Higher-order Fourier Analysis and Applications

Hamed Hatami[1] Pooya Hatami[2] Shachar Lovett[3]

April 17, 2018

[1]McGill University. hatami@cs.mcgill.ca
[2]DIMACS & IAS pooyahat@math.ias.edu
[3]UC San Diego lovett@cs.ucsd.edu

Contents

Chapter 1

Introduction

The purpose of this text is to provide an introduction to the field of higher-order Fourier analysis with an emphasis on its applications to theoretical computer science. Higher-order Fourier analysis is an extension of the classical Fourier analysis. It was initiated by the seminal work of Gowers [35] on a new proof for Szemerédi's theorem, and has been developed by several mathematicians over the past few decades in order to study problems in an area of mathematics called additive combinatorics, which is primarily concerned with linear patterns such as arithmetic progressions in subsets of integers. While most of the developments in additive combinatorics were focused on the group $\mathbb{Z}$, it was quickly noticed that the analogous questions and results for the group $\mathbb{F}_2^n$ are of great importance to theoretical computer scientists as they are related to basic concepts in areas such as property testing and coding theory.

Classical Fourier analysis is a powerful tool that studies functions by expanding them in terms of the Fourier characters, which are "linear phase functions" such as $n \mapsto e^{-\frac{2\pi i}{N}n}$ for the group $\mathbb{Z}_N$, or $(x_1, \ldots, x_n) \mapsto (-1)^{\sum a_j x_j}$ for the group $\mathbb{F}_2^n$. Note that n and $\sum a_j x_j$ are both linear functions. Fourier analysis has been extremely successful in the study of certain linear patterns such as three-term arithmetic progressions. For example, if the number of three-term arithmetic progressions in a subset $A \subseteq \mathbb{Z}_N$ deviates from the expected number of them in a random subset of $\mathbb{Z}_N$ with the same cardinality as A, then A must have significant correlation with a linear phase function. In other words, the characteristic function of A must have a large non-principal Fourier coefficient. Roth [63] used these ideas to show that every subset of integers of positive upper density contains an arithmetic progression of length 3. However, classical Fourier analysis seems to be inadequate in detecting more complex linear patterns such as four-term or longer arithmetic progressions. Indeed, one can easily construct dense sets $A \subseteq \mathbb{Z}_N$ that do not have significant correlation with any linear phase function, and nevertheless do not contain the number of four-term arithmetic progressions that one expects by considering random subsets of the same cardinality. Hence in order to generalize Roth's theorem to arithmetic progressions of arbitrary length, Szemerédi [72, 73] departed from the Fourier analytic approach and appealed to purely combinatorial ideas. However, his proof of this major result, originally conjectured by Erdös and Turán [25], provided poor quantitative bounds on the minimal density that guarantees the existence of the arithmetic progressions of the desired length. Later Furstenberg [29] developed an ergodic-theoretic framework and gave a new proof for Szemerédi's theorem, but his proof was still qualitative. His theory is further developed by - to name a few - Host, Kra, Ziegler, Bergelson, Tao (See e.g. [49], [84], and [9, 78]), and there are important parallels between this theory and higher-order Fourier analysis. Indeed some of the terms that are commonly used in higher-order Fourier analysis such as "phase functions" or "factors" are ergodic theoretic terms.

Generalizing Roth's original proof and obtaining good quantitative bounds for Szemerédi's theorem remained a challenge until finally Gowers [35] discovered that the essential idea to overcome the obstacles described above is to consider higher-order phase functions. His proof laid the foundation for the area of higher-order Fourier analysis, where one studies a function by approximating it by a linear combination of few higher-order phase functions. Although the idea of using higher-order phase functions already appears in Gowers's work [35], it was not until more than fifteen years later that some of the major technical difficulties

in achieving a satisfactory theory of higher-order Fourier analysis have been resolved. By now, due to great contributions by prominent mathematicians such as Gowers, Green, Tao, Szegedy, Host, Kra and Ziegler (See [71] and [76] and the references there), there is a deep understanding of qualitative aspects of this theory. However, despite these major breakthroughs, still very little is known from a quantitative perspective as many of the proofs are based on soft analytic techniques, and obtaining efficient bounds is one of the major challenges in this area.

This survey will emphasize the applications of the theory of higher-order Fourier analysis to theoretical computer science, and to this end, we will present the foundations of this theory through such applications, in particular to the area of property testing. In the early nineties, it was noticed in [19, 5] that Fourier analysis can be used to design a very efficient algorithm that distinguishes linear functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ from functions that are far from being linear. This initiated the area of property testing, the study of algorithms that query their input a very small number of times and with high probability decide correctly whether their input satisfies a given property or is "far" from satisfying that property. It was soon noticed that generalizing the linearity test of [19, 5] to other properties such as the property of being a quadratic polynomial requires overcoming the same obstacles that one faces in an attempt to generalize Fourier analytic study of three-term arithmetic progressions to four-term arithmetic progressions. Hence in parallel to additive combinatorics, theoretical computer scientists have also been working on developing tools in higher-order Fourier analysis to tackle such problems. In fact some of the most basic results, such as the inverse theorem for the Gowers U^3 norm for the group $\mathbb{F}_2^n$, were first proved by Samorodnitsky [66] in the context of property testing for quadratic polynomials.

In Part I we discuss the linearity test of [19] and its generalization to higher degree polynomials. We will see how this naturally necessitates the development of a theory of higher-order Fourier analysis. In Part II we present the fundamental results of the theory of higher-order Fourier analysis. Since we are interested in the applications to theoretical computer science, we will only consider the group $\mathbb{F}_p^n$ where p is a fixed prime, and asymptotics are as n tends to infinity. Higher-order Fourier analysis for the group $\mathbb{Z}_N$, which is of more interest for number theoretic applications, shares the same basic ideas but differs on some technical aspects. For this group, the higher order phase functions, rather than being exponentials of polynomials, are the so called nilsequences. We refer the interested reader to [76] for more details. In Part III we use the tools developed in Part II to prove some general results about property testing for algebraic properties.

Throughout most of the text, we will consider fields of constant prime order, namely $\mathbb{F} = \mathbb{F}_p$ where p is a constant, and study functions from $\mathbb{F}_p^n$ to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{F}_p$ when n is growing. Our choice is mainly for simplicity of exposition, as there have been recent works that extend several of the tools from higher-order Fourier analysis to large or non-prime fields. We refer the interested reader to a paper of Bhattacharyya et. al. [11] for treatment of non-prime fields. In Chapter 8 we will discuss work of Bhowmick and Lovett [18] considering the case $\mathbb{F}_p^n$ when p is allowed to grow as a function of n.

Part I

Low Degree Testing

We start Part I by presenting the Fourier analytic linearity test of [19] in Chapter 2. In Chapter 3, we show how this test can be generalized from linear functions to polynomials of higher degree. The focus will be on the so-called "99% regime", where the test is designed to distinguish polynomials of a given degree d from the functions that are somewhat far from them, i.e. functions that do not match with any polynomial of degree d on more than $1-\epsilon$ fraction of the points for a small $\epsilon > 0$. In Chapter 4, we turn our attention to the more challenging case of the so-called "1% regime", where the test is supposed to distinguish functions that have some noticeable correlation with polynomials of degree d (i.e. more than a small constant $\epsilon > 0$) from functions that, similar to a typical random function, have essentially no correlation with any polynomial of degree d. Chapter 4 is focused on the case of quadratic polynomials. Polynomials of higher degrees are more complex and are discussed in Part II of the survey.

Chapter 2

Fourier analytic property testing

The field of property testing is the study of algorithms that query their input a very small number of times and with high probability decide correctly whether their input satisfies a given property or is "far" from satisfying that property. In this chapter we survey some of the earliest yet important results in the area of property testing that use classical Fourier analysis to design efficient tests for certain algebraic properties. In particular, we analyze the linearity test of [19] which together with [5] are often considered as the earliest explicit examples of property testing. These results inspired [65, 30] to formally define this field and initiate a systematic study of testable properties. A property is called *testable*, or sometimes *strongly testable* or *locally testable*, if the number of queries can be made independent of the size of the object without affecting the correctness probability. Perhaps surprisingly, it has been found that many natural properties satisfy this strong requirement. In this survey we will only discuss algebraic properties however we refer the reader to [27, 64, 62, 70] for a general overview of this field.

Let $[R]$ denote the set $\{1, \ldots, R\}$. Let $\mathbb{F}$ denote a finite field, and $\mathbb{F}^n$ denote an n-dimensional vector space over $\mathbb{F}$. Given two functions $f, g : \mathbb{F}^n \to [R]$, their *distance* is the fraction of points on which they disagree,

$$\mathrm{dist}(f, g) := \Pr_{x \in \mathbb{F}^n}[f(x) \neq g(x)],$$

where here and throughout $\Pr_{x \in \mathbb{F}^n}[\cdot]$ means the probability of an event given a uniform choice of $x \in \mathbb{F}^n$.

A *property* is a subset of functions, $\mathcal{P} \subset \{f : \mathbb{F}^n \to [R] : n \in \mathbb{Z}_{\geq 0}\}$. The distance of $f : \mathbb{F}^n \to [R]$ is *ε-far* from $\mathcal{P}$ is its minimal distance to a function in $\mathcal{P}$,

$$\mathrm{dist}(f, \mathcal{P}) := \min_{g \in \mathcal{P}} \mathrm{dist}(f, g),$$

where the minimum is taken over all functions $g \in \mathcal{P}$ defined over the same domain as f (namely, $g : \mathbb{F}^n \to [R]$). If $\mathrm{dist}(f, \mathcal{P}) \leq \varepsilon$ then we say that f is *ε-close* to $\mathcal{P}$, and otherwise we say that it is *ε-far* from $\mathcal{P}$.

Definition 2.0.1 (Testability with one-sided error)**.** *A property $\mathcal{P}$ is said to be* testable *with one-sided error if there are functions $q : (0,1) \to \mathbb{Z}_{>0}$, $\delta : (0,1) \to (0,1)$, and an algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{F}^n \to [R]$, makes at most $q(\varepsilon)$ queries to the oracle for f, always accepts if $f \in \mathcal{P}$ and rejects with probability at least $\delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. If, furthermore, q is a constant function, then $\mathcal{P}$ is said to be* proximity-obliviously testable (PO testable)*.*

In the above definition we only allow one-sided error. These are algorithms that do not err when the function satisfies the property. However, it is also natural to study tests with two-sided error, and indeed in Chapter 16 we will see examples of such tests.

The term proximity-oblivious testing was coined by Goldreich and Ron in [33]. Indeed as we will see in Section 2.1, the famous linearity test of Blum, Luby and Rubinfeld [19] is an example of a proximity oblivious test. It shows that the linearity of a function $f : \mathbb{F}^n \to \mathbb{F}$ is testable using only 3 queries. This test accepts if f is linear and rejects with probability $\Omega(\varepsilon)$ if f is ε-far from linear.

2.1 Linearity Testing

Linearity testing is a cornerstone of property testing, and in particular the birthplace of algebraic property testing. There are two equivalent definitions for linear functions. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *linear* if

(1) **Global definition.** $f(x) = \sum_{i=1}^n a_i x_i$ for some $a_i \in \mathbb{F}_2$.

(2) **Local definition.** For all $x, y \in \mathbb{F}_2^n$, $f(x+y) = f(x) + f(y)$.

Our goal will to be determine whether a given function f is linear, or whether it is far from it. The distance of a function f to the family of linear functions is given by

$$\min_{g:\mathbb{F}_2^n \to \mathbb{F}_2 \text{ linear}} \Pr[f(x) \neq g(x)],$$

where the probability is over a uniform choice of $x \in \mathbb{F}_2^n$. The BLR tester of Blum, Luby and Rubinfeld [19] uses the local definition of linearity: it chooses random x and y accepts the function if $f(x+y) = f(x)+f(y)$.

BLR test. With query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Choose $x, y \in \mathbb{F}_2^n$ independently and uniformly at random.
2. Query $f(x), f(y), f(x+y)$.
3. **Accept** if $f(x) + f(y) = f(x+y)$ and **Reject** otherwise.

It is clear that the tester always accepts linear functions. It turns out that the probability that the above tester rejects is directly related to the distance of f from linearity. We follow the proof of [8], which simplifies and improves the original analysis in [19]. The proof relies on classical Fourier analysis. As there are a number of good books on the topic, for example [60], we assume basic familiarity with classical Fourier analysis (concretely: the definition of Fourier coefficients, the Fourier inversion formula and Parseval's identity).

Theorem 2.1.1 ([8]). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function whose distance to linear functions is ε. Then the BLR test rejects f with probability between ε and 5ε.*

Proof. Let $F(x) = (-1)^{f(x)}$. We can express the acceptance probability of the test in terms of the Fourier coefficients of F, which to recall are given by the Fourier decomposition $F(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)(-1)^{\langle \alpha, x \rangle}$. Below, all probabilities and expectations are taken over uniformly chosen $x, y \in \mathbb{F}_2^n$.

$$\begin{aligned}
\Pr_{x,y}[f(x) + f(y) = f(x+y)] &= \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y}\left[F(x)F(y)F(x+y)\right] \\
&= \frac{1}{2} + \frac{1}{2} \sum_{\alpha,\beta,\gamma \in \mathbb{F}_2^n} \widehat{F}(\alpha)\widehat{F}(\beta)\widehat{F}(\gamma)\mathbb{E}_{x,y}\left[(-1)^{\langle \alpha, x\rangle + \langle \beta, y\rangle + \langle \gamma, x+y\rangle}\right] \\
&= \frac{1}{2} + \frac{1}{2} \sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^3.
\end{aligned}$$

Here, we used the fact that

$$\mathbb{E}_{x,y}\left[(-1)^{\langle \alpha, x\rangle + \langle \beta, y\rangle + \langle \gamma, x+y\rangle}\right] = \mathbb{E}_x\left[(-1)^{\langle \alpha+\gamma, x\rangle}\right]\mathbb{E}_y\left[(-1)^{\langle \beta+\gamma, x\rangle}\right]$$

is equal to 1 if $\alpha = \beta = \gamma$, and equal to 0 otherwise.

To prove the upper bound on the acceptance probability, we can bound

$$\Pr_{x,y}[f(x) + f(y) = f(x+y)] \leq \frac{1}{2}\left(1 + \max_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)\right),$$

where we applied Parseval's identity $\sum \widehat{F}(\alpha)^2 = \mathbb{E}[F(x)^2] = 1$. If f has distance ε from linear functions, then for any linear function $g(x) = \langle x, \alpha \rangle$ it holds that $\Pr[f(x) \neq g(x)] \geq \varepsilon$. Then for any $\alpha \in \mathbb{F}_2^n$,

$$\hat{F}(\alpha) = \mathbb{E}_x[F(x)(-1)^{\langle x,\alpha \rangle}] = \Pr[f(x) = g(x)] - \Pr[f(x) \neq g(x)] \leq 1 - 2\varepsilon.$$

We thus obtain that

$$\Pr_{x,y}[f(x) + f(y) = f(x+y)] \leq 1 - \varepsilon.$$

For the lower bound, if the distance of f to linear functions is ε, then there exists $\alpha^* \in \mathbb{F}_2^n$ such that $\widehat{F}(\alpha^*) = 1 - 2\varepsilon$. Thus $\widehat{F}(\alpha^*)^3 \geq 1 - 6\varepsilon$ and the contribution of all other terms is bounded by $\sum_{\alpha \neq \alpha^*} |\widehat{F}(\alpha)|^3 \leq \sum_{\alpha \neq \alpha^*} |\widehat{F}(\alpha)|^2 = 1 - \widehat{F}(\alpha^*)^2 \leq 4\varepsilon$. So

$$\Pr_{x,y}[f(x) + f(y) = f(x+y)] = \frac{1}{2}\left(1 + \widehat{F}(\alpha^*)^3 + \sum_{\alpha \neq \alpha^*} \widehat{F}(\alpha)^3\right) \geq 1 - 5\varepsilon.$$

□

2.2 Testing for affine linearity

A very related notion to linear functions is that of being an affine linear function. Again, there are two equivalent definitions. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *affine linear* if

(1) **Global definition.** $f(x) = a_0 + \sum_{i=1}^n a_i x_i$ for some $a_i \in \mathbb{F}_2$.

(2) **Local definition.** For all $x, y, z \in \mathbb{F}_2^n$, $f(x) + f(y) + f(z) = f(x + y + z)$.

Let $\mathcal{P}_1$ denote the family of affine linear functions (equivalently, functions whose degree as an $\mathbb{F}_2$-polynomial is at most 1). A very similar test to the BLR test can detect if a function is affine linear or far from it.

BLR test for affine linearity. With query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Choose $x, y, z \in \mathbb{F}_2^n$ independently and uniformly at random.
2. Query $f(x), f(y), f(z), f(x + y + z)$.
3. **Accept** if $f(x) + f(y) + f(x) = f(x + y + z)$ and **Reject** otherwise.

Theorem 2.2.1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function whose distance to affine linear functions is ε. Then the BLR test for affine linearity rejects f with probability between ε and 6ε.*

Proof. Let $F(x) = (-1)^{f(x)}$. We can express the acceptance probability of the test in terms of the Fourier coefficients of F as

$$\begin{aligned}\Pr_{x,y,z}[f(x) + f(y) + f(z) = f(x + y + z)] &= \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y,z}\left[F(x)F(y)F(z)F(x + y + z)\right] \\ &= \frac{1}{2} + \frac{1}{2}\sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^4.\end{aligned}$$

For the upper bound, we can apply Parseval's identity and bound

$$\Pr_{x,y,z}[f(x) + f(y) + f(z) = f(x + y + z)] \leq \frac{1}{2}\left(1 + \max_{\alpha \in \mathbb{F}_2^n} |\widehat{F}(\alpha)|^2\right) \leq \frac{1}{2}\left(1 + \max_{\alpha \in \mathbb{F}_2^n} |\widehat{F}(\alpha)|\right).$$

If f has distance ε from affine linear functions, then for any linear function $g(x) = \langle x, \alpha \rangle$ it holds that $\Pr[f(x) \neq g(x)] \geq \varepsilon$ and $\Pr[f(x) \neq g(x) + 1] \geq \varepsilon$. Then for any $\alpha \in \mathbb{F}_2^n$,

$$\left|\widehat{F}(\alpha)\right| = \left|\mathbb{E}_x[F(x)(-1)^{\langle x,\alpha\rangle}]\right| \leq 1 - 2\varepsilon.$$

Thus

$$\Pr_{x,y,z}[f(x) + f(y) + f(z) = f(x + y + z)] \leq 1 - \varepsilon.$$

For the lower bound, if the distance of f to affine linear functions is ε, then there exists $\alpha^* \in \mathbb{F}_2^n$ such that $|\widehat{F}(\alpha^*)| = 1 - 2\varepsilon$. By the same analysis as that of linearity testing, $\widehat{F}(\alpha^*)^4 \geq 1 - 8\varepsilon$ and $\sum_{\alpha \neq \alpha^*} \widehat{F}(\alpha)^4 \leq 4\varepsilon$. So

$$\Pr_{x,y,z}[f(x) + f(y) + f(z) = f(x + y + z)] = \frac{1}{2}\left(1 + \widehat{F}(\alpha^*)^4 + \sum_{\alpha \neq \alpha^*} \widehat{F}(\alpha)^4\right) \geq 1 - 6\varepsilon.$$

□

2.3 Limitations of Fourier analysis

The analysis of linearity testing makes use of the fact that functions that are close to linear, must have a large Fourier coefficient; and those that are close to affine linear, must have a large Fourier coefficient in absolute value. Unfortunately, this property does not carry over if we are interested in testing for higher-degree behavior, such as quadratic polynomials. As the following example shows, there are quadratic polynomials which have negligible Fourier coefficients. This will necessitate the introduction of higher-order analogs of Fourier analysis.

Claim 2.3.1. *Assume n is even and consider the quadratic polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ defined as*

$$f(x) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n.$$

Let $F(x) = (-1)^{f(x)}$. Then $|\widehat{F}(\alpha)| = 2^{-n/2}$ for all $\alpha \in \mathbb{F}_2^n$.

Proof. Fix $\alpha \in \mathbb{F}_2^n$. Then

$$\widehat{F}(\alpha) = \mathbb{E}_x\left[(-1)^{f(x)+\langle x,\alpha\rangle}\right] = \prod_{i=1}^{n/2} \mathbb{E}_{x_{2i-1},x_{2i}}\left[(-1)^{x_{2i-1}x_{2i}+\alpha_{2i-1}x_{2i-1}+\alpha_{2i}x_{2i}}\right].$$

The proof follows as for any $a, b \in \mathbb{F}_2$,

$$\mathbb{E}_{x,y\in\mathbb{F}_2}\left[(-1)^{xy+ax+by}\right] = (-1)^{ab}\mathbb{E}_{x,y\in\mathbb{F}_2}\left[(-1)^{(x+a)(y+b)}\right] = (-1)^{ab}\mathbb{E}_{x,y\in\mathbb{F}_2}\left[(-1)^{xy}\right] = \frac{(-1)^{ab}}{2}.$$

Thus,

$$\widehat{F}(\alpha) = 2^{-n/2} \cdot (-1)^{\alpha_1\alpha_2+\ldots+\alpha_{n-1}\alpha_n}.$$

□

Chapter 3

Low-degree Tests, the 99% Regime

In Section 2.1 we showed that linearity can be tested by a very simple test: sample a random pair $x, y \in \mathbb{F}_2^n$ and check if $f(x+y) = f(x) + f(y)$. This tempts one to suspect that similar tests would perform well even for more complex properties which similarly have a local definition. One such property is that of being a low-degree polynomial. Given query access to a function, we are interested in testing whether the function is a low-degree polynomial or far from any low-degree polynomial. The testability of low-degree polynomials was first proved by Alon et al. [2], where they showed that the local characterization of low-degree polynomials can be used to design a natural test for this task. The analysis was sharpened by Bhattacharyya et al. [17]. In this chapter, we focus on the "99% regime", where the goal is to distinguish functions very close to low-degree polynomials, from functions which are somewhat far from low-degree polynomials.

3.1 Basic properties of low-degree polynomials

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$, and $d \geq 1$ be an integer. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a polynomial of degree (at most) d if one of the following two equivalent conditions hold:

(1) **Global definition.** $f(x) = \sum_{S \subset [n], |S| \leq d} a_S \prod_{i \in S} x_i$ for some $a_S \in \mathbb{F}_2$.

(2) **Local definition.** For all $x, y_1, \cdots, y_{d+1} \in \mathbb{F}_2^n$, it holds that $\sum_{S \subseteq [d+1]} f\left(x + \sum_{i \in S} y_i\right) = 0$.

We denote by $\mathcal{P}_d$ the family of all polynomials of degree *at most* d. In order to prove that the two definitions for degree d polynomials agree, we need to first prove a few basic facts about polynomials.

A very useful notion is that of directional derivatives. For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $y \in \mathbb{F}_2^n$, the directional derivative of f in direction y, denoted $D_y f : \mathbb{F}_2^n \to \mathbb{F}_2$, is defined as

$$D_y f(x) := f(x+y) + f(x).$$

Note that D_y is a linear operator, that is $D_y(f+g) = D_y f + D_y g$. Iterated derivatives are defined as $D_{y_1,\ldots,y_k} f := D_{y_1} \ldots D_{y_k} f$. It is straightforward to verify that

$$D_{y_1,\ldots,y_k} f(x) = \sum_{S \subseteq [k]} f\left(x + \sum_{i \in S} y_i\right).$$

Then, the local definition of f being a degree d polynomial is equivalent to $D_{y_1,\ldots,y_{d+1}} f = 0$ for all $y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$. We next prove some facts about derivatives.

Claim 3.1.1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d (according to the global definition). Then $D_y f$ is a polynomial of degree at most $d-1$ (according to the global definition).*

Proof. It suffices to prove the claim for monomials of degree d. Let $m(x) = \prod_{i=1}^{d} x_i$. Then

$$D_y m(x) = m(x+y) + m(x) = \prod_{i=1}^{d}(x_i + y_i) + \prod_{i=1}^{d} x_i = \sum_{S \subsetneq [d]} \prod_{i \in S} x_i \prod_{i \notin S} y_i.$$

All the x-monomials in $D_y m$ have degree at most $d-1$, and hence $D_y m$ is a polynomial of degree at most $d-1$. □

Claim 3.1.2. *Let $m(x) = \prod_{i=1}^{d} x_i$. Then for any $y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$,*

1. *$D_{y_1,\ldots,y_d} m(x) = \sum_\pi \prod_{i=1}^{d} y_{i,\pi(i)}$, where π ranges over all permutations of $\{1,\ldots,d\}$. In particular, it is independent of x.*
2. *$D_{y_1,\ldots,y_{d+1}} m = 0$.*

Proof. The second claim follows immediately from the first, as any derivative of a constant function is zero. We prove the first claim by induction on d. We have that

$$D_{y_1,\ldots,y_d} m(x) = D_{y_1,\ldots,y_{d-1}} D_{y_d} m(x) = D_{y_1,\ldots,y_{d-1}} \left(\sum_{S \subsetneq [d]} \prod_{i \in S} x_i \prod_{i \notin S} y_{d,i} \right).$$

All the monomials in $D_{y_d} m(x)$ of degree less than $d-1$ will be annihilated by taking $d-1$ derivatives. Hence

$$D_{y_1,\ldots,y_d} m(x) = D_{y_1,\ldots,y_{d-1}} \left(\sum_{j \in d} y_{d,j} \prod_{i \in [d] \setminus \{j\}} x_i \right).$$

By induction, we have for any $j \in [d]$ that

$$D_{y_1,\ldots,y_{d-1}} \left(\prod_{i \in [d] \setminus \{j\}} x_i \right) = \sum_\sigma \prod_{i \in [d] \setminus \{j\}} y_{i,\sigma(i)},$$

where σ enumerates all one-to-one functions from $[d-1]$ to $[d] \setminus \{j\}$. The claim follows by linearity. □

We now prove that the global and local definitions for being a polynomial of degree at most d are equivalent.

Lemma 3.1.3. *The global and local definitions of being a degree d polynomial are equivalent.*

Proof. Let $f(x)$ be a polynomial which has degree exactly d according to the global definition. We need to show that its degree according to the local definition is also exactly d. That is, we need to show that $D_{y_1,\ldots,y_{d+1}} f = 0$ for all $y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$, but that $D_{y_1,\ldots,y_d} f \neq 0$ for some $y_1, \ldots, y_d \in \mathbb{F}_2^n$.

The first claim follows immediately from Claim 3.1.1. For the latter claim, let $f(x) = \sum_{S \subset [n], |S| \leq d} a_S \prod_{i \in S} x_i$ where $a_S \neq 0$ for some S with $|S| = d$. By Claim 3.1.2,

$$D_{y_1,\ldots,y_d} f(x) = \sum_{S \subset [n], |S| = d} a_S \sum_\pi \prod_{i=1}^{d} y_{i,\pi(i)}.$$

This is a nonzero polynomial (according to the global definition) in the variables $\{y_{i,j} : i \in [d], j \in [n]\}$, as all the monomials $\prod_{i=1}^{d} y_{i,\pi(i)}$ are distinct and some have nonzero coefficient. Hence, there must exist a value for $y_1, \ldots, y_d$ which makes it nonzero. For example, if $S = \{j_1, \ldots, j_d\}$ satisfies $a_S \neq 0$, then we can choose $y_{i,j_i} = 1$ and $y_{i,j} = 0$ for all $j \neq j_i$. □

We will also need the basic fact that the minimal distance of $\mathcal{P}_d$ is 2^{-d}. For a proof see e.g. [59].

Claim 3.1.4. *Let $f, g \in \mathcal{P}_d$ be distinct polynomials of degree d. Then $\Pr[f(x) \neq g(x)] \geq 2^{-d}$.*

3.2 Low-degree testing

The question of testing low-degree polynomials was first studied by Alon, Kaufman, Krivelevich, Litsyn and Ron [2], where they showed that an appropriate extension of the linearity test also works for low-degree polynomials. The original test defined in [2] is specific to polynomials of degree at most d with $f(0) = 0$, but with minor changes it extends to the more natural family of all polynomials of degree at most d, which is the test presented below.

AKKLR(d) **test.** With query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Choose $x, y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$ independently and uniformly at random.
2. Query $f(x + \sum_{i \in S} y_i)$ for all $S \subseteq [d+1]$.
3. **Accept** if $\sum_{S \subseteq [d+1]} f(x + \sum_{i \in S} y_i) = 0$ and **Reject** otherwise.

The reader can verify that this is indeed an extension of the BLR test for affine linear functions, which correspond to $d = 1$. There, we tested if $f(x) + f(y) + f(z) = f(x + y + z)$ for uniformly chosen $x, y, z \in \mathbb{F}_2^n$. Here, we test that $f(x) + f(x + y_1) + f(x + y_2) + f(x + y_1 + y_2) = 0$ for uniformly chosen $x, y_1, y_2 \in \mathbb{F}_2^n$. The tests are identical as can be seen by setting $y_1 = x + y, y_2 = x + z$. We denote by $\delta_d(f)$ the distance of a function f from degree d polynomials,

$$\delta_d(f) := \mathrm{dist}(f, \mathcal{P}_d).$$

Theorem 3.2.1 ([2])**.** *The* AKKLR(d) *test rejects every function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ with probability at least $\Omega\left(\frac{\delta_d(f)}{d \cdot 2^d}\right)$.*

If we want to reject functions which are ε-far from $\mathcal{P}_d$ with constant probability, then we can simply repeat the basic AKKLR(d) test $O(\frac{d}{\epsilon} \cdot 2^d)$ times. This gives a tester with query complexity $O(\frac{d}{\epsilon} \cdot 4^d)$. It was shown in [2] that any tester for $\mathcal{P}_d$ must make at least $\Omega(2^d + \frac{1}{\epsilon})$ queries, leaving a quadratic gap between their upper and lower-bounds. More recently, Bhattacharyya, Kopparty, Schoenebeck, Sudan and Zuckerman [17] gave a tighter analysis of this tester and closed this gap.

Theorem 3.2.2. *[17] Fix $1 \leq d \leq n$. For all functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr\left[\textit{AKKLR}(d) \textit{ test rejects } f\right] \geq c \cdot \min\{2^d \delta_d(f), 1\},$$

where $c > 0$ is an absolute constant.

Given the improved analysis, if one wishes to reject functions which are ϵ-far from $\mathcal{P}_d$ with constant probability, a tester can repeat the AKKLR(d) test only $O(1 + \frac{1}{\epsilon 2^d})$ times, obtaining an asymptotically optimal query complexity of $O(2^d + \frac{1}{\epsilon})$. We present the analysis of [17] in the next section.

3.3 Analysis of the AKKLR test

In this section we present a proof of Theorem 3.2.2. First, we describe a variant of the test which is easier to analyze, and is essentially equivalent to the AKKLR(d) test.

$(d+1)$**-flat test.** With query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Pick A, a random affine $(d+1)$-dimensional subspace of $\mathbb{F}_2^n$.
2. Query $f|_A$, the restriction of f to A.
3. Reject if $f|_A$ is not a degree d polynomial, or equivalently if $\sum_{x \in A} f(x) \neq 0$.

Theorem 3.2.2 follows from the following theorem analyzing the $(d+1)$-flat test.

Theorem 3.3.1. *Fix $1 \le d \le n$. For all functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr\left[(d+1)\text{-flat test rejects } f\right] \ge c \min\{2^d \delta_d(f), 1\},$$

where $c > 0$ is an absolute constant.

First we show how Theorem 3.2.2 follows from Theorem 3.3.1.

Proof of Theorem 3.2.2 given Theorem 3.3.1. We may assume that $n > d+1$. Let $y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$ be uniformly chosen. The probability that $y_1, \ldots, y_{d+1}$ are linearly independent is at least $1 - 2^{d+1-n}$. Conditioned on this event, they span a uniform $(d+1)$-flat. Hence,

$$\Pr[\text{AKKLR}(d) \text{ test rejects } f] \ge \Pr[(d+1)\text{-flat test rejects } f](1 - 2^{d+1-n}) \ge (c/2)\min(2^d \delta_d(f), 1).$$

□

The proof of Theorem 3.3.1 is split into two parts, based on proximity of f to degree d polynomials. We first analyze the case where $\delta_d(f)$ is small.

Lemma 3.3.2 (Small proximity to degree d polynomials)**.** *For any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr\left[(d+1)\text{-flat test rejects } f\right] \ge 2^{d+1}\delta_d(f)(1 - 2^{d+1}\delta_d(f)).$$

In particular, if $\delta_d(f) \le 2^{-(d+2)}$, then

$$\Pr\left[(d+1)\text{-flat test rejects } f\right] \ge 2^d \delta_d(f).$$

Proof. The main idea is to lower-bound the probability that in the $(d+1)$-flat A, there is exactly one point where f and the closest polynomial of degree d differ. Note that in this case, the tester rejects f. Let $g \in \mathcal{P}_d$ be a polynomial of degree d such that $\Pr[f(x) = g(x)] = \delta_d(f)$. Consider a random $(d+1)$-dimensional affine subspace A, generated by picking a random $x \in \mathbb{F}_2^n$ and a full-rank matrix $M \in \mathbb{F}_2^{n \times (d+1)}$ and letting $A = x + M\mathbb{F}_2^{d+1} = \{x + My \mid y \in \mathbb{F}_2^{d+1}\}$.

For $a \in \mathbb{F}_2^{d+1}$, let E_a be the event that $f(x+Ma) \ne g(x+Ma)$, and let F_a be the event that $f(x+Ma) \ne g(x+Ma)$ and for every $b \ne a$, $f(x+Mb) = g(x+Mb)$. Note that if F_a occurs then the tester rejects f, and that the events F_a are pairwise disjoint. Hence

$$\Pr\left[(d+1)\text{-flat test rejects } f\right] \ge \Pr[\cup F_a] = \sum_{a \in \mathbb{F}_2^{d+1}} \Pr[F_a].$$

Next, for $a \in \mathbb{F}_2^{d+1}$ we have that

$$\Pr[F_a] \ge \Pr[E_a] - \sum_{b \ne a} \Pr[E_a \cap E_b].$$

Moreover $\Pr[E_a] = \delta_d(f)$ and $\Pr[E_a \cap E_b] = \Pr[E_a]\Pr[E_b] = \delta_d(f)^2$, as the events E_a, E_b are pairwise independent (since M has full rank). So

$$\Pr\left[(d+1)\text{-flat test rejects } f\right] \ge 2^{d+1}\delta_d(f)(1 - 2^{d+1}\delta_d(f)).$$

□

So, from now on we consider the case where f is at least $2^{-(d+2)}$-far from $\mathcal{P}_d$, and show that in such a case, the $(d+1)$-flat test rejects f with constant probability (independent of d, n). It will be useful to introduce yet another variant of the tester, called the k-flat test, for $k \ge d+1$, where the tester first samples a k-flat and then tests if f is a degree d polynomial when restricted to A.

k-flat test. With query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Pick A, a random affine k-dimensional subspace of $\mathbb{F}_2^n$.
2. Query $f|_A$, the restriction of f to A.
3. Reject if $f|_A$ is not a degree d polynomial.

The following lemma allows us to analyze the rejection probability of the k-flat test for $k = d + c$ where $c = O(1)$.

Lemma 3.3.3. *For every n and $k \geq k' \geq d + 1$, and every $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we have*

$$\Pr[k'\text{-flat test rejects } f] \geq \Pr[k\text{-flat test rejects } f] \cdot 2^{-(k-k')}.$$

Proof. It suffices to prove the lemma for $k' = k - 1$, as the general case follows by an iterative application. Let A be a random k-flat and let A' be a random $k-1$ flat inside A. We will show that if the k-flat tester rejects f, namely if $f|_A$ has degree $> d$, then with probability at least $1/2$ it holds that $f|_{A'}$ also has degree $> d$, and hence reject by the $(k-1)$-flat tester.

So, fix A and assume towards contradiction that $f|_A$ has degree $> d$, but that for a strict majority of $(k-1)$-flats (namely, hyperplanes) $H \subset A$ it holds that $f|_H$ has degree $\leq d$. First, this implies that there must exist two disjoint hyperplanes H', H'' such that $f|_{H'}, f|_{H''}$ have degree d. By interpolation, this implies that $f|_A$ has degree at most $d+1$. Decompose $f|_A(x) = f_{d+1}(x) + f_{\leq d}(x)$ where $f_{d+1}(x)$ is the homogeneous part of f of degree $d+1$. On any hyperplane H where $f|_H$ has degree $\leq d$ it then holds that $f_{d+1}|_H$ has degree $\leq d$. The number of such hyperplanes is by assumption more than half of all hyperplanes, namely $> \frac{1}{2}2(2^k - 1)$, which is at least 2^k. Hence, there must exist k linearly independent hyperplanes $H_1, \ldots, H_k$ on which $f|_{H_i}$ has degree $\leq d$. We will show that this implies that f_{d+1} has degree $\leq d$, which implies that $f|_A$ has degree $\leq d$, contradicting our assumption. To see that, apply an affine linear transformation mapping H_i to $\{x : x_i = 0\}$. Then f_{d+1} has degree $\leq d$ whenever we set some variable $x_i = 0$. As $k \geq d + 2$, this can only happen if f_{d+1} has degree $\leq d$. □

So, it suffices to prove that if $\delta_d(f) \geq 2^{-(d+2)}$ then the k-flat tester rejects it with constant probability for some $k = d + O(1)$. This is exactly what the next lemma shows.

Lemma 3.3.4. *For any $0 < \beta < 1/24$ there exist absolute constants $\gamma, \varepsilon, c > 0$ such that the following holds for any $n \geq k \geq d + c$. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function with $\delta_d(f) \geq \beta 2^{-d}$. Then*

$$\Pr[k\text{-flat test rejects } f] \geq \varepsilon + \gamma \cdot 2^{d-n}.$$

Proof. The proof will require choices of $\beta < 1/24$, $\varepsilon < 1/8$, $\gamma \geq 72$ and $2^c \geq \max\{\frac{4\gamma}{1-8\varepsilon}, \frac{\gamma}{1-\varepsilon}, \frac{2}{\beta}\}$. Fix $k \geq d + c$. We apply induction on $n \geq k$. For the base case of $n = k$, the k-flat test rejects f with probability 1. By our choice of parameters, $1 > \varepsilon + \gamma \cdot 2^{d-n}$.

Let $\mathcal{H}$ denote the set of all hyperplanes of $\mathbb{F}_2^n$ and let $N := |\mathcal{H}| = 2(2^n - 1)$. Also, let $\mathcal{H}^*$ be the set of all hyperplanes $A \in \mathcal{H}$ such that $\delta_d(f|_A) < \beta 2^{-d}$, and let $K := |\mathcal{H}^*|$. We have

$$\Pr[k\text{-flat test rejects } f] = \mathbb{E}_{A \in \mathcal{H}} \Pr[k\text{-flat test rejects } f|_A].$$

By the induction hypothesis for every $A \in \mathcal{H} \setminus \mathcal{H}^*$, the k-flat test rejects $f|_A$ with probability at least $\varepsilon + \gamma \cdot \frac{2^d}{2^{n-1}}$, and thus

$$\Pr[k\text{-flat test rejects } f|_A] \geq \left(1 - \frac{K}{N}\right)\left(\varepsilon + \gamma \cdot \frac{2^d}{2^{n-1}}\right) \geq \varepsilon + \gamma \cdot \frac{2^d}{2^{n-1}} - \frac{K}{N}.$$

If $K \leq \gamma 2^d$ then we are done. So, assume from now on that $K > \gamma 2^d$. We show below in Theorem 3.3.5 that as long as $\beta < 1/4$, this implies that

$$\delta_d(f) \leq \frac{3}{2}\beta 2^{-d} + \frac{9}{\gamma 2^d} \leq 2^{-(d+2)},$$

where the second inequality uses $\beta < 1/24$ and $\gamma \geq 72$. At this point Theorem 3.3.2 asserts that the k-flat test (in fact, even the $(d+1)$-flat test), will reject f with probability at least

$$2^d \delta_d(f) \geq \beta \geq \varepsilon + \gamma 2^{-c} \geq \varepsilon + \gamma \cdot 2^{d-n}.$$

□

Lemma 3.3.5. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Assume that there exist distinct hyperplanes $A_1, \cdots, A_K$ such that $\delta_d(f|_{A_i}) \leq \alpha$. If $K > 2^{d+1}$ and $\alpha < 2^{-(d+2)}$, then*

$$\delta_d(f) \leq \frac{3}{2}\alpha + \frac{9}{K}.$$

Proof. For each A_i, let P_i be a degree d polynomial α-close to $f|_{A_i}$.

Claim 3.3.6. *If $4\alpha < 2^{-d}$, then for every i, j, $P_i|_{A_i \cap A_j} = P_j|_{A_i \cap A_j}$.*

Proof. The claim is vacuously true if $A_i = \overline{A_j}$. Otherwise, $|A_i \cap A_j| = |A_i|/2 = |A_j|/2$ and

$$\operatorname{dist}(f|_{A_i \cap A_j}, P_i|_{A_i \cap A_j}) \leq 2\alpha$$

and similarly for P_j, and thus $\operatorname{dist}(P_i|_{A_i \cap A_j}, P_j|_{A_i \cap A_j}) \leq 4\alpha < 2^{-d}$. The claim now follows because $P_i|_{A_i \cap A_j}, P_j|_{A_i \cap A_j}$ are degree $\leq d$ polynomials, and by Claim 3.1.4 the minimum distance of distinct degree d polynomials is 2^{-d}. □

Observe that there are at least $\ell = \lfloor \log_2(K+1) \rfloor > d$ linearly independent hyperplanes among $A_1, \cdots, A_K$. Without loss of generality assume that $A_1, \cdots, A_\ell$ are linearly independent, and that $A_i = \{x \in \mathbb{F}_2^n \mid x_i = 0\}$, by applying an appropriate affine transformation on $\mathbb{F}_2^n$. This way, for $i \in [\ell]$, P_i naturally corresponds to a polynomial over $\mathbb{F}_2^n$ which does not depend on x_i. The idea of the proof now is to glue $P_1, \cdots, P_\ell$ together to get a polynomial P close to f.

To this end, decompose $x = (x_1, \ldots, x_\ell, y)$ where $x_i \in \mathbb{F}_2$ and $y \in \mathbb{F}_2^{n-\ell}$. For every i, P_i can be decomposed as

$$P_i(x_1, \cdots, x_\ell, y) = \sum_{S \subseteq [\ell]} P_{i,S}(y) \prod_{j \in S} x_j,$$

where $P_{i,S}$ is a polynomial of degree at most $d - |S|$ which depends only on y. In particular $P_{i,S} = 0$ if $|S| > d$. Note that, if P_i does not depend on x_j and $j \in S$, then $P_{i,S} = 0$. The following claim follows immediately from Theorem 3.3.6.

