that $\epsilon = o_{p,k;n\to\infty}(1)$. □

# Chapter 6

# Quantitative bounds for the $U^4$-inverse theorem over low characteristic finite fields

A central problem in additive combinatorics is to understand the inverse theory of the Gowers uniformity norms. Given a 1-bounded function $f\colon G \to \mathbb{C}$ with large Gowers uniformity norm where $G$ is a finite abelian group, the goal is to show that $f$ must correlate with some structured object.

This problem has primarily received attention in the cases that $G$ is a cyclic group of prime order or a finite field vector space. For $G = \mathbb{Z}/N\mathbb{Z}$ the structured objects require the theory of nilsequences to discribe, as proved by Green, Tao, and Ziegler [GTZ11, GTZ12]. For finite field vector spaces the situation is known to be somewhat simpler. For $p \geq k$, a 1-bounded function $f\colon \mathbb{F}_p^n \to \mathbb{C}$ has large $U^{k+1}$-norm if and only if it correlates with a (classical) polynomial phase function, i.e., $e^{2\pi i P(x)/p}$ for a polynomial $P\colon \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most $k$, as shown by Bergelson, Tao, and Ziegler [BTZ10, TZ10] (see also [BSST21] for a discussion of the $p = k$ case). For $p < k$, one needs the theory of non-classical polynomials to describe the structured objects [TZ12].

The original proofs of these inverse theorems give extremely bad or even ineffective quantitative bounds. Recently Manners proved quantitative bounds for the $U^{k+1}$-inverse theorem over $\mathbb{Z}/N\mathbb{Z}$ [Man18] and Gowers and Milićević proved quantitative bounds for the

$U^{k+1}$-inverse theorem over $\mathbb{F}_p^n$ when $p > k$ [GM17, GM20].

The goal of this chapter is to build upon the work of Gowers and Milićević to give quantitative bounds for the $U^4$-inverse theorem over $\mathbb{F}_p^n$ in the low-characteristic regime $p = 2, 3$. Our proof has two main steps: *symmetrization* and *integration* of low-characteristic trilinear forms. We are able to solve the integration problem for all $k$-linear forms, but the symmetrization problem we are only able to solve for trilinear forms.

This chapter is based on the paper [Tid21].

# 6.1 Introduction

## 6.1.1 Statement of main result

In this chapter we always use $V$ to denote a finite-dimensional $\mathbb{F}_p$-vector space and we write $\omega = e^{2\pi i/p}$. Given a function $f\colon V \to \mathbb{C}$ and a shift $h \in V$, we write $\partial_h f(x) = f(x+h)\overline{f(x)}$ for the multiplicative derivative.

**Definition 6.1.1.** Given a function $f\colon V \to \mathbb{C}$ and $d \geq 2$, the **Gowers uniformity norm** $\|f\|_{U^d}$ is defined by

$$\|f\|_{U^d}^{2^d} = \mathbb{E}_{x,h_1,\dots,h_d \in V}(\partial_{h_1} \cdots \partial_{h_d} f)(x).$$

This definition has many useful properties, including being a well-defined norm for all $d \geq 2$. See [TZ12, Lemma B.1] for more properties and references about the Gowers uniformity norms.

Consider a function $f\colon V \to G$ where $G$ is an abelian group. Given a shift $h \in V$, we write $\Delta_h f(x) = f(x+h) - f(x)$ for the additive derivative. It follows from the definition of the Gowers uniformity norm that a 1-bounded function $f\colon V \to \mathbb{C}$ satisfies $\|f\|_{U^d} \leq 1$ with equality if and only if $f = e^{2\pi i P}$ where $P\colon V \to \mathbb{R}/\mathbb{Z}$ satisfies $\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_d}P(x) = 0$.

**Definition 6.1.2.** A **non-classical polynomial** of degree at most $k$ is a map $P\colon V \to \mathbb{R}/\mathbb{Z}$ that satisfies

$$(\Delta_{h_1} \cdots \Delta_{h_{k+1}} P)(x) = 0$$

for all $h_1, \dots, h_{k+1}, x \in V$. We write $\mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$ for the set of non-classical polynomials of degree at most $k$.

A classical polynomial is a map $P\colon V \to \mathbb{F}_p$ satisfying the same definition. Thus we use $\mathrm{Poly}_{\leqslant k}(V \to \mathbb{F}_p)$ to denote the set of classical polynomials of degree at most $k$. In this chapter we commonly abuse notation to identify $\mathbb{F}_p$ with $\{0, 1/p, \dots, (p-1)/p\} \subset \mathbb{R}/\mathbb{Z}$ when convenient. In this way we consider $\mathrm{Poly}_{\leqslant k}(V \to \mathbb{F}_p)$ to be a subset of $\mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$.

**Theorem 6.1.3.** *Fix a prime $p$ and $\delta > 0$. There exists*

$$1/\epsilon = \exp\left(\exp\left(\exp\left(O_p\left(\log(1/\delta)^{O(1)}\right)\right)\right)\right)$$

*such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Given a function $f\colon V \to \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^4} > \delta$, there exists a non-classical cubic polynomial $P \in \mathrm{Poly}_{\leqslant 3}(V \to \mathbb{R}/\mathbb{Z})$ such that*

$$\left|\mathbb{E}_{x \in V} f(x) e^{-2\pi i P(x)}\right| \geq \epsilon.$$

*Furthermore, if $p \geq 3$, the polynomial can be taken to be classical.*

This theorem was proved with ineffective bounds by Tao and Ziegler [TZ12, Theorem 1.10] and with effective bounds for $p \geq 5$ by Gowers and Milićević [GM20, Theorem 7]. The bounds we prove are of the same shape as Gowers and Milićević's result.

### 6.1.2 Proof strategy

We start by summarizing the Gowers-Milićević proof of the $U^4$-inverse theorem for $p \geq 5$ and then explain where the difficulties arise when $p < 5$.

We are given a function $f\colon \mathbb{F}_p^n \to \mathbb{C}$ which is 1-bounded and satisfies $\|f\|_{U^4} > \delta$. By expanding the definition of the $U^4$-norm and applying the $U^2$-inverse theorem, this implies

$$\delta^{16} < \|f\|_{U^4}^{16} = \mathbb{E}_{a,b,x} \|\partial_a \partial_b f(x)\|_{U^2}^4 \leq \mathbb{E}_{a,b} \|\widehat{\partial_a \partial_b f}\|_\infty^2.$$

By an application of Markov's inequality, this implies that there is a fairly dense set $A \subset V^2$ and a function $\phi\colon A \to V$ such that $|\widehat{\partial_a \partial_b f}(\phi(a,b))|$ is large for each $(a,b) \in A$.

One can show that the function $\phi$ has some weak bilinearity properties. The main bulk of the work done by Gowers and Milićević is to show that this local bilinearity can be upgraded to some global structure. Namely, they prove that $\phi$ agrees with a biaffine map $\Phi\colon V^2 \to V$ on a fairly dense set. This immediately implies that there is a triaffine form $T\colon V^3 \to \mathbb{F}_p$ that satisfies

$$\left|\mathbb{E}_{a,b,c,x} \partial_a \partial_b \partial_c f(x) \omega^{T(a,b,c)}\right| \geq \eta. \tag{6.1}$$

This part of the proof has no assumptions on the characteristic $p$.

To finish the proof, we would like to find a cubic polynomial $P$ such that $\|f\omega^P\|_{U^3}$ is large. From this point we would be able to conclude by a single application of the $U^3$-inverse theorem. We expand

$$\|f\omega^P\|_{U^3}^8 = \mathbb{E}_{a,b,c,x} \partial_a \partial_b \partial_c f(x) \omega^{\Delta_a \Delta_b \Delta_c P(x)}.$$

Thus the object $\Delta_a \Delta_b \Delta_c P(x)$ naturally appears. We call this object the total derivative of $P$. The total derivative of $P$ clearly does not depend on $x$, and is symmetric and trilinear in $a, b, c$.

Thus our next step is to massage Eq. (6.1) to turn the triaffine form $T$ into a symmetric trilinear form. For simplicity of exposition, let us assume that $T$ is a trilinear. (The parts of $T$ that are not trilinear can often be removed by an appropriate application of the Cauchy-Schwarz inequality.)