Claim 3.3.7. *For every $S \subseteq [\ell]$ and $i, j \in [\ell] \setminus S$, $P_{i,S}(y) = P_{j,S}(y)$.*

We can now formally glue together the polynomials P_i. For $S \subsetneq [\ell]$, define P_S to be $P_{i,S}$ for some $i \in [\ell] \setminus S$. The above claim asserts that P_S is unique. Now define the degree d polynomial P over $\mathbb{F}_2^n$ as follows

$$P(x_1, \ldots, x_\ell, y) = \sum_{S \subsetneq [\ell]} P_S(y) \prod_{i \in S} x_i.$$

Claim 3.3.8. *For every $i \in [K]$, $P|_{A_i} = P_i|_{A_i}$.*

Proof. The claim is easy to see for $i \in [\ell]$, since the coefficients of the two polynomials P and P_i are identically the same when $x_i = 0$. Now assume that $i \in K \setminus [\ell]$. First note that for any $j \in [\ell]$ and $x \in A_i \cap A_j$, it follows from Theorem 3.3.6 that $P_i(x) = P_j(x) = P(x)$. Thus $P_i|_{(\cup_{j \in [\ell]} A_j) \cap A_i} = P|_{(\cup_{j \in [\ell]} A_j) \cap A_i}$. Now, note that since $\ell > d$ then

$$\frac{\left| A_i \cap \left(\bigcup_{j \in [\ell]} A_j \right) \right|}{|A_i|} \geq 1 - 2^{-\ell} > 1 - 2^{-d}.$$

The claim now follows from the minimum distance between degree d polynomials (Claim 3.1.4). □

Next, we will show that P is close to f. Since $A_1, \cdots, A_K$ do not necessarily cover $\mathbb{F}_2^n$ uniformly, we show that they do so approximately uniformly. Let

$$\text{BAD} := \{z \in \mathbb{F}_2^n \mid z \text{ is contained in less than } K/3 \text{ of the hyperplanes } A_1, \cdots, A_K\}.$$

Claim 3.3.9. $|\text{BAD}| \leq 2^n \frac{9}{K}$.

Proof. Let $z \in \mathbb{F}_2^n$ be uniformly chosen. The random variables $1_{z \in A_i}$ each have probability $1/2$, and they are pairwise independent. Let $S = \sum_{i=1}^{k} 1_{z \in A_i}$. Then $\mathbb{E}[S] = K/2$ and $\text{Var}[S] = K/4$. By Chebyshev's inequality,

$$\Pr[z \in \text{BAD}] = \Pr[S \leq K/3] \leq \Pr[|S - \mathbb{E}[S]| \leq K/6] \leq \frac{K/4}{(K/6)^2} = \frac{9}{K}.$$

□

The following claim now finishes the proof of the lemma.

Claim 3.3.10. $\text{dist}(f, P) \leq \frac{3}{2}\alpha + \frac{|\text{BAD}|}{2^n} \leq \frac{3}{2} + \frac{9}{K}$.

Proof. Pick $z \in \mathbb{F}_2^n$ and $i \in [K]$ uniformly at random. Then

$$\Pr_{i,z}[z \in A_i \wedge f(z) \neq P_i(z)] = \frac{1}{2}\mathbb{E}_i\left[\Pr_{z \in A_i}[f(z) \neq P_i(z)]\right] \leq \frac{\alpha}{2}.$$

On the other hand, since $P|_{A_i} = P_i|_{A_i}$, we have that

$$\begin{aligned}
\Pr_{i,z}[z \in A_i \wedge f(z) \neq P_i(z)] &= \Pr_{i,z}[z \in A_i \wedge f(z) \neq P(z)] \\
&\geq \Pr_{i,z}[z \in A_i \wedge f(z) \neq P(z) \wedge z \notin \text{BAD}] \\
&= \Pr_{z}[f(z) \neq P(z) \wedge z \notin \text{BAD}] \cdot \Pr_{i,z}[z \in A_i | f(z) \neq P(z) \wedge z \notin \text{BAD}] \\
&\geq \Pr_{z}[f(z) \neq P(z) \wedge z \notin \text{BAD}] \cdot \frac{1}{3} \\
&\geq \frac{1}{3}\left(\text{dist}(f, P) - \frac{|\text{BAD}|}{2^n}\right).
\end{aligned}$$

□

□

Theorem 3.2.2 now directly follows from Theorem 3.3.2, Theorem 3.3.3, and Theorem 3.3.4.

3.4 Implications for the Gowers norms

We will introduce the Gowers uniformity norm in later chapters. For now, it suffices to know that for $f : \mathbb{F}_2^n \to \mathbb{F}_2$, its $(d+1)$-th Gowers norm is defined as

$$\|(-1)^f\|_{U^{d+1}} = \left(\Pr[\text{AKKLR}(d) \text{ accepts } f] - \Pr[\text{AKKLR}(d) \text{ rejects } f]\right)^{1/2^{d+1}}.$$

It is easy to see that this expression is bounded between 0 and 1, and is equal to 1 if and only if f is a polynomial of degree d. Theorem 3.2.2 shows that if the Gowers norm is close to 1, then f is close to a degree d polynomial.

Theorem 3.4.1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Then for any $1 \leq d \leq n$, if $\|f\|_{U^{d+1}} \geq 1 - \varepsilon$ then $\delta_d(f) \leq c\varepsilon$, where $c > 0$ is an absolute constant.*

Proof. If $\|f\|_{U^{d+1}} \geq 1-\varepsilon$ then by definition, $2\Pr[\text{AKKLR}(d) \text{ rejects } f] \leq 1-(1-\varepsilon)^{2^{d+1}} \leq 2^{d+1}\varepsilon$. Hence by Theorem 3.2.2, we have $\delta_d(f) \leq c\varepsilon$. □

Theorem 3.4.1 becomes trivial when $\varepsilon > 1/c$. In particular it says nothing about the case where $\|(-1)^f\|_{U^{d+1}}$ is bounded away from zero, e.g. $\|(-1)^f\|_{U^{d+1}} > 1/3$. One expects such functions to have some structure as a typical random function will satisfy $\|(-1)^f\|_{U^{d+1}} = o(1)$ with high probability. In the next section we will see a result due to Samorodnitsky [66] and independently Green and Tao [40] relating the Gowers U^3 norm to proximity to quadratic polynomials even in the regime where the distance is close to $1/2$. More precisely, they prove that if $\|(-1)^f\|_{U^3}$ is bounded away from 0, then f must have significant correlation with a quadratic polynomial. This immediately gives a tester for proximity to quadratics in the 1% regime.

Chapter 4

Low-degree Tests, the 1% Regime

In Chapter 3 we presented the AKKLR test for proximity to degree-d polynomials. The analysis in Theorem 3.2.2 shows that the test is effective in the 99% regime: it can distinguish functions which are very close to degree-d polynomials from those which are somewhat far. Concretely, its rejection probability is $c \cdot \min(\delta_d(f)2^d, 1)$ for some absolute constant $c > 0$. In this chapter, we will focus on the 1% regime, where we have the more ambitious goal of detecting any non-trivial agreement with degree-d polynomials. Concretely, we would like to distinguish between functions with $\delta_d(f) \leq \frac{1}{2} - \varepsilon$ from functions with $\delta_d(f) = \frac{1}{2} - o(1)$.

Compare to the 99% regime, the 1% regime is much more complex. In the 99% regime, we are interested in functions which are close to polynomials of degree d. Hence so much of the polynomial structure is remained in these functions that by taking only few samples, we can obtain a good understanding of the global structure of the function. Indeed, as we saw in Chapter 3, the proof of the AKKLR test in the 99% regime was based on such ideas. However in the 1% regime, we are concerned with functions with very little structure. For example, consider a degree d polynomial (think of it as a completely structured function), and obtain a new function from it by retaining the values of 1% of the points and assigning random values to the rest. Thus, this function is likely to agree with the polynomial on about 50.5% fraction of the inputs, only slightly more than a random function, which is likely to agree with the polynomial on about 50% of the inputs. While in this blurry picture much of the structure of the original polynomial is lost, some of it is still maintained since a typical random function cannot agree with any polynomial of degree d on 50.5% of the points. However, detecting this structure seems to be much harder as now if we sample the function, we will typically receive random values that are not related to the original polynomial.

It was conjectured independently by Samorodnitsky [66] and Green and Tao [40] that the AKKLR test can be used to detect such weak structures, and thus distinguish between functions with $\delta_d(f) \leq \frac{1}{2} - \varepsilon$ from functions with $\delta_d(f) = \frac{1}{2} - o(1)$.

Conjecture 4.0.1 (AKKLR test, 1% regime)**.** *Fix $d \geq 1$. For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ let $p_d(f) = \Pr[AKKLR(d) \text{ accepts } f]$. Then*

- ***(Completeness)** If $\delta_d(f) \leq \frac{1}{2} - \varepsilon$ then $p_d(f) \geq \frac{1}{2} + \alpha(\varepsilon)$, where $\alpha(\varepsilon) > 0$.*
- ***(Soundness)** If $p_d(f) \geq \frac{1}{2} + \varepsilon$ then $\delta_d(f) \leq \frac{1}{2} - \beta(\varepsilon)$, where $\beta(\varepsilon) > 0$.*

We will see that completeness holds for any $d \geq 1$ with $\alpha(\varepsilon) = \Omega(\varepsilon^{2^{d+1}})$. However, soundness turns out to be more intricate. For $d = 1$ it follows by a relatively simple extension of linearity testing. For $d = 2$ it also holds, but the proof is much more involved. It was accomplished independently by Samorodnitsky [66] and Green and Tao [40], where the analysis relies on tools from additive combinatorics. For $d = 3$ it turns out to be false! The counter-example was discovered independently by Lovett, Meshulam and Samorodnitsky [57] and Green and Tao [41], and will be presented in Chapter 5.

4.1 Completeness

We prove that any function which has a distance noticeably smaller than $\frac{1}{2}$ from degree d polynomials, is accepted by the AKKLR test with probability noticeably larger that $\frac{1}{2}$.

Theorem 4.1.1. *Fix $d \geq 1$. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ with $\delta_d(f) = \frac{1}{2} - \varepsilon$. Then*

$$\Pr[\textit{AKKLR}(d) \textit{ accepts } f] \geq \frac{1}{2}\left(1 + (2\varepsilon)^{2^d}\right).$$

Proof. Let $P(x)$ be a polynomial of degree d so that $\mathrm{dist}(f, P) = \frac{1}{2} - \varepsilon$ and hence $\mathbb{E}[(-1)^{(f+P)(x)}] = 2\varepsilon$. To recall, the AKKLR test samples $x, y_1, \ldots, y_{d+1} \in \mathbb{F}_2^n$ independently and accepts f if $\sum_{S\subseteq[d+1]} f\left(x + \sum_{i\in S} y_i\right) = 0$. The test always accepts degree-d polynomials, as it holds that $\sum_{S\subseteq[d+1]} P\left(x + \sum_{i\in S} y_i\right) \equiv 0$. Thus, we can reformulate the acceptance probability of the test as

$$\begin{aligned}
\Pr[\text{AKKLR}(d) \text{ accepts } f] &= \Pr_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}_2^n}\left[\sum_{S\subseteq[d+1]} f\left(x + \sum_{i\in S} y_i\right) = 0\right] \\
&= \Pr_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}_2^n}\left[\sum_{S\subseteq[d+1]} (f+P)\left(x + \sum_{i\in S} y_i\right) = 0\right] \\
&= \frac{1}{2}\left(1 + \mathbb{E}_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}_2^n}\left[(-1)^{D_{y_1,\ldots,y_{d+1}}(f+P)(x)}\right]\right).
\end{aligned}$$

The lemma then follows from the following claim: for any function $g : \mathbb{F}_2^n \to \mathbb{F}_2$ and any $k \geq 1$, it holds that

$$\mathbb{E}_{x,y_1,\ldots,y_k\in\mathbb{F}_2^n}\left[(-1)^{D_{y_1,\ldots,y_k}g(x)}\right] \geq \left(\mathbb{E}_x\left[(-1)^{g(x)}\right]\right)^{2^k}.$$

The claim holds for $k = 1$ as

$$\mathbb{E}_{x,y}\left[(-1)^{D_y g(x)}\right] = \mathbb{E}_{x,y}\left[(-1)^{g(x)+g(x+y)}\right] = \mathbb{E}_{x,y}\left[(-1)^{g(x)+g(y)}\right] = \left(\mathbb{E}_x\left[(-1)^{g(x)}\right]\right)^2,$$

and for $k > 1$ by induction:

$$\begin{aligned}
\mathbb{E}_{x,y_1,\ldots,y_k}\left[(-1)^{D_{y_1,\ldots,y_k}g(x)}\right] &= \mathbb{E}_{y_1,\ldots,y_{k-1}}\left[\mathbb{E}_{x,y_k}\left[(-1)^{D_{y_k}D_{y_1,\ldots,y_{k-1}}g(x)}\right]\right] \\
&= \mathbb{E}_{y_1,\ldots,y_{k-1}}\left(\mathbb{E}_x\left[(-1)^{D_{y_1,\ldots,y_{k-1}}g(x)}\right]\right)^2 \\
&\geq \left(\mathbb{E}_{x,y_1,\ldots,y_{k-1}}\left[(-1)^{D_{y_1,\ldots,y_{k-1}}g(x)}\right]\right)^2 \\
&\geq \left(\mathbb{E}_x\left[(-1)^{g(x)}\right]\right)^{2^k},
\end{aligned}$$

where the first inequality follows by the Cauchy-Schwartz inequality, and the second by induction. □

4.2 Soundness for $d = 1$

We show that if the AKKLR(1) test accepts a function with probability noticeably larger than $\frac{1}{2}$, then the function has a non-trivial correlation with some linear polynomial. Equivalently, it has a noticeable Fourier coefficient.

Theorem 4.2.1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Assume that $\Pr[\textit{AKKLR}(1) \textit{ accepts } f] \geq \frac{1}{2} + \varepsilon$. Then $\delta_1(f) \leq \frac{1}{2} - \sqrt{\varepsilon/2}$.*

Proof. As in the proof of Theorem 2.2.1, let $F(x) = (-1)^{f(x)}$, where we expand the acceptance probability of the test in terms of the Fourier coefficients of F,

$$\Pr[\text{AKKLR(1) accepts } f] = \Pr_{x,y,z \in \mathbb{F}_2^n}[f(x) + f(y) + f(z) = f(x+y+z)] = \frac{1}{2} + \frac{1}{2} \sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^4.$$

Then

$$2\varepsilon \leq \sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^4 \leq \|\widehat{F}\|_\infty^2 \sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^2 = \|\widehat{F}\|_\infty^2.$$

Thus $\|\widehat{F}\|_\infty \geq \sqrt{2\varepsilon}$. To conclude observe that $\delta_1(f) = \frac{1}{2}(1 - \|\widehat{F}\|_\infty)$. □

Corollary 4.2.2. *For any $\varepsilon > 0$ there is a test which makes $O(1/\varepsilon^8)$ queries to a function f and distinguishes, with high probability, between functions with $\delta_1(f) \leq \frac{1}{2} + \varepsilon$ and functions with $\delta_1(f) \geq \frac{1}{2} - o(1)$.*

Proof. Consider the AKKLR(1) test. If $\delta_1(f) \leq \frac{1}{2} + \varepsilon$, then by Theorem 4.1.1 it accepts f with probability at least $\frac{1}{2}(1 + (2\varepsilon)^4)$. If, on the other hand, $\delta_1(f) \geq \frac{1}{2} - o(1)$, then by Theorem 4.2.1 it accepts f with probability at most $\frac{1}{2}(1 + o(1))$. Thus, if we repeat the test $\Omega(1/\varepsilon^8)$ times, then with high probability we can distinguish the two cases. □

4.3 Soundness for $d = 2$

We show that if the AKKLR(2) test accepts a function with probability noticeably larger than $\frac{1}{2}$, then the function has a non-trivial correlation with some quadratic polynomial. We follow Samorodnitsky [66] below.

Theorem 4.3.1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Assume that $\Pr[AKKLR(2)$ accepts $f] \geq \frac{1}{2} + \varepsilon$. Then $\delta_2(f) \leq \frac{1}{2} - \beta(\varepsilon)$, where $\beta(\varepsilon) \geq \exp(-c \cdot \log(1/\varepsilon)^4)$ for an absolute constant $c > 0$.*

Fix a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ for the remainder of this section, which we assume is accepted by the AKKLR(2) test with probability at least $\frac{1}{2} + \varepsilon$. We may assume that $n \geq c \cdot \log(1/\varepsilon)^4$, as otherwise the conclusion of Theorem 4.3.1 is trivial. Let $F(x) = (-1)^{f(x)}$ and $F_y(x) = (-1)^{D_y f(x)} = F(x)F(x+y)$.

In order to build intuition, consider first the case that $f(x) = x^t Dx$ is a quadratic polynomial. Then $F_y(x) = (-1)^{\langle x, (D+D^t)y \rangle}$ and hence F_y has only one nonzero Fourier coefficient, which is at $(D + D^t)y$. Note that $D + D^t$ is a symmetric matrix with a zero diagonal. We will show that obtaining this, even approximately, implies that f is correlated with a quadratic polynomial.

Lemma 4.3.2. *Assume there exists an $n \times n$ symmetric matrix A with a zero diagonal such that*

$$\mathbb{E}_y \left[\widehat{F_y}(Ay)^2 \right] \geq \gamma.$$

Then there exists a quadratic polynomial $q(x)$ such that

$$\operatorname{dist}(f, q) \leq \frac{1}{2} - \frac{1}{2}\sqrt{\gamma}.$$

In order to prove Theorem 4.3.2 we need a few technical claims. For functions $G, H : \mathbb{F}_2^n \to \mathbb{R}$ define $GH(x) = G(x)H(x)$ and $(G * H)(x) = \mathbb{E}_y G(y)H(x+y)$. It is a standard fact in classical Fourier analysis that $\widehat{G * H}(\alpha) = \widehat{G}(\alpha)\widehat{H}(\alpha)$.

Claim 4.3.3. *Let $G : \mathbb{F}_2^n \to \mathbb{R}$. Let $G_y(x) = G(x)G(x+y)$. Then $(G_y * G_y)(x) = (G_x * G_x)(y)$.*

Proof. We have

$$(G_y * G_y)(x) = \mathbb{E}_{s \in \mathbb{F}_2^n} \left[G_y(s) G_y(x+s) \right] = \mathbb{E}_{s \in \mathbb{F}_2^n} \left[G(s)G(s+y)G(x+s)G(x+s+y) \right].$$

The claim follows as the RHS is symmetric to swapping x and y. □

Claim 4.3.4. *Let $G, H : \mathbb{F}_2^n \to \mathbb{R}$. Let $G_y(x) = G(x)G(x+y)$ and $H_y(x) = H(x)H(x+y)$. Then*

$$\mathbb{E}_y\left[\left(\mathbb{E}_x\left[G_y(x)H_y(x)\right]\right)^2\right] = \sum_\alpha \widehat{GH}(\alpha)^4.$$

Proof.

$$\begin{aligned}
\mathbb{E}_y\left[\left(\mathbb{E}_x\left[G_y(x)H_y(x)\right]\right)^2\right] &= \mathbb{E}_{y,x,x'}\left[G_y(x)H_y(x)G_y(x')H_y(x')\right] \\
&= \mathbb{E}_{y,x,x'}\left[G(x)G(x+y)H(x)H(x+y)G(x')G(x'+y)H(x')H(x'+y)\right] \\
&= \mathbb{E}_{y,x,x'}\left[GH(x)GH(x+y)GH(x')GH(x'+y)\right] \\
&= \mathbb{E}_y\left[\left(\mathbb{E}_x\left[GH(x)GH(x+y)\right]\right)^2\right] = \mathbb{E}_y\left[(GH*GH)(y)^2\right] \\
&= \sum_\alpha \widehat{GH*GH}(\alpha)^2 = \sum_\alpha \widehat{GH}(\alpha)^4.
\end{aligned}$$

□

Proof of Lemma 4.3.2. Decompose $A = D + D^t$ where D is upper triangular and let $G(x) = (-1)^{x^t Dx}$. Note that $G_y(x) = G(x)G(x+y) = (-1)^{\langle x, Ay\rangle}$. We will show that FG has a noticeable Fourier coefficient. By Claim 4.3.4 and our assumption,

$$\sum_\alpha \widehat{FG}(\alpha)^4 = \mathbb{E}_y\left[\left(\mathbb{E}_x\left[F_y(x)G_y(x)\right]\right)^2\right] = \mathbb{E}_y\left[\left(\mathbb{E}_x\left[F_y(x)(-1)^{\langle x,Ay\rangle}\right]\right)^2\right] = \mathbb{E}_y\left[\widehat{F_y}(Ay)^2\right] \geq \gamma.$$

This implies that FG has a large Fourier coefficient, as

$$\sum_\alpha \widehat{FG}(\alpha)^4 \leq \|\widehat{FG}\|_\infty^2 \sum_\alpha \widehat{FG}(\alpha)^2 = \|\widehat{FG}\|_\infty^2.$$

So, there exists $\alpha \in \mathbb{F}_2^n$ and $c \in \mathbb{F}_2$ such that

$$\mathbb{E}_x\left[F(x)G(x)(-1)^{\langle x,\alpha\rangle + c}\right] = |\widehat{FG}(\alpha)| \geq \sqrt{\gamma}.$$

So, the quadratic polynomial $q(x) = x^t Dx + \langle x, \alpha\rangle + c$ satisfies $\mathrm{dist}(f, q) \leq \frac{1}{2} - \frac{1}{2}\sqrt{\gamma}$. □

So, our goal from now on is to show the existence of such a matrix A for which $\widehat{F_y}(Ay)$ is noticeable for a typical y. As a first step, we show that for a typical y, F_y has some noticeable Fourier coefficients. Formally, we show that the 4th moments of $\widehat{F_y}$ are noticeable.

Claim 4.3.5. $\mathbb{E}_{y\in\mathbb{F}_2^n}\left[\sum_{\alpha\in\mathbb{F}_2^n} \widehat{F_y}(\alpha)^4\right] \geq 2\varepsilon$.

Proof. Let $f_y = D_y f$. We can express

$$\begin{aligned}
\Pr[\text{AKKLR(2) accepts } f] &= \Pr_{x,y_1,y_2,y_3\in\mathbb{F}_2^n}\left[\sum_{S\subseteq[3]} f\left(x + \sum_{i\in S} y_i\right) = 0\right] \\
&= \mathbb{E}_{y\in\mathbb{F}_2^n} \Pr_{x,y_1,y_2\in\mathbb{F}_2^n}\left[f_y(x) + f_y(x+y_1) + f_y(x+y_2) + f_y(x+y_1+y_2) = 0\right].
\end{aligned}$$

As in the proof of Theorem 4.2.1, we can express the inner probability as

$$\Pr_{x,y_1,y_2\in\mathbb{F}_2^n}\left[f_y(x) + f_y(x+y_1) + f_y(x+y_2) + f_y(x+y_1+y_2) = 0\right] = \frac{1}{2} + \frac{1}{2}\sum_{\alpha\in\mathbb{F}_2^n} \widehat{F_y}(\alpha)^4.$$

Thus

$$\Pr[\text{AKKLR(2) accepts } f] = \frac{1}{2}\left(1 + \mathbb{E}_{y\in\mathbb{F}_2^n}\left[\sum_{\alpha\in\mathbb{F}_2^n}\widehat{F_y}(\alpha)^4\right]\right).$$

The claim follows from our assumption that $\Pr[\text{AKKLR(2) accepts } f] \geq \frac{1}{2} + \epsilon$. □

The next step is to show that the noticeable Fourier coefficients of $\widehat{F_y}$ have an approximate linear structure. The following lemma plays an important role in this.

Lemma 4.3.6. $\mathbb{E}_{y,z\in\mathbb{F}_2^n}\left[\sum_{\alpha,\beta\in\mathbb{F}_2^n}\widehat{F_y}(\alpha)^2\widehat{F_z}(\beta)^2\widehat{F_{y+z}}(\alpha+\beta)^2\right] = \mathbb{E}_{y\in\mathbb{F}_2^n}\left[\sum_{\alpha\in\mathbb{F}_2^n}\widehat{F_y}(\alpha)^6\right]$.

Proof. Note that $\widehat{F_y}(\alpha)^2 = \mathbb{E}_{u,u'\in\mathbb{F}_2^n}F_y(u)F_y(u')(-1)^{\langle u+u',\alpha\rangle}$. Using this, we can expand the LHS in the lemma statement as

$$\begin{aligned}
\text{LHS} &= \mathbb{E}_{y,z}\left[\sum_{\alpha,\beta}\widehat{F_y}(\alpha)^2\widehat{F_z}(\beta)^2\widehat{F_{y+z}}(\alpha+\beta)^2\right] \\
&= \mathbb{E}_{y,z}\sum_{\alpha,\beta}\Bigg[\mathbb{E}_{u,u'}F_y(u)F_y(u')(-1)^{\langle u+u',\alpha\rangle}\cdot\mathbb{E}_{v,v'}F_z(v)F_z(v')(-1)^{\langle v+v',\beta\rangle}\cdot \\
&\qquad\qquad \mathbb{E}_{w,w'}F_{y+z}(w)F_{y+z}(w')\cdot(-1)^{\langle w+w',\alpha+\beta\rangle}\Bigg].
\end{aligned}$$

Next, note that

$$\begin{aligned}
&\sum_{\alpha,\beta}(-1)^{\langle u+u',\alpha\rangle}(-1)^{\langle v+v',\beta\rangle}(-1)^{\langle w+w',\alpha+\beta\rangle} \\
&= \sum_{\alpha}(-1)^{\langle u+u'+w+w',\alpha\rangle}\sum_{\beta}(-1)^{\langle v+v'+w+w',\beta\rangle} \\
&= 2^{2n}\cdot 1_{[u+u'=v+v'=w+w']}.
\end{aligned}$$

We can thus restrict our attention to the case that $u+u' = v+v' = w+w' = s$ for some $s\in\mathbb{F}_2^n$, and obtain a simplified expression for the LHS as

$$\begin{aligned}
\text{LHS} &= \mathbb{E}_{y,z,s}\left[\mathbb{E}_u F_y(u)F_y(u+s)\cdot\mathbb{E}_v F_z(v)F_z(v+s)\cdot\mathbb{E}_w F_{y+z}(w)F_{y+z}(w+s)\right] \\
&= \mathbb{E}_{y,z,s}\left[(F_y * F_y)(s)(F_z * F_z)(s)(F_{y+z} * F_{y+z})(s)\right] \\
&= \mathbb{E}_{y,z,s}\left[(F_s * F_s)(y)(F_s * F_s)(z)(F_s * F_s)(y+z)\right] \\
&= \mathbb{E}_s\left[\sum_{\alpha}\widehat{F_s * F_s}(\alpha)^3\right] \\
&= \mathbb{E}_s\left[\sum_{\alpha}\widehat{F_s}(\alpha)^6\right].
\end{aligned}$$

□

As a corollary, we obtain that the noticeable Fourier coefficients of $\widehat{F_y}$ have an approximate linear structure, at least locally.

Corollary 4.3.7. $\mathbb{E}_{y,z\in\mathbb{F}_2^n}\left[\sum_{\alpha,\beta\in\mathbb{F}_2^n}\widehat{F_y}(\alpha)^2\widehat{F_z}(\beta)^2\widehat{F_{y+z}}(\alpha+\beta)^2\right] \geq 4\varepsilon^2$.

Proof. By Theorem 4.3.5 we have $\mathbb{E}_y \sum_\alpha \widehat{F_y}(\alpha)^4 \geq 2\varepsilon$. By the Cauchy-Schwartz inequality

$$\sum_\alpha \widehat{F_y}(\alpha)^4 = \sum_\alpha \widehat{F_y}(\alpha) \cdot \widehat{F_y}(\alpha)^3 \leq \sqrt{\sum_\alpha \widehat{F_y}(\alpha)^2 \cdot \sum_\alpha \widehat{F_y}(\alpha)^6} = \sqrt{\sum_\alpha \widehat{F_y}(\alpha)^6}.$$

Thus we have

$$\mathbb{E}_y \sum_\alpha \widehat{F_y}(\alpha)^6 \geq \mathbb{E}_y \left(\sum_\alpha \widehat{F_y}(\alpha)^4 \right)^2 \geq \left(\mathbb{E}_y \sum_\alpha \widehat{F_y}(\alpha)^4 \right)^2 \geq 4\varepsilon^2.$$

The corollary now follows by Theorem 4.3.6. □

The next step is to show that a typical $\widehat{F_x}$ has a noticeable Fourier coefficient at $\phi(x)$, where ϕ is approximately linear. This will be the starting point for finding an actual linear map for which this holds.

Lemma 4.3.8. *There exists $X \subset \mathbb{F}_2^n$ and a map $\phi : X \to \mathbb{F}_2^n$ such that*

(i) $\widehat{F_x}(\phi(x))^2 \geq \varepsilon^2$ *for all* $x \in X$.

(ii) $\Pr_{x,y \in \mathbb{F}_2^n}[x, y, x+y \in X \wedge \phi(x) + \phi(y) = \phi(x+y)] \geq \varepsilon^2/2$.

Proof. Define a random function $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ by picking, independently for each $y \in \mathbb{F}_2^n$, $\phi(y) = \alpha$ with probability $\widehat{F_y}(\alpha)^2$. Note that this is indeed a probability distribution as by Parseval's identity $\sum_\alpha \widehat{F_y}(\alpha)^2 = 1$ for each y. For $\delta = \varepsilon^2$ define

$$L(\phi) := \Pr_{x,y \in \mathbb{F}_2^n} \left[\phi(x+y) = \phi(x) + \phi(y),\ \widehat{F_x}(\phi(x))^2 \geq \delta,\ \widehat{F_y}(\phi(y))^2 \geq \delta,\ \widehat{F_{x+y}}(\phi(x+y))^2 \geq \delta \right].$$

It suffices to show that $\mathbb{E}_\phi L(\phi) \geq \varepsilon^2/2$. Indeed for a choice of ϕ which attains the bound, by taking

$$X = \{x \in \mathbb{F}_2^n : \widehat{F_x}(\phi(x))^2 \geq \delta\},$$

the desired items (i),(ii) will follow by definition. In order to compute $\mathbb{E}_\phi L(\phi)$, it will be useful to define the set

$$\Lambda := \left\{ (\alpha, \beta) \in \mathbb{F}_2^{2n} : \widehat{F_x}(\alpha)^2 \geq \delta,\ \widehat{F_y}(\beta)^2 \geq \delta,\ \widehat{F_{x+y}}(\alpha+\beta)^2 \geq \delta \right\}.$$

Then

$$\mathbb{E}_\phi L(\phi) = \mathbb{E}_{x,y} \sum_{(\alpha,\beta) \in \Lambda} \Pr_\phi[\phi(x) = \alpha, \phi(y) = \beta, \phi(x+y) = \alpha + \beta]$$

Next, by definition of ϕ, if $x, y, x+y$ are all distinct then

$$\begin{aligned} \Pr_\phi[\phi(x) = \alpha, \phi(y) = \beta, \phi(x+y) = \alpha+\beta] &= \Pr_\phi[\phi(x) = \alpha] \Pr_\phi[\phi(y) = \beta] \Pr_\phi[\phi(x+y) = \alpha+\beta] \\ &= \widehat{F_x}(\alpha)^2 \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha+\beta)^2. \end{aligned}$$

The probability that $x, y, x+y$ are not all distinct is $O(2^{-n})$. Thus

$$\mathbb{E}_\phi L(\phi) \geq \mathbb{E}_{x,y} \sum_{(\alpha,\beta) \in \Lambda} \widehat{F_x}(\alpha)^2 \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha+\beta)^2 - O(2^{-n}).$$

We would like to complete the sum over all α, β in order to apply Theorem 4.3.7. The sum over α for which $\widehat{F_x}(\alpha)^2 < \delta$ can be bounded by

$$\begin{aligned} \mathbb{E}_{x,y} \sum_{\alpha,\beta : \widehat{F_x}(\alpha)^2 < \delta} \widehat{F_x}(\alpha)^2 \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha+\beta)^2 &\leq \delta \cdot \mathbb{E}_{x,y} \sum_{\alpha,\beta} \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha+\beta)^2 \\ &= \delta \cdot \mathbb{E}_{x,y} \sum_{\alpha,\beta} \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha)^2 \\ &= \delta, \end{aligned}$$

where we used Parseval's identity. Similarly, we can bound the sum over α, β for which $\widehat{F_y}(\beta)^2 < \delta$ or $\widehat{F_{x+y}}(\alpha+\beta)^2 < \delta$. We thus have

$$\begin{aligned}\mathbb{E}_\phi L(\phi) &\geq \mathbb{E}_{x,y} \sum_{\alpha,\beta \in \mathbb{F}_2^n} \widehat{F_x}(\alpha)^2 \widehat{F_y}(\beta)^2 \widehat{F_{x+y}}(\alpha+\beta)^2 - 3\delta - O(2^{-n}) \\ &\geq 4\varepsilon^2 - 3\delta - O(2^{-n}) \geq \varepsilon^2 - O(2^{-n}) \geq \varepsilon^2/2,\end{aligned}$$

where in the last inequality we used our assumption that n is large enough. □

Next, we show that ϕ can be approximated by a linear map. To that end, we will need several results in additive combinatorics. For a set $S \subseteq \mathbb{F}_2^n$ its sumset is $S+S = \{s+s' : s, s' \in S\}$. The first result is the Balog-Szemerédi-Gowers theorem (abbreviated BSG theorem) [7, 35]. It states that if many pairs in S have their sum in a small set, then there is a large subset of S with a small sumset. The proof can be found in the original paper of Gowers [35] (see also [77]). A simplified proof for the case of $G = \mathbb{F}_2^n$ can be found in an exposition of Viola [80].

Theorem 4.3.9 (BSG theorem [7, 35])**.** *Let G be an Abelian group and let $S \subseteq G$. If*

$$\Pr_{s,s' \in S}[s+s' \in S] \geq \varepsilon,$$

then there exists $S' \subset S$ of size $|S'| \geq c\varepsilon^2|S|$ such that $|S'+S'| \leq c^{-1}\varepsilon^{-5}|A|$. Here, $c > 0$ is an absolute constant.

The other ingredient required is the structure of sets S for which $S+S$ is not much larger than S. Here, the best result to date is by Sanders [67]. See also the survey by Sanders [68] for more details, and the exposition by Lovett [56] giving a simplified proof for $G = \mathbb{F}_2^n$. Below, we present the result specialized to the case of $G = \mathbb{F}_2^n$.

Theorem 4.3.10 ([67])**.** *Let $S \subset \mathbb{F}_2^n$ be a set such that $|S+S| \leq K|S|$. Then there exists an affine linear subspace $V \subset \mathbb{F}_2^n$ of size $|V| \leq |S|$ such that*

$$|S \cap V| \geq \exp(-c\log^4 K)|S|,$$

where $c > 0$ is an absolute constant.

With these theorems at our disposal, let $S = \{(x, \phi(x)) : x \in X\} \subset \mathbb{F}_2^{2n}$ be the graph of ϕ. By Theorem 4.3.8 we have that $\Pr_{s_1,s_2 \in S}[s_1+s_2 \in S] \geq \varepsilon^2/2$. By Theorem 4.3.9, there exists a subset $S' \subset S$ of size $|S'| \geq c'\varepsilon^4$ such that $|S'+S'| \leq c''\varepsilon^{-10}|S'|$. By Theorem 4.3.10, there exists an affine linear subspace $V \subset \mathbb{F}_2^{2n}$ such that $S' \cap V$ is large and $|V| \leq |S'|$. We will only use the fact that $|V| \leq 2^n$ and that $|S \cap V| \geq \eta 2^n$ where $\eta = \exp(-c'''\log^4(1/\varepsilon))$ (here, $c', c'', c''' > 0$ are unspecified absolute constants). Next, we show that this implies the existence of a linear map which approximates ϕ.

Lemma 4.3.11. *There exist an $n \times n$ matrix A and a vector $b \in \mathbb{F}_2^n$ such that*

$$\Pr_{x \in S}[\phi(x) = Ax + b] \geq \eta^2.$$

Proof. Let $U = \{x : (x,y) \in V\}$ be the projection of V to the first n coordinates. As $|S \cap V| \geq \eta 2^n$ and S contains at most one element (x,y) for each $x \in \mathbb{F}_2^n$, it must be that $|U| \geq \eta 2^n$. Thus, we can decompose V as the disjoint union of subspaces of the form $V_i = \{(x, \ell_i(x)) : x \in U\}$, where $\ell_i : U \to \mathbb{F}_2^n$ are affine linear maps, and $1 \leq i \leq |V/U|$. Choose $\ell = \ell_i$ which maximizes $|S \cap V_i|$. For this choice, we get $|S \cap V_i| \geq \eta |S \cap V| \geq \eta^2 2^n$. If $|U| < \mathbb{F}_2^n$ then complete ℓ arbitrarily to a linear map from $\mathbb{F}_2^n$. The lemma follows as the fact that ℓ is an affine linear map implies $\ell(x) = Ax + b$. □

Corollary 4.3.12. $\nu := \mathbb{E}_{y \in \mathbb{F}_2^n}[\widehat{F_y}(Ay+b)^2] \geq \eta^3\varepsilon^2$.

Proof. For any $y \in S$ we have $\widehat{F_y}(\phi(y))^2 \geq \varepsilon^2$. There are $\eta^2|S|$ elements $y \in S$ for which $\phi(y) = Ay + b$. Finally, $|S| \geq |S \cap V| \geq \eta 2^n$. □

To conclude, we would want to apply Lemma 4.3.2. To do so, we need to show that we can take (i) A to be symmetric with zero diagonal; and (ii) $b = 0$. The following claim will be useful for both. For $M \in \{A, A^t\}$, define $R_M : \mathbb{F}_2^n \to \mathbb{R}$ as $R_M(z) := \sum_{y \in \mathbb{F}_2^n} \widehat{F_y}(My + z)^2$. First, we prove a general claim on the Fourier coefficients of R_M.

Claim 4.3.13. $\widehat{R_M}(\alpha) = \widehat{F_\alpha}(M^t\alpha)^2$.

Proof.

$$\begin{aligned}\widehat{R_M}(\alpha) &= \mathbb{E}_z \sum_y \widehat{F_y}(My+z)^2(-1)^{\langle \alpha, z\rangle} = \mathbb{E}_z \sum_y \mathbb{E}_{u,u'} F_y(u)F_y(u')(-1)^{\langle u+u', My+z\rangle + \langle \alpha, z\rangle} \\ &= \sum_y \mathbb{E}_{u,u'} F_y(u)F_y(u')(-1)^{\langle u+u', My\rangle} \cdot \mathbb{E}_z(-1)^{\langle u+u'+\alpha, z\rangle}\end{aligned}$$

The last term is 2^{-n} if $u' = u + \alpha$, and is zero otherwise. Thus, we may restrict to the case that $u' = u + \alpha$ and get

$$\begin{aligned}\widehat{R_M}(\alpha) &= 2^{-n} \sum_y \mathbb{E}_u F_y(u)F_y(u+\alpha)(-1)^{\langle \alpha, My\rangle} \\ &= \mathbb{E}_y \left[\mathbb{E}_u F_y(u)F_y(u+\alpha)(-1)^{\langle \alpha, My\rangle}\right] \\ &= \mathbb{E}_y \left[(F_y * F_y)(\alpha)(-1)^{\langle \alpha, My\rangle}\right] \\ &= \mathbb{E}_y \left[(F_\alpha * F_\alpha)(y)(-1)^{\langle \alpha, My\rangle}\right] \\ &= \widehat{F_\alpha * F_\alpha}(M^t\alpha) = \left(\widehat{F_\alpha}(M^t\alpha)\right)^2.\end{aligned}$$

□

Next we show that we may assume $b = 0$.

Claim 4.3.14. $\mathbb{E}_{y \in \mathbb{F}_2^n}[\widehat{F_y}(Ay)^2] \geq \nu$.

Proof. Let $R = R_A$. Since the Fourier coefficients of R are all non-negative, we have that $R(0) \geq R(z)$ for all $z \in \mathbb{F}_2^n$, as

$$R(0) = \sum_\alpha \widehat{R}(\alpha) \geq \sum_\alpha \widehat{R}(\alpha)(-1)^{\langle \alpha, z\rangle} = R(z).$$

In particular,

$$\mathbb{E}_{y \in \mathbb{F}_2^n}[\widehat{F_y}(Ay)^2] = R(0) \geq R(b) = \mathbb{E}_{y \in \mathbb{F}_2^n}[\widehat{F_y}(Ay+b)^2] = \nu.$$

□

We continue with showing that A may be taken to be symmetric with zero diagonal. The following claim will be useful.

Claim 4.3.15. *Let $y, \alpha \in \mathbb{F}_2^n$ with $\langle \alpha, y\rangle = 1$. Then $\widehat{F_y}(\alpha) = 0$.*

Proof. $\widehat{F_y}(\alpha) = \mathbb{E}_x\left[F(x)F(x+y)(-1)^{\langle \alpha, x\rangle}\right] = -\mathbb{E}_x\left[F(x)F(x+y)(-1)^{\langle \alpha, x+y\rangle}\right] = -\widehat{F_y}(\alpha)$. □

We next show that many large Fourier coefficients are supported on y for which $Ay = A^t y$.

Claim 4.3.16. $\mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(Ay)^2 \cdot 1_{Ay=A^ty}] \geq \nu^2$.