**Symmetrization:** We start with an argument based on an idea of Green and Tao. By several applications of the Cauchy-Schwarz inequality, Eq. (6.1) implies that $T(a,b,c)$ is "close" to $T(b,a,c)$, more precisely, we can prove that the map $(a,b,c) \mapsto T(a,b,c) - T(b,a,c)$ has bounded rank, and similarly for all other permutations of $a, b, c$.[1]

To conclude the symmetrization step, we define

$$S(a,b,c) = \frac{1}{6}\left(T(a,b,c) + T(a,c,b) + T(b,a,c) + T(b,c,a) + T(c,a,b) + T(c,b,a)\right). \tag{6.2}$$

[1]See Section 6.2 for the definition of tensor rank that we use.

Clearly $S$ is a symmetric trilinear form. Furthermore, by the Green-Tao argument, we know that $\mathrm{rank}(T - S) \ll_\eta 1$.

**Integration:** We now integrate the symmetric trilinear form $S$ to a cubic polynomial $P$. Namely, define

$$P(x) = \frac{1}{6} S(x, x, x). \tag{6.3}$$

One can easily check that $\Delta_a \Delta_b \Delta_c P = S(a, b, c)$.

Starting from Eq. (6.1), with some involved but not particularly interesting additional inequalities, we can conclude that $\|f\omega^P\|_{U^3} \gg_\eta 1$. Applying the $U^3$-inverse theorem lets us complete the proof.

### 6.1.3 Low characteristic

Now we explain the difficulties that appear in characteristic $p < 5$. Fortunately the bulk of the proof, up to Eq. (6.1), works in arbitrary characteristic. The Green-Tao argument also goes through, showing that $T$ is close to each of its permutations. However, the remainder of the symmetrization step and the integration step, which were essentially trivial in high characteristic, fail badly when $p < 5$. In particular, Eq. (6.2) and Eq. (6.3) both involve dividing by 6.

For $p \geq 5$ we have the convenient classification that $S$ satisfies $S(a, b, c) = \Delta_a \Delta_b \Delta_c P$ for some cubic polynomial $P$ if and only if $S$ is a symmetric trilinear form. However, this is no longer the case in low characteristic.

Following Tao and Ziegler, we define a classical symmetric trilinear form (or CSM for short) to be a trilinear form $S \colon V^3 \to \mathbb{F}_p$ that satisfies $S(a, b, c) = \Delta_a \Delta_b \Delta_c P$ for some classical cubic polynomial $P$. As part of their proof of the inverse theorem Tao and Ziegler classified CSMs. In the same vein, we define a non-classical symmetric trilinear form (or nCSM for short) to be $S \colon V^3 \to \mathbb{F}_p$ that satisfies $S(a, b, c) = \Delta_a \Delta_b \Delta_c P$ for some non-classical cubic polynomial $P$. One problem we solve in this chapter is the classification of non-classical symmetric multilinear forms.

In particular, we show that if $S$ is a symmetric trilinear form and also has some further symmetries which will be defined in Section 6.3, then we can integrate $S$ to a non-classical

cubic polynomial $P$ (i.e., $\Delta_a\Delta_b\Delta_c P = S(a,b,c)$). We are able to classify non-classical symmetric $k$-linear forms for all $k$, which solves the integration problem necessary for the $U^{k+1}$-inverse theorem for all $k$.

For the symmetrization problem, we have a trilinear form $T$ which we know is close in rank to all of its permutations. We wish to find, for $p = 2$, an nCSM $S$ such that $\operatorname{rank}(T - S) \ll 1$ and, for $p = 3$, a CSM $S$ such that $\operatorname{rank}(T - S) \ll 1$.

Even the problem of finding a symmetric multilinear $S$ that is close to $T$ is quite hard and we do not know how to solve it for general $k$. When $k = 3$, the hypothesis that $T$ is close to all of its permutations implies that there is a subspace $U \leq V$ of bounded codimension so that $T|_U$ is symmetric. This lets us easily find a symmetric trilinear $S'$ such that $\operatorname{rank}(T - S') \ll 1$. For $p = 3$, turning $S'$ into a CSM is easy. For $p = 2$, we need to turn $S'$ into an nCSM which takes some more work, but we are able to do.

**Structure:** We give the necessary preliminaries in Section 6.2. We solve the integration problem for all $k$ in Section 6.3 and the symmetrization problem for $k = 3$ in Section 6.4. We combine these tools to prove our main result in Section 6.5. In Section 6.6 we give some conjectures.

## 6.2 Preliminaries

We gave a local definition of non-classical polynomials in Section 6.1. There is an equivalent global definition known. (Also see [TZ12, Lemma 1.7] for some more basic facts about non-classical polynomials.)

**Lemma 6.2.1** ([TZ12, Lemma 1.7(iii)])**.** *$P\colon \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ is a non-classical polynomial of degree at most $k$ if and only if it can be expressed in the form*

$$P(x_1,\dots,x_n) = \alpha + \sum_{\substack{0\leq i_1,\dots,i_n<p,\, j\geq 0:\\ 0<i_1+\cdots+i_n\leq k-j(p-1)}} \frac{c_{i_1,\dots,i_n,j}|x_1|^{i_1}\cdots|x_n|^{i_n}}{p^{j+1}} \pmod 1,$$

*for some $\alpha \in \mathbb{R}/\mathbb{Z}$ and coefficients $c_{i_1,\dots,i_n,j} \in \{0,\dots,p-1\}$ where $|\cdot|$ is the standard map $\mathbb{F}_p \to \{0,\dots,p-1\}$. Furthermore, this representation is unique.*

**Definition 6.2.2.** Given a $k$-linear form $T\colon V^k \to \mathbb{F}_p$ we use two different notions for the rank of $T$.

The **analytic rank** of $T$, denoted $\operatorname{arank}(T)$, is defined by the equation

$$p^{-\operatorname{arank}(T)} = \mathbb{E}_{h_1,\dots,h_k \in V}\left[\omega^{T(h_1,\dots,h_k)}\right].$$

The **partition rank** of $T$, denote $\operatorname{prank}(T)$, is the length of the shortest decomposition of $T$ as a sum of partition rank 1 forms. We define $T\colon V^k \to \mathbb{F}_p$ to have partition rank 1 if we can write $T(h_1,\dots,h_k) = R(h_I)S(h_{[k]\setminus I})$ where $R\colon V^I \to \mathbb{F}_p$ and $S\colon V^{[k]\setminus I} \to \mathbb{F}_p$ are multilinear forms and $\emptyset \neq I \subsetneq [k]$.

These definitions of rank enjoy many useful properties. One which will be important to us is the subadditivity of rank. Obviously $\operatorname{prank}(S+T) \leq \operatorname{prank}(S) + \operatorname{prank}(T)$. For analytic rank, we also have $\operatorname{arank}(S+T) \leq \operatorname{arank}(S) + \operatorname{arank}(T)$, but the proof is more subtle [Lov19, Theorem 1.5].

For $k = 2$, the partition rank and analytic rank coincide and both are equal to the usual notion of matrix rank. For general $k$ we have the following bounds.

**Theorem 6.2.3.** *There are constants $C_{p,k}$ and $D_k$ such that for a $k$-linear form $T\colon V^k \to \mathbb{F}_p$ we have the bounds*

$$\operatorname{arank}(T) \leq \operatorname{prank}(T) \leq C_{p,k} \operatorname{arank}(T)^{D_k}.$$

The lower bound is due to Lovett [Lov19] and the upper bound independently by Milićević [Mil19] and Janzer [Jan20]. Note that a recent result of Cohen and Moshkovitz and independently Adiprasito, Kazhdan, and Ziegler improves the upper bound to linear when $k = 3$ and $p > 2$ [CM21, AKZ21].

## 6.3 Non-classical symmetric multilinear forms

**Definition 6.3.1.** A **non-classical symmetric multilinear form** is a map $T\colon V^k \to \mathbb{F}_p$ that is:

- (multilinear) fixing all the variables but one, the map $h_i \mapsto T(h_1,\dots,h_k)$ is a linear map;

- (symmetric) $T(h_1, \ldots, h_k)$ is invariant under permutations of the $k$ variables $h_1, \ldots, h_k$;
- (non-classical) $T(h_1, \ldots, h_1, h_2, \ldots, h_{k-p+1}) = T(h_1, h_2, \ldots, h_2, h_3, \ldots, h_{k-p+1})$ where the variable $h_1$ appears $p$ times in the first expression and $h_2$ appears $p$ times in the second expression.