Proof. Let $G(x) = (-1)^{x^tAx}$ and $R = R_{A^t}$. By Theorem 4.3.13, $\widehat{R}(y) = \widehat{F_y}(Ay)^2$. By Theorem 4.3.15, if $\widehat{F_y}(Ay) \neq 0$ then necessarily $\langle Ay, y\rangle = 0$, which is equivalent to $G(y) = 1$. Thus

$$G(y)\widehat{R}(y) = \widehat{R}(y).$$

This implies that

$$\nu = \mathbb{E}_y\widehat{R}(y) = \mathbb{E}_y G(y)\widehat{R}(y) = \mathbb{E}_z\widehat{G}(z)R(z).$$

Recall that by definition $R(z) \geq 0$. Moreover, by Parseval's identity

$$\mathbb{E}_z R(z) = 2^{-n}\sum_{y,z}\widehat{F_y}(My+z)^2 = 1.$$

Thus by Jensen inequality,

$$\mathbb{E}_z\widehat{G}(z)^2R(z) \geq \left(\mathbb{E}_z\widehat{G}(z)R(z)\right)^2 = \nu^2.$$

On the other hand,

$$\mathbb{E}_z\widehat{G}(z)^2R(z) = \mathbb{E}_y(G*G)(y)\widehat{R}(y).$$

Now,

$$(G*G)(y) = \mathbb{E}_z(-1)^{z^tAz+(z+y)^tA(z+y)} = \mathbb{E}_z(-1)^{\langle z,(A+A^t)y\rangle+y^tAy} = 1_{Ay=A^ty}G(y).$$

We thus obtained

$$\mathbb{E}_y\left[1_{Ay=A^ty}\widehat{F_y}(Ay)^2\right] = \mathbb{E}_y\left[1_{Ay=A^ty}\widehat{R}(y)\right] = \mathbb{E}_y\left[1_{Ay=A^ty}G(y)\widehat{R}(y)\right] = \mathbb{E}_y\left[(G*G)(y)\widehat{R}(y) \geq \nu^2\right].$$

□

Let $W := \{x : Ax = A^tx\}$ be a linear subspace of $\mathbb{F}_2^n$. Restricted to W, A is symmetric. Thus, we can find a symmetric matrix A' such that $A'x = Ax$ for all $x \in W$. We have that

$$\mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(A'y)^2] \geq \mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(Ay)^2 \cdot 1_{y\in W}] \geq \nu^2.$$

It remains to deal with the diagonal of A'. Let $v \in \mathbb{F}_2^n$ be the diagonal of A', and define

$$A'' = A' + vv^t.$$

Clearly, A'' is symmetric with zero diagonal. The follow claim is the last step in the proof.

Claim 4.3.17. $\mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(A''y)^2] \geq \mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(A'y)^2]$.

Proof. We have $A''y = A'y + \langle y, v\rangle vv^t$. If $\langle y, v\rangle = 0$ then $A''y = A'y$. If $\langle y, v\rangle = 1$ then $y^tA'y = \langle y, v\rangle = 1$ and hence by Theorem 4.3.15, $\widehat{F_y}(A'y) = 0$. □

So, we have constructed a symmetric matrix with zero diagonal A'' for which

$$\mathbb{E}_{y\in\mathbb{F}_2^n}[\widehat{F_y}(A''y)^2] \geq \nu^2.$$

By Theorem 4.3.2 this implies the existence of a quadratic polynomial q whose distance from f is noticeably less than $1/2$. Concretely,

$$\mathrm{dist}(f, q) \leq \frac{1}{2} - \frac{\nu}{2}.$$

Chapter 5

Gowers Norms, the Inverse Gowers Conjecture, and its Failure

To recall, the probability that the AKKLR(d) test accepts a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is given by

$$\Pr[\text{AKKLR}(d) \text{ accepts } f] = \Pr_{x,y_1,\ldots,y_{d+1} \in \mathbb{F}_2^n} \left[\sum_{S \subseteq [d+1]} f\left(x + \sum_{i \in S} y_i \right) = 0 \right].$$

The AKKLR test is intimately related to Gowers norms, which we now define.

5.1 Gowers norms

In its most general form, the Gowers norms are defined for functions $F : G \to \mathbb{C}$, where G is a finite Abelian group (as a motivating example, consider the case of $G = \mathbb{F}_2^n$ and $F(x) = (-1)^{f(x)}$). The (multiplicative) derivative of F in direction $y \in G$ is given by $\Delta_y F(x) = F(x+y)\overline{F(x)}$. Note that if $F(x) = (-1)^{f(x)}$ then $\Delta_y F = (-1)^{f(x+y)+f(x)} = (-1)^{D_y f}$. Iterative derivatives are defined as $\Delta_{y_1,\ldots,y_d} F = \Delta_{y_1} \ldots \Delta_{y_d} F$. The *Gowers norm* of order d for F is defined as the expected d-th multiplicative derivative of F in d random directions at a random point.

Definition 5.1.1 (Gowers norm)**.** *Let G be a finite Abelian group, $d \geq 1$. Given a function $F : G \to \mathbb{C}$, the* Gowers norm of order d *for F is given by*

$$\begin{aligned} \|F\|_{U^d} &= \left| \mathbb{E}_{y_1,\ldots,y_d,x \in G} \left[(\Delta_{y_1} \Delta_{y_2} \cdots \Delta_{y_d} F)(x) \right] \right|^{1/2^d} \\ &= \left| \mathbb{E}_{y_1,\ldots,y_d,x \in G} \left[\prod_{S \subseteq [d]} \mathcal{C}^{d-|S|} F(x + \sum_{i \in S} y_i) \right] \right|^{1/2^d}, \end{aligned}$$

where $\mathcal{C}$ is the conjugation operator $\mathcal{C}(z) = \overline{z}$.

Note that as $\|F\|_{U^1} = |\mathbb{E}[F]|$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it turns out that $\|\cdot\|_{U^d}$ is indeed a norm [35]. It satisfies the following inequality, known as the Gowers Cauchy-Schwarz inequality.

Lemma 5.1.2 (Gowers Cauchy-Schwarz [35])**.** *Consider a family of functions $F_S : G \to \mathbb{C}$, where $S \subseteq [d]$. Then*

$$\left| \mathbb{E}_{x,y_1,\ldots,y_d \in \mathbb{F}^n} \left[\prod_{S \subseteq [d]} F_S(x + \sum_{i \in S} y_i) \right] \right| \leq \prod_{S \subseteq [d]} \|F_S\|_{U^d}.$$

In this survey, we will only consider the case of $G = \mathbb{F}^n$ where $\mathbb{F}$ is a finite field. Gowers norms are directly related to the acceptance probability of the AKKLR test. If $f : \mathbb{F}_2^n \to \mathbb{F}_2$ then

$$\|(-1)^f\|_{U^d}^{2^d} = \Pr[\text{AKKLR}(d) \text{ accepts } f] - \Pr[\text{AKKLR}(d) \text{ reject } f]$$

In particular, if $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a polynomial of degree $\leq d$, then $\|(-1)^f\|_{U^{d+1}} = 1$. We can rephrase Theorem 4.1.1, Theorem 4.2.1 and Theorem 4.3.1 as a direct theorem and inverse theorems for the Gowers norm for functions $F : \mathbb{F}_2^n \to \{-1, 1\}$. The results and proofs generalize to bounded functions $F : \mathbb{F}_p^n \to \mathbb{C}$, where $p \geq 2$ is a fixed prime. For a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ let $e(P) = \exp(2\pi i/p \cdot P(x))$. For two functions $F, G : \mathbb{F}_p^n \to \mathbb{C}$ define their inner product as $\langle F, G\rangle = \mathbb{E}_x F(x)\overline{G(x)}$.

Theorem 5.1.3. *Let $p \geq 2$ be a fixed prime. Let $F : \mathbb{F}_p^n \to \mathbb{C}$ with $\|F\|_\infty \leq 1$. Then*

- ***(Direct theorem)** For any polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree $\leq d-1$, it holds that $|\langle F, e(P)\rangle| \leq \|F\|_{U^d}$.*
- ***(Inverse theorem, $d = 1$)** If $\|F\|_{U^2} \geq \varepsilon$, then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree $d = 1$ such that $|\langle F, e(P)\rangle| \geq \varepsilon^2$.*
- ***(Inverse theorem, $d = 2$)** If $\|F\|_{U^3} \geq \varepsilon$, then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree $d = 2$ such that $|\langle F, e(P)\rangle| \geq \delta$ where $\delta = \delta(p, \varepsilon) > 0$.*

The (original) inverse Gowers conjecture, independently due to Samorodnitsky [66] and Green and Tao [40], states that the inverse theorem should hold for any $d \geq 1$.

Conjecture 5.1.4. *Let $p \geq 2$ be a fixed prime, and let $d \geq 1$. Let $F : \mathbb{F}_p^n \to \mathbb{C}$ with $\|F\|_\infty \leq 1$. If $\|F\|_{U^{d+1}} \geq \varepsilon$ then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree $\leq d$ such that $|\langle F, e(P)\rangle| \geq \delta$ where $\delta = \delta(p, d, \varepsilon)$.*

As we will shortly see, Theorem 5.1.4 is false, even for $p = 2, d = 3$. To salvage the conjecture, we will need to revise our notion of polynomials. These are the so-called nonclassical polynomials, which will be introduced in Chapter 6.

5.2 The counter-example

We give a counter example to Theorem 5.1.4 for $p = 2, d = 3$. It combines the bounds obtained in [57] and [41]. The example is the degree 4 symmetric polynomial S_4. In this section, all of the functions are defined on $\mathbb{F}_2^n$.

Definition 5.2.1 (Symmetric Polynomials)**.** *Let $k \geq 1$ be an integer. The elementary symmetric polynomial of degree k over n variables is denoted by S_k and is defined as*

$$S_k(x_1, \ldots, x_n) = \sum_{S \subseteq [n], |S| = k} \prod_{i \in S} x_i.$$

Theorem 5.2.2 ([57] and [41])**.** *Let $n \geq 1$ be large enough. Then*

$$\|(-1)^{S_4}\|_{U^4}^{16} = \frac{1}{8} + \mathrm{O}(2^{-n/2}),$$

but for any polynomial $Q : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree ≤ 3 it holds that

$$\left| \Pr_x[S_4(x) = Q(x)] - \frac{1}{2} \right| \leq \exp(-cn),$$

for some absolute constant $c > 0$.

5.2.1 U^4-norm of S_4

Our goal is to prove

$$\|(-1)^{S_4}\|_{U^4}^{16} = \frac{1}{8} + \mathrm{O}(2^{-n/2}).$$

Define the symmetric bilinear form $B : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ by

$$B(a,b) = \sum_{i,j\in[n]:i\neq j} a_i b_j,$$

for $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n)$ in $\mathbb{F}_2^n$. It is easy to verify the following identity regarding the fourth additive derivatives of S_4.

$$D_a D_b D_c D_d S_4(x) = B(a,b)B(c,d) + B(a,c)B(b,d) + B(a,d)B(b,c).$$

Consequently

$$\|(-1)^{S_4}\|_{U^4}^{16} = \mathbb{E}_{a,b,c,d\in\mathbb{F}_2^n}(-1)^{B(a,b)B(c,d)+B(a,c)B(b,d)+B(a,d)B(b,c)}. \tag{5.1}$$

In order to understand the above quantity, we need to understand the distribution of

$$B_6(a,b,c,d) := (B(a,b), B(a,c), B(a,d), B(b,c), B(b,d), B(c,d))$$

over $\mathbb{F}_2^6$, when $a, b, c, d \in \mathbb{F}_2^n$ are chosen uniformly and independently at random. The next lemma shows that for large n, they are essentially independently and uniformly chosen.

Lemma 5.2.3. *For every $\tau \in \mathbb{F}_2^6$,*

$$\Pr_{a,b,c,d\in\mathbb{F}_2^n}[B_6(a,b,c,d) = \tau] = \frac{1}{2^6} \pm \mathrm{O}(2^{-n/2}).$$

Proof. By standard Fourier analysis, it is sufficient to show that for every collection of $\lambda_{ab}, \lambda_{ac}, \lambda_{ad}, \lambda_{bc}, \lambda_{bd}, \lambda_{cd} \in \mathbb{F}_2$, not all zero,

$$\mathbb{E}_{a,b,c,d\in\mathbb{F}_2^n}\left[(-1)^{\lambda_{ab}B(a,b)+\lambda_{ac}B(a,c)+\lambda_{ad}B(a,d)+\lambda_{bc}B(b,c)+\lambda_{bd}B(b,d)+\lambda_{cd}B(c,d)}\right] = \mathrm{O}(2^{-n/2}).$$

We may assume $\lambda_{ab} = 1$ by symmetry. It suffices to show that for every $c, d \in \mathbb{F}_2^n$,

$$\mathbb{E}_{a,b\in\mathbb{F}_2^n}\left[(-1)^{B(a,b)+\lambda_{ac}B(a,c)+\lambda_{ad}B(a,d)+\lambda_{bc}B(b,c)+\lambda_{bd}B(b,d)+\lambda_{cd}B(c,d)}\right] = \mathrm{O}(2^{-n/2}).$$

Since B is a bilinear form, $B(\cdot, c)$ and $B(\cdot, d)$ are linear forms and we can rewrite the above quantity as

$$\mathbb{E}_{a,b\in\mathbb{F}_2^n}\left[(-1)^{B(a,b)+L_1(a)+L_2(b)}\right],$$

where $L_1 := \lambda_{ac}B(\cdot, c) + \lambda_{ad}B(\cdot, d)$ and $L_2 := \lambda_{bc}B(\cdot, c) + \lambda_{bd}B(\cdot, d)$ are two linear forms. The term $L_2(b)$ can be removed by an application of the Cauchy-Schwarz inequality on a, and we obtain

$$\left|\mathbb{E}_{a,b\in\mathbb{F}_2^n}\left[(-1)^{B(a,b)+L_1(a)+L_2(b)}\right]\right|^2 \leq \mathbb{E}_{a,a',b\in\mathbb{F}_2^n}\left[(-1)^{B(a+a',b)+L_1(a+a')}\right].$$

Now observe that by the definition of B, $\mathbb{E}_b\left[(-1)^{B(a+a',b)+L_1(a+a')}\right] = 0$ whenever $a \neq a'$. Thus we have

$$\mathbb{E}_{a,a',b\in\mathbb{F}_2^n}\left[(-1)^{B(a+a',b)+L_1(a+a')}\right] = \Pr_{a,a'\in\mathbb{F}_2^n}[a = a'] = 2^{-n},$$

as was desired. □

Theorem 5.2.3 implies that the joint distribution of $B_6(a,b,c,d)$ is $O(2^{-n/2})$ close in statistical distance to uniform over $\mathbb{F}_2^6$. Thus, if we let $\tau = (\tau_{ab}, \tau_{ac}, \tau_{ad}, \tau_{bc}, \tau_{bd}, \tau_{cd}) \in \mathbb{F}_2^6$ to be uniformly chosen, then we can approximate the U^4 norm of S_4 by

$$\begin{aligned}\|(-1)^{S_4}\|_{U^4}^{16} &= \mathbb{E}_{a,b,c,d\in\mathbb{F}_2^n}\left[(-1)^{B(a,b)B(c,d)+B(a,c)B(b,d)+B(a,d)B(b,c)}\right] \\ &= \mathbb{E}_{\tau\in\mathbb{F}_2^6}\left[(-1)^{\tau_{ab}\tau_{cd}+\tau_{ac}\tau_{bd}+\tau_{ad}\tau_{bc}} + \mathrm{O}(2^{-n/2})\right] \\ &= \left(\mathbb{E}_{u,v\in\mathbb{F}_2}\left[(-1)^{uv}\right]\right)^3 + \mathrm{O}(2^{-n/2}) = \frac{1}{8} + \mathrm{O}(2^{-n/2}).\end{aligned}$$

5.2.2 Bounds on correlation with cubic polynomials

The exponential bound on the correlation with cubic polynomials was obtained in [57], the proof of which is involved. We instead present the proof of a weaker bound from [41] which is still sufficient to refute Theorem 5.1.4. Green and Tao [41] obtain the following bound using a modification of a clever Ramsey-theoretic argument by Alon and Beigel [1].

Theorem 5.2.4 ([41])**.** *Let $n \geq 1$ be large enough. Then for any polynomial $Q : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most 3,*

$$\left|\Pr_x[S_4(x) = Q(x)] - \frac{1}{2}\right| \leq \exp(-c\log\log\log n),$$

for some absolute constant $c > 0$.

We first show that S_4 has no correlation with symmetric cubic polynomials.

Lemma 5.2.5. *For any $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$,*

$$\mathbb{E}_{x\in\mathbb{F}_2^n}\left[(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0}\right] \leq \exp(-cn)$$

for some absolute constant $c > 0$.

Proof. Let $|x|$ denote the hamming weight of x, that is the number of nonzero coordinates in x. Let $|x| = \sum_i b_i(x)2^i$ be its binary representation with $b_i(x) \in \{0,1\}$. By Lucas' theorem on binomial coefficients [58],

$$S_1(x) = b_0(x), \quad S_2(x) = b_1(x), \quad S_3(x) = b_0(x)b_1(x), \quad S_4(x) = b_2(x).$$

Thus

$$\mathbb{E}_{x\in\mathbb{F}_2^n}\left[(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0}\right] = \mathbb{E}_{x\in\mathbb{F}_2^n}\left[(-1)^{b_2(x)+c_3b_0b_1(x)+c_2b_1(x)+c_1b_0(x)+c_0}\right].$$

Note that if $b_0(x), b_1(x), b_2(x)$ were uniformly distributed in $\{0,1\}^3$, then the average would be zero. To conclude, we need to show that they are close to uniformly distributed. Equivalently, we will show that $|x| \bmod 8 = \sum_{i=0}^2 b_i(x)2^i$ is close to uniformly distributed in $\mathbb{Z}_8$. For any $a \in \mathbb{Z}_8$ we have

$$\begin{aligned}\Pr_{x\in\mathbb{F}_2^n}[|x| \bmod 8 = a] &= 2^{-n}\sum_{k=0}^n \binom{n}{k} 1_{k \bmod 8 = a} \\ &= 2^{-n}\sum_{k=0}^n \binom{n}{k} \frac{\sum_{r=0}^7 e^{2\pi i r(k-a)/8}}{8} \\ &= \frac{1}{8}\sum_{r=0}^7 e^{-2\pi i r a/8}\left(\frac{1+e^{-2\pi i r/8}}{2}\right)^n.\end{aligned}$$

The term corresponding for $r = 0$ is equal to $1/8$. For any $r \neq 0$, we can bound its corresponding term contribution by $|(1 + e^{-2\pi ir/8})/2|^n \leq \exp(-cn)$, for some $c > 0$. Thus

$$\left| \Pr_{x \in \mathbb{F}_2^n}[|x| \bmod 8 = a] - \frac{1}{8} \right| \leq \exp(-cn).$$

□

The proof of Theorem 5.2.4 follows from a Ramsey-type argument, which allows to reduce any polynomial to a symmetric polynomial with fewer variables.

Lemma 5.2.6. *Let $Q : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree ≤ 3. Then there exists a symmetric polynomial $Q' : \mathbb{F}_2^{n'} \to \mathbb{F}_2$ of degree ≤ 3, with $n' = \Omega(\log\log\log n)$, such that*

$$\Pr_{x \in \mathbb{F}_2^{n'}}[Q'(x) = S_4(x)] \geq \Pr_{x \in \mathbb{F}_2^n}[Q(x) = S_4(x)].$$

Proof. Let Q be a degree 3 polynomial. Decompose

$$Q(x) = \sum q_{i,j,k} x_i x_j x_k + Q_2(x),$$

where Q_2 is a quadratic polynomial. Let H be a 3-uniform hypergraph with $V(H) = [n]$ and $E(H) = \{(i,j,k) : q_{i,j,k} = 1\}$. By the hypergraph Ramsey theorem [23, 22] there exists a set $A \subset [n]$ of size $|A| \geq \Omega(\log\log n)$ such that A is either a clique or an independent set for H. For $x \in \mathbb{F}_2^A, a \in \mathbb{F}_2^{[n]\setminus A}$ let $S_4(x,a)$ and $Q(x,a)$ be the polynomials with the appropriate inputs. By an averaging argument, there exists an assignment $a \in \mathbb{F}_2^{[n]\setminus A}$ such that

$$\Pr_{x \in \mathbb{F}_2^A}[S_4(x,a) = Q(x,a)] \geq \Pr_{x \in \mathbb{F}_2^n}[S_4(x) = Q(x)].$$

Note that $S_4(x,a)$ is a symmetric polynomial in x of degree 4, whose homogenous degree 4 part is $S_4(x)$; and $Q(x,a)$ is a polynomial of degree ≤ 3, whose homogeneous degree 3 part is symmetric, equal either to 0 or to $S_3(x)$. Next, we make the quadratic part symmetric. Let

$$Q_2(x,a) = \sum_{i,j \in A} q'_{i,j} x_i x_j + Q_1(x),$$

where $Q_1(x)$ is a linear polynomial. Applying a similar argument, let G be a graph with $V(G) = A$ and $E(G) = \{(i,j) \in A : q'_{i,j} = 1\}$. By Ramsey's theorem for graphs [24], there exists a subset $B \subset A$ of size $|B| \geq \Omega(\log|A|)$ which is either a clique or an independent set. Thus, there exists an assignment $b \in \mathbb{F}_2^{A\setminus B}$ for which

$$\Pr_{x \in \mathbb{F}_2^B}[S_4(x,a,b) = Q(x,a,b)] \geq \Pr_{x \in \mathbb{F}_2^A}[S_4(x,a) = Q(x,a)] \geq \Pr_{x \in \mathbb{F}_2^n}[S_4(x) = Q(x)].$$

Note that $Q(x,a,b)$ has homogeneous parts of degrees 3 and 2 both being symmetric. To conclude, let

$$Q_1(x,a,b) = \sum_{i \in B} q''_i x_i + q'''.$$

Let $C \subset B$ be a set for which q''_i for $i \in B$ are all equal, where $|C| \geq |B|/2$. There exists an assignment $c \in \mathbb{F}_2^{B\setminus C}$ such that

$$\Pr_{x \in \mathbb{F}_2^C}[S_4(x,a,b,c) = Q(x,a,b)] \geq \Pr_{x \in \mathbb{F}_2^n}[S_4(x) = Q(x)].$$

The lemma follows as $S_4(x,a,b,c) + Q(x,a,b)$ is a symmetric polynomial of degree 4 in $|C| = \Omega(\log\log\log n)$ variables, whose homogeneous part of degree 4 is equal to $S_4(x)$. □

Proof of Theorem 5.2.4. Let $Q(x)$ be a cubic polynomial which maximizes $|\Pr[Q(x) = S_4(x)] - 1/2|$. By possibly replacing Q with $Q + 1$, we may assume that $\Pr[Q(x) = S_4(x)] \geq 1/2$. By Theorem 5.2.6, there exists a symmetric cubic polynomial Q' on $n' = \Omega(\log\log\log n)$ variables such that $\Pr[Q(x) = S_4(x)] \leq \Pr[Q'(x) = S_4(x)]$. By Theorem 5.2.5, $\Pr[Q'(x) = S_4(x)] \leq 1/2 + \exp(-cn')$. □

Part II

Higher Order Fourier Analysis

In Part II of the survey we will introduce the foundations of higher-order Fourier analysis in detail. In classical Fourier analysis characters are the exponentials of linear functions. Since a typical random function $f : \mathbb{F}_p^n \to [-1,1]$ has very small correlation with all the Fourier characters, it is natural to consider the characters as highly structured functions. This leads to a "structure" versus "pseudorandomness" dichotomy in which functions that have small correlation with all the Fourier characters are regarded as pseudorandom functions, and the ones that have noticeable correlation with a Fourier character are considered to be somewhat structured.

Consequently, the large Fourier coefficients of $f : \mathbb{F}_p^n \to [-1,1]$ constitute the structured part of f, while the small Fourier coefficients correspond to its pseudorandom part. This can be formulated as an "approximate structure theorem" that says that, we can decompose any function f as $f = f_1 + f_2$, where f_1 is structured as it is a linear combination of few Fourier characters, and f_2 is pseudorandom in that all of the Fourier coefficients of f_2 are small. Such approximate structure theorems are very useful as for many problems such as estimating the probability of the success of linearity test, or estimating the density of 3-term arithmetic progressions in a set. The reason is that it is possible to treat f_2 as a small random noise, and extract the desired information about f from the highly structured part, namely f_1.

In order to be able to discard f_2 safely, we need to deduce from the fact that f_2 does not have a noticeable correlation with any Fourier character that it does not contribute significantly to the estimated parameter. In other words, we need an inverse theorem which would state that significant contribution implies significant correlation with a character. Although, at first glance, it might appear that one needs to establish a separate inverse theorem for every problem that one wishes to study via this theory, fortunately, this is not the case. It suffices to prove an inverse theorem for the Gowers U^2 norm. It turns out that for many problems, a few applications of the classical Cauchy-Schwarz inequality show that f_2 can be discarded if $\|f_2\|_{U^2}$ is small.

The structure versus pseudo-randomness dichotomy discussed above is not sufficiently strong to be applicable to more complex linear structures. In other words discarding f_2, even when $\|f_2\|_{U^2}$ is very small, can have a significant affect on certain parameters such as the probability of success of AKKLR(d) test for $d > 1$, or the density of four-term arithmetic progressions in a set. To resolve this issue, Gowers [35] defined higher uniformity norms U^d, and found a new proof for Szemerédi's theorem by partially extending the above program for these norms. Indeed, as we shall see in Chapter 11, for every linear pattern, there exists a d such that if $\|f_2\|_{U^d}$ is sufficiently small, then for the purpose of estimating the density of that pattern, the pseudo-random f_2 can be discarded without having a significant effect on the estimate. However, this fact by itself is not very useful without an inverse theorem for U^d norm as one needs an inverse theorem to extract a structured part from every function with noticeable U^d norm.

In Theorem 5.1.3 we established an inverse theorem for the U^3 norm. This theorem lays the foundation of quadratic Fourier analysis as it implies that every function $f : \mathbb{F}_p^n \to [-1,1]$ can be decomposed into a structured part that is a linear combination of the exponentials of a few quadratic characters and a pseudo-random part that has small U^3 norm. Unfortunately, as we saw in Theorem 5.2.2 its most natural generalization Theorem 5.1.4 to higher uniformity norms is false. In other words, classical polynomials are not the right structure for higher uniformity norms. Tao and Ziegler [79] proved that Theorem 5.1.4 can be fixed by replacing classical polynomials with a generalization of classical polynomials, which we will refer to as non-classical polynomials.

We introduce non-classical polynomials and state Tao and Ziegler's inverse theorem in Chapter 6. We then proceed to develop the theory of higher order Fourier analysis in the remainder of this part. Concretely, in Chapter 7 we introduce the notions of rank, regularity and uniformity for polynomials, which are the basis for the structure vs pseudo-randomness dichotomy discussed above. In Chapter 9 we develop the decomposition theorems. In Chapter 10 we develop a notion of homogeneous non-classical polynomials, which is useful in certain applications. In Chapter 11 we apply these to the study of linear patterns in sets. A few technical proofs from these chapters are deferred to Chapter 12.

Chapter 6

Nonclassical Polynomials, and the Inverse Gowers Theorem

In Theorem 5.2.2 we saw that the most natural generalization of the U^2 and U^3 inverse theorem to higher uniformity norms is false. That is for $d \geq 4$, there exist bounded functions that have large U^d norm but do not have significant correlation with the exponential of any polynomial of degree $d-1$. In this chapter, we address this issue by introducing an extension of classical polynomials, called *nonclassical polynomials*. We will show that an inverse theorem for Gowers uniformity norms of any order holds for nonclassical polynomials. Below we will give some intuition on how we can arrive to the definition of nonclassical polynomials.

Let $d \geq 0$ be an integer. Similar to the situation over finite fields, there are both a global and a local definition for a low degree real-valued polynomial A real-valued function $P : \mathbb{R}^n \to \mathbb{R}$ being a polynomial of (total) degree $\leq d$ can be defined in two equivalent ways:

(1) **Global definition.** P is a polynomial of degree $\leq d$ if it can be written as

$$P(x_1, \cdots, x_n) = \sum_{\substack{i_1, \cdots, i_n \geq 0 \\ i_1 + \cdots + i_n \leq d}} c_{i_1, \cdots, i_n} x_1^{i_1} \cdots x_n^{i_n},$$

with coefficients $c_{i_1, \cdots, i_n} \in \mathbb{R}$.

(2) **Local definition.** P is a polynomial of degree $\leq d$ if it is $d+1$ times differentiable and its $(d+1)$-th derivative vanishes everywhere.

It is easy to see by linearity of the derivative operator that (i) implies (ii). For the other direction, one can use the Taylor series expansion to go from the local to the global condition, and show that the two definitions above are equivalent. Both these definitions can be extended to the finite characteristic setting, i.e. when $P : \mathbb{F}^n \to G$ for a finite field $\mathbb{F}$ and an Abelian group G. The global definition extends in a straightforward manner, and the local definition uses the notion of additive directional derivatives.

Definition 6.0.1 (Polynomials over finite fields (local definition))**.** *For an integer $d \geq 0$, a function $P : \mathbb{F}^n \to G$ is said to be a* polynomial of degree $\leq d$ *if for all $y_1, \ldots, y_{d+1}, x \in \mathbb{F}^n$, it holds that*

$$(D_{y_1} \cdots D_{y_{d+1}} P)(x) = 0,$$

where $D_y P(x) = P(x+y) - P(x)$ is the additive derivative of P with direction y evaluated at x. The degree *of P is the smallest d for which the above holds.*

It follows simply from the definition that for any direction $y \in \mathbb{F}^n$, $\deg(D_y P) < \deg(P)$. In the "classical" case of polynomials $P : \mathbb{F}^n \to \mathbb{F}$, it is a well-known fact that the global and local definitions coincide. However, the situation is different in more general groups. One can already suspect this from the fact that division by

$d!$ may not be possible in G, and hence one cannot make use of the Taylor expansion to go from the local definition to the global one.

When the range of P is the torus $\mathbb{R}/\mathbb{Z}$, it turns out that the global definition must be refined to the "nonclassical polynomials", which may have monomials that are different from the classical case. This phenomenon was first noted by Tao and Ziegler [79] in the study of Gowers norms, where it was also proved that Theorem 5.1.4 can be modified to hold by replacing classical polynomials with nonclassical polynomials.

6.1 Nonclassical polynomials

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the torus (namely, the group of addition modulo 1). Let $e : \mathbb{T} \to \mathbb{C}$ be given by $e(x) = \exp(2\pi i x)$. Fix a prime finite field $\mathbb{F} = \mathbb{F}_p$ for the remainder of this chapter. Nonclassical polynomials arise when studying functions $P : \mathbb{F}^n \to \mathbb{T}$ and their phase functions $f = e(P) : \mathbb{F}^n \to \mathbb{C}$.

Definition 6.1.1 (Nonclassical Polynomials)**.** *For an integer $d \geq 0$, a function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a* nonclassical polynomial of degree $\leq d$ *(or simply a* polynomial of degree $\leq d$*) if for all $y_1, \ldots, y_{d+1}, x \in \mathbb{F}^n$, it holds that*

$$(D_{y_1} \cdots D_{y_{d+1}} P)(x) = 0. \tag{6.1}$$

The degree *of P is the smallest d for which the above holds. A function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a* classical polynomial of degree $\leq d$ *if it is a nonclassical polynomial of degree $\leq d$ whose image is contained in $\frac{1}{p}\mathbb{Z}/\mathbb{Z}$.*

Denote by $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$, $\mathrm{Poly}_d(\mathbb{F}^n \to \mathbb{T})$ *and* $\mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$*, the set of all nonclassical polynomials over $\mathbb{F}^n$, all nonclassical polynomials of degree d and all nonclassical polynomials of degree $\leq d$, respectively.*

The following lemma of Tao and Ziegler [79] shows that a *classical polynomial* P of degree d must always be of the form $x \mapsto \frac{|Q(x)|}{p}$, where $Q : \mathbb{F}^n \to \mathbb{F}$ is a polynomial (in the usual sense) of degree d, and $|\cdot|$ is the canonical map from $\mathbb{F} = \mathbb{F}_p$ to $\{0, 1, \ldots, p-1\}$. Moreover, it provides a global characterization of the structure of nonclassical polynomials.

Lemma 6.1.2 (Lemma 1.7 in [79])**.** *A function $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if and only if P can be represented as*

$$P(x_1, \ldots, x_n) = \alpha + \sum_{\substack{0 \leq d_1, \ldots, d_n < p;\ k \geq 0: \\ 0 < d_1 + \ldots + d_n \leq d - k(p-1)}} \frac{c_{d_1,\ldots,d_n,k} |x_1|^{d_1} \cdots |x_n|^{d_n}}{p^{k+1}} \ \mathit{mod}\ 1,$$

for a unique choice of $c_{d_1,\ldots,d_n,k} \in \{0, 1, \ldots, p-1\}$ and $\alpha \in \mathbb{T}$. The element α is called the shift *of P, and the largest integer k such that there exist $d_1, \ldots, d_n$ for which $c_{d_1,\ldots,d_n,k} \neq 0$ is called the* depth *of P. A depth-k polynomial P takes values in a coset of the subgroup $\mathbb{U}_{k+1} := \frac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}$. Classical polynomials correspond to polynomials with 0 shift and 0 depth.*

In many cases, for the sake of brevity, we will omit writing "mod 1" in the description of the defined nonclassical polynomials. We do not include the proof of Theorem 6.1.2 in this text. For a proof we refer the reader to either the original paper [79] or to a blog post of Tao [75] describing the proof.

Example 6.1.3. *Consider the univariate function $f : \mathbb{F}_2 \to \mathbb{T}$ given by $f(x) = \frac{|x|}{4}$ mod 1. Is is a nonclassical polynomial of degree 2. To see this, we compute its derivatives $D_y f$ for $y \in \mathbb{F}_2$. Clearly $D_0 f = 0$ for any univariate function, so it suffices to compute $D_1 f$. One can verify that*

$$D_1 f(x) = f(x \oplus 1) - f(x) = \begin{cases} 1/4 & x = 0 \\ -1/4 & x = 1 \end{cases} = \frac{1}{4} + \frac{|x|}{2} \ \mathit{mod}\ 1$$

and

$$D_1 D_1 f(x) = \frac{1}{2}, \qquad D_1 D_1 D_1 f(x) = 0.$$

Thus, taking 3 derivatives annihilates f, but not 2 derivatives. So by definition it is a degree 2 polynomial. Similarly, one can show that $\frac{|x|}{2^k}$ is a nonclassical polynomial of degree k.

Remark 6.1.4. *An equivalent definition of nonclassical polynomials is via functions which map $\mathbb{F}_p^n$ to the ring $\mathbb{Z}_{p^{k+1}}$. Concretely, if $f : \mathbb{F}_p^n \to \mathbb{T}$ is a nonclassical polynomial of degree d and depth k, then its takes values in $\frac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}$. Thus, we can write f as*

$$f(x) = \frac{F(x)}{p^{k+1}} \ mod\ 1,$$

where $F : \mathbb{F}_p^n \to \mathbb{Z}_{p^{k+1}}$. One can verify that F is a polynomial. However, note that the classical definition of degree for F (based on monomials) does not coincide with our definition of degree for f (based on derivatives). For example, in the univariate case $F(x) = x$ corresponds to $f(x) = \frac{x}{p^{k+1}}$ which has degree $1 + (p-1)k$, while $F(x) = px$ corresponds to $f(x) = \frac{x}{p^k}$ which has degree $1 + (p-1)(k-1)$.

Note that Theorem 6.1.2 immediately implies the following important corollary. Below we use the following standard shorthand, which follows as $\mathbb{T}$ is an Abelian group. For $n \in \mathbb{Z}$ and $x \in \mathbb{T}$, nx stands for $x + \cdots + x$ if $n \geq 0$ and $-x - \cdots - x$ otherwise, where there are $|n|$ terms in both expressions.

Corollary 6.1.5. *Let $Q : \mathbb{F}_p^n \to \mathbb{T}$ be a polynomial of degree d and depth k. Then*

(1) If $\lambda \in \mathbb{Z}$ is co-prime to p then λQ also has degree d and depth k.

(2) pQ has degree $\max(d - p + 1, 0)$ and depth $k - 1$. In other words, if Q is classical, then pQ vanishes, and otherwise, its degree decreases by $p - 1$ and its depth by 1.

(3) $p^{k+1}Q = 0$. This implies that if $\lambda \equiv \lambda' \ mod\ p^{k+1}$ then $\lambda Q = \lambda' Q$.

For convenience of exposition, henceforth we will assume that the shifts of all polynomials are zero. This can be done without affecting any of the results presented in this text. Under this assumption, all polynomials of depth k take values in $\mathbb{U}_{k+1}$.

6.2 The inverse theorem for Gowers norms

Gowers norms, which were introduced by Gowers [35], play an important role in additive combinatorics, more specifically in the study of polynomials of bounded degree. We have seen in the direct theorem for Gowers norms (Theorem 4.1.1; see also Theorem 5.1.3, direct theorem) that correlation with polynomials implies large Gowers norm. The proof generalizes directly to include nonclassical polynomials.

Theorem 6.2.1 (Direct Theorem for Gowers Norm)**.** *Let $f : \mathbb{F}^n \to \mathbb{C}$ be a function and $d \geq 1$ an integer. Then for every degree-d nonclassical polynomial $P : \mathbb{F}^n \to \mathbb{T}$,*

$$|\langle f, e(P)\rangle| \leq \|f\|_{U^{d+1}}.$$

The theorem follows from the monotonicity of the Gowers norms.

Claim 6.2.2. *Let $f : \mathbb{F}^n \to \mathbb{C}$ and $d \geq 1$. Then $\|f\|_{U^{d+1}} \geq \|f\|_{U^d}$.*

Proof. We have

$$\begin{aligned}
\|f\|_{U^{d+1}}^{2^{d+1}} &= \mathbb{E}_{x,y_1,\ldots,y_{d+1}}\left[\Delta_{y_1,\ldots,y_{d+1}} f(x)\right] \\
&= \mathbb{E}_{y_1,\ldots,y_d}\mathbb{E}_{x,y_{d+1}}\left[\Delta_{y_1,\ldots,y_d} f(x + y_{d+1}) \cdot \overline{\Delta_{y_1,\ldots,y_d} f(x)}\right] \\
&= \mathbb{E}_{y_1,\ldots,y_d}\left|\mathbb{E}_x\left[\Delta_{y_1,\ldots,y_d} f(x)\right]\right|^2 \\
&\geq \left|\mathbb{E}_{x,y_1,\ldots,y_d}\left[\Delta_{y_1,\ldots,y_d} f(x)\right]\right|^2 \\
&= \|f\|_{U^d}^{2^{d+1}}.
\end{aligned}$$

□

Proof of Theorem 6.2.1. Let $g(x) = f(x)e(P(x))$. Then

$$|\langle f, e(P)\rangle| = \|g\|_{U^1} \leq \|g\|_{U^{d+1}}.$$

Now, for any $y_1, \ldots, y_{d+1} \in \mathbb{F}^n$ it holds that

$$\Delta_{y_1,\ldots,y_{d+1}} g(x) = \Delta_{y_1,\ldots,y_{d+1}} f(x) \cdot e(D_{y_1,\ldots,y_{d+1}} P(x)) = \Delta_{y_1,\ldots,y_{d+1}} f(x),$$

and hence

$$\|g\|_{U^{d+1}}^{2^{d+1}} = \mathbb{E}_{y_1,\ldots,y_{d+1}\in\mathbb{F}^n}\left[\Delta_{y_1,\ldots,y_{d+1}} g(x)\right] = \mathbb{E}_{y_1,\ldots,y_{d+1}\in\mathbb{F}^n}\left[\Delta_{y_1,\ldots,y_{d+1}} f(x)\right] = \|f\|_{U^{d+1}}^{2^{d+1}}.$$

□

As we showed in Section 5.2, the inverse direction is false, when we restrict ourselves to classical polynomials. However, Tao and Ziegler [79] proved that it is true if we include nonclassical polynomials. The inverse theorem only applies to bounded functions. Let $\mathbb{D} := \{z \in \mathbb{C} : |z| \leq 1\}$ be the unit disk in the complex plane.

Theorem 6.2.3 (Inverse Theorem for Gowers Norm (Theorem 1.11 of [79]))**.** *Fix $d \geq 1$ an integer and $\varepsilon > 0$. There exists an $\delta = \delta_{6.2.3}(\mathbb{F}, d, \varepsilon)$ such that the following holds. For every function $f : \mathbb{F}^n \to \mathbb{D}$ with $\|f\|_{U^{d+1}} \geq \varepsilon$, there exists a polynomial $P \in \mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$ that is δ-correlated with f, that is*

$$|\langle f, e(P)\rangle| \geq \delta.$$

It is easy to see that for every degree d nonclassical polynomial P, $\|e(P)\|_{U^{d+1}} = 1$. Theorem 6.2.1 and Theorem 6.2.3 provide a robust version of this statement, showing that the Gowers norm of a function f is large if and only if it contains some low degree structure, namely if f correlates with a low degree polynomial. We also note that Theorem 6.2.3 only shows existence of a constant $\delta > 0$ for every ϵ and finding reasonable quantitative bounds or limitations to such bounds is a fascinating problem which to this day remains unsolved. In the case of quadratics $d = 1$, we have $\|f\|_{U^2} = \|\hat{f}\|_4$ which gives $\delta = \Omega(\epsilon^2)$. The case of $d = 2$ is already nontrivial and the best known lower-bound for δ is quasi-polynomial in ϵ, which follows from the work of Sanders on the Bogolyubov-Ruzsa conjecture [67] (see also [55, 56]).

Chapter 7

Rank, Regularity, and Other Notions of Uniformity

Consider a function $f : \mathbb{F}_p^n \to [-1, 1]$ and a positive integer d. As we shall see in Chapter 9 the inverse theorems (Theorem 5.1.3 and Theorem 6.2.3) allow us to approximate f in the U^{d+1} norm by a linear combination of a few higher-order *phase functions*, which are exponentials of nonclassical polynomials of degree d. We will think of this as an order-d Fourier expansion of f, and regard it as the structured part of f. Unlike classical Fourier expansion, higher-order Fourier expansions are not unique by any means. This is simply because there are too many polynomials of degree d, (there are asymptotically $2^{O(n^d)}$ classical degree $\leq d$ polynomials), and thus they cannot form a linear basis for the space of functions whose dimension is p^n. However, this might be disappointing as one of the most important and useful properties of the classical Fourier characters is that they form an orthonormal basis. While it is not possible to achieve this orthogonality in an exact way, we can still hope to obtain an approximate version of it by approximating f with a linear combination of a few higher-order phase functions which are pairwise almost orthogonal. We will see how to accomplish this through the notions of *rank* and *regularity*, which are the topic of this chapter. We will present these topics for the general case of an arbitrary fixed value of d, and unfortunately this leads to many technicalities that might be overwhelming for the readers who see these materials for the first time. Some of these technicalities are unnecessary for the case of the quadratic Fourier analysis, i.e. $d = 2$, and many definitions and statements are more intuitive and familiar in this case, and for example the definition of rank coincides with the familiar notion of rank from linear algebra. Thus we recommend those readers who find this chapter too technical and unintuitive to first consult the excellent lecture notes of Ben Green on quadratic Fourier analysis [39].

A property of Fourier characters is that they behave like a random function, for example their value is uniformly distributed when the input is drawn uniformly at random. Having this in mind, we can demand several such random-like behaviors from higher-degree polynomials. We refer to such properties as *regularity*. One property of random functions is that they cannot be expressed as a function of few low-degree polynomials. We will capture this property by defining the notion of *rank* of a polynomial, intended to capture its complexity according to lower degree polynomials.

Definition 7.0.1 (Rank of a polynomial)**.** *Given a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ and an integer $d \geq 1$, the* d-rank *of P, denoted $\mathsf{rank}_d(P)$, is defined to be the smallest integer r such that there exist polynomials $Q_1, \ldots, Q_r : \mathbb{F}^n \to \mathbb{T}$ of degree $\leq d-1$ and a function $\Gamma : \mathbb{T}^r \to \mathbb{T}$ satisfying $P(x) = \Gamma(Q_1(x), \ldots, Q_r(x))$. If $d = 1$, then 1-rank is defined to be ∞ if P is non-constant, and 0 otherwise.*

The rank *of a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is its $\deg(P)$-rank. We say that P is r-regular if $\mathsf{rank}(P) \geq r$.*

Note that for an integer $1 \leq \lambda \leq p-1$, $\mathrm{rank}(P) = \mathrm{rank}(\lambda P)$. A high-rank polynomial of degree d is, intuitively, a "generic" degree-d polynomial; there are no unexpected ways to decompose it into lower degree polynomials. For future use, we record here a simple lemma stating that restrictions of high rank polynomials to hyperplanes generally preserve degree and high rank.

Lemma 7.0.2. *Suppose $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree d and rank $\geq r$, where $r > p + 1$. Let A be a hyperplane in $\mathbb{F}^n$, and denote by P' the restriction of P to A. Then, P' is a polynomial of degree d and rank $\geq r - p$, unless $d = 1$ and P is constant on A.*

Proof. For the case $d = 1$, we can check directly that either P' is constant or else, P' is a non-constant degree-1 polynomial and so has rank infinity.

So, assume $d > 1$. By applying an affine transformation, we can assume without loss of generality that A is the hyperplane $\{x : x_1 = 0\}$. Let $\pi : \mathbb{F}^n \to \mathbb{F}^{n-1}$ be the projection to A given by $\pi(x_1, x_2, \ldots, x_n) = (0, x_2, \ldots, x_n)$. Let $P'' = P - P' \circ \pi$. Clearly, P'' is zero on A. For $a \in \mathbb{F} \setminus \{0\}$, let $h_a = (a, 0, \ldots, 0) \in \mathbb{F}^n$. Note that $D_{h_a} P''$ is of degree $\leq d - 1$ and that $(D_{h_a} P'')(y) = P''(y + h_a)$ for all $y \in A$. Hence, for every $a \in \mathbb{F} \setminus \{0\}$, P'' on $A + h_a$ agrees with a polynomial Q_a of degree $\leq d - 1$. So, for a function $\Gamma : \mathbb{T}^{p+1} \to \mathbb{T}$, we can write $P(x) = \Gamma(|x_1|/p, P'(x), Q_1(x), Q_2(x), \ldots, Q_{p-1}(x))$, where $|x_1|/p, Q_1, \ldots, Q_{p-1}$ are of degree $\leq d - 1$.

Now, if P' itself is of degree $d-1$, then P is of rank $\leq p+1 < r$, a contradiction. If P' is of rank $< r - p$, then again P is of rank $< r - p + p = r$, a contradiction. □

7.1 Polynomial factors

Next, we will formalize the notion of a generic collection of polynomials. Intuitively, it should mean that there are no unexpected algebraic dependencies among the polynomials. First, we need to set up some notation.

Definition 7.1.1 (Factors)**.** *If X is a finite set, then by a* factor $\mathcal{B}$ *we simply mean a partition of X into finitely many parts called* atoms*.*

A finite collection of functions $\phi_1, \ldots, \phi_C$ from X to some other space Y naturally define a factor $\mathcal{B} = \mathcal{B}_{\phi_1,\ldots,\phi_C}$ whose atoms are sets of the form $\{x \in X : (\phi_1(x), \ldots, \phi_C(x)) = (y_1, \ldots, y_C)\}$ for some $(y_1, \ldots, y_C) \in Y^C$. By an abuse of notation we also use $\mathcal{B}$ to denote the map $x \mapsto (\phi_1(x), \ldots, \phi_C(x))$, thus also identifying the atom containing x with $(\phi_1(x), \ldots, \phi_C(x))$.

Definition 7.1.2 (Polynomial factors)**.** *If $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ is a sequence of polynomials, then the factor $\mathcal{B}_{P_1,\ldots,P_C}$ is called a* polynomial factor*.*

The complexity *of $\mathcal{B}$, denoted $|\mathcal{B}| := C$, is the number of defining polynomials. The* degree *of $\mathcal{B}$ is the maximum degree among its defining polynomials $P_1, \ldots, P_C$. If $P_1, \ldots, P_C$ are of depths $k_1, \ldots, k_C$, respectively, then the number of atoms of $\mathcal{B}$ is at most $\|\mathcal{B}\| := \prod_{i=1}^{C} p^{k_i+1}$.*

Note that it makes a difference whether we define atoms according to the ordered set of evaluations $(P_1(x), \ldots, P_C(x))$ rather than the multiset $\{P_1(x), \ldots, P_C(x)\}$, and we choose the former in our definition.

We say that a function $f : \mathbb{F}^n \to \mathbb{T}$ is *measurable* in a polynomial factor $\mathcal{B}_{P_1,\ldots,P_C}$ (or *$\mathcal{B}$-measurable* in short) if there exists a $\Gamma : \mathbb{T}^C \to \mathbb{T}$ such that

$$f = \Gamma(P_1, \ldots, P_C),$$

or in other words the value of $f(x)$ can be determined by knowing only the values of $P_1(x), \ldots, P_C(x)$. Note that, here Γ is an arbitrary map with no restriction on its degree or complexity. Equivalently, we say that f is *$\mathcal{B}$-measurable* if f is constant over each atom of $\mathcal{B}$.

Example 7.1.3. *The function $f(x) := \frac{x_1 x_2 + x_1 x_3}{2} + \frac{x_3}{4}$ is measurable in the factor defined by $P_1 = \frac{x_1}{2}$, $P_2 = \frac{x_2 + x_3}{2}$ and $P_3 = \frac{x_3}{4}$.*

Next we define conditional expectation over a factor, which results in a function that is constant on each atom of the factor.

Definition 7.1.4 (Conditional Expectation over Factors)**.** *Let $\mathcal{B}$ be a polynomial factor defined by $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$. For $f : \mathbb{F}^n \to \mathbb{C}$, the conditional expectation of f with respect to $\mathcal{B}$, denoted $\mathbb{E}[f|\mathcal{B}] : \mathbb{F}^n \to \mathbb{C}$, is*

$$\mathbb{E}[f|\mathcal{B}](x) := \mathbb{E}_{y \in \mathbb{F}^n}\left[f(y)\middle|P_1(y) = P_1(x), \ldots, P_C(y) = P_C(x)\right].$$

Namely, $\mathbb{E}[f|\mathcal{B}]$ is constant on every atom of $\mathcal{B}$, and takes the average value of f over this atom.

Note that $\mathbb{E}[f|\mathcal{B}]$ is $\mathcal{B}$-measurable. The following is a simple observation stating that $E[f|\mathcal{B}]$ has the same correlation as f with any $\mathcal{B}$-measurable function.

Remark 7.1.5. *Let $f : \mathbb{F}^n \to \mathbb{C}$. Let $\mathcal{B}$ be a polynomial factor defined by polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$, and let $g : \mathbb{F}^n \to \mathbb{C}$ be a $\mathcal{B}$-measurable function. Then*

$$\langle f, g\rangle = \langle \mathbb{E}(f|\mathcal{B}), g\rangle.$$

Finally, we define the rank of a polynomial factor. We require that every nonzero linear combination of the polynomials which define the factor has high rank. Recall that by Theorem 6.1.5, if $P : \mathbb{F}_p^n \to \mathbb{T}$ is a polynomial of depth k then for $\lambda \in \mathbb{Z}$, the depth of λP depends only on $\lambda \bmod p^{k+1}$.

Definition 7.1.6 (Rank of a Factor)**.** *Let $\mathcal{B}$ be a polynomial factor defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$. The rank of $\mathcal{B}$ is the least integer r, for which there exists $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \bmod p^{k_1+1}, \ldots, \lambda_C \bmod p^{k_C+1}) \neq 0^C$, such that $\mathrm{rank}_d(\sum_{i=1}^C \lambda_i P_i) \leq r$, where $d = \max_i \deg(\lambda_i P_i)$.*

Given a polynomial factor $\mathcal{B}$ and a function $r : \mathbb{N} \to \mathbb{N}$, we say that $\mathcal{B}$ is r-regular if the rank of $\mathcal{B}$ is larger than $r(|\mathcal{B}|)$.

Notice that by the definition of rank, for a degree-d polynomial P of depth k we have

$$\mathrm{rank}(\{P\}) = \min\left\{\mathrm{rank}_d(P), \mathrm{rank}_{d-(p-1)}(pP), \ldots, \mathrm{rank}_{d-k(p-1)}(p^k P)\right\},$$

where $\{P\}$ is a polynomial factor consisting of one polynomial P.

Regular factors indeed do behave like a generic collection of polynomials, and we will establish this in a precise sense in Section 7.4. Thus, given any factor $\mathcal{B}$ that is not regular, it is often useful to *regularize* $\mathcal{B}$, that is to find a refinement $\mathcal{B}'$ of $\mathcal{B}$ that is regular up to our desires. Various regularity lemmas for polynomials will be discussed in Section 7.5

7.2 Analytic measures of uniformity

Regularity defined by the notion of rank is an algebraic/combinatorial notion of pseudorandomness. There are several cases where an analytic notion would be much more useful. In many cases, what is needed from the notion of regularity is that the polynomials defining the factor, when evaluated jointly on a uniform input, would behave as independent random variables. This can be equivalently expressed as the condition that any nonzero linear combination of the polynomials is almost uniformly distributed. We will accomplish this through the analytic notion of bias.

Definition 7.2.1 (Bias)**.** *The bias of a function $f : \mathbb{F}^n \to \mathbb{T}$ is defined to be*

$$\mathrm{bias}(f) := \mathbb{E}_{x \in \mathbb{F}^n}\left[e\left(f(x)\right)\right].$$

Note that a function f that takes every value from $\mathbb{U}_k$ equally likely, will satisfy $\mathrm{bias}(f) = 0$, and that a random polynomial will have bias very close to 0, and thus having small bias can be potentially used as an analytic notion of regularity.

It turns out that the bias and the rank of a polynomial are closely related. The following theorem was first proven for the case of $d < |\mathbb{F}|$ by Green and Tao [41], and then extended to the general case by Kaufman and Lovett [51].

Theorem 7.2.2 ([41, 51])**.** *Fix a prime finite field $\mathbb{F}$, an integer $d \geq 1$ and a real $\varepsilon > 0$. There exists $r = r_{7.2.2}(\mathbb{F}, d, \varepsilon)$ such that the following is true. If $P : \mathbb{F}^n \to \mathbb{T}$ is a degree-d polynomial with rank greater than r, then $|\mathbb{E}_x[e(P(x))]| < \varepsilon$.*

Kaufman and Lovett originally proved Theorem 7.2.2 for classical polynomials. However, their proof extends to nonclassical ones without modification. Note that $r_{7.2.2}(\mathbb{F}, d, \varepsilon)$ does not depend on the dimension n of $\mathbb{F}^n$, but depends in an unspecified way to $\mathbb{F}, d, \varepsilon$. The original proof had terrible dependency on all three parameters (Ackerman-type), and hence only applies for constant-size prime finite fields, constant degree and constant bias.

Remark 7.2.3. *Theorem 7.2.2 was extended in two ways by subsequent works. Bhattacharyya and Bhowmick [11] extended it to constant-size non-prime finite fields. Bhowmick and Lovett [18] refined the dependency on the field size and bias to be polynomial. This allows the theorem to be applicable for large finite fields, of size possibly growing with n. We present this in Chapter 8. In this section, we restrict our attention to constant-size prime finite fields.*

Next, motivated by Theorem 7.2.2 we define unbiasedness for polynomial factors.

Definition 7.2.4 (Unbiased Factor)**.** *Let $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function. A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$ is said to be ε-unbiased if for every collection $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \bmod p^{k_1+1}, \ldots, \lambda_C \bmod p^{k_C+1}) \neq 0^C$, it holds that*

$$\left| \mathbb{E}_x \left[e\left(\sum_i \lambda_i P_i(x) \right) \right] \right| < \varepsilon(|\mathcal{B}|).$$

Using Gowers norms, one can define the following analytic notion of uniformity for polynomials which is stronger than unbiasedness.

Definition 7.2.5 (Uniformity)**.** *Let $\varepsilon > 0$ be a real. A degree-d polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is said to be ε-uniform if*

$$\|e(P)\|_{U^d} < \varepsilon.$$

Tao and Ziegler [79] used Theorem 7.2.2 to show that high rank polynomials have small Gowers norm.

Theorem 7.2.6 (Theorem 1.20 of [79])**.** *Fix a prime finite field $\mathbb{F}$, an integer $d \geq 1$ and $\varepsilon > 0$. There exists $r = r_{7.2.6}(\mathbb{F}, d, \varepsilon)$ such that the following is true. For every nonclassical polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leq d$, if $\|e(P)\|_{U^d} \geq \varepsilon$, then $\operatorname{rank}_d(P) \leq r$.*

This immediately implies that a high-rank polynomial is also uniform in the sense of Theorem 7.2.5.

Corollary 7.2.7. *Let $\mathbb{F}, d, \varepsilon$ and $r(\mathbb{F}, d, \varepsilon)$ be as in Theorem 7.2.6. Every r-regular polynomial P of degree d is also ε-uniform.*

The next claim, which is a standard application of Fourier analysis, shows that the converse of this is true at least qualitatively.

Claim 7.2.8. *Fix a prime finite field $\mathbb{F}$, and integers $d, r \geq 1$. There exists $\varepsilon = \varepsilon_{7.2.8}(\mathbb{F}, d, r)$ such that the following is true. For every $P : \mathbb{F}^n \to \mathbb{T}$, if $\operatorname{rank}_d(P) \leq r$ then $\|e(P)\|_{U^d} \geq \varepsilon$.*

Proof. We will show that low rank functions cannot be very uniform. Let $\mathbb{F} = \mathbb{F}_p$. Assume that $P(x) = \Gamma(Q_1(x), \ldots, Q_r(x))$, where each $Q_i : \mathbb{F}_p^n \to \mathbb{T}$ is a polynomial of degree $\leq d-1$ and depth k_i, and where $\Gamma : \prod_{i=1}^r \mathbb{U}_{k_i+1} \to \mathbb{T}$. Shorthand $G := \prod_{i=1}^r \mathbb{U}_{k_i+1}$ and let $G' := \prod_{i=1}^r \mathbb{Z}_{p^{k_i+1}}$ be the dual group. The Fourier decomposition of $e(\Gamma) : G \to \mathbb{C}$ is given by

$$e(\Gamma(z)) = \sum_{\alpha \in G'} \widehat{\Gamma}(\alpha) e(\langle \alpha, z \rangle),$$

where $|\widehat{\Gamma}(\alpha)| \leq 1$ for all α. Let $Q_\alpha := \sum \alpha_i Q_i(x)$ for $\alpha \in G'$, which is a polynomial of degree $\leq d-1$. Then

$$e(P(x)) = \sum_{\alpha \in G'} \widehat{\Gamma}(\alpha) e(Q_\alpha(x)).$$

Thus

$$1 = |\langle e(P(x)), e(P(x))\rangle| \leq \sum_{\alpha \in G'} |\langle e(P(x)), e(Q_\alpha(x))\rangle|,$$

which shows that for some $\alpha^* \in G'$,

$$|\langle e(P), e(Q_{\alpha^*})\rangle| \geq \frac{1}{|G|}.$$

Next, observe that by the direct theorem for the Gowers norm (Theorem 6.2.1), as Q_{α^*} has degree $\leq d-1$ we have

$$\|e(P)\|_{U^d} \geq |\langle e(P), e(Q_{\alpha^*})\rangle| \geq |G|^{-1}.$$

To conclude, a simple calculation gives that

$$|G| = p^{(k_1+1)+\ldots+(k_r+1)} \leq p^{(1+\lceil (d-1)/(p-1)\rceil)r}$$

as $k_i(p-1) \leq d-1$. Thus we conclude that

$$\|e(P)\|_{U^d} \geq \varepsilon(\mathbb{F}_p, d, r) := p^{-(1+\lceil (d-1)/(p-1)\rceil)r}.$$

□

It is straightforward to extend the notion of uniformity from a single polynomial (Theorem 7.2.5) to a polynomial factor.

Definition 7.2.9 (Uniform Factor)**.** *Let $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function. A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$ is said to be ϵ-uniform if for every collection $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \bmod p^{k_1+1}, \ldots, \lambda_C \bmod p^{k_C+1}) \neq 0^C$, it holds that*

$$\left\| e\left(\sum_i \lambda_i P_i\right)\right\|_{U^d} < \varepsilon(|\mathcal{B}|),$$

where $d = \max_i \deg(\lambda_i P_i)$.

Remark 7.2.10 (Equivalence between regularity and uniformity)**.** *Similar to Theorem 7.2.7 it also follows from Theorem 7.2.6 that an r-regular degree-d factor $\mathcal{B}$ is also ε-uniform when $r = r_{7.2.6}(\mathbb{F}, d, \varepsilon)$ is as in Theorem 7.2.6. By Theorem 7.2.8 the converse of this also holds and we have an approximate equivalence between regularity and uniformity.*

7.3 The derivative polynomial

Let $P : \mathbb{F}^n \to \mathbb{T}$ be a nonclassical polynomial of degree d. The order-d Gowers norm of $e(P)$, which we have seen to control its correlation with lower degree polynomials, can be expressed as the bias of the *derivative polynomial* of P.

Definition 7.3.1 (Derivative Polynomial)**.** *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-d polynomial. The derivative polynomial $\partial P : (\mathbb{F}^n)^d \to \mathbb{T}$ of P is defined as*

$$\partial P(h_1, \ldots, h_d) := D_{h_1} \cdots D_{h_d} P(0),$$

where $h_1, \ldots, h_d \in \mathbb{F}^n$.

Remark 7.3.2. *Note that we could equivalently define $\partial P(h_1, \ldots, h_d) := D_{h_1} \cdots D_{h_d} P(y)$ for an arbitrary $y \in \mathbb{F}^n$. This is due to the fact that $D_{h_1} \cdots D_{h_d} P$ is a degree 0 function, which is a constant function.*

A closed form for ∂P, applying the derivatives iteratively, is

$$\partial P(h_1, \ldots, h_d) = \sum_{S \subseteq [d]} (-1)^{d-|S|} P\left(\sum_{i \in S} h_i\right).$$

It is easy to verify that

$$\|e(P)\|_{U_d}^{2^d} = \text{bias}(\partial P).$$

The following lemma shows some useful properties of the derivative polynomial.

Lemma 7.3.3. *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-d polynomial. Then*

(i) $\partial P(h_1, \ldots, h_d)$ is a classical homogeneous polynomial of degree d.

(ii) $\partial P(h_1, \ldots, h_d)$ is invariant under permutations of $h_1, \ldots, h_d$.

(iii) $\partial P(h_1, \ldots, h_d)$ is linear in each of $h_1, \ldots, h_d$.

(iv) For any $x \in \mathbb{F}^n$, $D_{h_1} \cdots D_{h_d} P(x) = \partial P(h_1, \ldots, h_d)$.

Proof. The proof follows by the properties of the additive derivative D_h. By definition, $\partial P(h_1, \ldots, h_d)$ is a (possibly nonclassical) polynomial of degree d. Item (i) follows since ∂P is annihilated by multiplication by p, since

$$p\partial P(h_1, h_2, \ldots, h_d) = pD_{h_1} \cdots D_{h_d}(P)(0) = D_{h_1} \cdots D_{h_d}(pP)(0) = 0,$$

as we have $\deg(pP) = \max(d-(p-1), 0) < d$ (see Theorem 6.1.5). Item (ii) holds since $D_h D_{h'} Q = D_{h'} D_h Q$ for every function $Q : \mathbb{F}^n \to \mathbb{T}$ and every $h, h' \in \mathbb{F}^n$. Item (iii) holds as every monomial of ∂P must depend on each of the variable sets h_i, as for example if we set $h_1 = 0$ then $\partial P(0, h_2, \ldots, h_d) = 0$ since $D_0 Q = 0$ for any function Q. As $\deg(\partial P) \leq d$, each monomial must contain exactly one variable from each h_i. Hence ∂P is linear in each h_i and has the form

$$\partial P(h_1, \ldots, h_d) = \sum_{i_1, \ldots, i_d = 1}^{n} c_{i_1, \ldots, i_d} \prod_{j=1}^{d} (h_j)_{i_j},$$

where $c_{i_1, \ldots, i_d}$ depends only on the multiset $\{i_1, \ldots, i_d\}$. Item (iv) follows as

$$D_{h_1} \cdots D_{h_d} P(x + h) - D_{h_1} \cdots D_{h_d} P(x) = D_{h_1} \cdots D_{h_d} D_h P(x) = 0.$$

□

Corollary 7.3.4. *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a polynomial. Then P has low rank if and only if its derivative polynomial is biased.*

Proof. The corollary follows from Theorem 7.2.6 and Theorem 7.2.8, since

$$\|e(P)\|_{U^d} = \mathbb{E}_{h_1, \ldots, h_d, x}\left[e(D_{h_1} \cdots D_{h_d} P(x))\right] = \mathbb{E}_{h_1, \ldots, h_d}\left[e(D_{h_1} \cdots D_{h_d} P(0))\right] = \text{bias}(\partial P).$$

□

7.4 Equidistribution of regular factors

In this section, we make precise the intuition that a high-rank collection of polynomials, evaluated on a joint input, behaves close to a collection of independent random variables. The key technical tool is the connection between the combinatorial notion of rank and the analytic notion of bias (Theorem 7.2.2).

Using a standard observation that relates the bias of a function to its distribution on its range, Theorem 7.2.2 implies the following.

Lemma 7.4.1 (Size of atoms)**.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d \geq 1$, complexity C, and rank $r_{7.2.2}(\mathbb{F}, d, \varepsilon)$, defined by a tuple of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ having respective depths $k_1, \ldots, k_C$. Suppose $b = (b_1, \ldots, b_C) \in \mathbb{U}_{k_1+1} \times \cdots \times \mathbb{U}_{k_C+1}$. Then*

$$\Pr_x[\mathcal{B}(x) = b] = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.$$

In particular, for $\varepsilon < \frac{1}{\|\mathcal{B}\|}$, $\mathcal{B}(x)$ attains every possible value in its range and thus has $\|\mathcal{B}\|$ atoms.

Proof. We can express the fraction of inputs in an atom b as

$$\begin{aligned}
\Pr_x[\mathcal{B}(x) = b] &= \mathbb{E}_x\left[\prod_{i=1}^{C} 1_{[P_i(x)=b_i]}\right] \\
&= \mathbb{E}_x\left[\prod_{i=1}^{C} \frac{1}{p^{k_i+1}} \sum_{\lambda_i=0}^{p^{k_i+1}-1} e\left(\lambda_i(P_i(x) - b_i)\right)\right] \\
&= \prod_i p^{-(k_i+1)} \cdot \sum_{(\lambda_1,\ldots,\lambda_C) \in \prod_i [0, p^{k_i+1}-1]} \mathbb{E}_x\left[e\left(\sum_{i=1}^{C} \lambda_i(P_i(x) - b_i)\right)\right].
\end{aligned}$$

The second equality uses the fact that $P_i(x) - b_i$ is in $\mathbb{U}_{k_i+1}$ and that for every nonzero $x \in \mathbb{U}_{k_i+1}$, $\sum_{\lambda=0}^{p^{k+1}-1} e(\lambda x) = 0$. By our assumption that $\mathcal{B}$ is r-regular, for every $(\lambda_1, \ldots, \lambda_C) \neq \vec{0}$ we have that $\sum_{i=1}^{C} \lambda_i P_i(x)$ has rank at least r, which by Theorem 7.2.2 implies that its bias is at most ε. Thus

$$\Pr_x[\mathcal{B}(x) = b] = \prod_i p^{-(k_i+1)} \cdot \left(1 \pm \varepsilon \prod_i p^{k_i+1}\right) = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.$$

□

An almost identical proof implies a similar statement for unbiased factors instead of regular factors.

Lemma 7.4.2 (Equidistribution for unbiased factors)**.** *Suppose that $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ is a decreasing function. Let $\mathcal{B}$ be an ε-unbiased factor of degree $d \geq 1$, defined by a tuple of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ having respective depths $k_1, \ldots, k_C$. Suppose $b = (b_1, \ldots, b_C) \in \mathbb{U}_{k_1+1} \times \cdots \times \mathbb{U}_{k_C+1}$. Then*

$$\Pr_{x \in \mathbb{F}^n}[\mathcal{B}(x) = b] = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon(|\mathcal{B}|).$$

7.5 Regularization of factors

Due to the generic properties of regular factors, it is often useful to *refine* a given polynomial factor to a regular one. We will first formally define what we mean by refining a polynomial factor.

Definition 7.5.1 (Refinement)**.** *A factor $\mathcal{B}'$ is called a* refinement *of $\mathcal{B}$, and denoted $\mathcal{B}' \succeq \mathcal{B}$, if the induced partition by $\mathcal{B}'$ is a combinatorial refinement of the partition induced by $\mathcal{B}$. In other words, if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

One needs to be careful about distinguishing between two types of refinements.

Definition 7.5.2 (Semantic and syntactic refinements)**.** *$\mathcal{B}'$ is called a* syntactic refinement *of $\mathcal{B}$, and denoted $\mathcal{B}' \succeq_{syn} \mathcal{B}$, if the sequence of polynomials defining $\mathcal{B}'$ extends that of $\mathcal{B}$. It is called a* semantic refinement*, and denoted $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if the induced partition is a combinatorial refinement of the partition induced by $\mathcal{B}$. In other words, if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

Remark 7.5.3. *Clearly, being a syntactic refinement is stronger than being a semantic refinement. But observe that if $\mathcal{B}'$ is a semantic refinement of $\mathcal{B}$, then there exists a syntactic refinement $\mathcal{B}''$ of $\mathcal{B}$ that induces the same partition of $\mathbb{F}^n$ as $\mathcal{B}'$, and for which $|\mathcal{B}''| \leq |\mathcal{B}'| + |\mathcal{B}|$. To construct $\mathcal{B}''$, simply add the defining polynomials of $\mathcal{B}$ to those of $\mathcal{B}'$.*

The following lemma by Green and Tao [41] shows that every classical polynomial factor can be refined to a regular factor. The basic idea is simple: if some polynomial has low rank, decompose it to a few lower degree polynomials, and repeat. Formally, it follows by transfinite induction on the number of polynomials of each degree that define the polynomial factor.

Lemma 7.5.4 (Regularity lemma for classical polynomials [41])**.** *Fix a prime finite field $\mathbb{F}$, an integer $d \geq 1$ and a non-decreasing function $r : \mathbb{N} \to \mathbb{N}$. There exists a function $C_{7.5.4}^{\mathbb{F},d,r} : \mathbb{N} \to \mathbb{N}$ for which the following holds.*

Suppose $\mathcal{B}$ is a polynomial factor of degree $\leq d$ and complexity C, defined by classical polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{F}$. Then there exists an r-regular factor $\mathcal{B}'$ that semantically refines $\mathcal{B}$, where $\mathcal{B}'$ has degree $\leq d$ and complexity $C' \leq C_{7.5.4}^{(\mathbb{F},d,r)}(C)$.

Furthermore, if $\mathcal{B}$ is itself a syntactic refinement of some $\mathcal{B}_0$ that has rank $> r(C')$, then $\mathcal{B}'$ can be taken to be a syntactic refinement of $\mathcal{B}_0$.

Proof. Let $\mathcal{B}$ be a polynomial factor defined by classical polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{F}$ of degree at most d. Define $M(\mathcal{B}) := (M_d, \ldots, M_1) \in \mathbb{N}^d$ where M_i is the number of polynomials of degree i among $P_1, \ldots, P_C$. Note that $M_1 + \ldots + M_d = C = |\mathcal{B}|$. We define the lexicographic order on $\mathbb{N}^d$, where $M > M'$ if $M_i > M'_i$ for some $1 \leq i \leq d$ and $M_j = M'_j$ for all $j > i$. The proof will be by transfinite induction on M under the lexicographic order. Namely, we will apply the fact that $\mathbb{N}^d$ under the lexicographic order is Noetherian, that is, it does not contain an infinite decreasing sequence.

Let $\mathcal{B}$ be a polynomial factor, and assume that $\mathcal{B}$ is not r-regular. Then by definition, some linear combination of the polynomials that define $\mathcal{B}$ has rank less than $r(C)$. Let $P(x) = \sum \lambda_i P_i(x)$ with $\lambda_i \in \mathbb{F}$, not all zero, such that $\mathrm{rank}(P) \leq r(C)$. By definition of rank, we can decompose P as a function of $r(C)$ lower degree polynomials. That is,

$$P(x) = \Gamma(Q_1(x), \ldots, Q_{r(C)}(x))$$

where $\deg(Q_i) \leq \deg(P) - 1$ and where $\Gamma : \mathbb{F}^{r(C)} \to \mathbb{F}$ is some function. Let $i^* \in [C]$ be chosen so that $\lambda_{i^*} \neq 0$ and $e = \deg(P_{i^*})$ is maximal. In particular, $\deg(P_{i^*}) \geq \deg(P)$. We can express P_{i^*} as a linear combination of $\{P_i : i \neq i^*\}$ and $Q_1, \ldots, Q_{r(C)}$. Define $\mathcal{B}_1 = \mathcal{B} \setminus \{P_{i^*}\} \cup \{Q_1, \ldots, Q_{r(C)}\}$. We claim that:

(i) $\mathcal{B}_1$ is a semantic refinement of $\mathcal{B}$.

(ii) $M(\mathcal{B}_1) < M(\mathcal{B})$.

The first item follows by definition, and the second since in order to construct $\mathcal{B}_1$, we removed one polynomial of degree e from $\mathcal{B}$ and added $r(|\mathcal{B}|)$ many lower degree polynomials. If $\mathcal{B}_1$ is still not r-regular, we can repeat this process and obtain $\mathcal{B}_2$ which is a semantic refinement of $\mathcal{B}_1$, and so on. By transfinite induction, this process must halt after a finite number of steps. More formally, the number of steps depends only on $M(\mathcal{B})$ as in the argument above, we can provide an upper bound M' which depends only on $M(\mathcal{B})$ such that $M(\mathcal{B}_1) \leq M' < M(\mathcal{B})$, by taking $M' = (M_d, \ldots, M_{e+1}, M_e - 1, M_{e-1} + r(|\mathcal{B}|), \ldots, M_1 + r(|\mathcal{B}|))$.

For the final part, assume that $\mathcal{B}$ is a syntactic refinement of a polynomial factor $\mathcal{B}_0$, defined without loss of generality by the $C_0 = |\mathcal{B}_0|$ polynomials $P_1, \ldots, P_{C_0} : \mathbb{F}^n \to \mathbb{F}$. In the regularization process, we

will attempt to choose $P_{i^*} \notin \{P_1, \ldots, P_{C_0}\}$ if possible. If we can achieve this at every step, then at the end we retain all the original polynomials that define $\mathcal{B}_0$, and hence we obtain a syntactic refinement of $\mathcal{B}_0$. If we fail, then at some stage we obtained a polynomial factor $\mathcal{B}''$ for which some linear combination of polynomials from $P_1, \ldots, P_{C_0}$ has rank at most $r(|\mathcal{B}''|) \leq r(C')$. This contradicts our assumption that $\mathcal{B}_0$ has high rank. □

The bounds obtained on $C^{\mathbb{F},d,r}_{7.5.4}$ have Ackermann-type dependence on the degree d, even when $r(\cdot)$ is a "reasonable" function. As such, it gives nontrivial results only for constant degrees. The extension of Theorem 7.5.4 to nonclassical polynomials is more involved, and was proved by Tao and Ziegler [79] as part of their proof of the inverse Gowers theorem (Theorem 6.2.3).

Theorem 7.5.5 (Regularity lemma for nonclassical polynomials [79])**.** *Fix a prime finite field $\mathbb{F}$, an integer $d \geq 1$ and a non-decreasing function $r : \mathbb{N} \to \mathbb{N}$. There exists a function $C^{\mathbb{F},d,r}_{7.5.5} : \mathbb{N} \to \mathbb{N}$ for which the following holds.*

Suppose $\mathcal{B}$ is a polynomial factor of degree $\leq d$ and complexity C, defined by nonclassical polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$. Then there exists an r-regular factor $\mathcal{B}'$ that semantically refines $\mathcal{B}$, where $\mathcal{B}'$ has degree $\leq d$ and complexity $C' \leq C^{(\mathbb{F},d,r)}_{7.5.4}(C)$.

Furthermore, if $\mathcal{B}$ is itself a syntactic refinement of some $\mathcal{B}_0$ that has rank $> r(C')$, then $\mathcal{B}'$ can be taken to be a syntactic refinement of $\mathcal{B}_0$.

We sketch the proof of Theorem 7.5.5. The basic approach is the same as that of Theorem 7.5.4. Let $\mathcal{B}$ be a polynomial factor defined by nonclassical polynomials of degree $\leq d$. It is said to be *extended* if whenever $P \in \mathcal{B}$, then either $pP = 0$ or $pP \in \mathcal{B}$. Clearly, every polynomial factor can be made extended by adding the appropriate polynomials to it. If $\mathcal{B}$ is extended and defined by polynomials $P_1, \ldots, P_C$, then for it to have rank r, it suffices if $\mathrm{rank}(\sum \lambda_i P_i) \geq r$ for all $\lambda_i \in \{0, \ldots, p-1\}$, not all zero. The main idea in the proof of Theorem 7.5.5 is to apply the same inductive argument as in the proof of Theorem 7.5.4, while maintaining an extended factor throughout the proof. However, this raises the following challenge: assume that P_i has low rank, and we wish to replace it with a few lower degree polynomials. However, assume that also $P_i = pP_j$, for some other polynomial P_j defining the factor. Then, we must remove P_j from the factor as well. If P_j also had low rank, that would suffice. However, this need not be true. What is true is that P_j can be decomposed as a function of a few polynomials, each of which either has lower degree or the same degree and lower depth as that of P_j. We summarize this in a lemma below, for which we omit the proof. With that lemma in hand, the proof goes through as before, except that now we need to keep track of the number of polynomials of a given depth and a given degree.

Lemma 7.5.6 ([79])**.** *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a polynomial of degree d and rank r. Assume that $P = pQ$, where Q is a polynomial of degree $d + (p-1)$. Then $Q(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$, where for each i, either $\deg(Q_i) < \deg(Q)$ or $\deg(Q_i) = \deg(Q)$ and depth$(Q_i) <$ depth(Q). Furthermore, $c \leq C_{7.5.6}(\mathbb{F}, d, r)$.*

7.6 Strong equidistribution of regular factors

One of the important and useful properties of the classical Fourier characters is that they form an orthonormal basis. Theorem 7.2.2, Theorem 7.4.1 and Theorem 7.4.2 provide an approximate version of this phenomenon that is useful for several applications. However, for certain applications we need a stronger notion of orthogonality, one where each polynomial of the factor is evaluated not just on a single input, but on a collection of inputs given by linear forms.

A linear form in k variables is a vector $L = (\lambda_1, \ldots, \lambda_k) \in \mathbb{F}^k$. It defines a map $L : (\mathbb{F}_n)^k \to \mathbb{F}_n$ by $L(x_1, \ldots, x_k) = \sum \lambda_i x_i$. Often, we would need to analyze averages of f evaluated on several linear forms with joint variables. Consider for example the BLR linearity test, explored in Chapter 2. Given a function $P : \mathbb{F}_2^n \to \mathbb{F}_2$, let $f(x) = (-1)^{P(x)}$. The acceptance probability of the BLR linearity test on P is equal to

$$\frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y \in \mathbb{F}^n}\big[f(x)f(y)f(x+y)\big].$$

Thus, in order to analyze it, we needed to evaluate f on three related inputs: $x, y, x+y$.

The general form is as follows. Let $L_1, \ldots, L_m \in \mathbb{F}^k$ be linear forms, where to recall $L_i : (\mathbb{F}^n)^k \to \mathbb{F}^n$. The goal is to analyze, for $f : \mathbb{F}^n \to \mathbb{C}$, averages of the form

$$\mathbb{E}_{X \in (\mathbb{F}^n)^k} \big[f(L_1(X)) \cdots f(L_m(X)) \big].$$

The first step in such an analysis will be to find a suitable polynomial factor $\mathcal{B}$, such that we can replace f with $\mathbb{E}[f|\mathcal{B}]$ without affecting the average by much. Let us assume that we have already done so. Then we can specialize our treatment to functions of the form $f(x) = \Gamma(P_1(x), \ldots, P_C(x))$, where $P_i : \mathbb{F}^n \to \mathbb{T}$ are bounded degree polynomials. This in turn would require to analyze the joint distribution of $P_1, \ldots, P_C$ on the linear forms $L_1(X), \ldots, L_m(X)$. That is, we would need to understand the distribution of the random variable

$$A_{\mathcal{B},\mathcal{L}}(X) := \begin{pmatrix} P_1(L_1(X)) & P_2(L_1(X)) & \ldots & P_C(L_1(X)) \\ P_1(L_2(X)) & P_2(L_2(X)) & \ldots & P_C(L_2(X)) \\ \vdots & & & \vdots \\ P_1(L_m(X)) & P_2(L_m(X)) & \ldots & P_C(L_m(X)) \end{pmatrix},$$

where $X \in (\mathbb{F}^n)^k$ is uniformly distributed, $\mathcal{B} = \{P_1, \ldots, P_C\}$ is a family of nonclassical polynomials and $\mathcal{L} = \{L_1, \ldots, L_m\} \subset \mathbb{F}^k$ is a system of linear forms.

Our analysis so far (see Theorem 7.4.1) have shown that if we restrict our attention to a single linear form, namely a single row of $A_{\mathcal{B},\mathcal{L}}(X)$, then if we assume that $\mathcal{B}$ is regular enough, then its values are distributed close to uniform in their range $\prod_{j=1}^{C} \mathbb{U}_{k_j+1}$. Thus, a first guess might be that if $\mathcal{B}$ is regular enough, then $A_{\mathcal{B},\mathcal{L}}$ should be uniformly distributed over all the $m \times C$ matrices $A \in \prod_{i=1}^{m} \prod_{j=1}^{C} \mathbb{U}_{k_j+1}$.

However, this is false. The reason is that nonclassical polynomials of a given degree satisfy various linear identities, governed by the fact that they are annihilated by taking enough derivatives. For example, if P_1 is a linear polynomial then $P_1(x_1+x_2+x_3) - P_1(x_1+x_2) - P_1(x_1+x_3) + P_1(x_1) = 0$ holds, and hence the four random variables $(P_1(x_1+x_2+x_3), P_1(x_1+x_2), P_1(x_1+x_3), P_1(x_1))$ are far from independent. Thus, the rows of $A_{\mathcal{B},\mathcal{L}}(X)$ satisfy certain linear constraints. As we will shortly see, strong equidistribution implies that $A_{\mathcal{B},\mathcal{L}}(X)$ is close to uniform on all values that satisfy the necessary linear constraints. In particular, the columns of $A_{\mathcal{B},\mathcal{L}}(X)$ are nearly independent, with each column being uniform modulo the required linear dependencies.

The development of the strong equidistribution theorem occurred in several steps. Hatami and Lovett [46] established a strong near-equidistribution for factors of classical polynomials, when the characteristic of the field $\mathbb{F}$ is greater than the degree of the polynomial factor. Bhattacharyya *et al.* [13] later extended the result to the general characteristic case, but under an extra assumption that the system of linear forms is affine, i.e. there is a variable that appears with coefficient 1 in all the linear forms. In [16] the same proof technique was used to replace the condition that each L_i is affine, with the condition that all the coefficients of L_i are in $\{0,1\}$. Finally, in [44] the joint distribution of the matrix $A_{\mathcal{B},\mathcal{L}}$ was fully characterized without any extra assumptions on the linear forms. Instead, it requires that the polynomials are *homogeneous*.

We formally define and analyze homogeneous nonclassical in Chapter 10. For the time being, it is sufficient to note that each nonclassical polynomial $P(x)$ of degree d can be decomposed as as sum of homogeneous nonclassical polynomials of the degrees $1, \ldots, d$; and that in the regularization process (as in Theorem 7.5.5) one can require that the resulting polynomials are all homogeneous. Thus, one can always assume that a polynomial factor is defined by homogeneous nonclassical polynomials, and these can be made high rank by regularization.

Theorem 7.6.1 (Near orthogonality over linear forms)**.** *Let $\mathbb{F}$ be a prime field, $d \geq 1$, $\varepsilon > 0$. Let $\{L_1, \ldots, L_m\}$ be a system of linear forms. Let $\mathcal{B} = \{P_1, \ldots, P_C\}$ be a polynomial factor of degree at most d and $\mathrm{rank}(\mathcal{B}) > r_{7.6.1}(\mathbb{F}, d, \varepsilon)$. Assume furthermore that each P_i is a nonclassical homogeneous polynomial (defined in Chapter 10). For every set of coefficients $\Lambda = \{\lambda_{i,j} \in \mathbb{Z} : i \in [C], j \in [m]\}$, let*

$$P_\Lambda(x) := \sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(x)).$$

Then one of the following two cases holds:

- $P_\Lambda \equiv 0$. *In this case, for every* $i \in [C]$, *it holds that* $\sum_{j\in[m]} \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ *for any nonclassical homogeneous polynomials* $Q_i : \mathbb{F}^n \to \mathbb{T}$ *with* $\deg(Q_i) = \deg(P_i)$ *and* $depth(Q_i) \leq depth(P_i)$.
- $P_\Lambda \not\equiv 0$. *In this case,* $|\mathbb{E}[e(P_\Lambda)]| < \varepsilon$.

We defer the proof of Theorem 7.6.1 to Section 12.1, after we cover some more necessary background material.

Next, we state a special version of Theorem 7.6.1 useful in some applications. A system of linear forms $L_1, \ldots, L_m$ is said to be an *affine constraint* if $L_{i,1} = 1$ for all $i \in [m]$. Such systems arise naturally in the study of affine invariant properties, which we discuss in Chapter 13. Here, we state a special case of Theorem 7.6.1 for affine constraints, where the polynomials are not required to be homogeneous. For the proof see [13].