We write $\mathrm{nCSM}^k(V)$ for the space of non-classical symmetric multilinear forms $T\colon V^k \to \mathbb{F}_p$.

Non-classical symmetric multilinear forms were implicitly introduced by Tao and Ziegler in their proof of the inverse theorem [TZ12]. There they study the related notion of classical symmetric multilinear forms which satisfy the more restrictive condition that

$$T(h_1, \ldots, h_1, h_2, \ldots, h_{k-p+1}) = 0.$$

**Definition 6.3.2.** For $P \in \mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$, define the **total derivative** of $P$ to be the map $d^k P\colon V^k \to \mathbb{F}_p$ defined by

$$d^k P(h_1, \ldots, h_k) = (\Delta_{h_1} \cdots \Delta_{h_k} P)(0).$$

Note that the right-hand side of this equation lies in $\{0, 1/p, \ldots, (p-1)/p\} \subset \mathbb{R}/\mathbb{Z}$ which we identify with $\mathbb{F}_p$. Also since $P$ is a non-classical polynomial of degree at most $k$ we have $d^k P(h_1, \ldots, h_k) = (\Delta_{h_1} \cdots \Delta_{h_k} P)(x)$ for all $h_1, \ldots, h_k, x \in V$.

Tao and Ziegler prove that the total derivatives of classical polynomials are exactly CSMs.

**Proposition 6.3.3** ([TZ12, Lemma 4.5])**.** *For $k \geq 1$ and a classical polynomial $P \in \mathrm{Poly}_{\leqslant k}(V \to \mathbb{F}_p)$, the total derivative $d^k P$ lies in $\mathrm{CSM}^k(V)$. Furthermore this map is surjective. In other words, we have the short exact sequence*

$$0 \to \mathrm{Poly}_{\leqslant k-1}(V \to \mathbb{F}_p) \to \mathrm{Poly}_{\leqslant k}(V \to \mathbb{F}_p) \xrightarrow{d^k} \mathrm{CSM}^k(V) \to 0.$$

They also show that the total derivative of a non-classical polynomial is an nCSM.

**Proposition 6.3.4** ([TZ12])**.** *For $P \in \mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$, the total derivative $d^k P$ lies in $\mathrm{nCSM}^k(V)$.*

Our integration result gives the implication in the opposite implication.

**Proposition 6.3.5** (Integrating nCSMs)**.** *For $k \geq 1$ and $T \in \mathrm{nCSM}^k(V)$, there exists a non-classical polynomial $P \in \mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$ such that $d^k P = T$. In other words, we have the short exact sequence*

$$0 \to \mathrm{Poly}_{\leqslant k-1}(V \to \mathbb{R}/\mathbb{Z}) \to \mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z}) \xrightarrow{d^k} \mathrm{nCSM}^k(V) \to 0.$$

*Proof.* First note that $\ker(d^k) = \mathrm{Poly}_{\leqslant k-1}(V \to \mathbb{R}/\mathbb{Z})$ since $\Delta_{h_1} \cdots \Delta_{h_k} P$ is identically zero for all $h_1, \ldots, h_k$ if and only if $P$ is a non-classical polynomial of degree at most $k-1$.

We show that $d^k$ is surjective in an inexplicit manner. First define $\overline{\mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})}$ to be $\mathrm{Poly}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})$ under the equivalence relation $[P] = [P + \alpha]$ for $\alpha \in \mathbb{R}/\mathbb{Z}$. Note that for $k \geq 1$, the total derivative $d^k$ acts on this quotient space. We will prove that

$$\frac{|\overline{\mathrm{Poly}}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})|}{|\overline{\mathrm{Poly}}_{\leqslant k-1}(V \to \mathbb{R}/\mathbb{Z})|} = |\,\mathrm{nCSM}^k(V)|. \tag{6.4}$$

This suffices to complete the proof.

Choose an isomorphism $V \cong \mathbb{F}_p^n$. By Lemma 6.2.1, we see that $|\overline{\mathrm{Poly}}_{\leqslant k}(V \to \mathbb{R}/\mathbb{Z})| = p^{C'_k}$ where $C'_k$ is defined to be the number of tuples $(i_1, \ldots, i_n, j)$ where $i_1, \ldots, i_n \in \{0, \ldots, p-1\}$ and $j \geq 0$ satisfy $0 < i_1 + \cdots + i_n \leq k - j(p-1)$. Then the left-hand side of Eq. (6.4) is $p^{C'_k - C'_{k-1}}$. Define $C_k = C'_k - C'_{k-1}$. We can easily see that $C_k$ is the number of tuples $(i_1, \ldots, i_n, j)$ where $i_1, \ldots, i_n \in \{0, \ldots, p-1\}$ and $0 \leq j \leq (k-1)/(p-1)$ satisfy $i_1 + \cdots + i_n = k - j(p-1)$.

Now we compute $|\,\mathrm{nCSM}^k(V)|$. Clearly any multilinear form $T \colon V^k \to \mathbb{F}_p$ can be expressed uniquely as

$$T(x_1, \ldots, x_k) = \sum_{j_1, \ldots, j_k \in [n]} c_{j_1, \ldots, j_k} x_{1, j_1} \cdots x_{k, j_k}$$

where $c_{j_1, \ldots, j_k} \in \mathbb{F}_p$. For a tuple $\boldsymbol{j} = (j_1, \ldots, j_k)$, define $i(\boldsymbol{j}) = (i_1, \ldots, i_n)$ such that $i_\ell$ is the number of $j_1, \ldots, j_k$ which are equal to $\ell$. With this notation, we can say that a multilinear form $T$ is symmetric if and only if $c_{\boldsymbol{j}_1} = c_{\boldsymbol{j}_2}$ whenever $i(\boldsymbol{j}_1) = i(\boldsymbol{j}_2)$.

Finally define $i'(\boldsymbol{j}) = (i'_1, \ldots, i'_n)$ where $i'_\ell = 0$ if $i_\ell = 0$ and otherwise $i'_\ell \in \{1, \ldots, p-1\}$ satisfies $i'_\ell \equiv i_\ell \pmod{p-1}$. We can see that a multilinear form $T$ is an nCSM if and only if $c_{\boldsymbol{j}_1} = c_{\boldsymbol{j}_2}$ whenever $i'(\boldsymbol{j}_1) = i'(\boldsymbol{j}_2)$. Thus $|\,\mathrm{nCSM}^k(V)| = p^{D_k}$ where $D_k$ is the number of

tuples $\boldsymbol{i}' \in \{0, \dots, p-1\}^n$ such that there exists $\boldsymbol{j} \in [n]^k$ such that $\boldsymbol{i}' = i'(\boldsymbol{j})$. Clearly the only constraint on such a tuple $\boldsymbol{i}' \in \{0, \dots, p-1\}^n$ is that $s = i'_1 + \cdots i'_n$ satisfies $0 < s \leq n$ and $s \equiv n \pmod{p-1}$. Thus we see that $C_k = D_k$, proving the desired result. □

## 6.4 Symmetrization in low characteristic

For a $k$-linear form $T\colon V^k \to \mathbb{F}_p$ and a permutation $\pi \in \mathfrak{S}_k$, write $T_\pi\colon V^k \to \mathbb{F}_p$ for

$$T_\pi(x_1, \dots, x_k) = T(x_{\pi(1)}, \dots, x_{\pi(k)}).$$

We have one tool that lets us show that forms are close to symmetric. The following result is based on an idea of Green and Tao.