Theorem 7.6.2 (Near orthogonality over affine constraints)**.** *Let* $\mathbb{F}$ *be a prime field,* $d \geq 1$, $\varepsilon > 0$. *Let* $\{L_1, \ldots, L_m\}$ *be an affine constraint. Let* $\mathcal{B} = \{P_1, \ldots, P_c\}$ *be a polynomial factor of degree at most* d *and* $\mathrm{rank}(\mathcal{B}) > r_{7.6.2}(\mathbb{F}, d, \varepsilon)$. *For every set of coefficients* $\Lambda = \{\lambda_{i,j} \in \mathbb{Z} : i \in [c], j \in [m]\}$, *let*

$$P_\Lambda(x) := \sum_{i=1}^{c} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(x)).$$

Then one of the following two cases holds:

- $P_\Lambda \equiv 0$. *In this case, for every* $i \in [C]$, *it holds that* $\sum_{j\in[m]} \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ *for any nonclassical polynomials* $Q_i : \mathbb{F}^n \to \mathbb{T}$ *with* $\deg(Q_i) \leq \deg(P_i)$ *and* $depth(Q_i) \leq depth(P_i)$.
- $P_\Lambda \not\equiv 0$. *In this case,* $|\mathbb{E}[e(P_\Lambda)]| < \varepsilon$.

In the next section, as a corollary of Theorem 7.6.1, we determine the distribution of $A_{\mathcal{B},\mathcal{L}}$ when $\mathcal{B}$ is of sufficiently high enough rank.

7.7 The joint distribution of high rank polynomials over linear forms

In order to understand the joint distribution of polynomials over linear forms, we first need to understand the necessary linear constraints. The following definition formalizes this.

Definition 7.7.1. *Let* $\mathbb{F} = \mathbb{F}_p$ *be a prime field. Given a system of linear forms* $L_1, \ldots, L_m$ *over* $\mathbb{F}$ *in* ℓ *variables. and integers* $d, k > 0$, *the* (d,k)-dependency set *of* $L_1, \ldots, L_m$ *is the set of tuples* $(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}_{p^{k+1}}^m$ *such that* $\sum_{i=1}^m \lambda_i P(L_i(x_1, \ldots, x_\ell)) \equiv 0$ *for every homogeneous polynomial* $P : \mathbb{F}^n \to \mathbb{T}$ *of degree* d *and depth* k.

Observe that the (d,k)-dependency set of $L_1, \ldots, L_m$ is closed under addition, and hence is a subgroup of $\mathbb{Z}_{p^{k+1}}^m$. We do not currently have an explicit description of it, except for the implicit definition given in Theorem 7.7.1.

Theorem 7.6.1 says that if $\mathcal{B}$ is a regular factor, $P_\Lambda \equiv 0$ exactly when the first condition holds. In other words:

Corollary 7.7.2. *Let* $L_1, \ldots, L_m$ *be a system of linear forms over* $\mathbb{F} = \mathbb{F}_p$ *in* ℓ *variables. Fix an integer* $c > 0$, *tuples* $(d_1, \ldots, d_c) \in \mathbb{Z}_{>0}^c$ *and* $(k_1, \ldots, k_c) \in \mathbb{Z}_{\geq 0}^c$. *For* $i \in [c]$, *let* Λ_i *be the* (d_i, k_i)*-dependency set of* $L_1, \ldots, L_m$.

Then, for every polynomial factor $\mathcal{B}$ defined by homogeneous polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$, where P_i has degree d_i and depth k_i, and $\mathcal{B}$ has rank $> r_{7.2.6}\left(\mathbb{F}, \max_i d_i, \frac{1}{2}\right)$, it is the case that a tuple $(\lambda_{i,j})_{i\in[c],j\in[m]}$ satisfies

$$\sum_{i=1}^{c}\sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$$

if and only if for every $i \in [c]$,

$$(\lambda_{i,1} \bmod p^{k_i+1}, \ldots, \lambda_{i,m} \bmod p^{k_i+1}) \in \Lambda_i.$$

Proof. The "if" direction is obvious. For the "only if" direction, we use Theorem 7.6.1 to conclude that if $\sum_{i,j} \lambda_{i,j} P_i(L_j(\cdot)) \equiv 0$, it must be that for every $i \in [c]$, $\sum_j \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ for every homogeneous polynomial Q_i with degree d_i and depth k_i. This is equivalent to saying $(\lambda_{i,1} \bmod p^{k_i+1}, \ldots, \lambda_{i,m} \bmod p^{k_i+1}) \in \Lambda_i$. □

Remark 7.7.3. *For large characteristic fields, Hatami and Lovett [46] showed that the analog of Theorem 7.7.2 is true even without the rank condition.*

The joint distribution of $(P_i(L_j(x_1, \ldots, x_\ell)) : i \in [c], j \in [m])$ is only going to be supported on atoms which respect the constraints imposed by dependency sets. This is obvious: if P is a polynomial of degree d and depth k, such that $P(L_j(x_1, \ldots, x_\ell)) = b_j$, then for every $(\lambda_1, \ldots, \lambda_m)$ in the (d,k)-dependency set of $(L_1, \ldots, L_m)$ it must be the case that $\sum_j \lambda_j b_j = 0$. We call atoms which respect this constraint for all P_i in a factor, *consistent*.

Definition 7.7.4 (Consistency)**.** *Let $\mathbb{F}$ be a prime field. Let $L_1, \ldots, L_m$ be a system of linear forms over $\mathbb{F}$ in ℓ variables. A sequence of elements $b_1, \ldots, b_m \in \mathbb{T}$ are said to be* (d,k)-consistent with $L_1, \ldots, L_m$ *if $b_1, \ldots, b_m \in \mathbb{U}_{k+1}$ and for every tuple $(\lambda_1, \ldots, \lambda_m)$ in the (d,k)-dependency set of $L_1, \ldots, L_m$, it holds that $\sum_{i=1}^m \lambda_i b_i = 0$.*

Given vectors $\mathbf{d} = (d_1, \ldots, d_c) \in \mathbb{Z}^c_{>0}$ and $\mathbf{k} = (k_1, \ldots, k_c) \in \mathbb{Z}^c_{\geq 0}$, a sequence of vectors $b_1, \ldots, b_m \in \mathbb{T}^c$ are said to be $(\mathbf{d}, \mathbf{k})$-consistent with $L_1, \ldots, L_m$ *if for every $i \in [c]$, the elements $b_{1,i}, \ldots, b_{m,i}$ are (d_i, k_i)-consistent with $L_1, \ldots, L_m$.*

If $\mathcal{B}$ is a polynomial factor, the term $\mathcal{B}$-consistent with $L_1, \ldots, L_m$ *is a synonym for* $(\mathbf{d}, \mathbf{k})$-consistent with $L_1, \ldots, L_m$ *where $\mathbf{d} = (d_1, \ldots, d_c)$ and $\mathbf{k} = (k_1, \ldots, k_c)$ are respectively the degree and depth sequences of polynomials defining $\mathcal{B}$.*

Theorem 7.6.1 implies that for a polynomial factor of large enough rank, the joint distribution $(P_i(L_j(x_1, \ldots, x_\ell))$ is equi-distributed on all consistent values.

Theorem 7.7.5. *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$ and rank at least $r_{7.2.2}(\mathbb{F}, d, \varepsilon)$. Assume that $\mathcal{B}$ is defined by a sequence of homogeneous polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ having respective degrees $d_1, \ldots, d_c$ and respective depths $k_1, \ldots, k_c$. Let $L_1, \ldots, L_m$ be a system of linear forms over $\mathbb{F}$ in ℓ variables, and let Λ_i be the (d_i, k_i)-dependency set of $L_1, \ldots, L_m$.*

Suppose $b_1, \ldots, b_m \in \mathbb{T}^c$ are atoms of $\mathcal{B}$ that are $\mathcal{B}$-consistent with $L_1, \ldots, L_m$. We denote $b_i = (b_{i,1}, \ldots, b_{i,c})$. Then

$$\Pr_{x_1,\ldots,x_\ell \in \mathbb{F}^n}\left[P_i(L_j(x_1, \ldots, x_\ell)) = b_{j,i} \ \forall i \in [c], j \in [m]\right] = \frac{\prod_{i=1}^c |\Lambda_i|}{\|\mathcal{B}\|^m} \pm \varepsilon.$$

Proof. The proof is similar to that of Theorem 7.4.1.

$$\begin{aligned}
&\Pr_{x_1,\ldots,x_\ell}\left[P_i(L_j(x_1,\ldots,x_\ell)) = b_{j,i}\ \forall i \in [c], \forall j \in [m]\right] \\
&= \mathbb{E}_{x_1,\ldots,x_\ell}\left[\prod_{i\in[c],j\in[m]} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} e\left(\lambda_{i,j}(P_i(L_j(x_1,\ldots,x_\ell)) - b_{j,i})\right)\right] \\
&= \left(\prod_{i\in[c]} p^{-(k_i+1)}\right)^m \cdot \\
&\sum_{(\lambda_{i,j})\in\prod_{i,j}[0,p^{k_i+1}-1]} e\left(-\sum_{i,j}\lambda_{i,j}b_{j,i}\right)\mathbb{E}\left[e\left(\sum_{i,j}\lambda_{i,j}P_i(L_j(x_1,\ldots,x_\ell)\right)\right] \\
&= p^{-m\sum_{i=1}^c (k_i+1)} \cdot \left(\prod_{i=1}^c |\Lambda_i| \ \pm\ \varepsilon p^{m\sum_{i=1}^c (k_i+1)}\right),
\end{aligned}$$

where the last equality follows because by Theorem 7.7.2, $\sum_{i,j}\lambda_{i,j}P_i(L_j(\cdot))$ is identically zero for $\prod_i |\Lambda_i|$ many tuples $(\lambda_{i,j})$ and, in that case, $\sum_{i,j}\lambda_{i,j}b_{j,i} = 0$ because of the consistency requirement. For any other tuple $(\lambda_{i,j})$, the expectation in the third line is bounded by ε in absolute value. □

Chapter 8

Bias vs low rank in large fields

Theorem 7.2.2 exhibited the "bias implies low rank" phenomena: every biased polynomial has low rank. Concretely, if $P : \mathbb{F}^n \to \mathbb{F}$ is a degree d polynomial, and $|\mathbb{E}_x[e(P(x))]| \geq \varepsilon$, then $\mathrm{rank}(P) \leq r_{7.2.2}(\mathbb{F}, d, \varepsilon)$. In this section, we describe an improved theorem due to Bhowmick and Lovett [18] which allows for large fields (of size growing with n) as well as for errors which are polynomially small in the field order.

Theorem 8.0.1. *Let $d, s \in \mathbb{N}$. Let $\mathbb{F}$ be a finite field of characteristic $char(\mathbb{F}) > d$. Let $P : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. Suppose that $|\mathbb{E}_{x\in\mathbb{F}^n}[e(P(x))]| \geq |\mathbb{F}|^{-s}$. Then,* $\mathrm{rank}(P) \leq r_{8.0.1}(d, s)$.

Observe that here $r_{8.0.1}(d, s)$ does not depend on the field. Also, the error is allowed to be polynomially small in the field size. For concreteness, we focus on the proof where $\mathbb{F}$ is a prime field, where we assume throughout that $|\mathbb{F}| > d$. For the proof for non-prime fields we refer the interested reader to the original paper [18].

8.1 Bias implies low rank approximation

We start by showing that biased polynomials can be approximated by a few lower degree polynomials.

Lemma 8.1.1. *Let $d, s, t \in \mathbb{N}$. Let $P : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. Suppose that $|\mathbb{E}_{x\in\mathbb{F}^n}[e(P(x))]| \geq |\mathbb{F}|^{-s}$. Then, there exist polynomials $Q_1, \cdots, Q_c : \mathbb{F}^n \to \mathbb{F}$ of degree at most $d-1$, where $c = c(d, s, t) = \binom{d+t+2s+3}{d}$, and a function $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that*

$$\Pr_{x\in\mathbb{F}^n}[P(x) \neq \Gamma(Q_1(x), \cdots, Q_c(x))] \leq |\mathbb{F}|^{-t}.$$

We prove Theorem 8.1.1 in this section. Fix a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d, and let $\mu := \mathbb{E}_{x\in\mathbb{F}^n}[e(P(x))]$ be its bias, where we assume $|\mu| \geq |\mathbb{F}|^{-s}$. We begin with the following claim.

Claim 8.1.2. *For all $x \in \mathbb{F}^n$,*

$$\mu \cdot e(-P(x)) = \mathbb{E}_{y\in\mathbb{F}^n}[e(D_yP(x))].$$

Proof. We have $\mathbb{E}_{y\in\mathbb{F}^n}[e(D_yP(x))] = \mathbb{E}_{y\in\mathbb{F}^n}[e(P(x+y))e(-P(x))] = \mathbb{E}_{y\in\mathbb{F}^n}[e(P(y))] \cdot e(-P(x)) = \mu \cdot e(-P(x))$. □

Fix $x \in \mathbb{F}^n$. Pick $z = (z_1, \ldots z_k) \in (\mathbb{F}^n)^k$ uniformly for some k to be specified later. For $a \in \mathbb{F}^k, z \in (\mathbb{F}^n)^k$, we shorthand $a \cdot z = \sum_{i=1}^k a_i z_i \in \mathbb{F}^n$. For $a \in \mathbb{F}^k \setminus \{0\}$, let $W_{x,a}(z)$ be the random variable (over the choice of z) defined as

$$W_{x,a}(z) := e(D_{a\cdot z}P(x)).$$

For $a \neq 0^k$, we have

$$\mathbb{E}_{z\in(\mathbb{F}^n)^k}[W_{x,a}(z)] = \mathbb{E}_{y\in\mathbb{F}^n}[e(D_yP(x))].$$

Also, observe that for distinct $\alpha, \beta \in \mathbb{F}$,

$$|e(\alpha) - e(\beta)| \geq |\mathbb{F}|^{-1}.$$

We have the following.

Claim 8.1.3. *Fix $x \in \mathbb{F}^n$. If for $z \in (\mathbb{F}^n)^k$ it holds that*

$$\left| \frac{1}{|\mathbb{F}|^k - 1} \sum_{a \in \mathbb{F}^k \setminus \{0^k\}} W_{x,a}(z) - \mathbb{E}_y \left[e\left(D_y P(x) \right) \right] \right| \leq \frac{1}{2|\mathbb{F}|^{s+1}},$$

then

$$P(x) = \Gamma_0 \left(D_{a \cdot z} P(x) : a \in \mathbb{F}^k \setminus \{0\} \right)$$

where $\Gamma_0 : \mathbb{F}^{|\mathbb{F}|^k - 1} \to \mathbb{F}$ is defined as

$$\Gamma_0(v_1, \ldots v_{|\mathbb{F}|^k - 1}) = \arg\min_{\alpha \in \mathbb{F}} \left| \frac{1}{|\mathbb{F}|^k - 1} \sum_{i=1}^{|\mathbb{F}|^k - 1} e(v_i) - e(-\alpha)\mu \right|.$$

Proof. Since $|e(\alpha) - e(\beta)| \geq |\mathbb{F}|^{-1}$ for distinct $\alpha, \beta \in \mathbb{F}$ and as $|\mu| \geq |\mathbb{F}|^{-s}$, then by the assumption of the claim, and Theorem 8.1.2

$$\Gamma_0 \left(D_{a \cdot z} P(x) : a \in \mathbb{F}^k \setminus \{0\} \right) = P(x).$$

□

Observe that the random variables $\{W_{x,a}(z) : a \in \mathbb{F}^k \setminus \{0\}\}$ are pairwise independent. That is, for any distinct a, a' the random variable $(W_{x,a}(z), W_{x,a'}(z))$ is uniformly distributed in $\mathbb{F}^2$, when $z \in \mathbb{F}^k$ is uniformly chosen. Thus, we can apply Chebychev's inequality and obtain that, for $k = t + 2s + 3$, it holds that

$$\Pr_{z \in (\mathbb{F}^n)^k} \left[\left| \frac{1}{|\mathbb{F}|^k - 1} \sum_{a \neq 0} W_{x,a}(z) - \mathbb{E}_y \left[e\left(D_y P(x) \right) \right] \right| \geq \frac{1}{2|\mathbb{F}|^{s+1}} \right] \leq \frac{4|\mathbb{F}|^{2s+2}}{|\mathbb{F}|^k - 1} \leq \frac{1}{|\mathbb{F}|^t}. \tag{8.1}$$

Thus, for all $x \in \mathbb{F}^n$,

$$\Pr_{z \in (\mathbb{F}^n)^k} [\Gamma_0 \left(D_{a \cdot z} P(x) : a \in \mathbb{F}^k \setminus \{0\} \right) = P(x)] \geq 1 - |\mathbb{F}|^{-t}.$$

Therefore, by an averaging argument, there exists a choice of $z \in (\mathbb{F}^n)^k$ for which

$$\Pr_{x \in \mathbb{F}^n} [\Gamma_0 \left(D_{a \cdot z} P(x) : a \in \mathbb{F}^k \setminus \{0\} \right) = P(x)] \geq 1 - |\mathbb{F}|^{-t}. \tag{8.2}$$

We now prove that we only need a constant number of derivatives in order to approximate P, instead of a number which is polynomial in $|\mathbb{F}|$. Let $|\cdot| : \mathbb{F} \to \mathbb{N}$ be the canonical map $|x| = x$ for $x = \{0, 1, \ldots, |\mathbb{F}| - 1\}$.

Claim 8.1.4. *Let $B := \{b \in \mathbb{F}^k : \sum_{j=1}^k |b_j| \leq d\}$. Then for any $a \in \mathbb{F}^k$,*

$$D_{a \cdot z} P(x) = \sum_{b \in B} \lambda_{a,b} D_{b \cdot z} P(x)$$

for some $\lambda_{a,b} \in \mathbb{F}$.

Proof. Let $|a| = \sum_{i=1}^k |a_i|$. We prove the claim by induction on $|a|$. If $|a| \leq d$, the claim is straightforward, so assume $|a| > d$. As P is a degree d polynomial, we have for any $m > d$ and $y_1, \ldots, y_m \in \mathbb{F}^n$ that

$$D_{y_1} \ldots D_{y_m} P \equiv 0.$$

This translates to

$$\sum_{c\in\{0,1\}^m} (-1)^{\sum c_i} P\left(x + \sum c_i y_i\right) = 0.$$

As the sum of the coefficients is zero, this implies

$$\sum_{c\in\{0,1\}^m} (-1)^{\sum c_i} \left(P\left(x + \sum c_i y_i\right) - P(x)\right) = \sum_{c\in\{0,1\}^m} (-1)^{\sum c_i} D_{c\cdot y} P(x) = 0,$$

for $y = (y_1, \ldots, y_m)$. Apply this for $m = |a|$ and $y_1, \ldots, y_m$ set to z_1 repeated a_1 times, z_2 repeated a_2 times, up to z_k repeated a_k times. Then we obtain that

$$\sum_{a'\le a} \nu_{a'} D_{a'\cdot z} P(x) = 0,$$

where the sum is over all $a' \in \mathbb{F}^k$ such that $|a'_i| \le |a_i|$ for all $1 \le i \le k$, and $\nu_{a'} = (-1)^{|a'|} \prod_{i=1}^k \binom{a_i}{a'_i}$. In particular, $\nu_a \in \{-1, +1\}$. This implies that $D_{a\cdot z} P(x)$ is a linear combination of $D_{a'\cdot z} P(x)$ for $a' \in \mathbb{F}^k$ with $|a'| < |a|$. The claim now follows by applying the induction hypothesis. □

This concludes the proof of Theorem 8.1.1. By Theorem 8.1.4 we may rewrite

$$\Gamma_0\left(D_{a\cdot z} P(x) : a \in \mathbb{F}^k \setminus \{0\}\right) = \Gamma\left(D_{a\cdot z} P(x) : a \in B\right),$$

where $c = |B|$ and $\Gamma : \mathbb{F}^c \to \mathbb{F}$ is defined as $\Gamma(v_b : b \in B) = \Gamma_0\left(\left(\sum_{b\in B} \lambda_{a,b} v_b\right) : a \in \mathbb{F}^k \setminus \{0\}\right)$. Let $Q_1, \ldots, Q_c$ in the statement of Theorem 8.1.1 be $\{D_{a\cdot z} P(x) : a \in B\}$. Clearly, these are polynomials of degree $\le d-1$. To conclude, we need to bound $c = |B|$. We have

$$|B| = \sum_{i=1}^{d} \binom{k}{d} \le \binom{d+k}{d}.$$

Theorem 8.1.1 follows as we set $k = t + 2s + 3$.

8.2 Bias implies low rank exact computation

We prove Theorem 8.0.1 in this section. The proof is by induction on the degree d and follows along the lines of Theorem 1.7 in [41]. We sketch the proof below.

We first show that Theorem 8.0.1 implies a similar theorem, where instead of assuming that a polynomial is biased, we assume that it has a noticeable Gowers uniformity norm.

Lemma 8.2.1 (Large Gowers norm implies low rank)**.** *Suppose Theorem 8.0.1 is true up to degree d. Let $P : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. Suppose that $||e(P)||_{U^d} \ge |\mathbb{F}|^{-s}$. Then* $\mathrm{rank}(P) \le r_{(8.2.1)}(d, s)$.

Proof. We have

$$|\mathbb{E}_{x,y_1,\ldots,y_d\in\mathbb{F}^n}\left[e\left(D_{y_1,\ldots,y_d} P(x)\right)\right]| = ||e(P)||_{U^d}^{2^d} \ge |\mathbb{F}|^{-2^d s}.$$

Define $Q : \mathbb{F}^{n(d+1)} \to \mathbb{F}$ as

$$Q(x, y_1, \ldots, y_d) := D_{y_1,\ldots,y_d} P(x).$$

By Theorem 8.0.1,

$$\mathrm{rank}(Q) \le r_{(8.0.1)}(d, 2^d s).$$

Applying the Taylor approximation theorem to P, as we assume $d < |\mathbb{F}|$, we have

$$P(x) = \frac{1}{d!} D_{x,\ldots,x} P(0) + R(x) = \frac{1}{d!} Q(0, x, \ldots, x) + R(x),$$

where R is a polynomial of degree $\le d-1$. We conclude that $\mathrm{rank}(P) \le \mathrm{rank}(Q) + 1$. This concludes the lemma by setting $r_{(8.2.1)}(d, s) = r_{(8.0.1)}(d, s2^d) + 1$. □

The next lemma shows that a regular factor has atoms of roughly equal size. We shorthand $\alpha \pm \delta$ to denote any element in the interval $[\alpha - \delta, \alpha + \delta]$.

Lemma 8.2.2 (Size of atoms)**.** *Suppose Theorem 8.0.1 is true up to degree d. Let $\mathcal{B} = \{Q_1, \ldots, Q_c\}$ be a polynomial factor of degree at most d. Given $s \in \mathbb{N}$, assume that $\mathcal{B}$ has rank at least $r_{(8.0.1)}(d, s)$. Then for every $b \in \mathbb{F}^c$,*

$$\Pr_{x \in \mathbb{F}^n}[\mathcal{B}(x) = b] = \frac{1}{|\mathbb{F}|^c} \pm \frac{1}{|\mathbb{F}|^s}.$$

Proof. For any $b \in \mathbb{F}^c$,

$$\begin{aligned}
\Pr[\mathcal{B}(x) = b] &= \frac{1}{|\mathbb{F}|^c} \sum_{a \in \mathbb{F}^c} \mathbb{E}_x \left[e\left(\sum_i a_i (Q_i(x) - b_i) \right) \right] \\
&= \frac{1}{|\mathbb{F}|^c} \pm \frac{1}{|\mathbb{F}|^c} \sum_{0 \neq a \in \mathbb{F}^c} \left| \mathbb{E}_x \left[e\left(\sum_i a_i Q_i(x) \right) \right] \right| \\
&= \frac{1}{|\mathbb{F}|^c} \pm \frac{1}{|\mathbb{F}|^s}
\end{aligned}$$

The last line follows because of the following argument. Suppose for some $a \neq 0$, $|\mathbb{E}_x [e (\sum_i a_i Q_i(x))]| > \frac{1}{|\mathbb{F}|^s}$. Then by Theorem 8.0.1, $\mathrm{rank}(\sum_i a_i Q_i) \leq r_{(8.0.1)}(d, s)$. This contradicts the assumption on the rank of $\mathcal{B}$. □

Recall that a linear form $L = (\ell_1, \ldots, \ell_k) \in \mathbb{F}^k$ is said to be *affine* if $\ell_1 = 1$. Next, we argue that a high rank factor evaluated on a system of affine linear forms is near orthogonal.

Lemma 8.2.3 (Near orthogonality of affine linear forms)**.** *Suppose Theorem 8.0.1 is true up to degree d. Let $c, d, k, s, m \in \mathbb{N}$. Let $\mathcal{B} = \{Q_1, \ldots, Q_c\}$ be a polynomial factor of degree at most d. Assume $\mathcal{B}$ has rank at least $r_{(8.2.3)}(d, k, s)$. Let $(L_1, \ldots, L_m)$ be a system of affine linear forms on k variables. Let $\Lambda = (\lambda_{ij})_{i \in [c], j \in [m]}$ be a tuple of integers. Define*

$$Q_\Lambda(x_1, \ldots, x_k) = \sum_{i \in [c], j \in [m]} \lambda_{ij} Q_i(L_j(x_1, \ldots, x_k)).$$

Then one of the following is true.

1. *$Q_\Lambda \equiv 0$. Moreover, for every $i \in [c]$, it holds that $\sum_{j=1}^m \lambda_{ij} Q_i(L_j(\cdot)) \equiv 0$ for all Q_i of degree at most d.*
2. *$Q_\Lambda \not\equiv 0$. Moreover, $|\mathbb{E}[e (Q_\Lambda(x_1, \ldots, x_k))]| \leq |\mathbb{F}|^{-s}$.*

The proof is identical to the proof of Theorem 7.6.2, specialized to only allow classical polynomials, and taking care of the improved dependence on $|\mathbb{F}|$ guaranteed by Theorem 8.2.1. As a corollary, we state the above result for the case of parallelepipeds, which will be needed for the inductive proof of Theorem 8.0.1. We first set up some notations, following Section 4 in [41].

Let $\mathcal{B} = \{Q_1, \ldots, Q_c\}$ be a polynomial factor of degree at most d. We assume $\mathcal{B}$ has rank at least $r_{(8.2.1)}(d, s)$. For $i \in [d]$, let M_i denote the number of polynomials in $\mathcal{B}$ of degree exactly equal to i. Let $\Sigma := \otimes_{i \in [d]} \mathbb{F}^{M_i}$.

Definition 8.2.4 (Faces and lower faces)**.** *Let $k \in \mathbb{N}$ and $0 \leq k' \leq k$. A set $F \subseteq \{0,1\}^k$ is called a* face *of dimension k' if*

$$F = \{b \in \{0,1\}^k : b_i = \delta_i, i \in I\},$$

where $I \subseteq [k]$, $|I| = k - k'$ and $\delta_i \in \{0,1\}$. If $\delta_i = 0$ for all $i \in I$, then F is called a lower face*.*

We will consider vectors $v \in \Sigma^{\{0,1\}^k} \cong |\mathbb{F}|^{(\sum_{i\in[d]}|M_i|)2^k}$ indexed as $v(i,j,w)$ for $i \in [d], j \in [M_i], w \in \{0,1\}^k$.

Definition 8.2.5 (Face vectors and parallelepiped constraints)**.** *Let $i_0 \in [d]$, $j_0 \in [M_{i_0}]$ and $F \subseteq \{0,1\}^k$. Let $v(i_0, j_0, F) \in \Sigma^{\{0,1\}^k}$ be given by $v(i,j,\omega) = (-1)^{|\omega|}$ if $i = i_0, j = j_0$ and $\omega \in F$ and zero otherwise. This is called a face vector. If F is a lower face, then it corresponds to a lower face vector. If $\dim(F) \geq i_0 + 1$, then it is a relevant face (lower face) vector. A vector $t \in \Sigma^{\{0,1\}^k}$ satisfies the parallelepiped constraints if it is orthogonal to all the relevant lower face vectors.*

Let $\Sigma_0 \subseteq \Sigma^{\{0,1\}^k}$ be the subspace of vectors satisfying the parallelepiped constraints. Below we use the following shorthand: $\binom{k}{\leq i} := \sum_{j=0}^{i}\binom{k}{j}$.

Claim 8.2.6 (Dimension of Σ_0, Lemma 4.4 [41])**.** *Let $d < k$. Then,*

$$\dim(\Sigma_0) = \sum_{i=1}^{d} M_i \binom{k}{\leq i}.$$

Lemma 8.2.7 (Equidistribution of parallelepipeds)**.** *Suppose Theorem 8.0.1 is true up to degree d. Given $s, d < k \in \mathbb{N}$, let $\mathcal{B}$ be a polynomial factor of rank at least $r_{(8.2.7)}(k,s)$ defined by polynomials $Q_1, \ldots, Q_c : \mathbb{F}^n \to \mathbb{F}$ of degree at most d. Then for every $t \in \Sigma_0$ and every $x \in \mathbb{F}^n$ such that $\mathcal{B}(x) = t(0)$, it holds that*

$$\Pr_{y\in(\mathbb{F}^n)^k}\left[\mathcal{B}(x + \omega \cdot y) = t(\omega)\ \forall \omega \in \{0,1\}^k\right] = \frac{1}{|\mathbb{F}|^{\sum_{i=1}^{d} M_i \binom{k}{\leq i}}} \pm \frac{1}{|\mathbb{F}|^s}.$$

Proof. This immediately follows from the dimension of Σ_0 (Theorem 8.2.6) and Theorem 8.2.3 applied to the parallelepiped. □

Proof of Theorem 8.0.1. The base case of $d = 1$ is trivial. Indeed, if a linear polynomial $P : \mathbb{F}^n \to \mathbb{F}$ satisfies $|\mathbb{E}[e(P(x))]| \geq |\mathbb{F}|^{-s}$, then by orthogonality of linear polynomials, we have $P(x)$ is a constant and hence has rank 0.

So, suppose that the hypothesis is true for degrees up to $d-1$, and we will prove it for d. Let $t \in \mathbb{N}$ depending on d be specified later. Recall that we assume that $|\mathbb{E}[e(P(x))]| \geq |\mathbb{F}|^{-s}$. By Theorem 8.1.1, there exists a polynomial factor $\mathcal{B} = \{Q_1, \ldots Q_c\}$ of degree $d-1$ where $c = c(d,s,t)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that

$$\Pr_{x\in\mathbb{F}^n}[P(x) \neq \Gamma(Q_1(x), \cdots, Q_c(x))] \leq |\mathbb{F}|^{-t}.$$

Let $r : \mathbb{N} \to \mathbb{N}$ be a growth function that depends on d and will be specified later. Regularize $\mathcal{B}$ to an r-regular polynomial factor $\mathcal{B}' = \{Q'_1, \ldots, Q'_{c'}\}$, $c' \leq C^{(7.5.4)}_{r,d}(c)$. Thus, we have for an appropriate $\Gamma' : \mathbb{F}^{c'} \to \mathbb{F}$ that

$$\Pr_{x\in\mathbb{F}^n}[P(x) \neq \Gamma'(Q'_1(x), \cdots, Q'_{c'}(x))] \leq |\mathbb{F}|^{-t}.$$

In the rest of the proof, we prove that P is $\mathcal{B}'$-measurable. This will conclude the proof.

We will assume that $r(j) \geq r_{(8.2.2)}(d, 2t + j)$ for all $j \in \mathbb{N}$. Say that an atom A of $\mathcal{B}'$ is *good* if

$$\Pr_{x\in A}[P(x) \neq \Gamma'(Q'_1(x), \cdots, Q'_{c'}(x))] \leq |\mathbb{F}|^{-t/4}.$$

By Markov's inequality and Theorem 8.2.2, for at least $1 - |\mathbb{F}|^{-t/4}$ fraction of the atoms A of $\mathcal{B}'$ are good. The first step is to prove that on such atoms, P is constant. Fix a good atom A and let $A' \subseteq A$ be the set where $P(x) = \Gamma'(Q'_1(x), \cdots, Q'_{c'}(x))$. Observe that $\Gamma'(Q'_1(x), \cdots, Q'_{c'}(x))$ is constant on A, and hence $P(x) = c_A$ for some constant c_A for all $x \in A'$.

Lemma 8.2.8. *Let t be large enough depending on d. Let $x \in A$ be arbitrary. Then there exists $h \in (\mathbb{F}^n)^{d+1}$ such that $x + \omega \cdot h \in A'$ for all $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$.*

We omit the proof of Theorem 8.2.8, and note for the interested reader that it is identical to the proof of Lemma 5.2 in [41]. Continuing, since P is a degree d polynomial, we have

$$\sum_{\omega \in \{0,1\}^{d+1}} (-1)^{|\omega|} P(x + \omega \cdot h) = 0.$$

Now, by Theorem 8.2.8, we have $P(x + \omega \cdot h) = c_A$ for all $\omega \neq 0$. Thus also $P(x) = c_A$, which means that $P(x)$ is constant on the entire atom A.

This finishes the first step. Thus, we have for $1 - |\mathbb{F}|^{-t/4}$ fraction of the atoms A of $\mathcal{B}$, i.e the good atoms, that $P(x) = c_A$ for all $x \in A$. The final step shows that for any arbitrary atom A, there exist good atoms A_ω, $\omega \in \{0,1\}^{d+1} \setminus \{0\}^{d+1}$ such that the vector $t = \mathcal{B}(A_\omega) \in \Sigma^{\{0,1\}^{d+1}}$ satisfies the parallelepiped constraints. It is enough to find one parallelepiped for which $x + \omega \cdot h$ lie in good atoms for all $\omega \neq 0$. Indeed, let $x \in A$ be arbitrary. Pick $h_1, \ldots, h_{d+1} \in \mathbb{F}^n$ randomly. The probability that for a fixed $\omega \neq 0$, $x + \omega \cdot h$ lies in a good atom is at least $1 - |\mathbb{F}|^{-t/4} > 1 - 2^{-2d}$ for t large enough. The result now follows by a union bound over $\omega \in \{0,1\}^{d+1}$. □

Chapter 9

Decomposition Theorems

"Decomposition theorems" [36, 74, 41] are important consequences of inverse theorems. They allow to decompose an arbitrary function into a "structured" part and a "pseudorandom" part. The structured part is simple enough to be analyzed directly (often when assuming regularity and equidistribution as a result of regularization lemmas), and the pseudorandomness requirements are usually set up so that the pseudorandom part has little effect on the analysis and can often be ignored as small noise. We refer the interested reader to [36] and [61] for a detailed discussion of decomposition theorems of this type and how the finite-dimensional Hahn-Banach theorem can be used to give short and transparent proofs of many results of these kinds.

9.1 Basic decomposition theorem

The following decomposition theorem is a direct consequence of the inverse theorem for Gowers norms, Theorem 6.2.3. Recall that $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$ denotes the unit disk in the complex plane.

Theorem 9.1.1 (Basic decomposition theorem)**.** *Let $\mathbb{F}$ be a prime finite field, $d \geq 1, \varepsilon > 0$ and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary function. Every function $f : \mathbb{F}^n \to \mathbb{D}$ can be decomposed as*

$$f = g + h,$$

where

(i) $g = \mathbb{E}[f|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor of degree at most d and complexity $C \leq C_{9.1.1}(\mathbb{F}, d, r, \varepsilon)$.

(ii) $\|h\|_{U^{d+1}} < \varepsilon$.

Furthermore, if we are given an initial polynomial factor $\mathcal{B}_1$ of degree at most d, then we may assume that $\mathcal{B}$ is a refinement of $\mathcal{B}_1$, and in which case we have that $C = |\mathcal{B}| \leq C_{9.1.1}(\mathbb{F}, d, r, \varepsilon, |\mathcal{B}_1|)$.

Proof. We define a sequence $\mathcal{B}_1, \mathcal{B}_2, \ldots$ of r-regular polynomial factors of degree at most d, where if $\mathcal{B}_1$ is not given then we take $\mathcal{B}_1 = \emptyset$. Let $g_i = \mathbb{E}[f|\mathcal{B}_i]$ and $h_i = f - g_i$. We will show that for a bounded value of i we have $\|h_i\|_{U^{d+1}} < \varepsilon$ in which case we are done, as we can take $g = g_i, h = h_i$.

So, consider $i \geq 1$ with $\|h_i\|_{U^{d+1}} \geq \varepsilon$. By Theorem 6.2.3, there exists a polynomial $P_i : \mathbb{F}^n \to \mathbb{T}$ of degree $\leq d$ such that $|\langle h_i, e(P_i)\rangle| \geq \delta$, where $\delta = \delta_{6.2.3}(\mathbb{F}, d, \varepsilon)$. In this case, let $\mathcal{B}_i' = \mathcal{B}_i \cup \{P_i\}$ and let $\mathcal{B}_{i+1}$ be an r-regular factor refining $\mathcal{B}_i'$, as given by Theorem 7.5.5. Note that $|\mathcal{B}_{i+1}|$ is bounded by a function of $|\mathcal{B}_i|, d, r$. Concretely, $|\mathcal{B}_{i+1}| \leq C_{7.5.5}^{\mathbb{F},d,r}(|\mathcal{B}_i| + 1)$. In order to show that this process terminates after a bounded number of steps, we will show that $\|g_{i+1}\|_2^2 \geq \|g_i\|_2^2 + \delta^2$, and hence the process must terminate after at most $1/\delta^2$ steps.

To see that, first note that $\langle g_i, h_i \rangle = 0$, as the average of $h_i = f - \mathbb{E}[f|\mathcal{B}_i]$ in any atom of $\mathcal{B}_i$ is zero, while g_i is $\mathcal{B}_i$-measurable, and hence is constant on any atom of $\mathcal{B}_i$. In particular, this implies that $\|g_i\|_2^2 + \|h_i\|_2^2 = \|f\|_2^2 \leq 1$. Next, let $g_i' = \mathbb{E}[f|\mathcal{B}_i']$ and $h_i' = f - g_i'$. As $\mathcal{B}_i'$ is a refinement of $\mathcal{B}_i$ we also have that $\langle g_i, g_i' - g_i \rangle = 0$. Thus

$$\|g_i'\|_2^2 = \|g_i\|_2^2 + \|g_i' - g_i\|_2^2 = \|g_i\|_2^2 + \|h_i' - h_i\|_2^2.$$

Next, note that $|\langle h_i, e(P_i) \rangle| \geq \delta$ but $\langle h_i', e(P_i) \rangle = 0$, since in any atom of $\mathcal{B}_i'$ we have that P_i is constant and the average of h_i' is zero. Thus

$$\|h_i' - h_i\|_2^2 = \mathbb{E}_x |h_i'(x) - h(x)|^2 = \mathbb{E}_x |(h_i'(x) - h(x))e(P_i(x))|^2 \geq |\mathbb{E}_x \left[(h_i'(x) - h(x))e(P_i(x))\right]|^2 \geq \delta^2.$$

To conclude the proof, note that as $\mathcal{B}_{i+1}$ is a refinement of $\mathcal{B}_i'$ we have that $\|g_{i+1}\|_2^2 = \|g_i'\|_2^2 + \|g_{i+1} - g_i'\|_2^2 \geq \|g_i'\|_2^2$. □

9.2 Higher-order Fourier expansion

Let $f : \mathbb{F}^n \to \mathbb{D}$ be a function, $\mathcal{B}$ a polynomial factor and let $g = \mathbb{E}[f|\mathcal{B}]$. It is possible to express g as a function of the polynomials which define $\mathcal{B}$. Assume that $\mathcal{B}$ is defined by polynomials $\{P_1, \ldots, P_C\}$. Then as g is constant on the atoms of $\mathcal{B}$, we can express

$$g(x) = \Gamma(P_1(x), \ldots, P_C(x)),$$

for some function $\Gamma : \mathbb{T}^c \to \mathbb{T}$. More concretely, assume that P_i has depth k_i, and hence takes values in $\mathbb{U}_{k_i+1}$. Let $G = \prod_{i=1}^C \mathbb{U}_{k_i+1}$ and let $G' = \prod_{i=1}^C \mathbb{Z}_{p^{k_i+1}}$ be the dual group. Applying standard Fourier decomposition to $\Gamma : G \to \mathbb{C}$ gives

$$\Gamma(z) = \sum_{\alpha \in G'} \hat{\Gamma}(\alpha) e(\langle \alpha, z \rangle).$$

As $\|g\|_\infty \leq \|f\|_\infty \leq 1$ we have $\|\Gamma\|_\infty \leq 1$. Hence by Parseval's identity, $\sum |\hat{\Gamma}(\alpha)|^2 \leq 1$. We can apply the same decomposition to g. Define polynomials $P_\alpha(x) = \sum \alpha_i P_i(x)$ for $\alpha \in G'$. Then the higher-order Fourier expansion of g is given by

$$g(x) = \sum_{\alpha \in G'} \hat{\Gamma}(\alpha) e(P_\alpha(x)).$$

Note that the polynomials P_α play the role of characters in standard Fourier analysis. In standard Fourier analysis, characters are orthogonal, which plays an important role. If we assume that the polynomial factor $\mathcal{B}$ has a high enough rank, then the same approximately holds for higher-order Fourier expansions.

Claim 9.2.1. *Assume that* $\mathrm{rank}(\mathcal{B}) \geq r_{9.2.1}(\mathbb{F}, d, \varepsilon, |\mathcal{B}|)$. *Then*

$$\left| \langle g, e(P_\alpha) \rangle - \hat{\Gamma}(\alpha) \right| \leq \varepsilon.$$

Proof. Assume that $\mathcal{B}$ is r-regular for $r = r_{7.2.2}(\mathbb{F}, d, \varepsilon/|G|)$, where note that $|G| \leq p^{\lceil (d-1)/(p-1) \rceil |\mathcal{B}|}$. By the regularity assumption, $\mathrm{bias}(P_{\alpha'}) \leq \varepsilon/|G|$ for all $\alpha' \neq 0$. Thus

$$\langle g, e(P_\alpha) \rangle = \mathbb{E}_x \left[g(x) e(-P_\alpha(x)) \right] = \sum_{\alpha'} \widehat{\Gamma}(\alpha') \mathbb{E}_x \left[e(P_{\alpha'}(x) - P_\alpha(x)) \right] = \sum_{\alpha'} \widehat{\Gamma}(\alpha') \mathrm{bias}(P_{\alpha' - \alpha})$$

and

$$\left| \langle g, e(P_\alpha) \rangle - \hat{\Gamma}(\alpha) \right| \leq \sum_{\alpha' \neq \alpha} |\widehat{\Gamma}(\alpha')| |\mathrm{bias}(P_{\alpha' - \alpha})| \leq \varepsilon.$$

□

Via a similar analysis we can obtain an approximate version of Plancheral's theorem when the expansion has high rank.

9.3 Strong decomposition theorems

In many applications, once we decompose $f = g + h$ where $g = \mathbb{E}[f|\mathcal{B}]$ and $\|h\|_{U^{d+1}} < \varepsilon$, then it is necessary to make ε small in terms of the complexity of $\mathcal{B}$. It turns out that this is possible, if we allow another L_2 error term.

Theorem 9.3.1 (Strong decomposition theorem)**.** *Let $\mathbb{F}$ be a prime finite field, $d \geq 1, \delta > 0$ and let $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ and $r : \mathbb{N} \to \mathbb{N}$ be arbitrary functions. Any function $f : \mathbb{F}^n \to \mathbb{D}$ can be decomposed as*

$$f = f_1 + f_2 + f_3$$

where

(i) $f_1 = \mathbb{E}[f|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor of degree at most d and complexity $C \leq C_{9.1.1}(\mathbb{F}, d, r, \varepsilon, \delta)$.

(ii) $\|f_2\|_{U^{d+1}} < \varepsilon(|\mathcal{B}|)$ and $\|f_2\|_\infty \leq 2$.

(iii) $\|f_3\|_2 \leq \delta$.

Furthermore, if we are given an initial polynomial factor $\mathcal{B}_1$ of degree at most d, then we may assume that $\mathcal{B}$ is a refinement of $\mathcal{B}_1$, in which case we have that $C = |\mathcal{B}| \leq C_{9.1.1}(\mathbb{F}, d, r, \varepsilon, \delta, |\mathcal{B}_1|)$.

Proof. Define a sequence $\mathcal{B}_1, \mathcal{B}_2, \ldots$ of polynomial factors as follows. The first factor $\mathcal{B}_1$ is given as input, and otherwise let $\mathcal{B}_1 = \emptyset$. To obtain $\mathcal{B}_{i+1}$, apply Theorem 9.1.1 with initial factor $\mathcal{B}_i$ and error $\varepsilon_i = \varepsilon(|\mathcal{B}_i|)$. Define as before $g_i = \mathbb{E}[f|\mathcal{B}_i]$ and $h_i = f - g_i$. Note that as $\mathcal{B}_{i+1}$ is a refinement of $\mathcal{B}_i$ we have that $\langle g_i, g_{i+1} - g_i \rangle = 0$ and hence $\|g_{i+1}\|_2^2 = \|g_i\|_2^2 + \|g_{i+1} - g_i\|_2^2$. Let $i \leq 1/\delta^2$ be minimal such that $\|g_{i+1}\|_2^2 \leq \|g_i\|_2^2 + \delta^2$. We then take

$$\mathcal{B} = \mathcal{B}_i, \quad f_1 = g_i, \quad f_2 = h_{i+1}, \quad f_3 = h_i - h_{i+1}.$$

Note that by definition, $f_1 = \mathbb{E}[f|\mathcal{B}]$ and $\|f_2\|_{U^{d+1}} \leq \varepsilon(|\mathcal{B}|)$. It is also simple to verify that $\|f_2\|_\infty \leq \|f\|_\infty + \|g_{i+1}\|_\infty \leq 2$. Finally by our construction we have that $\|f_3\|_2^2 = \|h_i - h_{i+1}\|_2^2 = \|g_i - g_{i+1}\|_2^2 = \|g_{i+1}\|_2^2 - \|g_i\|_2^2 \leq \delta^2$. □

If we specialize Theorem 9.3.2 to Boolean functions, we can have more control on the ranges of f_1, f_2, f_3.

Theorem 9.3.2 (Strong decomposition theorem, Boolean functions)**.** *Let $f : \mathbb{F}^n \to \{0, 1\}$. Under the same conditions as that of Theorem 9.3.1, we obtain the same decomposition $f = f_1 + f_2 + f_3$, which furthermore satisfies that f_1 and $f_1 + f_3$ take values in $[0, 1]$; and f_2 and f_3 take values in $[-1, 1]$.*

Proof. We have $f_1 = g_i = \mathbb{E}[f|\mathcal{B}_i]$ and $f_1 + f_3 = g_{i+1} = \mathbb{E}[f|\mathcal{B}_{i+1}]$, which guarantees that they take values in $[0, 1]$. We have $f_2 = f - \mathbb{E}[f|\mathcal{B}_i]$ and $f_3 = \mathbb{E}[f|\mathcal{B}_i] - \mathbb{E}[f|\mathcal{B}_{i+1}]$, which guarantees that they take values in $[-1, 1]$. □

Theorem 9.3.1 can also be extended to decompose multiple functions with a single polynomial factor, which is useful in certain applications. The proof is identical and is omitted.

Theorem 9.3.3 (Strong decomposition theorem, multiple functions)**.** *Let $\mathbb{F}$ be a prime finite field, $d \geq 1, m \geq 1, \delta > 0$ and let $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ and $r : \mathbb{N} \to \mathbb{N}$ be arbitrary functions. Any m functions $f^{(1)}, \ldots, f^{(m)} : \mathbb{F}^n \to \mathbb{D}$ can be decomposed as*

$$f^{(i)} = f_1^{(i)} + f_2^{(i)} + f_3^{(i)}$$

where

- *$f_1^{(i)} = \mathbb{E}[f^{(i)}|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor of degree at most d and complexity $C \leq C_{9.3.3}(\mathbb{F}, d, m, r, \varepsilon, \delta)$.*
- *$\|f_2^{(i)}\|_{U^{d+1}} < \varepsilon(|\mathcal{B}|)$ and $\|f_2^{(i)}\|_\infty \leq 2$.*
- *$\|f_3^{(i)}\|_2 \leq \delta$.*

Furthermore, if we are given an initial polynomial factor $\mathcal{B}_1$ of degree at most d, then we may assume that $\mathcal{B}$ is a refinement of $\mathcal{B}_1$, and in which case we have that $C = |\mathcal{B}| \leq C_{9.3.3}(\mathbb{F}, d, m, r, \varepsilon, \delta, |\mathcal{B}_1|)$.

9.4 Sub-atom selection

It turns out that the strong decomposition theorems are not sufficiently strong for certain applications in algebraic property testing. These require to control all the error terms, including the L_2 error term, by a function of the complexity of the polynomial factor. This is impossible if we have just one polynomial factor, but becomes possible when working in parallel with two factors: a polynomial factor and a refinement of it.

First, we need to define the notion of a polynomial factor representing another polynomial factor for a function. As all the applications of these tools apply to Boolean functions, we specialize the treatment below for Boolean functions, and note that it can be easily extended to bounded complex functions.

Definition 9.4.1 (Polynomial factor represents another factor)**.** *Let $f : \mathbb{F}^n \to \{0,1\}$, $\mathcal{B}$ a polynomial factor and $\mathcal{B}'$ a refinement of it. For $0 < \zeta < 1$, we say that $\mathcal{B}'$ ζ-represents $\mathcal{B}$ with respect to f, if for at most ζ fraction of atoms c of $\mathcal{B}$, more than ζ fraction of the atoms c' of $\mathcal{B}'$ lying inside c satisfy $|\mathbb{E}[f|c] - \mathbb{E}[f|c']| > \zeta$.*

We now state the following "Two-level decomposition theorem" from [14] (it is referred to as "super decomposition theorem" in [14]).

Theorem 9.4.2 (Two-level decomposition theorem; Theorem 4.9 of [14])**.** *Let $d \geq 1, \zeta > 0$ and let $\varepsilon, \delta : \mathbb{N} \to \mathbb{R}^+$ and $r : \mathbb{N} \to \mathbb{N}$ be arbitrary functions. Given any function $f : \mathbb{F}^n \to \{0,1\}$ there exists a polynomial factor $\mathcal{B}$ of degree d, and a refinement of it $\mathcal{B}'$ of degree d, both r-regular and of complexity at most $C_{9.4.2}(\mathbb{F}, d, r, \varepsilon, \delta, \zeta)$, such that the following holds. We can decompose*

$$f = f_1 + f_2 + f_3$$

where

(i) $f_1 = \mathbb{E}[f|\mathcal{B}']$.

(ii) $\|f_2\|_{U^{d+1}} < \varepsilon(|\mathcal{B}'|)$.

(iii) $\|f_3\|_2 \leq \delta(|\mathcal{B}|)$.

(iv) f_1 *and* $f_1 + f_3$ *take values in* $[0,1]$*; and* f_2, f_3 *take values in* $[-1,1]$.

(v) $\mathcal{B}'$ *ζ-represents* $\mathcal{B}$ *with respect to* f.

Proof. Define a sequence $\mathcal{B}_1, \mathcal{B}_2, \ldots$ of polynomial factors as follows, where initially $\mathcal{B}_1$ is empty, and $\mathcal{B}_{i+1}$ is obtained by applying Theorem 9.3.2 with the parameters d, r, ε and $\delta_i = \delta(|\mathcal{B}_i|)$. If $\mathcal{B}_{i+1}$ ζ-represents $\mathcal{B}_i$ with respect to f, then the theorem follows with $\mathcal{B} = \mathcal{B}_i, \mathcal{B}' = \mathcal{B}_{i+1}$, and f_1, f_2, f_3 as given in the decomposition with $\mathcal{B}_{i+1}$. If not, then one can verify that

$$\mathbb{E}[f|\mathcal{B}_{i+1}]^2 \geq \mathbb{E}[f|\mathcal{B}_i]^2 + \zeta^3.$$

Hence, for some $i \leq 1/\zeta^3$, the condition will hold. □

Although the above two-level decomposition theorem may be useful by itself for some applications, certain algebraic property testing applications require the ability to choose a sub-atom in $\mathcal{B}'$ inside each atom of $\mathcal{B}$, such that the error of f_3 on all sub-atoms is bounded, and most sub-atoms represent their atoms. Moreover, we would need this choice to be algebraically consistent.

To define this latter condition formally, assume that $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, and thus is defined by adding new polynomials $Q_1, \ldots, Q_{|\mathcal{B}'|-|\mathcal{B}|}$ to the polynomials defining $\mathcal{B}$. Thus, we can describe atoms of $\mathcal{B}'$ as (c, s), where $c \in \mathbb{T}^{|\mathcal{B}|}$ describes an atom of $\mathcal{B}$ and $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$. The choice of s allows to choose a sub-atom (c, s) of the finer factor within an atom c of the coarser partition.

We require that there exists a *fixed* $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$, such that

- For all atoms c in $\mathcal{B}$, the L^2 error term f_3 within the corresponding sub-atom is small.

- For most atoms c in $\mathcal{B}$, the sub-atom (c, s) represents the atom c, in the sense that $\mathbb{E}[f|c] \approx \mathbb{E}[f|(c, s)]$.

The following theorem formalizes this. It is presented for one function, but can be easily extended to allow multiple functions, in the same way that Theorem 9.3.3 extends Theorem 9.3.1 to multiple functions.

Theorem 9.4.3 (Subatom selection; Theorem 4.12 of [14])**.** *Let $f : \mathbb{F}^n \to \{0, 1\}$. Under the same conditions as in Theorem 9.4.2, there exists a polynomial factor $\mathcal{B}$ and a syntactic refinement of it $\mathcal{B}'$, whose atoms are indexed by (c, s) with $c \in \mathbb{T}^{|\mathcal{B}|}$, $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$, such that in addition to the guarantees of Theorem 9.4.2, there exists a choice of $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$ for which the following is true:*

(vi) For every atom c of $\mathcal{B}$, the sub-atom (c, s) of $\mathcal{B}'$ satisfies that

$$\mathbb{E}\left[|f_3(x)|^2 \mid \mathcal{B}'(x) = (c, s)\right] \leq \delta(|\mathcal{B}|)^2.$$

(vii) For at most a ζ fraction of atoms c in $\mathcal{B}$ it holds that

$$|\mathbb{E}[f|c] - \mathbb{E}[f|(c, s)]| > \zeta.$$

Proof sketch. The proof is very similar to the proof of Theorem 9.4.2. The fact that $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$ can be guaranteed by making sure that $\mathcal{B}$ has high enough rank. Choose s uniformly among the possible set of values. Condition (vi) holds with high probability by a union bound on all the $|\mathcal{B}|$ possible atoms, by choosing $\delta(|\mathcal{B}|) \ll 1/|\mathcal{B}|$ small enough. Condition (vii) follows by condition (v) and Markov inequality. □

Chapter 10

Homogeneous Nonclassical Polynomials

The main difficulty in dealing with fields of low characteristic is that in the higher-order Fourier expansions, instead of the exponentials of classical polynomials, one has to work with exponentials of nonclassical polynomials. These do not share many of the convenient properties of classical polynomials. To overcome these difficulties, we develop a theory of nonclassical homogeneous polynomials which will enable us to confine to this simpler class of nonclassical polynomials. Recall that the need for homogeneous nonclassical polynomials arises in Theorem 7.6.1, which characterizes the joint distribution of multiple high-rank polynomials evaluated in parallel on multiple linear forms.

A classical degree d polynomial $P : \mathbb{F}^n \to \mathbb{F}$ is called homogeneous if all of its monomials are of degree d. A simple consequence of homogeneity is that every homogeneous classical polynomial $P(x)$ satisfies $P(cx) = c^d P(x)$ for every $c \in \mathbb{F}$. It turns out that this consequence is the "right" way to generalize homogeneity for nonclassical polynomials.

Definition 10.0.1 (Homogeneity)**.** *A (nonclassical) polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is called homogeneous if, for every $c \in \mathbb{F}$, there exists $\sigma_c \in \mathbb{Z}$ such that $P(cx) = \sigma_c P(x)$ for all $x \in \mathbb{F}^n$.*

We start with a simple yet useful constraint on σ_c.

Claim 10.0.2. *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree d polynomial. Assume that $P(cx) = \sigma_c P(x)$ for some $\sigma_c \in \mathbb{Z}$. Then*

$$\sigma_c = |c|^{\deg(P)} \ mod\ p.$$

Proof. Recall that ∂P is the derivative polynomial of P (see Theorem 7.3.1). Define $Q(x) = P(cx) - \sigma_c P(x)$, where by assumption $Q = 0$. Then

$$0 = \partial Q(h_1, \ldots, h_d) = \partial P(ch_1, \ldots, ch_d) - \sigma_c \partial P(h_1, \ldots, h_d).$$

By Theorem 7.3.3 ∂P is a classical homogeneous polynomial of degree d which is linear in each of $h_1, \ldots, h_d$. Thus $\partial P(ch_1, \ldots, ch_d) = |c|^d \partial P(h_1, \ldots, h_d)$. So we obtained that

$$0 = (|c|^d - \sigma_c)\partial P(h_1, \ldots, h_d).$$

As ∂P is a nonzero classical polynomial this implies that $|c|^d - \sigma_c = 0 \bmod p$. □

Notice that for a polynomial P to be homogeneous, it suffices that for some generator ζ of F^* (the multiplicative group of the field $\mathbb{F}$) it holds that $P(\zeta x) = \sigma P(x)$. This is since any nonzero $c \in F$ can be represented as $c = \zeta^m$ for some $0 \le m \le p-1$, and hence $P(cx) = \sigma^m P(x)$.

If P has depth k, then we can assume that $\sigma \in \{0, \ldots, p^{k+1} - 1\}$ since $p^{k+1} P \equiv 0$. Obviously, the value of σ depends on the choice of ζ. However, the following lemma shows that for a fixed ζ, the value of σ is

uniquely determined for all homogeneous polynomials of degree d and depth k. Henceforth, we will denote this unique value by $\sigma(d,k)$.

Let $\mathbf{Z}_p$ denote the p-adic integers. We will show that there exists a choice of $\sigma(d) \in \mathbf{Z}_p$ such that $\sigma(d,k) = \sigma(d) \bmod p^{k+1}$ for all $k \geq 0$. This is related to the so-called Teichmüller characters (see e.g. Section 4.3 of [21]). However, in order to keep the presentation elementary, we avoid exploiting this connection.

Lemma 10.0.3. *Fix a prime finite field $\mathbb{F} = \mathbb{F}_p$ and a generator $\zeta \in \mathbb{F}^*$. For every $d \geq 1, k \geq 0$ there exists a unique $\sigma = \sigma(d,k) \in \{0,\ldots,p^{k+1}-1\}$, such that for every homogeneous polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d and depth k, it holds that $P(\zeta x) = \sigma P(x)$. Furthermore, there exists $\sigma(d) \in \mathbf{Z}_p$ such that $\sigma(d,k) = \sigma(d) \bmod p^{k+1}$.*

Proof. Let P be a homogeneous polynomial of degree d and depth k, and let $\sigma \in \{0,\ldots,p^{k+1}-1\}$ be such that $P(\zeta x) = \sigma P(x)$. By Theorem 10.0.2 we know that $\sigma = |\zeta|^d \bmod p$. Observe that as $\zeta^{p-1} = 1$ it holds that $P(x) = P(\zeta^{p-1}x) = \sigma^{p-1}P(x)$, from which and the assumption that P has depth k it follows that $\sigma^{p-1} \equiv 1 \bmod p^{k+1}$.

We claim that $\sigma \in \{0,\ldots,p^{k+1}-1\}$ is uniquely determined by these two properties:

(i) $\sigma = |\zeta|^d \bmod p$;

(ii) $\sigma^{p-1} = 1 \bmod p^{k+1}$.

Suppose to the contrary that there exist two nonzero values $\sigma_1, \sigma_2 \in \mathbb{Z}_{p^{k+1}}$ that satisfy the above two properties, and choose $t \in \mathbb{Z}_{p^{k+1}}$ such that $\sigma_1 = t\sigma_2$. It follows from (i) that $t = 1 \bmod p$ and from (ii) that $t^{p-1} = 1$. We will show that $t = 1$ is the only possible such value in $\mathbb{Z}_{p^{k+1}}$.

Let $a_1,\ldots,a_{p^k} \in \mathbb{Z}_{p^{k+1}}$ be all the possible solutions to $x = 1 \bmod p$ in $\mathbb{Z}_{p^{k+1}}$. Note that $ta_1,\ldots,ta_{p^k}$ is a permutation of the first sequence and thus

$$t^{p^k} \prod a_i = \prod a_i.$$

Consequently $t^{p^k} = 1 \bmod p^{k+1}$, which combined with $t^p = t$ implies $t = 1 \bmod p^{k+1}$.

For the last assertion, note that $\sigma(d,k)^{p-1} = 1 \bmod p^{k+1}$ implies $\sigma(d,k)^{p-1} = 1 \bmod p^{\ell+1}$ for every $\ell < k$. By the uniqueness of $\sigma(d,\ell)$, this implies that $\sigma(d,k) = \sigma(d,\ell) \bmod p^{\ell+1}$. We can thus take $\sigma(d) \in \mathbf{Z}_p$ given by the equations $\sigma(d) \bmod p^{k+1} = \sigma(d,k)$. □

Theorem 6.1.2 allows us to express every nonclassical polynomial as a linear span of monomials of the form $\frac{|x_1|^{d_1}\cdots|x_n|^{d_n}}{p^{k+1}}$. Unfortunately, unlike in the classical case, these monomials are not necessarily homogeneous, and for some applications it is important to express a polynomial as a linear span of homogeneous polynomials. We show that this is possible as homogeneous nonclassical polynomials form a basis for the space of nonclassical polynomials.

10.1 A homogeneous basis for nonclassical polynomials

In this section we will prove that homogeneous polynomials span the space of all nonclassical polynomials.

Theorem 10.1.1. *There is a basis for* $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$ *consisting only of homogeneous multivariate polynomials.*

This theorem allows us to make the extra assumption in decomposition theorems, that the resulting polynomial factor $\mathcal{B}$ consists only of homogeneous polynomials. In order to achieve that, we decompose each polynomial of degree d and depth k as a sum of homogeneous polynomials of degree d and depths $1,\ldots,k$. If one of these polynomials has low rank, we decompose it to a few lower degree polynomials, and repeat the process.

We start by proving the following simple observation.

Claim 10.1.2. *Let $P : \mathbb{F} \to \mathbb{T}$ be a univariate polynomial of degree d. Then for every $c \in \mathbb{F}\backslash\{0\}$,*

$$\deg\left(P(cx) - |c|^d P(x)\right) < d.$$

Proof. By Theorem 6.1.2 it suffices to prove the claim for a monomial $m(x) = \frac{|x|^s}{p^{k+1}}$ with $k(p-1)+s = d$. Note that $m(cx) - |c|^d m(x)$ takes values in $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$ as by Fermat's little theorem

$$|cx|^s - |c|^d|x|^s \equiv |x|^s|c|^s(1 - |c|^{k(p-1)}) \equiv 0 \bmod p.$$

It follows then from Theorem 6.1.2 that $m(cx) - |c|^d m(x)$ is of depth at most $k-1$, and hence

$$\deg\left(m(cx) - |c|^d m(x)\right) \leq (p-1)(k-1) + (p-1) < d. \tag{10.1}$$

□

It is not difficult to show that the above claim holds also for any multivariate polynomial $P : \mathbb{F}^n \to \mathbb{T}$. We provide a proof of this fact, although the univariate case suffices for our purposes in this section.

Claim 10.1.3. *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a multivariate polynomial of degree d. Then for every $c \in \mathbb{F}\backslash\{0\}$,*

$$\deg\left(P(cx) - |c|^d P(x)\right) < d.$$

Proof. Notice that the claim is trivial for classical polynomials, since in this case, if R denotes the homogeneous degree-d part of P, then $R(cx) - |c|^d R(x) = 0$. We prove the statement for nonclassical polynomials. Let $Q(x) := P(cx)$, and note $\deg(Q) = d$. We will inspect the derivative polynomial of Q. Recall from Theorem 7.3.1 and Theorem 7.3.3 that the derivative polynomial of Q,

$$\partial Q(y_1, \ldots, y_d) = D_{y_1} \cdots D_{y_d} Q(0),$$

is a degree-d classical homogeneous multi-linear polynomial which is invariant under permutations of $(y_1, \ldots, y_d)$. In particular

$$\begin{aligned}
|c|^{-d}\partial Q(y_1, \ldots, y_d) &= \partial Q(c^{-1}y_1, \ldots, c^{-1}y_d) = D_{c^{-1}y_1} D_{c^{-1}y_2} \cdots D_{c^{-1}y_d} Q(0) \\
&= \sum_{S \subseteq [d]} (-1)^{d-|S|} Q\left(c^{-1}\sum_{i \in S} y_i\right) = \sum_{S \subseteq [d]} (-1)^{d-|S|} P\left(\sum_{i \in S} y_i\right) \\
&= \partial P(y_1, \ldots, y_d).
\end{aligned}$$

This implies that $\partial(Q - |c|^d P) \equiv 0$ and thus $\deg(Q - |c|^d P) < d$. □

10.1.1 A homogeneous basis, the univariate case

First, we prove Theorem 10.1.1 for univariate polynomials.

Lemma 10.1.4. *There is a basis of homogeneous univariate polynomials for* $\mathrm{Poly}(\mathbb{F} \to \mathbb{T})$.

Proof. We will prove by induction on d that there is a basis $\{h_1, \ldots, h_d\}$ of homogeneous univariate polynomials for $\mathrm{Poly}_{\leq d}(\mathbb{F} \to \mathbb{T})$ for every d. Let ζ be a fixed generator of $\mathbb{F}^*$. For any degree $d > 0$, we will build a degree-d homogeneous polynomial $h_d(x)$ such that $h_d(\zeta x) = \sigma_d h_d(x)$ for some integer σ_d. The base case of $d \leq p-1$ is trivial as $\mathrm{Poly}_{\leq p-1}(\mathbb{F} \to \mathbb{T})$ consists of only classical polynomials, and those are spanned by $h_0(x) := \frac{1}{p}, h_1(x) := \frac{|x|}{p}, \ldots, h_{p-1}(x) := \frac{|x|^{p-1}}{p}$. Now suppose that $d = s + (p-1)(k-1)$ with $0 < s \leq p-1$, and $k > 1$. It suffices to show that the degree-d monomial $\frac{|x|^s}{p^k}$ can be expressed as a linear combination of homogeneous polynomials. Consider the function

$$f(x) := \frac{|\zeta x|^s}{p^k} - \frac{|\zeta|^d |x|^s}{p^k}.$$

Theorem 10.1.2 implies that $\deg(f) < d$. Using the induction hypothesis, we can express $f(x)$ as a linear combination of $\frac{|x|^s}{p^\ell}$ for $\ell = 0, \ldots, k-1$, and h_e for $e < d$ with $e \neq s \bmod (p-1)$:

$$f(x) = \sum_{\ell=1}^{k-1} a_\ell \frac{|x|^s}{p^\ell} + \sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} b_e h_e(x).$$

Set $A := |\zeta|^d + \sum_{\ell=1}^{k-1} a_\ell p^{k-\ell}$, so that

$$\frac{|\zeta x|^s}{p^k} - A\frac{|x|^s}{p^k} = \sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} b_e h_e(x). \tag{10.2}$$

By the induction hypothesis, for $e < d$, $h_e(\zeta x) = \sigma_e h(x)$ where $\sigma_e = |\zeta|^e \bmod p$, and thus as $A = |\zeta|^d \bmod p$, we have $\sigma_e \neq A \bmod p$ when $e \neq d \bmod (p-1)$. Consequently,

$$\begin{aligned}\sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} b_e h_e(x) &= \sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A}(\sigma_e - A)h_e(x) \\ &= \sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A}(h_e(\zeta x) - Ah_e(x)).\end{aligned}$$

Combing this with (10.2) we conclude that

$$h_d(x) := \frac{|x|^s}{p^k} - \sum_{\substack{e<d, \\ e \neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A} h_e(x),$$

satisfies

$$h_d(\zeta x) = Ah_d(x).$$

□

10.1.2 A homogeneous basis, the multivariate case

We are now ready to prove the main result of this section.

Theorem 10.1.1 (restated). *There is a basis for* $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$ *consisting only of homogeneous multivariate polynomials.*

Proof. We will show by induction on the degree d and k, that every degree d monomial of depth k can be written as a linear combination of homogeneous polynomials. The base case of $d < p$, $k = 0$ is trivial as such monomials are classical and thus homogeneous themselves. Consider a nonclassical monomial $m(x_1, \ldots, x_n) = \frac{|x_1|^{s_1} \cdots |x_n|^{s_n}}{p^{k+1}}$ of degree $d = s_1 + \cdots + s_n + (p-1)k$. For every $i \in [n]$ let $g_i(x_i) := h_{s_i+(p-1)k}(x_i)$ where $h_{s_i+(p-1)k}$ is the homogeneous univariate polynomial from Theorem 10.1.4.

Every g_i takes values in $\frac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}$, and thus corresponds to a polynomial $G_i : \mathbb{F} \to \mathbb{Z}_{p^{k+1}}$. That is,

$$g_i(x) = \frac{G_i(x)}{p^{k+1}} \bmod 1.$$

Define $H : \mathbb{F}^n \to \mathbb{Z}_{p^{k+1}}$ as

$$H(x_1, \ldots, x_n) := G_1(x_1) \cdots G_n(x_n),$$

and $h : \mathbb{F}^n \to \mathbb{T}$ as

$$h(x_1, \ldots, x_n) := \frac{H(x_1, \ldots, x_n)}{p^{k+1}}.$$

We claim that

(i) h is a homogeneous polynomial.

(ii) $\deg(h) = d$.

(iii) $\deg(h - m) \leq d - 1$.

Thus by induction m can be written as a sum of homogeneous polynomials. To conclude the proof we verify (i)-(iii).

To verify (i) note that as g_i is a homogeneous univariate polynomial it holds that $g_i(\zeta x_i) = \sigma_i g_i(x_i) \bmod 1$ for some $\sigma_i \in \mathbb{Z}$. Thus $G_i(\zeta x_i) = \sigma_i g_i(x_i) \bmod p^{k+1}$. This implies that $H(\zeta x) = (\prod \sigma_i) H(x) \bmod p^{k+1}$, and hence $h(\zeta x) = (\prod \sigma_i) h(x) \bmod 1$. To verify (ii) and (iii) note that by construction, $G_i(x_i) = x_i^{s_i} + P_i(x_i) + pQ_i(x_i)$ where $\deg(P_i) < s_i$. One can then verify that $H(x)$ contains the monomial $\prod x_i^{s_i}$, and all other monomials are of the form $p^e \prod x_i^{t_i}$, where $\sum t_i \leq d - 1 + (p-1)e$. Thus $h(x)$ contains a single monomial of degree d, which is $m(x)$, and all the remaining monomials have degree $\leq d-1$. Thus $\deg(h-m) \leq d-1$. $\square$

Chapter 11

Complexity of Systems of Linear Forms

Gowers norms control the density of linear patterns in subsets of an Abelian group. For example, given a linear pattern, for a sufficiently large d, if the characteristic function of two sets are close in the U^d norm, then they contain almost the same number of copies of this pattern. For example, if $A, B \subset \mathbb{F}^n$ are two subsets whose indicator functions $1_A, 1_B : \mathbb{F}^n \to \{0,1\}$ satisfy that $\|1_A - 1_B\|_{U^d}$ is negligible, then the number of $(d+1)$-term arithmetic progressions $x, x+y, x+2y, \ldots, x+dy$ that fall in A is approximately the same as the number that falls in B.

More generally, in order to study the density of linear patterns in a set, one can use a decomposition theorem to first decompose the function into a structured part and and a pseudorandom part, and then use this property of Gowers norms to discard the pseudorandom part and restrict the analysis to the simpler structured part. Since for smaller values of d, decomposition theorems provide a simpler structured part, a natural question arises:

Given a linear pattern, what is the smallest value of d for which the above statements hold?

Investigating such questions lead to various notions of complexity associated with a collection of linear forms. In this chapter we study such notions.

11.1 Cauchy-Schwarz complexity

Let A be a subset of $\mathbb{F}^n$ with the indicator function $\mathbf{1}_A : \mathbb{F}^n \to \{0,1\}$. Let $L_1, \ldots, L_m$ be linear forms over $\mathbb{F}$ in ℓ variables. Let $X \in (\mathbb{F}^n)^\ell$ be chosen uniformly at random. The probability that $L_1(X), \ldots, L_m(X)$ all fall in A, can be expressed as

$$\mathbb{E}_X \left[\mathbf{1}_A(L_1(X)) \cdots \mathbf{1}_A(L_m(X)) \right].$$

Roughly speaking, we say $A \subseteq \mathbb{F}^n$ is *pseudorandom* with regards to $\mathcal{L} = (L_1, \ldots, L_m)$ if

$$\mathbb{E}_X \left[\prod_{i=1}^m \mathbf{1}_A(L_i(X)) \right] \approx \left(\frac{|A|}{p^n} \right)^m;$$

That is if the probability that all $L_1(X), \ldots, L_m(X)$ fall in A is close to what we would expect if A was a random subset of $\mathbb{F}^n$ of cardinality $|A|$. Let $\alpha := |A|/|\mathbb{F}|^n$ be the density of A, and define $f := \mathbf{1}_A - \alpha$. We have

$$\mathbb{E}_X \left[\prod_{i=1}^m \mathbf{1}_A(L_i(X)) \right] = \mathbb{E}_X \left[\prod_{i=1}^m (\alpha + f(L_i(X))) \right] = \alpha^m + \sum_{S \subseteq [m], S \neq \emptyset} \alpha^{m-|S|} \mathbb{E}_X \left[\prod_{i \in S} f(L_i(X)) \right].$$

Therefore, a sufficient condition for A to be pseudorandom with regards to $\mathcal{L}$ is that $\mathbb{E}_X\left[\prod_{i\in S} f(L_i(X))\right]$ is negligible for all nonempty subsets $S \subseteq [m]$. Green and Tao [43] showed that a sufficient condition for this to occur is that $\|f\|_{U^{s+1}}$ is small enough, where s is the *Cauchy-Schwarz complexity* of the system of linear forms, defined below.

Definition 11.1.1 (Cauchy-Schwarz complexity [43])**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over a field $\mathbb{F}$ in ℓ variables. The* Cauchy-Schwarz complexity *of $\mathcal{L}$ is the minimal s such that the following holds. For every $1 \le i \le m$, we can partition $\{L_j\}_{j\in[m]\setminus\{i\}}$ into $s+1$ subsets, such that L_i does not belong to the linear span of any of the subsets.*

The reason for the term *Cauchy-Schwarz complexity* is the following lemma due to Green and Tao [43] whose proof is based on carefully chosen iterative applications of the Cauchy-Schwarz inequality.

Lemma 11.1.2 ([43], See also [37, Theorem 2.3])**.** *Let $f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{D}$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms in ℓ variables of Cauchy-Schwarz complexity s. Then*

$$\left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[\prod_{i=1}^m f_i(L_i(X))\right]\right| \le \min_{1\le i\le m} \|f_i\|_{U^{s+1}}.$$

Note that the Cauchy-Schwarz complexity of any system of m linear forms in which every two linear forms are linearly independent (i.e. one is not a multiple of the other) is at most $m-2$, since we can always partition $\{L_j\}_{j\in[m]\setminus\{i\}}$ into the $m-1$ singleton subsets.

As an example, we prove the statement made in the beginning of the chapter regarding arithmetic progressions.

Corollary 11.1.3. *Fix a prime finite field $\mathbb{F}$, $d \ge 2, \varepsilon > 0$. Let $A \subset \mathbb{F}^n$ of density $\alpha = |A|/|\mathbb{F}|^n$. Assume that $\|1_A - \alpha\|_{U^d} \le \varepsilon$. Then the number of $(d+1)$-term arithmetic progressions in A is*

$$\left|\Pr_{x,y\in\mathbb{F}^n}[x, x+y, \ldots, x+dy \in A] - \alpha^{d+1}\right| \le \varepsilon(1+\alpha)^{d+1}.$$

Proof. Let $f = 1_A - \alpha$. Consider the system of $m = d+1$ linear forms $L_i(x,y) = x + iy$ for $i = 0, \ldots, d$. Its Cauchy-Schwarz complexity is $s = d-1$. Theorem 11.1.2 gives that for every nonempty $S \subseteq [d+1]$,

$$\mathbb{E}_{x,y\in\mathbb{F}^n}\left[\prod_{i\in S} f(x+iy)\right] \le \|f\|_{U^{s+1}} \le \varepsilon.$$

Thus

$$\begin{aligned}\Pr_{x,y\in\mathbb{F}^n}[x, x+y, \ldots, x+dy \in A] &= \mathbb{E}_{x,y\in\mathbb{F}^n}\left[\prod_{i=0}^d 1_A(L_i(x,y))\right] \\ &= \alpha^{d+1} + \sum_{S\subseteq[d+1], S\neq\emptyset} \alpha^{d+1-|S|}\mathbb{E}_{x,y\in\mathbb{F}^n}\left[\prod_{i\in S} f(x+iy)\right]\end{aligned}$$

Thus

$$\left|\Pr_{x,y\in\mathbb{F}^n}[x, x+y, \ldots, x+dy \in A] - \alpha^{d+1}\right| \le \varepsilon \sum_{S\subseteq[d+1], S\neq\emptyset} \alpha^{d+1-|S|} \le \varepsilon(1+\alpha)^{d+1}.$$

□

11.2 The true complexity

The Cauchy-Schwarz complexity of $\mathcal{L}$ gives an upper bound on s, such that if $\|f\|_{U^{s+1}}$ is small enough for some function $f : \mathbb{F}^n \to \mathbb{D}$, then f is pseudorandom with regards to $\mathcal{L}$. Gowers and Wolf [37] defined the *true complexity* of a system of linear forms as the minimal s such that the above condition holds for all $f : \mathbb{F}^n \to \mathbb{D}$.

Definition 11.2.1 (True complexity [37])**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}$ in ℓ variables. The true complexity of $\mathcal{L}$ is the smallest $d \in \mathbb{N}$ with the following property. For every $\varepsilon > 0$ and $\alpha \geq 0$, there exists $\delta > 0$ such that if $f : \mathbb{F}^n \to \mathbb{D}$ is any function with $\|f - \alpha\|_{U^{d+1}} \leq \delta$, then*

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^\ell} \left[\prod_{i=1}^{m} f(L_i(X)) \right] - \alpha^m \right| \leq \varepsilon.$$

An obvious bound on the true complexity is the Cauchy-Schwarz complexity of the system. However, there are cases where this is not tight. Gowers and Wolf conjectured that the true complexity of a system of linear forms can be characterized by a simple linear algebraic condition: the smallest $d \geq 1$ such that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent, where the d-th tensor power of a linear form $L = (\lambda_1, \ldots, \lambda_\ell)$ is defined as

$$L^d = \left(\prod_{j=1}^{d} \lambda_{i_j} : i_1, \ldots, i_d \in [\ell] \right) \in \mathbb{F}^{\ell^d}.$$

Conjecture 11.2.2 (True complexity characterization [37]. Resolved below in Theorem 11.2.4)**.** *The true complexity of a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ is the smallest d such that $\{L_1^{d+1}, \ldots, L_m^{d+1}\}$ are linearly independent.*

Example 11.2.3. *Consider the collection of linear forms $L_1 = (1, 0, 0), L_2 = (1, 1, 0), L_3 = (1, 0, 1), L_4 = (1, 1, 1), L_5 = (1, 1, -1), L_6 = (1, -1, 1)$. It is easy to check that here the Cauchy-Schwarz complexity is 2, while as observed by Gowers and Wolf, the true complexity of this system of linear forms is 1.*

Gowers and Wolf later in [38, Theorem 6.1] verified their conjecture in the case where $|\mathbb{F}|$ is sufficiently large; more precisely when $|\mathbb{F}|$ is at least the Cauchy-Schwarz complexity of the system of linear form. Theorem 11.2.2 in its full generality was settled by Hatami, Hatami and Lovett in [45] through the following theorem.

Theorem 11.2.4 ([45])**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}$ in ℓ variables. Assume that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any collection of functions $f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{D}$ with $\min_{i \in [m]} \|f_i\|_{U^{d+1}} \leq \delta$, we have*

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^\ell} \left[\prod_{i=1}^{m} f_i(L_i(X)) \right] \right| \leq \varepsilon.$$

We defer the proof of Theorem 11.2.4 to Section 12.2.

Remark 11.2.5. *Note that if $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent, then for every $e \geq d$ the vectors $L_1^{e+1}, \ldots, L_m^{e+1}$ are also linearly independent. Thus, for a system of linear forms $L_1, \ldots, L_m$ it is natural to consider the smallest d for which $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent. By Theorem 11.2.4, this equals the true complexity of the system.*

To see the usefulness of Theorem 11.2.4, let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of true complexity d. Assume that we are given bounded functions $f_1, \ldots, f_m$. A version of Theorem 9.1.1 for multiple functions (a simpler version of Theorem 9.3.3), allows us to find a high rank polynomial factor $\mathcal{B}$ of degree d, such that we can decompose $f_i = g_i + h_i$ with $g_i = \mathbb{E}[f|\mathcal{B}]$ and $\|h_i\|_{U^{d+1}} \leq \delta$. The following corollary to Theorem 11.2.4 shows that we may simply replace f_i with g_i when counting linear patterns. In particular, choosing $f_1 = \cdots = f_m = f$ and $g_1 = \cdots = g_m = \alpha$ proves the Gowers-Wolf conjecture (Theorem 11.2.2) in its full generality.

Corollary 11.2.6. *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}$ in ℓ variables. Assume that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any functions $f_1, \ldots, f_m, g_1, \ldots, g_m : \mathbb{F}^n \to \mathbb{D}$ with $\|f_i - g_i\|_{U^{d+1}} \leq \delta$, we have*

$$\left| \mathbb{E}_X \left[\prod_{i=1}^m f_i(L_i(X)) \right] - \mathbb{E}_X \left[\prod_{i=1}^m g_i(L_i(X)) \right] \right| \leq \varepsilon.$$

Proof. Choosing $\delta = \delta(\epsilon')$ as in Theorem 11.2.4 for $\epsilon' := \epsilon/m$, we have

$$\begin{aligned}
\left| \mathbb{E}_X \left[\prod_{i=1}^m f_i(L_i(X)) \right] - \mathbb{E}_X \left[\prod_{i=1}^m g_i(L_i(X)) \right] \right| &= \left| \sum_{i=1}^m \mathbb{E}_X \left[(f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right] \right| \\
&\leq \sum_{i=1}^m \left| \mathbb{E}_X \left[(f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right] \right| \\
&\leq m \cdot \varepsilon' \leq \varepsilon,
\end{aligned}$$

where the second inequality follows from Theorem 11.2.4 since $\|f_i - g_i\|_{U^{d+1}} \leq \delta$. □

Chapter 12

Deferred Technical Proofs

We give in this chapter the proofs of Theorem 7.6.1 and Theorem 11.2.4. The proofs are somewhat technical, but they demonstrate the power of the techniques we have developed so far.

12.1 Near-orthogonality: Proof of Theorem 7.6.1

Theorem 7.6.1 (restated). *Let $\mathbb{F}$ be a prime field, $d \geq 1$, $\varepsilon > 0$. Let $\{L_1, \ldots, L_m\}$ be a system of linear forms. Let $\mathcal{B} = \{P_1, \ldots, P_C\}$ be a polynomial factor of degree at most d and $\mathrm{rank}(\mathcal{B}) > r_{7.6.1}(\mathbb{F}, d, \varepsilon)$. Assume furthermore that each P_i is a nonclassical homogeneous polynomial. For every set of coefficients $\Lambda = \{\lambda_{i,j} \in \mathbb{Z} : i \in [C], j \in [m]\}$, let*

$$P_\Lambda(x) := \sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(x)).$$

Then one of the following two cases holds:

- *$P_\Lambda \equiv 0$. In this case, for every $i \in [C]$, it holds that $\sum_{j \in [m]} \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ for any nonclassical homogeneous polynomials $Q_i : \mathbb{F}^n \to \mathbb{T}$ with $\deg(Q_i) = \deg(P_i)$ and $depth(Q_i) \leq depth(P_i)$.*
- *$P_\Lambda \not\equiv 0$. In this case, $|\mathbb{E}\left[e(P_\Lambda)\right]| < \varepsilon$.*

We present the proof from [45]. By a linear form we will always mean hear $L \in \mathbb{F}^\ell$, i.e. a linear form over a prime field $\mathbb{F} = \mathbb{F}_p$ in ℓ variables. For a linear form $L = (\lambda_1, \ldots, \lambda_\ell)$ define $|L| = \sum_{i=1}^{\ell} |\lambda_i|$.

Claim 12.1.1. *Let $d > 0$ be an integer, and $L = (\lambda_1, \ldots, \lambda_\ell) \in \mathbb{F}^\ell$ be a linear form. Then there exist linear forms $L_i = (\lambda_{i,1}, \ldots, \lambda_{i,\ell}) \in \mathbb{F}^\ell$ for $i = 1, \ldots, m$, and integer coefficients $a_1, \ldots, a_m \in \mathbb{Z}$ with $m \leq |\mathbb{F}|^\ell$ such that*

- *$P(L(X)) = \sum_{i=1}^{m} a_i P(L_i(X))$ for every degree-d polynomial $P : \mathbb{F}^n \to \mathbb{T}$;*
- *$|L_i| \leq d$ for every $i \in [m]$;*
- *$|\lambda_{i,j}| \leq |\lambda_j|$ for every $i \in [m]$ and $j \in [\ell]$.*

Proof. The proof proceeds by simplifying $P(L(X))$ using identities that are valid for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d.