**Lemma 6.4.1** (cf. [Gre07, Lemma 2.8])**.** *Let $A\colon V^2 \to \mathbb{F}_p$ be a bilinear form. Let $b_1, b_2, b_3\colon V \to \mathbb{C}$ be arbitrary 1-bounded functions. If*

$$\left|\mathbb{E}_{u,v \in V} b_1(u) b_2(v) b_3(u+v) \omega^{A(u,v)}\right| \geq \delta,$$

*then*

$$p^{-\operatorname{arank}(A - A_{(12)})} = \mathbb{E}_{u,v} \omega^{A(u,v) - A(v,u)} \geq \delta^8.$$

*Proof.* By an application of the Cauchy-Schwarz inequality and the 1-boundedness of $b_2$, we have

$$\delta^2 \leq \mathbb{E}_{u,u',v \in V} b_1(u) \overline{b_1(u')} b_3(u+v) \overline{b_3(u'+v)} \omega^{A(u-u',v)}.$$

We make the change of variables $u'' = u + u' + v$. Rearranging, we obtain

$$\begin{aligned}\delta^2 &\leq \mathbb{E}_{u,u',u''} b_1(u) \overline{b_1(u')} b_3(u''-u') \overline{b_3(u''-u)} \omega^{A(u-u',u''-u-u')} \\ &= \mathbb{E}_{u,u',u''} \left(b_1(u) \overline{b_3(u''-u)} \omega^{A(u,u''-u)}\right) \left(\overline{b_1(u')} b_3(u''-u') \omega^{A(-u',u''-u')}\right) \omega^{A(u',u) - A(u,u')}.\end{aligned}$$

Write $B = A - A_{(12)}$. By first averaging over $u''$ and then choosing $b'_1, b'_2\colon V \to \mathbb{C}$

appropriate 1-bounded functions, the above inequality becomes

$$\delta^2 \leq \left|\mathbb{E}_{u,u'} b_1'(u) b_2'(u') \omega^{B(u,u')}\right|.$$

By two more applications of the Cauchy-Schwarz inequality, we conclude

$$\delta^8 \leq \mathbb{E}_{u,v,u',v'} \omega^{B(u-v,u'-v')} = \mathbb{E}_{u,u'} \omega^{B(u,u')} = p^{-\operatorname{arank}(B)}. \qquad \square$$

**Corollary 6.4.2.** *Let $T\colon V^3 \to \mathbb{F}_p$ be a trilinear form. Suppose there are 1-bounded functions $b_1, \ldots, b_7\colon V \to \mathbb{C}$ such that*

$$\left|\mathbb{E}_{x,y,z\in V} b_1(x) b_2(y) b_3(z) b_4(x+y) b_5(x+z) b_6(y+z) b_7(x+y+z)(-1)^{T(x,y,z)}\right| \geq \delta.$$

*Then we have* $\operatorname{arank}(T - T_\pi) \leq 16 \log_p(1/\delta)$ *for all* $\pi \in \mathfrak{S}_3$.

*Proof.* Suppose $\pi = (12)$. By the triangle inequality, we see that there exists a function $\delta\colon V \to [0,1]$ and 1-bounded functions $b_{1,z}, b_{2,z}, b_{3,z}\colon V \to \mathbb{C}$ such that

$$\left|\mathbb{E}_{x,y} b_{1,z}(x) b_{2,z}(y) b_{3,z}(x+y) \omega^{T(x,y,z)}\right| \geq \delta(z)$$

for all $z \in V$ and additionally $\mathbb{E}_{z\in V}\delta(z) \geq \delta$.

By Lemma 6.4.1 for all $z \in V$,

$$\mathbb{E}_{x,y} \omega^{T(x,y,z)-T(y,x,z)} \geq \delta(z)^8.$$

Averaging over $z$ and applying convexity,

$$p^{-\operatorname{arank}(T-T_{(12)})} = \mathbb{E}_{x,y,z} \omega^{T(x,y,z)-T(y,x,z)} \geq \mathbb{E}_z \delta(z)^8 \geq \delta^8,$$

so $\operatorname{arank}(T - T(12)) \leq 8 \log_p(1/\delta)$. The same holds for all transpositions. Since every permutation in $\mathfrak{S}_3$ can be written as the product over two transpositions, we conclude by the subadditivity of analytic rank. $\square$

**Proposition 6.4.3.** *Let $T\colon V^3 \to \mathbb{F}_p$ be a trilinear form. Suppose that $T$ is close to symmetric in the sense that* $\operatorname{prank}(T - T_\pi) \le r$ *for all permutations $\pi \in \mathfrak{S}_3$. Then there exists a subspace $U \le V$ with* $\operatorname{codim} U \le 5r$ *such that $T|_U\colon U^3 \to \mathbb{F}_p$ is a symmetric trilinear form.*

*Proof.* By definition, there exists a partition rank decomposition of each $T - T_\pi$ (for $\pi \in \mathfrak{S}_3 \setminus \{\mathrm{id}\}$) as the sum of at most $r$ terms, each of which is the product of a linear form and a bilinear form. Let $U$ be the subspace of $V$ where all the linear forms vanish. Since there are at most $5r$ linear forms in this decomposition, we have the desired bound on the codimension of $U$. Furthermore, since $(T - T_\pi)|_U$ is identically zero for each $\pi \in \mathfrak{S}_3$, we see that $T|_U$ is symmetric. □

*Remark* 6.4.4. This result immediately implies that there exists a symmetric trilinear form $S\colon V^3 \to \mathbb{F}_p$ with $\operatorname{prank}(T - S) \le 15r$. To see this, pick an arbitrary decomposition $V = U \oplus W$. Then define $S(u_1 \oplus w_1, u_2 \oplus w_2, u_3 \oplus w_3) = T|_U(u_1, u_2, u_3)$. We have the bound $\operatorname{prank}(T - S) \le 15r$ since $T - S$ can be written as a linear combination of linear and bilinear forms where the linear forms are given by dotting with a basis element of $W$ and the bilinear forms are arbitrary.

Proposition 6.4.3 easily handles the initial symmetrization step. To complete the symmetrization, we need to find some $R \in \mathrm{nCSM}^3(V)$ such that $\operatorname{prank}(T - R)$ is small. For $p \ge 5$ this is trivial, since every symmetric trilinear form is an nCSM and actually a CSM.

For $p = 3$ every symmetric trilinear form is also an nCSM though not necessarily a CSM. We note however, that for $p = 3$ every nCSM is close to a CSM. This allows us to prove the $U^4$-inverse theorem over $\mathbb{F}_3^n$ with classical polynomials.

**Proposition 6.4.5.** *Let $T\colon V^3 \to \mathbb{F}_3$ be a symmetric trilinear form. Then there exists a codimension 1 subspace $U \le V$ such that $T|_U \in \mathrm{CSM}^3(U)$.*

*Proof.* The only additional condition that $T$ needs to satisfy to be a CSM is $T(x, x, x) = 0$. Thus if $U$ is any subspace of $V$ such that $T(x, x, x) = 0$ for all $x \in U$, we would have $T|_U \in \mathrm{CSM}^3(U)$.

Consider the map $x \mapsto T(x, x, x) \in \mathbb{F}_3$. Because $T$ is a symmetric trilinear form, this map is linear. Indeed, $T(x + y, x + y, x + y) = T(x, x, x) + 3T(x, x, y) + 3T(x, y, y) + T(y, y, y) =$

$T(x,x,x) + T(y,y,y)$. Defining $U$ to be the codimension 1 subspace on which $T(x,x,x)$ vanishes, we have the desired conclusion. □

**Corollary 6.4.6.** *Fix $p \geq 3$. Let $T\colon V^3 \to \mathbb{F}_p$ be a trilinear form that satisfies* $\operatorname{prank}(T - T_\pi) \leq r$ *for all permutations $\pi \in \mathfrak{S}$. Then there exists $S \in \mathrm{CSM}^3(V)$ such that* $\operatorname{prank}(T-S) \leq 15r + 3$.

For $p = 2$ the situation is more complicated. It is known, due to examples of Green and Tao [GT09] and Lovett, Meshulam, and Samorodnitsky [LMS11] that classical polynomials are not sufficient for the $U^4$-inverse theorem over $\mathbb{F}_2$. This means that we will not be able to find a CSM that is close to $T$. However the following argument shows that $T$ is close to an nCSM.

**Proposition 6.4.7.** *Let $T\colon V^3 \to \mathbb{F}_2$ be a symmetric trilinear form. Suppose there are 1-bounded functions $b_1, \dots, b_7\colon V \to \mathbb{C}$ such that*

$$\left|\mathbb{E}_{x,y,z\in V} b_1(x)b_2(y)b_3(z)b_4(x+y)b_5(x+z)b_6(y+z)b_7(x+y+z)(-1)^{T(x,y,z)}\right| \geq \delta.$$

*Then there exists a subspace $U \leq V$ satisfying* $\operatorname{codim} U \leq 8\log_2(1/\delta)$ *such that $T|_U \in \mathrm{nCSM}^3(U)$.*

*Proof.* Make the change of variables $z = x + w$. Then the hypothesis becomes

$$\begin{aligned}\delta &\leq \Big|\mathbb{E}_{x,y,w\in V} b_1(x)b_2(y)b_3(x+w)b_4(x+y)b_5(w) \\ &\qquad\qquad b_6(x+y+w)b_7(y+w)(-1)^{T(x,x,y)+T(x,y,w)}\Big| \\ &= \Big|\mathbb{E}_{w\in V} b_5(w)\mathbb{E}_{x,y\in V}(b_1(x)b_3(x+w))(b_2(y)b_7(y+w)) \\ &\qquad\qquad (b_4(x+y)b_6(x+y+w))(-1)^{T(x,x,y)+T(x,y,w)}\Big|.\end{aligned}$$

By averaging (and the 1-boundedness of $b_5$), there exists $w_0 \in V$ such that

$$\delta \leq \left|\mathbb{E}_{x,y}(b_1(x)b_3(x+w_0))(b_2(y)b_7(y+w_0))(b_4(x+y)b_6(x+y+w_0))(-1)^{T(x,x,y)+T(x,y,w_0)}\right|.$$

Consider the map $A\colon V^2 \to \mathbb{F}_2$ defined by $A(x,y) = T(x,x,y) + T(x,y,w_0)$. We claim

that $A$ is bilinear. Using the symmetry and trilinearity of $T$,

$$\begin{aligned} A(x+x',y) &= T(x+x',x+x',y) + T(x+x',y,w_0) \\ &= T(x,x,y) + 2T(x,x',y) + T(x',x',y) + T(x,y,w_0) + T(x',y,w_0) \\ &= A(x,y) + A(x',y). \end{aligned}$$

Similarly,

$$\begin{aligned} A(x,y+y') &= T(x,x,y+y') + T(x,y+y',w_0) \\ &= T(x,x,y) + T(x,x,y') + T(x,y,w_0) + T(x,y',w_0) \\ &= A(x,y) + A(x,y'). \end{aligned}$$

Since $A$ is bilinear, we can apply Lemma 6.4.1 to conclude that

$$\delta^8 \leq \mathbb{E}_{x,y}(-1)^{A(x,y)-A(y,x)}.$$

Note that

$$A(x,y) - A(y,x) = T(x,x,y) + T(x,y,w_0) - T(y,y,x) - T(y,x,w_0) = T(x,x,y) - T(x,y,y)$$

by the symmetry of $T$.

Define $B\colon V^2 \to \mathbb{F}_2$ by $B(x,y) = T(x,x,y) - T(x,y,y)$. Note that $B$ is bilinear. (This can be seen since $B(x,y) = A(x,y) - A(y,x)$ or by direct computation.) We have shown that $\operatorname{arank} B \leq 8\log_2(1/\delta)$. Let $U$ be the nullspace of $B$. This satisfies $\operatorname{codim} U = \operatorname{prank} B = \operatorname{arank} B \leq 8\log_2(1/\delta)$. Furthermore $B(x,y) = 0$ for all $x, y \in U$, which implies that $T|_U$ is an nCSM. □

This essentially implies the desired symmetrization result. To put the result in the form we need, we use the following application of the Cauchy-Schwarz inequality which will also appear again later in the chapter.

**Lemma 6.4.8.** *Let $\phi\colon V^3 \to \mathbb{F}_p$ be a triaffine form. Suppose there are 1-bounded functions*

$b_1, \dots, b_7 \colon V \to \mathbb{C}$ *such that*

$$\left|\mathbb{E}_{x,y,z \in V} b_1(x) b_2(y) b_3(z) b_4(x+y) b_5(x+z) b_6(y+z) b_7(x+y+z) \omega^{\phi(x,y,z)}\right| \geq \delta.$$

*Let* $T \colon V^3 \to \mathbb{F}_p$ *be the trilinear part of* $\phi$*. Then there exist 1-bounded functions* $b_1', \dots, b_7' \colon V \to \mathbb{C}$ *such that*

$$\left|\mathbb{E}_{x,y,z \in V} b_1'(x) b_2'(y) b_3'(z) b_4'(x+y) b_5'(x+z) b_6'(y+z) b_7'(x+y+z) \omega^{T(x,y,z)}\right| \geq \delta^8.$$

*Proof.* This result follows from three applications of the Cauchy-Schwarz inequality, one in each variable. We have

$$\begin{aligned}
\delta^2 &\leq \left|\mathbb{E}_{y,z} b_2(y) b_3(z) b_6(y+z) \mathbb{E}_x b_1(x) b_4(x+y) b_5(x+z) b_7(x+y+z) \omega^{\phi(x,y,z)}\right|^2 \\
&\leq \left(\mathbb{E}_{y,z} \left|b_2(y) b_3(z) b_6(y+z)\right|^2\right) \\
&\qquad \cdot \Big(\mathbb{E}_{x,x',y,z} b_1(x) \overline{b_1(x')} b_4(x+y) \overline{b_4(x'+y)} b_5(x+z) \overline{b_5(x'+z)} \\
&\qquad\quad \cdot b_7(x+y+z) \overline{b_7(x'+y+z)} \omega^{\phi(x-x',y,z)}\Big).
\end{aligned}$$

The 1-boundedness of $b_2, b_3, b_6$ lets us remove the first term.

Repeating two more times gives

$$\delta^8 \leq \left|\mathbb{E}_{x,x',y,y',z,z'} b(x, x', y, y', z, z') \omega^{\phi(x-x', y-y', z-z')}\right|$$

where $b$ is a product of the $b_7$'s evaluated at sums of the six variables. Note that $\phi(x-x', y-y', z-z') = T(x-x', y-y', z-z')$. Then by averaging, we can find $x_0, y_0, z_0$ such that

$$\delta^8 \leq \left|\mathbb{E}_{x,y,z} b(x+x_0, x_0, y+y_0, y_0, z+z_0, z_0) \omega^{T(x,y,z)}\right|.$$

Tracing through the proof, we can see that $b(x+x_0, x_0, y+y_0, y_0, z+z_0, z_0)$ has the desired factorization. □

**Corollary 6.4.9.** *Let* $T \colon V^3 \to \mathbb{F}_2$ *be a trilinear form. Suppose there are 1-bounded functions*

$b_1, \dots, b_7 \colon V \to \mathbb{C}$ *such that*

$$\left|\mathbb{E}_{x,y,z \in V} b_1(x) b_2(y) b_3(z) b_4(x+y) b_5(x+z) b_6(y+z) b_7(x+y+z) (-1)^{T(x,y,z)}\right| \geq \delta.$$

*Then there exists* $S \in \mathrm{nCSM}^3(V)$ *such that* $\mathrm{prank}(T - S) \leq 432 \log_2(1/\delta)$.

*Proof.* By Corollary 6.4.2, we have that $\mathrm{arank}(T - T_\pi) \leq 16 \log_2(1/\delta)$ for all $\pi \in \mathfrak{S}_3$. Then applying Proposition 6.4.3, there exists a subspace $U \leq V$ with $\operatorname{codim} U \leq 80 \log_2(1/\delta)$ such that $T|_U \colon U^3 \to \mathbb{F}_2$ is a symmetric trilinear form.

By averaging, there exist cosets $x_0 + U, y_0 + U, z_0 + U$ such that

$$\begin{aligned}\delta &\leq \left| \mathbb{E}_{\substack{x \in x_0 + U \\ y \in y_0 + U \\ z \in z_0 + U}} b_1(x) b_2(y) b_3(z) b_4(x+y) b_5(x+z) b_6(y+z) b_7(x+y+z) (-1)^{T(x,y,z)} \right| \\ &= \left|\mathbb{E}_{x,y,z \in U} b_1'(x) b_2'(y) b_3'(z) b_4'(x+y) b_5'(x+z) b_6'(y+z) b_7'(x+y+z) (-1)^{T(x_0+x, y_0+y, z_0+z)}\right|.\end{aligned}$$

Now by Lemma 6.4.8, this implies that

$$\delta^8 \leq \left|\mathbb{E}_{x,y,z \in U} b_1''(x) b_2''(y) b_3''(z) b_4''(x+y) b_5''(x+z) b_6''(y+z) b_7''(x+y+z) (-1)^{T(x,y,z)}\right|.$$

We can now apply Proposition 6.4.7 to find a subspace $W \leq U$ such that $\operatorname{codim} W \leq 144 \log_2(1/\delta)$ and such that $T|_W \in \mathrm{nCSM}^3(W)$. Pick an arbitrary decomposition $V = W \oplus W'$ and then define $S(w_1 \oplus w_1', w_2 \oplus w_2', w_3 \oplus w_3') = T|_W(w_1, w_2, w_3)$. Then $S \in \mathrm{nCSM}^3(V)$ and clearly $\mathrm{prank}(T - S) \leq 432 \log_2(1/\delta)$. □

## 6.5 Proof of main theorem

We now have the tools to prove the main theorem. We first record the result of Gowers and Milićević that we require.