We prove the statement by induction on $|L|$. For the base case $|L| \leq d$ there is nothing to prove. Consider $|L| > d$. Since P is of degree d then deriving it $d+1$ times yields the zero polynomial. As $|L| > d$, the same holds if we derive it $|L|$ times. So, for every choice of $y_1, \ldots, y_{|L|} \in \mathbb{F}^n$, we have

$$\sum_{S \subseteq [|L|]} (-1)^{|L|-|S|} P\left(\sum_{i \in S} y_i\right) = (D_{y_1} \ldots D_{y_{|L|}} P)(0) \equiv 0. \tag{12.1}$$

Let $X = (x_1, \ldots, x_\ell) \in (\mathbb{F}^n)^\ell$. Setting $|\lambda_i|$ of the vectors $y_1, \ldots, y_{|L|}$ to be equal to x_i for every $i \in [\ell]$, Equation (12.1) implies

$$P(L(X)) = \sum_i \alpha_i P(M_i(X)), \tag{12.2}$$

where $M_i = (\tau_{i,1}, \ldots, \tau_{i,\ell})$, $|M_i| \leq |L| - 1$ and for every $j \in [\ell]$, $|\tau_{i,j}| \leq |\lambda_j|$. Since $|M_i| \leq |L| - 1$, we can apply the induction hypothesis to obtain the identities

$$P(M_i(X)) = \sum_{j=1}^{m_j} a_{ij} P(M_{ij}(X))$$

where the linear forms $M_{ij} = (\rho_{ij1}, \ldots, \rho_{ij\ell})$ satisfy $|M_{ij}| \leq d$, and $|\rho_{ijk}| \leq |\tau_{ij}| \leq |\lambda_j|$ for every i, j, k. Substituting these identities in Equation (12.2) yields the desired result. □

The next claim shows that we can further simplify the expression given in Theorem 12.1.1. Let $\mathcal{L}_d \subseteq \mathbb{F}^\ell$ denote the set of nonzero linear forms L with $|L| \leq d$ and with the first (left-most) nonzero coefficient equal to 1, e.g. $(0, 1, 0, 2) \in \mathcal{L}_3$ but $(2, 1, 0, 0) \notin \mathcal{L}_3$.

Claim 12.1.2. *For any linear form $L \in \mathbb{F}^\ell$ and integer $d > 0$, there exists a collection of integer coefficients $\{a_{M,c} \in \mathbb{Z}\}_{M \in \mathcal{L}_d, c \in \mathbb{F}^*}$ such that for every degree-d polynomial $P : \mathbb{F}^n \to \mathbb{T}$,*

$$P(L(X)) = \sum_{M \in \mathcal{L}_d, c \in \mathbb{F}^*} a_{M,c} P(cM(X)). \tag{12.3}$$

Proof. Similar to the proof of Theorem 12.1.1 we simplify $P(L(X))$ using identities that are valid for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d.

We use induction on the number of nonzero entries of L. The case when L has only one nonzero entry is trivial. For the induction step, choose $c \in \mathbb{F}$ so that the leading nonzero coefficient of $L' = c \cdot L$ is equal to 1. Assume that $L' = (\lambda_1, \ldots, \lambda_\ell)$. If $|L'| \leq d$ we are done. Assume otherwise that $|L'| > d$. Applying Theorem 12.1.1 for the degree-d polynomial $R(x) := P(c^{-1}x)$ and the linear form L' we can write

$$P(L(X)) = P(c^{-1}L'(X)) = R(L'(X)) = \sum_i \beta_i R(M_i(X)) = \sum_i \beta_i P(c^{-1}M_i(X)), \tag{12.4}$$

where for every i, $\beta_i \in \mathbb{Z}$ and $M_i = (\lambda_{i,1}, \ldots, \lambda_{i,\ell})$ satisfies $|M_i| \leq d$; and $|\lambda_{i,j}| \leq |\lambda_j|$ for every $j \in [\ell]$. Let $\mathcal{I} = \{i : M_i \in \mathcal{L}_d\}$. Then

$$P(L(x)) = \sum_{i \in \mathcal{I}} \alpha_i P(c^{-1}M_i(X)) + \sum_{i \notin \mathcal{I}} \alpha_i P(c^{-1}M_i(X)). \tag{12.5}$$

In order to conclude the proof, we need to handle the sum over $i \notin \mathcal{I}$. Observe that if $i \notin \mathcal{I}$, then as the leading coefficient of L' is 1, then Theorem 12.1.1 implies in particular that M_i has smaller support than L. Thus, we may apply our induction hypothesis for $c^{-1}M_i$ for all $i \notin \mathcal{I}$, which concludes the proof. □

Theorem 12.1.2 applies to all polynomials of degree d. If we also specify the depth, then we can obtain a stronger statement. However, in this statement we shall need to assume that the polynomials are homogeneous.

Claim 12.1.3. *Let $d \geq 1, k \geq 0$, $L_1, \ldots, L_m \in \mathbb{F}^\ell$ be linear forms, and let $\{\lambda_i \in \mathbb{Z}\}_{i \in [m]}$ be integer coefficients. Then there exist integer coefficients $\{a_M \in \mathbb{Z}\}_{M \in \mathcal{L}_d}$ such that the following is true. For every homogeneous polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d and depth $\leq k$,*

- $\sum_{i=1}^m \lambda_i P(L_i(X)) \equiv \sum_{M \in \mathcal{L}_d} a_M P(M(X))$;
- *For every M with $a_M \neq 0$, we have $|M| \leq \deg(a_M P)$.*

Proof. The proof is similar to that of Theorem 12.1.2, except that now we repeatedly apply Theorem 12.1.2 to every term of the form $\lambda P(L(X))$ to express it as an integer linear combination of $P(cM(X))$ for $M \in \mathcal{L}_{\deg(\lambda P)}$ and $c \in \mathbb{F}^*$. We use the assumption that P is homogeneous of degree d to replace $P(cM(X))$ with $\sigma_c P(M(X))$, where if $c = \zeta^i$ for the fixed generator $\zeta \in \mathbb{F}^*$ then $\sigma_c = \sigma(d,k)^i$. Observe that the condition $a_M P \neq 0$ depends only on the depth of P (that is, if it is zero for some polynomial P of depth k, then it is also zero for all polynomials of depth $\leq k$). Thus, the above process depends only on the assumption that P is homogeneous, its degree d, and a bound on its depth k. By repeating this procedure we arrive at the desired expansion. □

We are now ready for the proof of Theorem 7.6.1. For a linear form $L = (\lambda_1, \ldots, \lambda_\ell)$, let $\mathrm{lc}(L)$ denote the index of its first nonzero entry, namely $\mathrm{lc}(L) := \min_{i:\lambda_i \neq 0} i$.

Proof of Theorem 7.6.1: Let d be the degree of the factor. For every $i \in [C]$, by Theorem 12.1.3 we have

$$\sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(X)) = \sum_{M \in \mathcal{L}_d} \lambda'_{i,M} P_i(M(X)) \tag{12.6}$$

for some integers $\lambda'_{i,M}$ such that if $\lambda'_{i,M} \neq 0$ then $|M| \leq \deg(\lambda'_{i,M} P_i)$. The simplifications of Theorem 12.1.3 depend only on the degrees and depths of the polynomials. Hence, if $\lambda'_{i,M} P_i \equiv 0$ for all $M \in \mathcal{L}_d$, then $(\lambda_{i,1}, \ldots, \lambda_{i,m})$ is in the (d_i, k'_i)-dependency set for all $k'_i \leq k_i$ (see Theorem 7.7.1 for the definition). This implies that $\sum_{j=1}^{m} \lambda_{i,j} Q_i(L_j(X))$ for all homogeneous polynomials Q_i of degree d_i and depth $\leq k_i$.

So to prove the theorem, it suffices to show that P_Λ has small bias if $\lambda'_{i,M} P_i \not\equiv 0$ for some $i \in [C]$ and $M \in \mathcal{L}_d$. Suppose this is true, and thus there exists a nonempty set $\mathcal{M} \subseteq \mathcal{L}_d$ such that

$$P_\Lambda(X) = \sum_{i \in [C], M \in \mathcal{M}} \lambda'_{i,M} P_i(M(X)),$$

and for every $M \in \mathcal{M}$, there is at least one index $i \in [C]$ for which $\lambda'_{i,M} P_i \not\equiv 0$. Choose $i^* \in [C]$ and $M^* \in \mathcal{M}$ in the following manner.

- First, let $M^* \in \mathcal{M}$ be such that $\mathrm{lc}(M^*) = \min_{M \in \mathcal{M}} \mathrm{lc}(M)$, and among these, $|M^*|$ is maximal.
- Then, let $i^* \in [C]$ be such that $\deg(\lambda'_{i^*,M^*} P_{i^*})$ is maximized.

Without loss of generality assume that $i^* = 1$, $\mathrm{lc}(M^*) = 1$, and let $d^* := \deg(\lambda'_{1,M^*} P_1) \leq d$. We claim that if $\sum_{j \in [m]} \lambda_{1,j} P_1(L_j(X))$ is not the zero polynomial, then $\deg(P_\Lambda) \geq d^*$, and moreover, P_Λ has small bias. We prove this by deriving P_Λ in specific directions in a manner that all the terms but $\lambda'_{1,M^*} P_1(M^*(X))$ vanish.

The following definition will be useful to that affect.

Definition 12.1.4 (Derivative according to pair)**.** *Let $P : (\mathbb{F}^n)^\ell \to \mathbb{T}$. For a vector $\alpha \in \mathbb{F}^\ell$ and an element $y \in \mathbb{F}^n$, the derivative of P according to the pair (α, y) is defined as*

$$D_{(\alpha,y)} P(x_1, \ldots, x_\ell) := P(x_1 + \alpha_1 y, \cdots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell). \tag{12.7}$$

For $P_i : \mathbb{F}^n \to \mathbb{T}$ and a linear form $M \in \mathbb{F}^\ell$, define $P_i \circ M : (\mathbb{F}^n)^\ell \to \mathbb{T}$ as

$$(P_i \circ M)(x_1, \ldots, x_\ell) := P_i(M(x_1, \ldots, x_\ell)).$$

Note that for every $M \in \mathcal{M}$,

$$\begin{aligned} D_{(\alpha,y)}(P_i \circ M)(x_1, \cdots, x_\ell) &= P_i(M(x_1, \ldots, x_\ell) + M(\alpha_1, \ldots, \alpha_\ell) y) - P_i(M(x_1, \ldots, x_\ell)) \\ &= (D_{\langle M, \alpha \rangle y} P_i)(M(x_1, \ldots, x_\ell)). \end{aligned}$$

Thus if α is chosen such that $\langle M, \alpha \rangle = 0$ then $D_{(\alpha,y)}(P_i \circ M) \equiv 0$. Thus, our goal is to carefully choose a set of directions that will annihilate all linear forms but M^*, which will greatly simplify the analysis.

Assume that $M^* = (w_1^*, \ldots, w_\ell^*)$, where $w_1^* = 1$. Let $t := |M^*|$. By Theorem 12.1.3 we have $t \leq d^*$, since if $\lambda'_{i,M^*} \neq 0$ then necessarily $|M^*| \leq \deg(\lambda'_{i,M^*} P_i) = d^*$. As a first step, take $\alpha_1 := e_1 = (1, 0, 0, \ldots, 0) \in \mathbb{F}^\ell$. Choose additional $t-1$ vectors as follows: for each $j = 2, \ldots, \ell$, and each $w = 1, \ldots, w_j^* - 1$, pick $\alpha_{j,w} := -we_1 + e_j = (-w, 0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{F}^\ell$, where the 1 is in the j-th coordinate. Observe that there are are indeed $|M^*| - 1 = t - 1$ such vectors, and for convenience number them as $\alpha_2, \ldots, \alpha_t \in \mathbb{F}^\ell$.

The following claim shows that deriving P_Λ iteratively according to the pairs $(\alpha_1, y_1), \ldots, (\alpha_t, y_t)$ annihilates all linear forms except M^*.

Claim 12.1.5.

$$D_{(\alpha_1,y_1)} \cdots D_{(\alpha_t,y_t)} P_\Lambda(X) = \left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_t\rangle y_t} \left(\sum_{i=1}^{C} \lambda'_{i,M^*} P_i \right) \right) (M^*(X)). \tag{12.8}$$

Proof. We need to show that for every $M = \mathcal{M} \setminus \{M^*\}$, there exists $i \in [t]$ such that $\langle M, \alpha_i \rangle = 0$. Assume that $M = (w_1, \ldots, w_\ell)$. If $w_1 = 0$ then $\langle M, \alpha_1 \rangle = 0$. Otherwise by our assumption $w_1 = 1$. As $|M^*|$ was chosen to be maximal among all M with $\mathrm{lc}(M) = \mathrm{lc}(M^*)$, there must exist some $j = 2, \ldots, \ell$ for which $w_j < w_j^*$. But then the appropriate $\alpha_i = \alpha_{j,w_j}$ satisfies $\langle M, \alpha_i \rangle = 0$. □

Next, we choose additional $d^* - t$ vectors $\alpha_{t+1}, \ldots, \alpha_{d^*} \in \mathbb{F}^\ell$ to only keep polynomials P_i in the sum for which $\lambda'_{i,M^*} P_i$ has maximal degree, namely d^*. To do so, simply choose $\alpha_{t+1} = \cdots = \alpha_{d^*} := e_1 \in \mathbb{F}^\ell$. The next claim shows that deriving P_Λ according to the pairs $(\alpha_1, y_1), \ldots, (\alpha_{d^*}, y_{d^*})$ annihilates all linear forms but M^*, and furthermore keeps only polynomials P_i where $\lambda'_{i,M^*} P_i$ has degree d.

Claim 12.1.6.

$$D_{(\alpha_1,y_1)} \cdots D_{(\alpha_{d^*},y_{d^*})} P_\Lambda(X) = \left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_{d^*}\rangle y_{d^*}} \sum_{\substack{i \in [C]: \\ \deg(\lambda'_{i,M^*} P_i) = d^*}} \lambda'_{i,M^*} P_i \right) (M^*(X)). \tag{12.9}$$

Proof. We already know by Theorem 12.1.5 that

$$D_{(\alpha_1,y_1)} \cdots D_{(\alpha_t,y_t)} P_\Lambda(X) = \left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_t\rangle y_t} \left(\sum_{i=1}^{C} \lambda'_{i,M^*} P_i \right) \right) (M^*(X)).$$

Thus, we just need to additionally derive this expression according to the pairs $(\alpha_{t+1}, y_{t+1}), \ldots, (\alpha_{d^*}, y_{d^*})$. We have

$$D_{(\alpha_1,y_1)} \cdots D_{(\alpha_{d^*},y_{d^*})} P_\Lambda(X) = \left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_{d^*}\rangle y_t} \left(\sum_{i=1}^{C} \lambda'_{i,M^*} P_i \right) \right) (M^*(X)).$$

The claim follows as by the choice of d^* we have $\deg(\lambda'_{i,M^*} P_i) \leq d^*$ for all i. If $\deg(\lambda'_{i,M^*} P_i) < d^*$ then the d^* derivatives $D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_{d^*}\rangle y_t}$ annihilate $\lambda'_{i,M^*} P_i$. Thus, we only retain the polynomials P_i for which $\lambda'_{i,M^*} P_i$ has maximal degree, namely d^*. □

Let $Q(x) := \sum_{i \in [C]: \deg(\lambda'_{i,M^*} P_i) = d^*} \lambda'_{i,M^*} P_i(x)$. Theorem 12.1.6 implies that

$$\begin{aligned} &\mathbb{E}_{y_1,\ldots,y_{d^*} \in \mathbb{F}^n, X \in (\mathbb{F}^n)^\ell} \left[e\left(D_{(\alpha_1,y_1)} \cdots D_{(\alpha_{d^*},y_{d^*})} P_\Lambda)(X) \right) \right] \\ &= \mathbb{E}_{y_1,\ldots,y_{d^*} \in \mathbb{F}^n, X \in (\mathbb{F}^n)^\ell} \left[e\left(\left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_{d^*}\rangle y_{d^*}} Q \right) (M^*(X)) \right) \right]. \end{aligned}$$

Next, define $x = M^*(X) \in \mathbb{F}^n$ and $z_j = \langle M^*, \alpha_j \rangle y_j \in \mathbb{F}^n$. It is simple to verify from our construction that $\langle M^*, \alpha_j \rangle \neq 0$ for all j. Thus, the joint distribution of $x, z_1, \ldots, z_{d^*} \in \mathbb{F}^n$ is uniform and independent, and hence

$$\begin{aligned} &\mathbb{E}_{y_1,\ldots,y_{d^*} \in \mathbb{F}^n, X \in (\mathbb{F}^n)^\ell} \left[e\left(D_{(\alpha_1,y_1)} \cdots D_{(\alpha_{d^*},y_{d^*})} P_\Lambda)(X) \right) \right] \\ &= \mathbb{E}_{x, z_1,\ldots,z_{d^*} \in \mathbb{F}^n} \left[e\left(\left(D_{z_1} \cdots D_{z_{d^*}} Q \right)(x) \right) \right] \\ &= \|Q\|_{U^{d^*}}^{2^{d^*}}. \end{aligned}$$

Recall that we assumed that $\mathcal{B}$ has high rank. Concretely, require $\mathcal{B}$ to have rank at least $r_{7.6.1}(\mathbb{F}, d, \varepsilon) := r_{7.2.6}(\mathbb{F}, d, \varepsilon)$ By definition, any linear combination of the polynomials defining $\mathcal{B}$ has such rank. In particular, this holds for Q, which by Theorem 7.2.6 implies that $\|Q\|_{U^{d^*}} \leq \varepsilon$. Thus

$$\mathbb{E}_{y_1,\ldots,y_{d^*},X}\left[e\left(D_{(\alpha_1,y_1)}\cdots D_{(\alpha_{d^*},y_{d^*})}P_\Lambda)(x_1,\ldots,x_\ell)\right)\right] \leq \epsilon^{2^d},$$

In order to conclude the proof, we need to relate the LHS of the above expression to the bias of P_Λ. This can be achieved via a repeated application of the Cauchy-Schwarz inequality. The following claim appeared first in [13].

Claim 12.1.7 ([13, Claim 3.4])**.** *For any nonzero $\alpha_1,\ldots,\alpha_{d^*} \in \mathbb{F}^\ell$,*

$$\mathbb{E}_{y_1,\ldots,y_{d^*}\in\mathbb{F}^n,X\in(\mathbb{F}^n)^\ell}\left[e\left((D_{(\alpha_1,y_1)}\cdots D_{(\alpha_{d^*},y_{d^*})}P_\Lambda)(X)\right)\right] \geq \left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[e\left(P_\Lambda(X)\right)\right]\right|^{2^d}.$$

Proof. It suffices to show that for any function $P:(\mathbb{F}^n)^\ell \to \mathbb{T}$ and nonzero $\alpha=(a_1,\ldots,a_\ell)\in\mathbb{F}^\ell$,

$$\left|\mathbb{E}_{y\in\mathbb{F}^n,X\in(\mathbb{F}^n)^\ell}\left[e\left((D_{\langle\alpha,y\rangle}P)(X)\right)\right]\right| \geq \left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[e\left(P(X)\right)\right]\right|^2.$$

Recall that $(D_{(\alpha,y)}P)(x_1,\ldots,x_\ell) = P(x_1+a_1y,\ldots,x_\ell+a_\ell y) - P(x_1,\ldots,x_\ell)$. Without loss of generality, suppose $a_1 \neq 0$. We make a change of coordinates so that α can be assumed to be $(1,0,\ldots,0)$. More precisely, define $P':(\mathbb{F}^n)^\ell\to\mathbb{T}$ as

$$P'(x_1,\ldots,x_\ell) := P\left(x_1, \frac{x_2+a_2x_1}{a_1}, \frac{x_3+a_3x_1}{a_1}, \ldots, \frac{x_\ell+a_\ell x_1}{a_1}\right).$$

Observe that with this definition,

$$P(x_1,\ldots,x_\ell) = P'(x_1, a_1x_2-a_2x_1, a_1x_3-a_3x_1,\ldots,a_1x_\ell-a_\ell x_1)$$

and

$$(D_{(\alpha,y)}P)(x_1,\ldots,x_\ell) = P'(x_1+a_1y, a_1x_2-a_2x_1,\ldots,a_1x_\ell-a_\ell x_1) - P'(x_1,a_1x_2-a_2x_1,\ldots,a_1x_\ell-a_\ell x_1).$$

Therefore

$$\begin{aligned}
&\mathbb{E}_{y,x_1,\ldots,x_\ell\in\mathbb{F}}\left[e\left((D_{(\alpha,y)}P)(x_1,\ldots,x_\ell)\right)\right]\\
&= \mathbb{E}_{y,x_1,\ldots,x_\ell\in\mathbb{F}}\left[e\left(P'(x_1+a_1y,a_1x_2-a_2x_1,\ldots,a_1x_\ell-a_\ell x_1) - P'(x_1,a_1x_2-a_2x_1,\ldots,a_1x_\ell-a_\ell x_1)\right)\right]\\
&= \mathbb{E}_{y,x_1,\ldots,x_\ell\in\mathbb{F}}\left[e\left(P'(x_1+a_1y,x_2,\ldots,x_\ell) - P'(x_1,x_2,\ldots,x_\ell)\right)\right]\\
&= \mathbb{E}_{x_2,\ldots,x_\ell\in\mathbb{F}}\left|\mathbb{E}_{x_1\in\mathbb{F}}\left[e\left(P'(x_1,x_2,\ldots,x_\ell)\right)\right]\right|^2
\end{aligned}$$

We can thus conclude the proof as, by the Cauchy-Schwartz inequality, it holds that

$$\begin{aligned}
\left|\mathbb{E}_{y,x_1,\ldots,x_\ell\in\mathbb{F}}\left[e\left((D_{(\alpha,y)}P)(x_1,\ldots,x_\ell)\right)\right]\right| &\leq \left|\mathbb{E}_{x_1,x_2,\ldots,x_\ell\in\mathbb{F}}\left[e\left(P'(x_1,x_2,\ldots,x_\ell)\right)\right]\right|^2\\
&= \left|\mathbb{E}_{x_1,x_2,\ldots,x_\ell\in\mathbb{F}}\left[e\left(P(x_1,x_2,\ldots,x_\ell)\right)\right]\right|^2.
\end{aligned}$$

□

□

Remark 12.1.8. *The proof of Theorem 7.6.1 also shows the following. Suppose, under the same conditions of Theorem 7.6.1, that for every polynomial P_i for $i\in[C]$ and any linear form L_j for $j\in[m]$, one the following two conditions holds:*

- $|L_j| \leq \deg(\lambda_{i,j}P_i)$*; or*
- $\lambda_{i,j}P_i = 0$.

Now, if $\lambda_{i,j}\equiv 0 \mod p^{depth(P_i)+1}$ for all $i\in[C], j\in[m]$, then clearly $\lambda_{i,j}P_i = 0$ for all i,j and hence $P_\Lambda = 0$. In all other cases, we obtain that $P_\lambda \neq 0$ and hence $\mathrm{bias}(P_\Lambda) < \varepsilon$*. The proof is identical to the proof of Theorem 7.6.1, except that there is no need to transform $\lambda_{i,j}$ to $\lambda'_{i,j}$ in the beginning of the proof.*

12.2 Proof of Theoren 11.2.4

Theorem 11.2.4 (restated). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms in ℓ variables, such that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. For every $\varepsilon > 0$, there exists $\delta > 0$ such that for any collection of functions $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$ with $\min_{i\in[m]} \|f_i\|_{U^{d+1}} \le \delta$, we have*

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[\prod_{i=1}^m f_i(L_i(X)) \right] \right| \le \varepsilon. \tag{12.10}$$

We present the proof from [45]. First, note that since $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent, it must be the case that $L_1, \dots, L_m$ are pairwise linearly independent. That is, it is not the case that $L_i = cL_j$ for some distinct $i, j \in [m]$ and $c \in \mathbb{F}$. Consequently, $\mathcal{L} = \{L_1, \dots, L_m\}$ is of finite Cauchy-Schwarz complexity s for some $s \le m - 2 < \infty$. Indeed, for every $i \in [m]$, the partition of $\{L_j\}_{j\in[m]\setminus\{i\}}$ into $m-1$ singletons satisfies the requirement of Theorem 11.1.1. The case when $s \le d$ follows from Theorem 11.1.2, thus we are left with the case when $s > d$. We will prove that $\|f_1\|_{U^{d+1}} \le \delta$ where δ is sufficiently small, implies Equation (12.10). The theorem then follows by symmetry in the choice of f_1.

As a first step, we decompose the functions as $f_i = g_i + h_i$, where $g_i = \mathbb{E}[f|\mathcal{B}]$ for a suitably chosen polynomial factor $\mathcal{B}$ of sufficiently high rank, and where h_i has negligible $(s+1)$-th Gowers norm. In order to do so, let $r : \mathbb{N} \to \mathbb{N}$ be a large enough rank bound to be chosen later. Apply a multifunction version of Theorem 9.1.1 along with Theorem 10.1.1 and Theorem 7.5.5, to obtain a simultaneous decomposition $f_i = g_i + h_i$ for all $i \in [m]$ where

1. $g_i = \mathbb{E}[f_i|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor, defined by homogeneous polynomials, of degree at most s and complexity $C \le C_{\max}(|\mathbb{F}|, s, m, \varepsilon, r(\cdot))$.
2. $\|h_i\|_{U^{s+1}} \le \frac{\varepsilon}{2m}$.

We will assume that the rank $r(C)$ is chosen large enough, so that $\mathcal{B}$ is $\nu(C)$-uniform for $\nu(C) > 0$ to be chosen later (See Theorem 7.2.10). We first show that we may replace f_i's in Equation (12.10) with g_i's.

Claim 12.2.1. $\left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[\prod_{i=1}^m f_i(L_i(X))\right] - \mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[\prod_{i=1}^m g_i(L_i(X))\right]\right| \le \frac{\varepsilon}{2}$.

Proof. We have

$$\begin{aligned}
\left| \mathbb{E}_X \left[\prod_{i=1}^m f_i(L_i(X)) \right] - \mathbb{E}_X \left[\prod_{i=1}^m g_i(L_i(X)) \right] \right| &= \left| \sum_{i=1}^m \mathbb{E}_X \left[h_i(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right] \right| \\
&\le \sum_{i=1}^m \left| \mathbb{E}_X \left[h_i(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right] \right| \\
&\le \sum_{i=1}^m \|h_i\|_{U^{s+1}} \le \frac{\varepsilon}{2},
\end{aligned}$$

where the second inequality follows from Theorem 11.1.2 as the Cauchy-Schwarz complexity of $\mathcal{L}$ is s. □

Thus it is sufficient to bound $\left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[\prod_{i=1}^m g_i(L_i(X))\right]\right| \le \varepsilon/2$. For each i, $g_i = \mathbb{E}[f_i|\mathcal{B}]$ and thus

$$g_i(x) = \Gamma_i(P_1(x), \dots, P_C(x)),$$

where $P_1, \dots, P_C$ are the nonclassical homogeneous polynomials of degree $\le s$ defining $\mathcal{B}$ and $\Gamma_i : \mathbb{T}^C \to \mathbb{D}$ is a function. Let $k_i = \operatorname{depth}(P_i)$ so that by Theorem 6.1.2, each P_i takes values in $\mathbb{U}_{k_i+1} = \frac{1}{p^{k_i+1}}\mathbb{Z}/\mathbb{Z}$. Let $\Sigma := \mathbb{Z}_{p^{k_1+1}} \times \cdots \times \mathbb{Z}_{p^{k_C+1}}$.

Using the Fourier transform on $\mathbb{U}_{k_1+1} \times \ldots \times \mathbb{U}_{k_C+1} \cong \Sigma$, for every $\tau = (\tau_1, \ldots, \tau_C) \in \Sigma$ we have

$$\Gamma_i(\tau) = \sum_{\Lambda=(\lambda_1,\ldots,\lambda_C)\in\Sigma} \widehat{\Gamma}_i(\Lambda) \cdot e\left(\sum_{j=1}^{C} \lambda_j \tau_j\right), \tag{12.11}$$

where

$$\widehat{\Gamma}_i(\Lambda) := \mathbb{E}_\tau\left[\Gamma_i(\tau) e\left(\sum_{j=1}^{C} \lambda_j \tau_j\right)\right]$$

is the Fourier coefficient of Γ_i corresponding to Λ. Observe that $|\widehat{\Gamma}_i(\Lambda)| \leq 1$. Consequently,

$$g_i(x) = \sum_{\Lambda=(\lambda_1,\ldots,\lambda_C)\in\Sigma} \widehat{\Gamma}_i(\Lambda) \cdot e\left(\sum_{j=1}^{C} \lambda_j P_j(x)\right). \tag{12.12}$$

Let $P_\Lambda := \sum_{j=1}^{C} \lambda_j P_j(x)$ for the sake of brevity so that we may write

$$\mathbb{E}_X\left[\prod_{i=1}^{m} g_i(L_i(X))\right] = \sum_{\Lambda_1,\ldots,\Lambda_m\in\Sigma} \left(\prod_{i=1}^{m} \widehat{\Gamma}_i(\Lambda_i)\right) \cdot \mathbb{E}_X\left[e\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right]. \tag{12.13}$$

We will show that each term in Equation (12.13) can be bounded by $\frac{\varepsilon}{2|\Sigma|^m}$, thus concluding the proof by the triangle inequality.

We first show that the terms for which $\deg(P_{\Lambda_1}) \leq d$ are small. Recall that by our assumption $\|f_1\|_{U^{d+1}} \leq \delta$, where δ is small enough to be determined later.

Claim 12.2.2. *Let $\Lambda_1 \in \Sigma$ be such that $\deg(P_{\Lambda_1}) \leq d$. Then*

$$\left|\widehat{\Gamma}_1(\Lambda_1)\right| \leq \delta + |\Sigma|\nu(C).$$

Proof. It follows from Equation (12.12) that

$$\widehat{\Gamma}_1(\Lambda_1) = \mathbb{E}_{x\in\mathbb{F}^n}\left[g_1(x) e\left(-P_{\Lambda_1}(x)\right)\right] - \sum_{\Lambda\neq\Lambda_1} \widehat{\Gamma}_1(\Lambda) \cdot \mathbb{E}_{x\in\mathbb{F}^n}\left[e\left(P_\Lambda(x) - P_{\Lambda_1}(x)\right)\right].$$

We first bound the first term.

$$|\mathbb{E}_x\left[g_1(x) e\left(-P_{\Lambda_1}(x)\right)\right]| = |\mathbb{E}_x\left[f_1(x) e\left(-P_{\Lambda_1}(x)\right)\right]| \leq \|f_1 e\left(-P_{\Lambda_1}\right)\|_{U^{d+1}} = \|f_1\|_{U^{d+1}} \leq \delta.$$

The first equality follows as $g_1 = \mathbb{E}[f_1|\mathcal{B}]$ and P_{Λ_1} is $\mathcal{B}$-measurable.

Next, we bound each summand in the second term. Fix $\Lambda \neq \Lambda_1$. Recall that the polynomial factor $\mathcal{B}$ is $\nu(C)$-uniform by our construction. In particular the bias of $P_\Lambda - P_{\Lambda_1} = P_{\Lambda-\Lambda_1}$ is at most $\nu(C)$. That is, $|\mathbb{E}_{x\in\mathbb{F}^n}\left[e\left(P_\Lambda(x) - P_{\Lambda_1}(x)\right)\right]| \leq \nu(C)$. As also $|\widehat{\Gamma}_1(\Lambda)| \leq 1$, each summand in the sum in the second term is bounded by $\nu(C)$. □

Theorem 12.2.2 allows us to bound the contribution of the terms in Equation (12.13) corresponding to tuples $(\Lambda_1, \ldots, \Lambda_m) \in \Sigma^m$ with $\deg(P_{\Lambda_1}) \leq d$.

$$\sum_{\substack{\Lambda_1,\ldots,\Lambda_m\in\Sigma:\\ \deg(P_{\Lambda_1})\leq d}} \left(\prod_{i=1}^{m} \widehat{\Gamma}_i(\Lambda_i)\right) \cdot \mathbb{E}_X\left[e\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right] \leq |\Sigma|^m\left(\delta + |\Sigma|\nu(C)\right). \tag{12.14}$$

Next, we bound the terms for which $\deg(P_{\Lambda_1}) > d$. We will need the following lemma.

Lemma 12.2.3. *Assume that $L_1^{d+1},\ldots,L_m^{d+1}$ are linearly independent, and let $(\Lambda_1,\ldots,\Lambda_m)\in\Sigma^m$ be such that $\deg(P_{\Lambda_1})\geq d+1$. Then*

$$\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) \not\equiv 0.$$

Before proving Theorem 12.2.3 let us first describe why it suffices to complete the proof of Theorem 11.2.4. Theorem 7.6.1 gives that, if we that the rank $r(\cdot)$ is chosen so that $r(C)\geq r_{7.6.1}(\mathbb{F},s,\delta)$, then under the conclusion of Theorem 12.2.3, we in fact have that

$$\left|\mathbb{E}_{X\in(\mathbb{F}^n)^\ell}\left[e\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right]\right| \leq \delta.$$

We may thus conclude that

$$\sum_{\substack{\Lambda_1,\ldots,\Lambda_m\in\Sigma:\\ \deg(P_{\Lambda_1})\geq d+1}} \left(\prod_{i=1}^{m}\widehat{\Gamma}_i(\Lambda_i)\right)\cdot\mathbb{E}_X\left[e\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right] \leq \delta|\Sigma|^m. \tag{12.15}$$

Combining (12.14) and (12.15) allows us to conclude that

$$\sum_{\Lambda_1,\ldots,\Lambda_m\in\Sigma} \left(\prod_{i=1}^{m}\widehat{\Gamma}_i(\Lambda_i)\right)\cdot\mathbb{E}_X\left[e\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right] \leq (\delta+|\Sigma|\nu(C))|\Sigma|^m.$$

To conclude we set parameters. Observe that $|\Sigma| = |\mathbb{F}|^{\sum_{i\in[C]}\mathrm{depth}(P_i)+1} \leq |\mathbb{F}|^{\sum_{i\in[C]}\deg(P_i)} \leq |\mathbb{F}|^{Cs}$. Thus, we may choose $r(\cdot)$ so that $\nu(C)|\Sigma|^{m+1}\leq\varepsilon/4$. To conclude, after choosing $r(\cdot)$ we have an upper bound on the complexity of $\mathcal{B}$, namely $C\leq C_{\max}=C_{\max}(|\mathbb{F}|,s,m,\varepsilon,r(\cdot))$. We set $\delta:=(\varepsilon/4)|\mathbb{F}|^{C_{\max}sm}$ and conclude the proof.

Thus, we are left with proving Theorem 12.2.3. This is the only place in the proof where we actually use the assumption that $L_1^{d+1},\ldots,L_m^{d+1}$ are linearly independent.

Proof of Theorem 12.2.3: Assume to the contrary that $\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\equiv 0$. Denoting the coordinates of Λ_i by $(\lambda_{i,1},\ldots,\lambda_{i,C})\in\Sigma^C$ we have

$$\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) = \sum_{i\in[m],j\in[C]} \lambda_{i,j}P_j(L_i(X)) \equiv 0.$$

We apply Theorem 7.6.1. This requires assuming that the polynomial factor $\mathcal{B}$ has high enough rank, which we can achieve by requiring that $r(C)\geq r_{7.6.1}(\mathbb{F},s,1/2)$, say. Thus, we for every $j\in[C]$ we must have

$$\sum_{i=1}^{m}\lambda_{i,j}P_j(L_i(X))\equiv 0. \tag{12.16}$$

In fact, we know more: for any nonclassical homogeneous polynomials Q_j with $\deg(Q_j)=\deg(P_j)$ and $\mathrm{depth}(Q_j)\leq\mathrm{depth}(P_j)$ it holds that

$$\sum_{i=1}^{m}\lambda_{i,j}Q_j(L_i(X))\equiv 0.$$

Next, as we assume that $\deg(P_{\Lambda_1})\geq d+1$, there must exist $j\in[C]$ such that $\deg(\lambda_{1,j}P_j)\geq d+1$. Let $j^*\in[C]$ be such that $d^*:=\deg(\lambda_{1,j^*}P_{j^*})$ is maximized. By our assumptions, $d^*\geq d+1$. Let $t\geq 0$ be the largest integer such that $p^t=|\mathbb{F}|^t$ divides λ_{i,j^*} for all $i\in[m]$. Define $\mu_i:=\lambda_{i,j^*}/p^t\in\mathbb{Z}$ and $P(x)=p^tP_{j^*}(x)$. Then

$$\sum_{i=1}^{m}\mu_i P(L_i(X)) = \sum_{i=1}^{m}\frac{\lambda_{i,j}}{p^t}p^tP_{j^*}(L_i(X))\equiv 0.$$

The polynomial P belongs to $\mathcal{B}$. As such, $\{P\}$ has rank at least $r(C)$ as well. Thus by another application of Theorem 7.6.1, for any homogeneous polynomial Q of degree $\deg(Q) = \deg(P)$ and $\text{depth}(Q) \leq \text{depth}(P)$ it holds that

$$\sum_{i=1}^{m} \mu_i Q(L_i(X)) \equiv 0. \tag{12.17}$$

In particular, this holds for any homogeneous classical polynomial Q of degree $D := \deg(P) \geq \deg(\lambda_{1,j} P_j) = d^* \geq d+1$. That is, if $R : \mathbb{F}^n \to \mathbb{F}$ is a classical homogeneous polynomial of degree D then, by setting $Q(x) = \frac{R(x)}{p}$ we get

$$\sum_{i=1}^{m} \mu_i R(L_i(X)) \equiv 0 \bmod p. \tag{12.18}$$

We apply it to the homogeneous classical monomial $R(z) = z_1 z_2 \dots z_D$. In order to compute $R(L_i(X))$, let $X = (x_1, \dots, x_\ell) \in (\mathbb{F}^n)^\ell$ where $x_i = (x_{i,1}, \dots, x_{i,n}) \in \mathbb{F}^n$. Let $L_i = (\lambda_{i,1}, \dots, \lambda_{i,\ell}) \in \mathbb{F}^\ell$. Then

$$R(L_i(X)) = R\left(\sum \lambda_i x_i\right) = \prod_{a=1}^{D} \left(\sum_{b=1}^{\ell} \lambda_{i,b} x_{b,a}\right).$$

In particular, for $b_1, \dots, b_D \in [\ell]$, the coefficient of the monomial $\prod_{a=1}^{D} x_{b_a,a}$ in $R(L_i(X))$ is $\prod_{a=1}^{D} \lambda_{i,b_a}$, which is the $(b_1, \dots, b_D)$ coefficient of L_i^D. Thus (12.18) implies that

$$\sum_{i=1}^{m} \mu_i L_i^D \equiv 0 \bmod p.$$

However, by our choice of t, there must exist $i^* \in [m]$ for which $p^t \lambda_{i^*,j^*}$ does not divide by p. That is, $\mu_i \neq 0 \bmod p$. But then we get that $L_1^D, \dots, L_m^D$ are linearly dependent. This contradicts the assumption that the assumption that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. □

This chapter is concerned with the algorithmic versions of regularity lemmas for polynomials over finite fields. The regularity lemmas proved by Green and Tao [41] and by Kaufman and Lovett [51] as discussed in Section 7.5 show that one can modify a given collection of polynomials $\mathcal{B} = \{P_1, \ldots, P_c\}$ into a regular collection $\mathcal{B}' = \{P_1, \ldots, P_{c'}\}$ of polynomials of same or lower degree. Here, we mean regularity in the sense of Theorem 7.1.6. These lemmas are central to higher-order Fourier analysis and have various applications. Bhattacharyya et. al. [16] studied algorithmic versions of these theorems and showed that analytic notions of regularity such as the ones defined in Section 7.2 allow for efficient algorithms.

12.3 A lemma of Bogdanov and Viola

A key first step in proofs of inverse theorems for Gowers norms and that biased polynomials have small rank is an elegant argument due to Bogdanov and Viola [20] proving that if a polynomial of degree d is *biased*, then it can be approximated by a bounded set of polynomials of lower degree. In Chapter 8 we saw a more efficient version of this lemma due to Lovett and Bhowmick [18]. It was observed in [16] that the Bogdanov-Viola lemma can be made algorithmic due to its probabilistic proof.

Lemma 12.3.1 (Algorithmic Bogdanov-Viola lemma [16])**.** *Let $d \geq 0$ be an integer, and $\delta, \sigma, \beta \in (0, 1]$ be parameters. There exists a randomized algorithm, that given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d with*

$$\mathrm{bias}(P) \geq \delta,$$

runs in time $\mathrm{O}_{\delta,\beta,\sigma}(n^d)$, and with probability $1 - \beta$ returns functions $\widetilde{P} : \mathbb{F}^n \to \mathbb{F}$ and $\Gamma : \mathbb{F}^C \to \mathbb{F}$, and a set of polynomials $P_1, \cdots, P_C$, where $C \leq \frac{|\mathbb{F}|^5}{\delta^2\sigma\beta}$ and $\deg(P_i) < d$ for all $i \in [C]$, for which

- $\Pr_x(P(x) \neq \widetilde{P}(x)) \leq \sigma$, *and*
- $\widetilde{P}(x) = \Gamma(P_1(x), \cdots, P_C(x))$.