**Theorem 6.5.1** ([GM17])**.** *Fix a prime* $p$ *and* $\delta > 0$*. There exists*

$$\epsilon = \left(\exp\left(\exp\left(\exp\left(O_p\left(\log(1/\delta)^{O(1)}\right)\right)\right)\right)\right)^{-1}$$

*such that the following holds. Let $V$ be a finite-dimensional $\mathbb{F}_p$-vector space. Given a function $f\colon V \to \mathbb{C}$ satisfying $\|f\|_\infty \le 1$ and $\|f\|_{U^4} > \delta$, there exists a triaffine form $\phi\colon V^3 \to \mathbb{F}_p$ such that*

$$\left|\mathbb{E}_{x,h_1,h_2,h_3 \in V}(\partial_{h_1}\partial_{h_2}\partial_{h_3} f)(x)\omega^{\phi(h_1,h_2,h_3)}\right| \ge \epsilon. \tag{6.5}$$

*Proof of Theorem 6.1.3.* Applying Theorem 6.5.1, we find a triaffine form $\phi\colon V^3 \to \mathbb{F}_p$ such that

$$\left|\mathbb{E}_{x,h_1,h_2,h_3 \in V}(\partial_{h_1}\partial_{h_2}\partial_{h_3} f)(x)\omega^{\phi(h_1,h_2,h_3)}\right| \ge \epsilon.$$

Let $T$ be the trilinear part of $\phi$. First average over $x$ and expand out the derivative to conclude that there exist 1-bounded functions $b_1, \dots, b_7\colon V \to \mathbb{C}$ such that

$$\left|\mathbb{E}_{x,y,z \in V} b_1(x)b_2(y)b_3(z)b_4(x+y)b_5(x+z)b_6(y+z)b_7(x+y+z)\omega^{\phi(x,y,z)}\right| \ge \epsilon.$$

Then applying Lemma 6.4.8 and then Corollary 6.4.2, we see that $\operatorname{arank}(T - T_\pi) \le 16\log_p(1/\epsilon)$ for all $\pi \in \mathfrak{S}_3$. By Theorem 6.2.3, we conclude that $\operatorname{prank}(T - T_\pi) \le \operatorname{poly}\log_p(1/\epsilon)$ for all $\pi \in \mathfrak{S}_3$.

Now for $p \ge 3$, we can apply Corollary 6.4.6 to find $S \in \mathrm{CSM}^3(V)$ such that $\operatorname{prank}(T - S) \le \operatorname{poly}\log_p(1/\epsilon)$. Then by Proposition 6.3.3, there exists a classical cubic polynomial $P \in \operatorname{Poly}_{\leqslant 3}(V \to \mathbb{F}_p)$ such that $d^3P = S$. For $p = 2$, we apply Corollary 6.4.9 to find $S \in \mathrm{nCSM}^3(V)$ such that $\operatorname{prank}(T - S) \le \operatorname{poly}\log_p(1/\epsilon)$. Then by Proposition 6.3.5, there exists a non-classical cubic polynomial $P \in \operatorname{Poly}_{\leqslant 3}(V \to \mathbb{R}/\mathbb{Z})$ such that $d^3P = S$.

Now in either case we have a decomposition $T - S = \sum_{i=1}^r \gamma_i$ where each $\gamma_i\colon V^3 \to \mathbb{F}_p$ is the product of a linear form in one of the variables with a bilinear form in the other two variables and $r = \operatorname{poly}\log_p(1/\epsilon)$. Write $\Gamma\colon V^3 \to \mathbb{F}_p^{2r}$ for the list of $2r$ linear and bilinear forms that factor the $\gamma_i$'s. For example, if $\gamma_1(x,y,z) = \alpha(x)\beta(y,z)$, then we define $\Gamma_1(x,y,z) = \alpha(x)$ and $\Gamma_2(x,y,z) = \beta(y,z)$.

Define $g = f\omega^P$. Note that $(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x) = (\partial_{h_1}\partial_{h_2}\partial_{h_3} f)(x)\omega^{S(h_1,h_2,h_3)}$. The inequality we started with can now be written as

$$\left|\mathbb{E}_{x,h_1,h_2,h_3} b(h_1,h_2,h_3)\omega^{\sum_{i=1}^r \gamma_i(h_1,h_2,h_3)}(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)\right| \ge \epsilon$$

where $b(h_1, h_2, h_3)$ is a product of linear and bilinear phase functions. We first remove the middle term.

By averaging, there exists $c \in \mathbb{F}_p^{2r}$ such that

$$|\mathbb{E}_{x,h_1,h_2,h_3} b(h_1, h_2, h_3) \mathbb{1}(\Gamma(h_1, h_2, h_3) = c)(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)| \geq \epsilon p^{-2r}.$$

Now by Fourier analysis, we can write

$$\mathbb{1}(\Gamma(h_1, h_2, h_3) = c) = \mathbb{E}_{\xi \in \mathbb{F}_p^{2r}} \omega^{(\Gamma(h_1,h_2,h_3)-c)\cdot\xi}.$$

Plugging this expression into the previous inequality, by averaging again, there exists $\xi_0 \in \mathbb{F}_p^{2r}$ such that

$$\left|\mathbb{E}_{x,h_1,h_2,h_3} b(h_1, h_2, h_3) \omega^{(\Gamma(h_1,h_2,h_3)-c)\cdot\xi_0} (\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)\right| \geq \epsilon p^{-2r}.$$

Since $\Gamma$ is a list of linear and bilinear forms, the middle term, $\omega^{(\Gamma(h_1,h_2,h_3)-c)\cdot\xi_0}$, is the product of linear and bilinear phase functions. Thus we conclude that

$$|\mathbb{E}_{x,h_1,h_2,h_3} b'(h_1, h_2, h_3)(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)| \geq \epsilon p^{-2r},$$

where $b'$ is the product of linear and bilinear phase functions.

Write

$$b'(h_1, h_2, h_3) = \omega^{\beta_1(h_2,h_3)+\beta_2(h_1,h_3)+\beta_3(h_1,h_2)+\alpha_1(h_1)+\alpha_2(h_2)+\alpha_3(h_3)}.$$

By averaging, there exist $x, h_1 \in V$ such that

$$|\mathbb{E}_{h_2,h_3} b'(h_1, h_2, h_3)(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)| \geq \epsilon p^{-2r}.$$

Expanding this out, we see that this inequality is of the correct form to apply Lemma 6.4.1 which then implies that $\operatorname{arank}(\beta_1 - (\beta_1)_{(12)}) \leq 8(2r + \log_p(1/\epsilon)) \leq 24r$. The same is true of $\beta_2, \beta_3$.

The next step is to find $\beta_1', \beta_2', \beta_3' \in \mathrm{nCSM}^2(V)$ such that $\operatorname{prank}(\beta_i - \beta_i') \leq 48r$. First note that for bilinear forms we have $\operatorname{arank}(\beta_i) = \operatorname{prank}(\beta_i)$. Thus there is a subspace $U \leq V$ of

codimension at most $24r$ such that $\beta_i - (\beta_i)_{(12)}$ vanishes on $U \times U$. This is the only condition we need to imply that $\beta_i|_U$ is an nCSM. Extending $\beta_i|_U$ to $\beta_i' \in \mathrm{nCSM}^2(V)$ arbitrarily, we have $\mathrm{prank}(\beta_i - \beta_i') \leq 48r$. Finally, by Proposition 6.3.3, there exist non-classical quadratic polynomials $Q_i \in \mathrm{Poly}_{\leqslant 2}(V \to \mathbb{R}/\mathbb{Z})$ such that $d^2 Q_i = \beta_i'$.

We repeat the same averaging and Fourier analysis argument from above to conclude that

$$\left|\mathbb{E}_{x,h_1,h_2,h_3} b''(h_1,h_2,h_3)\omega^{\beta_1'(h_2,h_3)+\beta_2'(h_1,h_3)+\beta_3'(h_1,h_2)}(\partial_{h_1}\partial_{h_2}\partial_{h_3} g)(x)\right| \geq \epsilon p^{-300r},$$

where $b''(h_1,h_2,h_3)$ is a product just of linear phase functions. Specifically let us write $b''(h_1,h_2,h_3) = \omega^{L_1(h_1)+L_2(h_2)+L_3(h_3)}$.

Now define

$$\begin{aligned}
g_{000} &= f\omega^P,\\
g_{001} &= f\omega^{P+Q_1},\\
g_{010} &= f\omega^{P+Q_2},\\
g_{011} &= f\omega^{P+Q_1+Q_2+L_3},\\
g_{100} &= f\omega^{P+Q_3},\\
g_{101} &= f\omega^{P+Q_1+Q_3+L_2},\\
g_{110} &= f\omega^{P+Q_2+Q_3+L_1},\\
g_{111} &= f\omega^{P+Q_1+Q_2+Q_3+L_1+L_2+L_3}.
\end{aligned}$$

One can verify that the inequality above can be rewritten as

$$\begin{aligned}
|\mathbb{E}_{x,h_1,h_2,h_3} g_{000}(x)\overline{g_{001}(x+h_1)g_{010}(x+h_2)}g_{011}(x+h_1+h_2)\overline{g_{100}(x+h_3)}&\\
\cdot g_{101}(x+h_1+h_3)g_{110}(x+h_2+h_3)\overline{g_{111}(x+h_1+h_2+h_3)}\Big| &\geq \epsilon p^{-300r}.
\end{aligned}$$

The left-hand side is in the correct form to apply the Gowers-Cauchy-Schwarz inequality [TZ12, Lemma B.1.(iv)], which implies that $\|f\omega^P\|_{U^3} \geq \epsilon p^{-300r} \geq p^{-O(\log_p(1/\epsilon)^{O(1)})}$. Finishing with a single application of the $U^3$-inverse theorem, we prove the desired result. □

## 6.6 Concluding remarks

The symmetrization results in this chapter are very specific to trilinear forms. We pose the following conjecture, which seems surprisingly difficult and may be of independent interest.

**Conjecture 6.6.1.** *Let $T\colon V^k \to \mathbb{F}_p$ be a $k$-linear form. Suppose that $T$ is close to symmetric in the sense that* $\operatorname{prank}(T - T_\pi) \leq r$ *for all permutations $\pi \in \mathfrak{S}_k$. Does there exists a symmetric $k$-linear form $S\colon V^k \to \mathbb{F}_p$ such that* $\operatorname{prank}(S - T) \leq r'$ *where $r'$ is a polynomial function of $r$?*

Note that this conjecture is only interesting if $p \leq k$, since for $p > k$, the symmetric form $S = \frac{1}{k!}\sum_\pi T_\pi$ can easily be seen to satisfy $\operatorname{prank}(S - T) \leq k!r$.

Combined with the recent work of Gowers and Milićević [GM20], Conjecture 6.6.1 is the missing ingredient which would suffice to give quantitative bounds on the $U^{p+1}$-inverse theorem. For the $U^{k+1}$-inverse theorem with $k > p$ there is additional symmetrization necessary to produce a non-classical symmetric multilinear form from a symmetric form.

**Conjecture 6.6.2.** *Let $T\colon V^k \to \mathbb{F}_p$ be a $k$-linear form. Suppose that $T$ satisfies*

$$\left|\mathbb{E}_{x_1,\dots,x_k} \prod_{I\subseteq[k]} b_I\left(\sum_{i\in I} x_i\right)\omega^{T(x_1,\dots,x_k)}\right| \geq \epsilon$$

*for some 1-bounded functions $b_I\colon V \to \mathbb{C}$. Does there exists $S \in \mathrm{nCSM}^k(V)$ such that* $\operatorname{prank}(S - T) \leq r'$ *where $r'$ is a polynomial function of* $\log(1/\epsilon)$*?*

Combined with the work of Gowers and Milićević and this chapter's integration result, Proposition 6.3.5, this result would give quantitative bounds on the $U^{k+1}$-inverse theorem over $\mathbb{F}_p^n$ for all $k, p$.

Very recently, Milićević [Mil21] disproved Conjecture 6.6.1. However it is still possible that Conjecture 6.6.2 holds, which would be sufficient to give quantitative bounds on the $U^{k+1}$-inverse theorem over $\mathbb{F}_p^n$ for all $k, p$.

# Bibliography


[AB01] Noga Alon and Richard Beigel, *Lower bounds for approximations by low degree polynomials over $\mathbb{Z}/m\mathbb{Z}$*, Proceedings 16th Annual IEEE Conference on Computational Complexity, IEEE, 2001, pp. 184–187.

[ABB21] Ethan Ackelsberg, Vitaly Bergelson, and Andrew Best, *Multiple recurrence and large intersections for abelian group actions*, 2021, arXiv:2101.02811v1.

[ADL+94] Noga Alon, Richard A. Duke, Hanno Lefmann, Vojtěch Rödl, and Raphael Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms **16** (1994), 80–109.

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2000), 451–476.

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron, *Testing Reed-Muller codes*, IEEE Trans. Inform. Theory **51** (2005), 4032–4039.

[AKZ21] Karim Adiprasito, David Kazhdan, and Tamar Ziegler, *On the Schmidt and analytic ranks for trilinear forms*, 2021, arXiv:2102.03659.

[AS08] Noga Alon and Asaf Shapira, *A characterization of the (natural) graph properties testable with one-sided error*, SIAM J. Comput. **37** (2008), 1703–1727.

[BCSX11] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie, *Testing linear-invariant non-linear properties*, Theory Comput. **7** (2011), 75–99.

[Beh46] Felix A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. **32** (1946), 331–332.

[Ber21] Aaron Berger, *Popular differences for corners in abelian groups*, Math. Proc. Cambridge Philos. Soc. **171** (2021), 207–225.

[BFH+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett, *Every locally characterized affine-invariant property is testable*, STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing, ACM, New York, 2013, pp. 429–435.

[BFL12] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett, *Testing low complexity affine-invariant properties*, Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2012, pp. 1337–1355.

[BGS15] Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira, *A unified framework for testing linear-invariant properties*, Random Structures Algorithms **46** (2015), 232–260.

[BHK05] Vitaly Bergelson, Bernard Host, and Bryna Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), 261–303, with an appendix by Imre Ruzsa.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld, *Self-testing/correcting with applications to numerical problems*, Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990), vol. 47, 1993, pp. 549–595.

[Bou86] Jean Bourgain, *A Szemerédi type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math. **54** (1986), 307–316.

[BSST21] Aaron Berger, Ashwin Sah, Mehtaab Sawhney, and Jonathan Tidor, *Non-classical polynomials and the inverse theorem*, Math. Proc. Cambridge Philos. Soc. (2021), 1–13.

[BSST22] Aaron Berger, Ashwin Sah, Mehtaab Sawhney, and Jonathan Tidor, *Popular differences for matrix patterns*, Trans. Amer. Math. Soc. **375** (2022), 2677–2704.

[BTZ10] Vitaly Bergelson, Terence Tao, and Tamar Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of* $\mathbb{F}_p^\infty$, Geom. Funct. Anal. **19** (2010), 1539–1596.

[CF12] David Conlon and Jacob Fox, *Bounds for graph regularity and removal lemmas*, Geom. Funct. Anal. **22** (2012), 1191–1256.

[CF13] David Conlon and Jacob Fox, *Graph removal lemmas*, Surveys in combinatorics 2013, London Math. Soc. Lecture Note Ser., vol. 409, Cambridge Univ. Press, Cambridge, 2013, pp. 1–49.

[Chu11] Qing Chu, *Multiple recurrence for two commuting transformations*, Ergodic Theory Dynam. Systems **31** (2011), 771–792.

[CM21] Alex Cohen and Guy Moshkovitz, *Structure vs. randomness for bilinear maps*, Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC 2021), ACM, New York, 2021, pp. 800–808.

[Deu75] Walter Deuber, *Partition theorems for abelian groups*, J. Combinatorial Theory Ser. A **19** (1975), 95–108.

[DS16] Sebastián Donoso and Wenbo Sun, *A pointwise cubic average for two commuting transformations*, Israel J. Math. **216** (2016), 657–678.

[FSS+20] Jacob Fox, Ashwin Sah, Mehtaab Sawhney, David Stoner, and Yufei Zhao, *Triforce and corners*, Math. Proc. Cambridge Philos. Soc. **169** (2020), 209–223.

[FTZ22] Jacob Fox, Jonathan Tidor, and Yufei Zhao, *Induced arithmetic removal: complexity 1 patterns over finite fields*, Israel J. Math. (2022), 1–38.

[Für95] Zoltán Füredi, *Extremal hypergraphs and combinatorial geometry*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 1343–1352.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron, *Property testing and its connection to learning and approximation*, J. ACM **45** (1998), 653–750.

[GM17] W. Timothy Gowers and Luka Milićević, *A quantitative inverse theorem for the $U^4$ norm over finite fields*, 2017, arXiv:1712.00241.

[GM20] W. Timothy Gowers and Luka Milićević, *An inverse theorem for Freiman multi-homomorphisms*, 2020, arXiv:2002.11667.

[Gow01] W. Timothy Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[Gow07] W. Timothy Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. **166** (2007), 897–946.

[GR71] Ronald L. Graham and Bruce L. Rothschild, *Ramsey's theorem for n-parameter sets*, Trans. Amer. Math. Soc. **159** (1971), 257–292.

[GR11] Oded Goldreich and Dana Ron, *On proximity-oblivious testing*, SIAM J. Comput. **40** (2011), 534–566.

[Gre05a] Ben Green, *Finite field models in additive combinatorics*, Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 1–27.

[Gre05b] Ben Green, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal. **15** (2005), 340–376.

[Gre07] Ben Green, *Montréal notes on quadratic Fourier analysis*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 69–102.

[GS15] Ben Green and Tom Sanders, *Fourier uniformity on subspaces*, 2015, arXiv:1510.08739.

[GT08a] Ben Green and Terence Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), 73–153.

[GT08b] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), 481–547.

[GT09] Ben Green and Terence Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, Contrib. Discrete Math. **4** (2009), 1–36.

[GT10a] Ben Green and Terence Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 261–334.

[GT10b] Ben Green and Terence Tao, *An equivalence between inverse sumset theorems and inverse conjectures for the $U^3$ norm*, Math. Proc. Cambridge Philos. Soc. **149** (2010), 1–19.

[GT10c] Ben Green and Terence Tao, *Linear equations in primes*, Ann. of Math. **171** (2010), 1753–1850.

[GTZ11] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers $U^4$-norm*, Glasg. Math. J. **53** (2011), 1–50.

[GTZ12] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. **176** (2012), 1231–1372.

[GW10] W. Timothy Gowers and Julia Wolf, *The true complexity of a system of linear equations*, Proc. Lond. Math. Soc. (3) **100** (2010), 155–176.

[GW11a] W. Timothy Gowers and Julia Wolf, *Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$*, Geom. Funct. Anal. **21** (2011), 36–69.

[GW11b] W. Timothy Gowers and Julia Wolf, *Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$*, Mathematika **57** (2011), 215–237.

[HHL16] Hamed Hatami, Pooya Hatami, and Shachar Lovett, *General systems of linear forms: equidistribution and true complexity*, Adv. Math. **292** (2016), 446–477.

[HHL19] Hamed Hatami, Pooya Hatami, and Shachar Lovett, *Higher-order Fourier analysis and applications*, Foundations and Trends in Theoretical Computer Science **13** (2019), 247–448.

[Jan20] Oliver Janzer, *Polynomial bound for the partition rank vs the analytic rank of tensors*, Discrete Anal. (2020), Paper No. 7, 18.

[Kov21] Vjekoslav Kovač, *Popular differences for right isosceles triangles*, 2021, arXiv:2101.12714.

[KS08] Tali Kaufman and Madhu Sudan, *Algebraic property testing: the role of invariance*, STOC'08, ACM, New York, 2008, pp. 403–412.

[KSV09] Daniel Král, Oriol Serra, and Lluís Vena, *A combinatorial proof of the removal lemma for groups*, J. Combin. Theory Ser. A **116** (2009), 971–978.

[KSV12] Daniel Kráľ, Oriol Serra, and Lluís Vena, *A removal lemma for systems of linear equations over finite fields*, Israel J. Math. **187** (2012), 193–207.

[LMS11] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky, *Inverse conjecture for the Gowers norm is false*, Theory Comput. **7** (2011), 131–145.

[Lov12] Shachar Lovett, *Equivalence of polynomial conjectures in additive combinatorics*, Combinatorica **32** (2012), 607–618.

[Lov19] Shachar Lovett, *The analytic rank of tensors and its applications*, Discrete Anal. (2019), Paper No. 7, 10.

[Man18] Freddie Manners, *Quantitative bounds in the inverse theorem for the Gowers $U^{s+1}$-norms over cyclic groups*, 2018, arXiv:1811.00718.

[Man21a] Matei Mandache, *A variant of the Corners theorem*, Math. Proc. Cambridge Philos. Soc. **171** (2021), 607–621.

[Man21b] Freddie Manners, *True complexity and iterated Cauchy–Schwarz*, 2021, arXiv:2109.05731.

[Mil19] Luka Milićević, *Polynomial bound for partition rank in terms of analytic rank*, Geom. Funct. Anal. **29** (2019), 1503–1530.

[Mil21] Luka Milićević, *Approximately Symmetric Forms Far From Being Exactly Symmetric*, 2021, arXiv:2112.14755.

[Pre15] Sean Prendiville, *Matrix progressions in multidimensional sets of integers*, Mathematika **61** (2015), 14–48.

[RNS+05] Vojtěch Rödl, Brendan Nagle, Jozef Skokan, Mathias Schacht, and Yoshiharu Kohayakawa, *The hypergraph regularity method and its applications*, Proc. Natl. Acad. Sci. USA **102** (2005), 8109–8113.

[Rot53] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

[RS78] Imre Z. Ruzsa and Endre Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam-New York, 1978, pp. 939–945.

[RS96] Ronitt Rubinfeld and Madhu Sudan, *Robust characterizations of polynomials with applications to program testing*, SIAM J. Comput. **25** (1996), 252–271.

[RS07] Vojtěch Rödl and Mathias Schacht, *Property testing in hypergraphs and the removal lemma [extended abstract]*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 488–495.

[Sam07] Alex Samorodnitsky, *Low-degree tests at large distances*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 506–515.

[San08] Tom Sanders, *A note on Freĭman's theorem in vector spaces*, Combin. Probab. Comput. **17** (2008), 297–305.

[Sha10] Asaf Shapira, *A proof of Green's conjecture regarding the removal properties of sets of linear equations*, J. Lond. Math. Soc. (2) **81** (2010), 355–373.

[She88] Saharon Shelah, *Primitive recursive bounds for van der Waerden numbers*, J. Amer. Math. Soc. **1** (1988), 683–697.

[Spe79] Joel H. Spencer, *Ramsey's theorem for spaces*, Trans. Amer. Math. Soc. **249** (1979), 363–371.

[SSZ21] Ashwin Sah, Mehtaab Sawhney, and Yufei Zhao, *Patterns without a popular difference*, Discrete Anal. (2021), Paper No. 8, 30.

[Tao14] Terence Tao, *A proof of Roth's theorem*, 2014, `https://terrytao.wordpress.com/2014/04/24/a-proof-of-roths-theorem/`.

[Tid21] Jonathan Tidor, *Quantitative bounds for the $U^4$-inverse theorem over low characteristic finite fields*, 2021, arXiv:2109.13108.

[TV06] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.

[TZ10] Terence Tao and Tamar Ziegler, *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, Anal. PDE **3** (2010), 1–20.

[TZ12] Terence Tao and Tamar Ziegler, *The inverse conjecture for the Gowers norm over finite fields in low characteristic*, Ann. Comb. **16** (2012), 121–188.

[TZ20] Jonathan Tidor and Yufei Zhao, *Testing linear-invariant properties*, 2020 IEEE 61st Annual Symposium on Foundations of Computer Science, IEEE Computer Soc., Los Alamitos, CA, 2020, pp. 1180–1190.

[Var59] Panayiotis Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360.

[vdC39] Johannes G. van der Corput, *Über Summen von Primzahlen und Primzahlquadraten*, Math. Ann. **116** (1939), 1–50.

[Vin37] Ivan M. Vinogradov, *The method of trigonometrical sums in the theory of numbers*, Trav. Inst. Math. Stekloff **23** (1937), 109.