Proof. The proof will be an adaptation of the proof from [41]. Given query access to the polynomial P, we can compute the explicit description of P in $O(n^d)$ queries. For every $a \in \mathbb{F}$ define the measure $\mu_a(t) := \Pr(P(x) = a + t)$. It is easy to see that if $\mathrm{bias}(P(x)) \geq \delta$ then, for every $a \neq b$,

$$\|\mu_a - \mu_b\|_\infty \geq \frac{4\delta}{|\mathbb{F}|}. \tag{12.19}$$

We will try to estimate each of these distributions. Let

$$\widetilde{\mu}_a(t) := \frac{1}{C} \sum_{1 \leq i \leq C} \mathbb{1}_{P(x_i)=a+t},$$

where $C > \frac{|\mathbb{F}|^5}{\delta \cdot \beta_1}$, and $x_1, x_2, \cdots, x_C \in \mathbb{F}^n$ are chosen uniformly at random. Therefore by an application of Chebyshev's inequality

$$\Pr\left(|\widetilde{\mu}_a(t) - \mu_a(t)| > \frac{\delta}{2|\mathbb{F}|^2}\right) < \frac{\beta_1}{|\mathbb{F}|},$$

for all $t \in \mathbb{F}$ and therefore

$$\Pr\left(\|\widetilde{\mu}_a - \mu_a\|_\infty > \frac{\delta}{2|\mathbb{F}|^2}\right) < \beta_1. \tag{12.20}$$

Now we will focus on approximating $P(x)$. Remember that $D_h P(x) = P(x + h) - P(x)$ is the additive derivative of $P(x)$ in direction h. We have

$$\Pr_h(D_h P(x) = r) = \Pr_h(P(x + h) - P(x) = r) = \mu_{P(x)}(r),$$

where $h \in \mathbb{F}^n$ is chosen uniformly at random. Let $\mathbf{h} = (h_1, \cdots, h_C) \in (\mathbb{F}^n)^C$ be chosen uniformly at random, where C is a sufficiently large constant to be chosen later. Define the corresponding "observed" distribution as

$$\mu_{\text{obs}}^{(x)}(t) := \frac{1}{C} \sum_{1 \leq i \leq C} \mathbb{1}_{D_{h_j} P(x) = t},$$

and let

$$\widetilde{P}_h(x) := \arg\min_{r \in \mathbb{F}} \|\widetilde{\mu}_r - \mu_{\text{obs}}^{(x)}\|_\infty.$$

Now choosing $C \geq \frac{|\mathbb{F}|^5}{\delta^2 \sigma \beta_2}$ follows

$$\Pr_h(\widetilde{P}_h(x) \neq P(x)) \leq \Pr_h\left(\|\mu_{\text{obs}}^{(x)} - \mu_{P(x)}\| \geq \frac{\delta}{|\mathbb{F}|}\right) \leq \sigma\beta_2, \tag{12.21}$$

where the first inequality follows from (12.19) and (12.20). Therefore

$$\Pr_{x,h}(\widetilde{P}_h(x) \neq P(x)) = \mathbb{E}_x \mathbb{E}_h \mathbb{1}_{\widetilde{P}_h(x) \neq P(x)} \leq \sigma\beta_2,$$

and thus

$$\Pr_h\left[\Pr_x\left(\widetilde{P}_h(x) \neq P(x)\right) \geq \sigma\right] \leq \beta_2.$$

Let $P_i := D_{h_i} P$, so that P_i is of degree $\leq d$ and $\widetilde{P}_h$ is a function of $P_1, \cdots, P_C$. Now setting $\beta_1 := \frac{\beta}{2|\mathbb{F}|^2}$ and $\beta_2 := \frac{\beta}{2}$ finishes the proof.

□

12.4 Algorithmic Regularity Lemmas

Theorem 12.3.1 implies an algorithmic analogue of the Theorem 7.5.4, with the caveat that the refinement is approximate. The proof is by an induction similar to that of Theorem 7.5.4 with the difference that at each step one has to control the errors that are introduced through the use of Theorem 12.3.1 and the probability of correctness.

Lemma 12.4.1 (Unbiased almost Refinement)**.** *Let $d \geq 1$ be an integer. Suppose $\gamma : \mathbb{N} \to \mathbb{R}^+$ is a decreasing function and $\sigma, \rho \in (0, 1]$. There is a randomized algorithm that given a factor $\mathcal{F}$ of degree d, runs in time $\mathrm{O}_{\gamma,\rho,\sigma,\dim(\mathcal{F})}(n^d)$ and with probability $1 - \rho$ returns a γ-unbiased factor $\mathcal{F}'$ with $\dim(\mathcal{F}') = \mathrm{O}_{\gamma,\rho,\sigma,\dim(\mathcal{F})}(1)$, such that $\mathcal{F}'$ is σ-close to being a refinement of $\mathcal{F}$.*

Proof. The proof idea is similar to that of Theorem 7.5.4 in the sense that we use the same type of induction. The difference is that at each step we will have to control the errors that we introduce and the probability of correctness. At all steps in the proof without loss of generality we will assume that the polynomials in the factor are linearly independent, because otherwise we can always detect such a linear combination in $\mathrm{O}_{\gamma,\rho,\sigma,\dim(\mathcal{F})}(n^d)$ time and remove a polynomial that can be written as a linear combination of the rest of the polynomials in the factor.

The base case for $d = 1$ is simple, a linearly independent set of non-constant linear polynomials is not biased at all, namely it is 0-unbiased. If $\mathcal{F}$ is γ-biased, then there exists a set of coefficients $\{c_{i,j} \in \mathbb{F}\}_{1 \leq i \leq d, 1 \leq j \leq M_i}$ such that

$$\text{bias}(\sum_{i,j} c_{i,j} P_{i,j}) \geq \gamma(\dim(\mathcal{F})).$$

To detect this, we will use the following algorithm:

We will estimate bias of each of the $|\mathbb{F}|^{\dim(\mathcal{F})}$ linear combinations and check whether it is greater than $\frac{3\gamma(\dim(\mathcal{F}))}{4}$. To do so, for each linear combination $\sum_{i,j} c_{i,j} P_{i,j}$ independently select a set of vectors $x_1, \cdots, x_C$

uniformly at random from $\mathbb{F}^n$, and let $\widetilde{\text{bias}}(\sum_{i,j} c_{i,j}P_{i,j}) := \left|\frac{1}{C}\sum_{\ell\in[C]} e_{\mathbb{F}}(y_\ell)\right|$, where $y_\ell = \sum_{i,j} c_{i,j}\cdot P_{i,j}(x_\ell)$. Choosing $C = \mathrm{O}_{\dim(\mathcal{F})}\left(\frac{1}{\gamma(\dim(\mathcal{F}))^2}\log(\frac{1}{\rho})\right)$, we can distinguish bias $\geq \gamma$ from bias $\leq \frac{\gamma}{2}$, with probability $1-\rho'$, where $\rho' := \frac{\rho}{4|\mathbb{F}|^{\dim(\mathcal{F})}}$. Let $\sum_{i,j} c_{i,j}P_{i,j}$ be such that the estimated bias was above $\frac{3\gamma(\dim(\mathcal{F}))}{4}$ and k be its degree. We will stop if there is no such linear combination or if the factor is of degree 1. Since by a union bound with probability at least $1-\frac{\rho}{4}$, $\text{bias}(\sum_{i,j} c_{i,j}P_{i,j}) \geq \frac{\gamma(\dim(\mathcal{F}))}{2}$, by Theorem 12.3.1 we can find, with probability $1-\frac{\rho}{4}$, a set of polynomials $Q_1,\cdots,Q_r$ of degree $k-1$ such that

- $\sum_{i,j} c_{i,j}P_{i,j}$ is $\frac{\sigma}{2}$-close to a function of $Q_1,\cdots,Q_r$,
- $r \leq \frac{16|\mathbb{F}|^5}{\gamma(\dim(\mathcal{F}))^2\cdot\sigma\cdot\rho}$.

We replace one polynomial of highest degree that appears in $\sum_{i,j} c_{i,j}P_{i,j}$ with polynomials $Q_1,\cdots,Q_r$.

We will prove by the induction that our algorithm satisfies the statement of the lemma. For the base case, if $\mathcal{F}$ is of degree 1, our algorithm does not refine $\mathcal{F}$ by design. Notice that since we have removed all linear dependencies, $\mathcal{F}$ is in fact 0-unbiased in this case.

Now given a factor $\mathcal{F}$, if $\mathcal{F}$ is γ-biased, then with probability $1-\rho'$ our algorithm will refine $\mathcal{F}$. With probability $1-\frac{\rho}{4}$ the linear combination used for the refinement is $\frac{\gamma(\dim(\mathcal{F}))}{2}$-biased. Let $\widetilde{\mathcal{F}}$ be the outcome of one step of our algorithm. With probability $1-\frac{\rho}{4}$, $\widetilde{\mathcal{F}}$ is $\frac{\sigma}{2}$-close to being a refinement of $\mathcal{F}$. Using the induction hypothesis with parameters $\gamma, \frac{\sigma}{2}, \frac{\rho}{4}$ we can find, with probability $1-\frac{\rho}{4}$, a γ-unbiased factor $\mathcal{F}'$ which is $\frac{\sigma}{2}$-close to being a refinement of $\widetilde{\mathcal{F}}$ and therefore, with probability at least $1-(\frac{\rho}{4}+\frac{\rho}{4}+\frac{\rho}{4}+\rho') > 1-\rho$, is σ-close to being a refinement of $\mathcal{F}$.

□

When the field order is large, this approximate refinement can be made exact if we work with uniform factors (see Theorem 7.2.9).

Lemma 12.4.2 (Uniform Refinement [16])**.** *Suppose $d < |\mathbb{F}|$ is a positive integer and $\rho \in (0,1]$ is a parameter. There is a randomized algorithm that, takes as input a factor $\mathcal{F}$ of degree d over $\mathbb{F}^n$, and a decreasing function $\gamma : \mathbb{N} \to \mathbb{R}^+$, runs in time $\mathrm{O}_{\rho,\gamma,|\mathcal{F}|}(n^d)$, and with probability $1-\rho$ outputs a γ-uniform factor $\widetilde{\mathcal{F}}$, where $\widetilde{\mathcal{F}}$ is a refinement of $\mathcal{F}$, is of same degree d, and $|\widetilde{\mathcal{F}}| \ll_{\sigma,\gamma,|\mathcal{F}|} 1$.*

We only sketch the proof of Theorem 12.4.2, which is by induction on the dimension vector of $\mathcal{F}$. For the induction step, one checks whether there is a linear combination of polynomials in $\mathcal{F}$ that has large Gowers norm. One then uses this to replace a polynomial P from $\mathcal{F}$ with a set of lower degree polynomials. To do this, we first approximate P with a few lower degree polynomials $Q_1,\cdots,Q_r$, then use the induction hypothesis to refine $\{Q_1,\cdots,Q_r\}$ to a uniform factor $\{\widetilde{Q}_1,\cdots,\widetilde{Q}_{r'}\}$ and use an argument similar to that seen in Section 8.2 that approximation by a sufficiently uniform factor implies exact computation to conclude that P is measurable in $\{\widetilde{Q}_1,\cdots,\widetilde{Q}_{r'}\}$.

A similar argument combined with Theorem 12.3.1 gives the following.

Lemma 12.4.3. *Suppose that an integer d satisfies $0 \leq d < |\mathbb{F}|$. Let $\delta,\sigma,\beta \in (0,1]$. There is a randomized algorithm that given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d such that $\text{bias}(P) \geq \delta > 0$, runs in $O_{\delta,\beta,\sigma}(n^d)$ and with probability $1-\beta$, returns a polynomial factor $\mathcal{F} = \{P_{i,j}\}_{1\leq i\leq d-1, 1\leq j\leq M_i}$ of degree $d-1$ and $|\mathcal{F}| = O_{\delta,\beta,\sigma}(1)$ such that P is measurable in $\mathcal{F}$.*

Algorithmic Regularity in Low Characteristic. Unfortunately, Gowers uniformity for polynomial factors fails to address "bias implies low rank" phenomena in the case when $\mathbb{F}$ has small characteristic. Kaufman and Lovett [51] introduce a stronger notion of regularity in order to handle the general case. This notion is rather technical and we omit introducing it here.

Algorithmic versions of the results in [51] were proved in [16] via a similar stronger notion of regular factors referred to as "strong unbiased factors". The reason for the need of this new notions is that uniform

factors (Theorem 7.2.9) fail to address fields of low characteristic, for the reason that previously in order to refine a factor to a uniform factor we made use of division by $d!$ which is not possible in fields with $|\mathbb{F}| \leq d$.

The main result of [16] in the low-characteristic setting is an algorithm for regularization of factors in low characteristic.

Lemma 12.4.4 (Strongly unbiased refinement (informal))**.** *Suppose that $\gamma : \mathbb{N} \to \mathbb{R}^+$ is a regularity parameter. There is a deterministic algorithm that given a factor $\mathcal{F} = \{P_1, \cdots, P_m\}$ of degree d, runs in $\mathrm{O}_\gamma(n^{O(d)})$, returns a strongly γ-unbiased degree $\leq d$ factor $\mathcal{F}' \succeq \mathcal{F}$.*

12.5 Algorithmic Inverse Theorem for Polynomials

From Theorem 12.4.3 one can deduce an algorithmic version of an inverse theorem for Gowers norm (Theorem 6.2.3) for polynomials in the high characteristic setting.

Theorem 12.5.1 (Algorithmic Inverse Theorem in High Characteristic)**.** *Suppose that $|\mathbb{F}| > d \geq 2$ and that $\epsilon, \beta \in (0, 1]$. There is an $\eta_{\epsilon,\beta,d} \in (0, 1]$, and a randomized algorithm that given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d with $\|e_{\mathbb{F}}(P(x))\|_{U^{k+1}} \geq \epsilon$, runs in $\mathrm{O}_{\delta,\beta}(n^d)$ and with probability $1 - \beta$, returns a polynomial Q of degree $\leq k$ such that*

$$|\langle e_{\mathbb{F}}(P), e_{\mathbb{F}}(Q)\rangle| \geq \eta.$$

This theorem follows from the following proposition via a basic Fourier analytic argument along with the Goldreich-Levin theorem from [32].

Proposition 12.5.2 (Computing Polynomials with High Gowers Norm)**.** *Suppose that $|\mathbb{F}| > d \geq 2$ and that $\delta, \beta \in (0, 1]$. There is a randomized algorithm that given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d with $\|e_{\mathbb{F}}(P(x))\|_{U^d} \geq \delta$, runs in $\mathrm{O}_{\delta,\beta}(n^d)$ and with probability $1-\beta$, returns a polynomial factor $\mathcal{F}$ of degree $d-1$ such that*

- *There is a function $\Gamma : \mathbb{F}^{|\mathcal{F}|} \to \mathbb{F}$ such that $P = \Gamma(\mathcal{F})$.*
- *$|\mathcal{F}| = O_{d,\delta,\beta}(1)$.*

Proof. Write $\partial^d P(h_1, \cdots, h_d) := D_{h_1} \cdots D_{h_d} P(x)$. Since P has degree d, $\partial^d P$ does not depend on x. From the definition of the U^d norm, we have

$$\mathrm{bias}(\partial^d P) = \|e(P)\|_{U^d}^{2^d} \geq \delta^{2^d}.$$

Applying Theorem 12.4.3 to $\partial^d P$, with probability $1 - \frac{\beta}{2}$, we can find a factor $\widetilde{\mathcal{F}}$ of degree $d-1$, such that $|\widetilde{\mathcal{F}}| = O_{\delta,\beta,d}(1)$ and $\partial^d P$ is measurable in $\widetilde{\mathcal{F}}$. It is easy to check that since $|\mathbb{F}| > d$, we have the following Taylor expansion

$$P(x) = \frac{1}{d!}\partial^d P(x, \cdots, x) + Q(x),$$

where Q is a polynomial of degree $\leq d-1$. We can find explicit description of P, and therefore Q in $O(n^d)$. Thus letting $\mathcal{F}' := \widetilde{\mathcal{F}} \cup \{Q\}$ finishes the proof. □

We are now ready to see the proof of Theorem 12.5.1.

Proof Theorem 12.5.1: By Theorem 12.5.2, with probability $1 - \frac{\beta}{3}$ we can find a polynomial factor $\widetilde{\mathcal{F}}$ of degree $d-1$ such that P is measurable in $\widetilde{\mathcal{F}}$. Let $\gamma : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function to be specified later. By Theorem 12.4.2, with probability $1 - \frac{\beta}{3}$, we can refine $\widetilde{\mathcal{F}}$ to a γ-uniform factor $\mathcal{F} = \{P_1, \cdots, P_m\}$ of same degree $d-1$, with $\dim(\mathcal{F}) = \mathrm{O}_{\gamma,\beta}(1)$. Since P is measurable in $\mathcal{F}$, there exists $\Gamma : \mathbb{F}^{\dim(\mathcal{F})} \to \mathbb{F}$ such that $P = \Gamma(\mathcal{F})$. Using the Fourier decomposition of $e_{\mathbb{F}}(\Gamma)$ we can write

$$f(x) := e_{\mathbb{F}}\big(P(x)\big) = \sum_{i=1}^{L} c_i \; e_{\mathbb{F}}\big(\langle \alpha^{(i)}, \mathcal{F}\rangle(x)\big), \tag{12.22}$$

where $L = |\mathbb{F}|^{\dim(\mathcal{F})} = O_{\gamma,\beta}(1)$, $\alpha^{(i)} \in \mathbb{F}^{\dim(\mathcal{F})}$, and

$$\langle \alpha^{(i)}, \mathcal{F} \rangle(x) := \sum_{j=1}^{m} \alpha_j^{(i)} \cdot P_j(x).$$

Notice that the terms in (12.22), unlike Fourier characters, are not orthogonal. But since the factor is γ-uniform, Theorem 7.6.2 ensures approximate orthogonality. Let $Q_i := \langle \alpha^{(i)}, \mathcal{F} \rangle$. Choose $\gamma(u) \leq \frac{\sigma}{|\mathbb{F}|^{2u}}$, so that $\gamma(\dim(\mathcal{F})) \leq \frac{\sigma}{L^2}$, where $\sigma := \frac{\varepsilon^{2^{k+1}}}{4}$. It follows from the near orthogonality of the terms in (12.22) by Theorem 7.6.2 that

$$|c_i - \langle f, e_{\mathbb{F}}(Q_i) \rangle| \leq \frac{\sigma}{L}, \tag{12.23}$$

and

$$\left| \|f\|_2^2 - \sum_{i=1}^{L} c_i^2 \right| \leq \sigma. \tag{12.24}$$

Claim 12.5.3. *There exists $\delta'(\epsilon, |\mathbb{F}|) \in (0,1]$ such that the following holds. Assume that f and $\mathcal{F}$ are as above. Then there is $i \in [L]$, for which $\deg(Q_i) \leq k$ and $\left|\langle f, e_{\mathbb{F}}(Q_i) \rangle\right| \geq \delta'$.*

Proof. We will induct on the degree of $\mathcal{F}$. Assume for the base case that $\mathcal{F}$ is of degree k, i.e. $d = k+1$, thus applying the following Cauchy-Schwarz inequality

$$\epsilon^{2^{k+1}} \leq \|f\|_{U^{k+1}}^{2^{k+1}} \leq \|f\|_2^2 \|f\|_\infty^{2^{k+1}-2}, \tag{12.25}$$

and (12.24) imply that there exists $i \in [L]$ such that $c_i^2 \geq \frac{\epsilon^{2^{k+1}} - \sigma}{L} = \frac{3\epsilon^{2^{k+1}}}{4L}$, which combined with (12.23) implies that $|\langle f, e_{\mathbb{F}}(Q_i) \rangle| \geq \frac{\epsilon^{2^{k+1}}}{2L}$.

Now for the induction step, assume that $d > k+1$. We will decompose (12.22) into two parts, first part consisting of the terms of degree $\leq k$ and the second part consisting of the terms of degree strictly higher than k. Namely, letting $S := \{i \in [L] : \deg(Q_i) \leq k\}$ we write $f = g + h$ where $g := \sum_{i \in S} c_i e_{\mathbb{F}}(Q_i)$ and $h := \sum_{i \in [L] \setminus S} c_i e_{\mathbb{F}}(Q_i)$. Notice that by the triangle inequality of Gowers norm, our choice of γ, and the fact that $\mathcal{F}$ is γ-uniform

$$\|h\|_{U^{k+1}} \leq \sum_{i \in [L] \setminus S} |c_i| \cdot \|e_{\mathbb{F}}(Q_i)\|_{U^{k+1}} \leq L \cdot \frac{\epsilon^{2^{k+1}}}{4L^2} = \frac{\epsilon^{2^{k+1}}}{4L},$$

and thus

$$\|g\|_{U^{k+1}} \geq \frac{\epsilon}{2}.$$

Now the claim follows by the base case. □

Let $\delta'(\epsilon, |\mathbb{F}|)$ be as in the above claim. We will use the following theorem of Goldreich and Levin [32] which gives an algorithm to find all the large Fourier coefficients of $e_{\mathbb{F}}(\Gamma)$.

Theorem 12.5.4 (Goldreich-Levin theorem [32])**.** *Let $\zeta, \rho \in (0,1]$. There is a randomized algorithm, which given oracle access to a function $\Gamma : \mathbb{F}^m \to \mathbb{F}$, runs in time $\mathrm{O}\big(m^2 \log m \cdot \mathrm{poly}(\frac{1}{\zeta}, \log(\frac{1}{\rho}))\big)$ and outputs a decomposition*

$$\Gamma = \sum_{i=1}^{\ell} b_i \cdot e_{\mathbb{F}}(\langle \eta_i, x \rangle) + \Gamma',$$

with the following guarantee:

- $\ell = \mathrm{O}(\frac{1}{\zeta^2})$.
- $\Pr\left[\exists i : \ |b_i - \hat{\Gamma}(\eta_i)| > \zeta/2\right] \leq \rho$.

- $\Pr\left[\forall \alpha \text{ such that } |\hat{f}(\alpha)| \geq \zeta, \exists i\ \eta_i = \alpha\right] \geq 1 - \rho$.

We will use the above theorem with parameters $\zeta := \frac{\delta'}{2}$ and $\rho := \frac{\beta}{3}$. By Claim 12.5.3 there is $i \in [L]$ such that $\hat{\Gamma}(\alpha) = c_i \geq \frac{3\delta'}{4}$. With probability $1 - \frac{\beta}{6}$ there is j such that $\eta_j = \alpha_i$ and with probability at least $1 - \frac{\beta}{6}$, $|b_j - c_i| \leq \frac{\zeta}{2} \leq \frac{\delta'}{4}$, and therefore $c_i \geq \frac{\delta'}{2}$.

By a union bound, adding up the probabilities of the errors, with probability at least $1 - \beta$, we find Q_i such that

$$\left|\langle f, e_{\mathbb{F}}(Q_i)\rangle\right| \geq \frac{\delta'}{4}.$$

□

12.6 Derandomization via PRGs for Polynomials

In this section we will discuss how all the algorithms that we have discussed above can be derandomized to efficient deterministic algorithms.

The key point here is that our only uses of randomness above have been in evaluating a collection of polynomials on set of inputs that are chosen uniformly at random from $\mathbb{F}^n$. We will use known efficient constructions of *pseudorandom generators* for polynomials [81, 54].

Definition 12.6.1. *A distribution D on $\mathbb{F}^n$ is said to ϵ-fool degree d polynomials in n variables over $\mathbb{F}^n$ if for every degree d polynomial $P : \mathbb{F}^n \to \mathbb{F}$,*

$$|\mathbb{E}_D[e(P(D))] - \mathbb{E}_U[e(P(U))]| \leq \epsilon,$$

where $U \in \mathbb{F}^n$ is uniformly distributed.

The following lemma shows that a distribution that fools a single polynomial can be used to fool a collection of polynomials with a slightly worse error term.

Lemma 12.6.2. *If D ϵ-fools degree d polynomials, then for every collection of C degree d polynomials $P_1, \cdots, P_C$,*

$$\sum_{b_1,\cdots,b_C \in \mathbb{F}^n} |\Pr_D[P_1(D) = b_1 \wedge \cdots \wedge P_C(D) = b_C] - \Pr_U[P_1(U) = b_1 \wedge \cdots \wedge P_C(U) = b_C]| \leq p^C \epsilon,$$

where $p = |\mathbb{F}|$.

We will use Viola's explicit construction of pseudorandom generators for degree d polynomials.

Theorem 12.6.3 ([81])**.** *There is an explicit generator $g : \mathbb{F}^s \to \mathbb{F}^n$ with $s = d \log_p n + O(d \cdot 2^d \cdot \log(1/\epsilon))$, such that for uniform $Z \in \mathbb{F}^s$, $g(Z)$ ϵ-fools degree d polynomials.*

Derandomized Bogdanov-Viola Lemma. In the proof of Theorem 12.3.1 we have used randomness in two steps.

First, we estimated μ_a in the statistical distance by sampling P on a constant number of randomly selected inputs. This step can be simply derandomized using Theorem 12.6.3, by evaluating P on $g(z)$ for $z \in \mathbb{F}^s$. This requires $2^s = O(n^d)$ queries to P.

Second, we chose $h = (h_1, \cdots, h_C) \in (\mathbb{F}^n)^C$ uniformly at random, and obtained a distribution μ_{obs} and $\widetilde{P}_h$ such that for every x

$$\Pr_{h_1,\cdots,h_C}(\widetilde{P}_h(x) \neq P(x)) \leq \Pr_h\left(\|\mu_{\text{obs}}^{(x)} - \mu_{P(x)}\| \geq \frac{\delta}{|\mathbb{F}|}\right) \leq \sigma\beta_2.$$

Suppose x is fixed, and consider polynomials $Q_i(h_1, \cdots, h_C) := P(x + h_i)$. By Theorem 12.6.2 and Theorem 12.6.3, there is a map $g : \mathbb{F}^s \to (\mathbb{F}^n)^C$ such that

$$\Pr_{Z \in \mathbb{F}^s}(\widetilde{P}_{g(Z)}(x) \neq P(x)) \leq \Pr_{Z \in \mathbb{F}^s}\left(\|\mu_{\text{obs}}^{(x)} - \mu_{P(x)}\| \geq \frac{\delta}{|\mathbb{F}|}\right) \leq \sigma\beta_2 + p^C \cdot \epsilon.$$

Therefore, choosing ϵ sufficiently small, there exists $Z \in \mathbb{F}^s$ such that

$$\Pr_x(\widetilde{P}_{g(Z)}(x) \neq P(x)) \leq \sigma\beta_2 + p^C \cdot \epsilon \leq \sigma. \tag{12.26}$$

Thus we can iterate through all the $p^s = \text{Poly}(n)$ choices of Z, and check whether Equation (12.26) holds for that value of Z. Note that Equation (12.26) can itself be deterministically decided in $\text{Poly}(n)$ time using pseudorandom generators for low-degree polynomials.

Derandomized Regularity Lemmas. Having access to a deterministic Bogdanov-Viola lemma, consequently Lemma 12.4.1, Theorem 12.4.2, Theorem 12.4.3, Theorem 12.4.4, Theorem 12.5.1 and Theorem 12.5.2 can also be made deterministic. The key observation is that all these results rely on the Bogdanov-Viola lemma and in addition use their randomness towards estimating bias or Gowers norm of a given polynomial. Theorem 12.6.3 can be used to derandomize any step that requires estimating bias of a low-degree polynomial, and Theorem 12.6.3 combined with Theorem 12.6.2 can be used in order to estimate the Gowers norm of a low-degree polynomial.

12.7 Algorithmic Decompositions

Given a positive integer k, a vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and a function $\Gamma : \mathbb{F}^k \to \mathbb{F}$, we say that a function $P : \mathbb{F}^n \to \mathbb{F}$ is $(k, \mathbf{\Delta}, \Gamma)$-*structured* if there exist polynomials $P_1, P_2, \ldots, P_k : \mathbb{F}^n \to \mathbb{F}$ with each $\deg(P_i) \leq \Delta_i$ such that for all $x \in \mathbb{F}^n$,

$$P(x) = \Gamma(P_1(x), P_2(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ are said to form a $(k, \mathbf{\Delta}, \Gamma)$-*decomposition* (or simply, a polynomial decomposition). For instance, an n-variate polynomial over the field $\mathbb{F}$ of total degree d factors non-trivially exactly when it is $(2, (d-1, d-1), \mathsf{prod})$-structured where $\mathsf{prod}(a, b) = a \cdot b$. Later in Chapter 15 we will see that the property of being $(k, \mathbf{\Delta}, \Gamma)$-structured is constant query testable with one-sided error even when P does not necessarily have bounded degree. However, the problem is harder if we ask for finding a decomposition realizing the $(k, \mathbf{\Delta}, \Gamma)$-structure.

Somewhat surprisingly, using the algorithmic regularity lemmas it can be shown that every degree-structural property can be decided in polynomial time.

Theorem 12.7.1 ([10])**.** *For every finite field $\mathbb{F}$ of prime order, positive integers $d < |\mathbb{F}|$, k, every vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and every function $\Gamma : \mathbb{F}^k \to \mathbb{F}$, there is a deterministic algorithm $\mathcal{A}_{k,\mathbf{\Delta},\Gamma}$ that takes as input a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d, runs in time polynomial in n, and outputs a $(k, \mathbf{\Delta}, \Gamma)$-decomposition of P if one exists while otherwise returning* NO*.*

The algorithm is quite simple. Given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$, we first use a corollary of Theorem 12.4.2 and Theorem 12.4.3 to write P as a function of a uniform polynomial factor $\{Q_1, \ldots, Q_m\}$, i.e. $P(x) = G(Q_1(x), \ldots, Q_m(x))$ where $m = O(1)$ and $G : \mathbb{F}^m \to \mathbb{F}$. Now, the proof technique of [10][1] shows that the only way P can have a $(k, \mathbf{\Delta}, \Gamma)$-decomposition is if there are functions $G_1, \ldots, G_k : \mathbb{T}^m \to \mathbb{T}$ such that $G(z_1, \ldots, z_m) = \Gamma(G_1(z_1, \ldots, z_m), \ldots, G_k(z_1, \ldots, z_m))$ and also, for every $i \in [k]$, $G_i(Q_1(x), \ldots, Q_m(x))$ is a polynomial of degree at most Δ_i. Since $m = O(1)$, there are only a constant number of possible $G_1, \ldots, G_k$, and so the whole algorithm runs in polynomial time.

We will use the following corollary of Theorem 12.4.2 and Theorem 7.2.10.

[1] A similar reasoning was used in [13] in their proof of testability of degree structural properties.

Theorem 12.7.2. *Suppose $d < |\mathbb{F}|$, $\rho \in (0,1)$ and $R : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a non-decreasing function. There is a function $C : \mathbb{Z}^+ \to \mathbb{Z}^+$ and an efficient algorithm that takes as input a polynomial factor $\mathcal{B}$ of degree $\leq d$ outputs an R-regular factor $\tilde{\mathcal{B}}$ where $\tilde{\mathcal{B}}$ is a refinement of $\mathcal{B}$, is of degree d, and $\|\tilde{\mathcal{B}}\| \leq C(|\mathcal{B}|)$. Additionally, if $\mathcal{B}$ is defined by polynomials $P_1, \cdots, P_m$, then we can find functions $\Gamma_1, \cdots, \Gamma_m$ such that $P_i(x) = \Gamma_i(\tilde{\mathcal{B}}(x))$ for every $i \in [m]$.*

Moreover, if $\mathcal{B}$ itself is a syntactic refinement of some polynomial factor $\mathcal{B}'$ of rank at least $R(|\mathcal{B}|) + 1$, then $\tilde{\mathcal{B}}$ will also be a syntactic refinement of $\mathcal{B}'$.

Proof of Theorem 12.7.1:

Let $R : \mathbb{Z}^+ \to \mathbb{Z}^+$ be chosen so that $R(m) = r(C(m+k)) + C(m+k) + |\mathbb{F}|$ for a function $r : \mathbb{Z}^+ \to \mathbb{Z}^+$ to be fixed later, where C is as in Theorem 12.7.2. Applying Theorem 12.7.2 to the factor defined by the single polynomial P, we find an R-regular polynomial factor $\mathcal{B}$ of degree d defined by polynomials $P_1, \cdots, P_C : \mathbb{F}^n \to \mathbb{F}$ such that $P(x) = G(\mathcal{B}(x))$ for some $G : \mathbb{F}^C \to \mathbb{F}$. Note that here $C = O(1)$. Note that for small $n \leq Cd$, we can decide whether f is (k, Δ, Γ)-structured by brute force in $O(1)$ time, and further find such a decomposition if it exists.

Suppose $n > Cd$. From each P_i, pick a monomial m_i with degree equal to $\deg(P_i)$. Since $n > Cd$, there exists $i_0 \in [n]$ such that x_{i_0} does not appear in any of the m_i's. Let $\mathcal{B}'$ be the polynomial factor defined by polynomials $P'_1, \cdots, P'_C$ are respectively the restrictions of $P_1, \cdots, P_C$ to $x_{i_0} = 0$. Moreover, by Theorem 7.0.2, $\mathcal{B}'$ is $(R - |\mathbb{F}|)$-regular.

Recursively, decide (k, Δ, Γ)-structure for the polynomial $P' := P|_{x_{i_0}=0}$ on $n-1$ variables. Note that

$$P'(x) = G(P'_1(x), \cdots, P'_C(x)).$$

It is easy to see that if P' does not admit (k, Δ, Γ)-structure, then the same is true for P. So, in this case we will safely output NO. Assume otherwise that we have found

$$P'(x) = \Gamma(S_1(x), \cdots, S_k(x)),$$

where $\deg(S_i) \leq \Delta_i$. We will show how in this case P is (k, Δ, Γ)-structured, and in fact we can find such a decomposition efficiently. We will use Theorem 12.7.2 to the factor defined by $P'_1, \cdots, P'_C, S_1, \cdots, S_k$ to find a refinement $\mathcal{B}'$ of rank at least $r(\mathcal{B}')$. Note that the setting of our parameters and Theorem 12.7.2 guarantee that $\mathcal{B}'$ is a syntactic refinement of $P'_1, \cdots, P'_C$. That is $\mathcal{B}'$ is defined by $P'_1, \cdots, P'_C$ and $S'_1, \cdots, S'_D$ where for each $i \in [k]$, $S_i(x) = G_i(P'_1(x), \cdots, P'_C(x), S'_1(x), \cdots, S'_D(x))$ for some function G_i. Thus we have for all x

$$G(P'_1(x), \cdots, P'_C(x)) = \Gamma\left(G_1\left(P'_1(x), \cdots, P'_C(x), S'_1(x), \cdots, S'_D(x)\right), \cdots, G_k\left(P'_1(x), \cdots, P'_C(x), S'_1(x), \cdots, S'_D(x)\right)\right)$$

Suppose r is set large enough so that by Theorem 7.4.1 all the atoms of $\mathcal{B}'$ are nonempty. This means that $\mathcal{B}'(x)$ acquires all possible values in its range $\mathbb{F}^{C+D}$. This allows us to deduce from the above equation that

$$G(a_1, \cdots, a_C) = \Gamma(G_1(a_1, \cdots, a_C, b_1, \cdots, b_D), \cdots, G_k(a_1, \cdots, a_C, b_1, \cdots, b_D)),$$

for all $a_1, \cdots, a_C, b_1, \cdots, b_D \in \mathbb{F}$. Define $Q_i(x) := G_i(P_1(x), \cdots, P_C(x), 0, \cdots, 0)$ for $i \in [k]$. We will use Theorem 15.1.2 which will be proved in Chapter 15, to show that for each $i \in [k]$, $\deg(Q_i) \leq \deg(S_i) \leq \Delta_i$. Thus we have

$$P(x) = \Gamma(Q_1(x), \cdots, Q_k(x)),$$

is a (k, Δ, Γ) decomposition of P.

The final algorithm is by recursively applying the above process until n is a constant, and building back the decomposition step by step. □

Part III

Algebraic Property Testing

In Part I of this text we discussed linearity testing and more generally tests for being a polynomial of a given degree. These are all instances of "algebraic properties", a general class that we are going to define shortly. It includes more complex properties such as being a product of two low-degree polynomials, or having sparse Fourier representation. In this part of the survey, we will use the tools developed in Part II to establish a general result showing that all algebraic properties that have local definitions are testable. We will focus on one-sided testable properties and for the treatment of the two-sided testable properties we refer the reader to [83] and [82] where nonstandard analysis is used to give a characterization.

Chapter 13

Algebraic Properties

13.1 Affine and linear invariance

Recall that $\mathbb{F} = \mathbb{F}_p$ for a fixed prime p. Let $[R] = \{1, \ldots, R\}$ be a constant sized set. A property of functions $f : \mathbb{F}^n \to [R]$ is simply a subset of all such functions, namely $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$. Typically, we assume that such property is defined for all $n \geq 1$.

Note that a generic property can completely ignore the algebraic structure of $\mathbb{F}^n$ and treat it as an arbitrary set of size $|\mathbb{F}|^n$. Hence any notion of "algebraic property" must require some assumptions on $\mathcal{P}$ that take the algebraic structure of the set $\mathbb{F}^n$ into account. This is typically done by requiring that the property $\mathcal{P}$ is invariant under certain algebraic transformations. Indeed this is analogous to other combinatorial settings such as graph properties, or hypergraph properties, where the properties are assumed to be invariant under permutation of the vertices.

We say that a property $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$ is *linear-invariant* if it is the case that for any $f \in \mathcal{P}$ and for any linear transformation $L : \mathbb{F}^n \to \mathbb{F}^n$, it holds that $f \circ L \in \mathcal{P}$, where $(f \circ L)(x) = f(L(x))$. Similarly, an *affine-invariant* property is closed under composition with affine transformations $A : \mathbb{F}^n \to \mathbb{F}^n$ (an affine transformation A is of the form $A(x) = Lx + c$ where L is linear and c is a constant). Both linear invariance and affine invariance are very natural candidates for defining algebraic properties.

The property of a function $f : \mathbb{F}^n \to \mathbb{F}$ being affine is testable by a simple reduction to the linearity testing [19], and is itself affine-invariant. Other well-studied examples of affine-invariant (and hence, linear-invariant) properties include Reed-Muller codes (in other words, bounded degree polynomials) [5, 6, 26, 65, 2] and Fourier sparsity [34]. In fact, affine invariance seems to be a common feature of most interesting properties that one would classify as algebraic. Kaufman and Sudan in [53] made explicit note of this phenomenon and initiated a general study of the testability of affine-invariant properties (see also [31]). In particular, they asked for necessary and sufficient conditions for the testability of affine-invariant properties.

13.2 Locally Characterized Properties

Let $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$. Let us first assume that $\mathcal{P}$ is testable in the strongest possible way, and see what this entails. To recall, we say that $\mathcal{P}$ is one-sided proximity-oblivious q-query testable if there is a local test which makes q queries to a function $f : \mathbb{F}^n \to [R]$, always accepts if $f \in \mathcal{P}$ and otherwise rejects f with probability related to the distance of f from $\mathcal{P}$. See Theorem 2.0.1 for a more formal definition.

So, assume $\mathcal{P}$ is one-sided proximity-oblivious q-query testable. Every such test has the following structure:

- Step 1: Randomly pick points $(x_1, \ldots, x_q) \in (\mathbb{F}^n)^q$ according to some joint distribution μ.
- Step 2: Query the values of $f(x_1), \ldots, f(x_q)$ and accordingly output "$f \in \mathcal{P}$" or "$f \notin \mathcal{P}$".

Suppose that the queries result in $(f(x_1), \ldots, f(x_q)) = \sigma \in [R]^q$. Since the algorithm is only allowed to make a one-sided error, if there is a function $g \in \mathcal{P}$ with $(g(x_1), \ldots, g(x_q)) = \sigma$, then the algorithm must return "$f \in \mathcal{P}$". On the other hand, if there is no such function $g \in \mathcal{P}$, then it can safely return "$f \notin \mathcal{P}$". In such a case we say that $\mathcal{C} = (x_1, x_2, \ldots, x_q; \sigma)$ forms a *q-local constraint* for $\mathcal{P}$ (equivalently, a q-local witness for non-membership in $\mathcal{P}$), as if $f \in \mathcal{P}$ then necessarily $(f(x_1), \ldots, f(x_q)) \neq \sigma$.

Definition 13.2.1 (Local constraint)**.** *A q-local constraint for a property $\mathcal{P} \subset \{\mathbb{F}^n \to [R]\}$ is $\mathcal{C} = (x_1, \ldots, x_q; \sigma)$, where $x_1, \ldots, x_q \in \mathbb{F}^n$, $\sigma \in [R]^q$ such that for all $g \in \mathcal{P}$ it holds that $(g(x_1), \ldots, g(x_q)) \neq \sigma$.*

Let $\mathcal{C} = (x_1, \ldots, x_q; \sigma)$ be a local constraint for $\mathcal{P}$. If $f : \mathbb{F}^n \to [R]$ satisfies that $(f(x_1), \ldots, f(x_q)) = \sigma$, then we say that f violates the constraint $\mathcal{C}$. In particular, we know that $f \notin \mathcal{P}$. Let $\mathcal{C}_1, \ldots, \mathcal{C}_m$ be the set of all the *q-local constraints* for the above test. One can rewrite the above test as in the following:

- Step 1: Randomly pick points $(x_1, \ldots, x_q) \in (\mathbb{F}^n)^q$ according to some joint distribution μ.
- Step 2: If $(x_1, \ldots, x_q; f(x_1), \ldots, f(x_q))$ equals some $\mathcal{C}_i$ then output "$f \notin \mathcal{P}$". Otherwise output "$f \in \mathcal{P}$".

Note that if $f \notin \mathcal{P}$, then according to Theorem 2.0.1, the tester must reject f with positive probability. Consequently, there must exist at least one q-local constraint among $\mathcal{C}_1, \ldots, \mathcal{C}_m$ that is violated by f. This provides a "local characterization" for $\mathcal{P}$. Namely,

$$\mathcal{P} = \{f : \mathbb{F}^n \to [R] \text{ that do not violate any of } \mathcal{C}_1, \ldots, \mathcal{C}_m\}.$$

The following definition formalizes this notion.

Definition 13.2.2 (Locally characterized property)**.** *A property $\mathcal{P} \subset \{F^n \to [R]\}$ is q-locally characterized if there exist q-local constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$ such that*

$$\mathcal{P} = \{f : \mathbb{F}^n \to [R] \textit{ that do not violate any of } \mathcal{C}_1, \ldots, \mathcal{C}_m\}.$$

We say $\mathcal{P}$ is *locally characterized* if it is q-locally characterized for some constant q. It follows from the above discussion that if $\mathcal{P}$ is one-sided proximity-oblivious q-query testable then it is also q-locally characterized.

We now give some examples of locally characterized affine-invariant properties. Consider the property of $f : \mathbb{F}^n \to \mathbb{F}$ being an affine function (namely, a degree 1 polynomial). It is 4-locally characterized because a function f is affine if and only if $f(x) - f(x+y) - f(x+z) + f(x+y+z) = 0$ for every $x, y, z \in \mathbb{F}^n$. Note that this characterization naturally suggests a 4-query test: pick random $x, y, z \in \mathbb{F}^n$ and check whether the identity holds or not for that choice of x, y, z. As we saw in Chapter 2, this is indeed a local test for the property of affine functions (we proved it for $\mathbb{F} = \mathbb{F}_2$ in Theorem 2.2.1, and the proof can be extended to any finite field).

More generally, consider the property of being a polynomial of degree at most d, for some fixed integer $d > 0$. If $f : \mathbb{F}^n \to \mathbb{F}$ has degree $\leq d$, then it is annihilated by taking $d+1$ directional derivatives (See Section 3.1). This implies that this property is also locally characterizable. Independent work of [52, 50] has obtained the optimal value of q for general fields $\mathbb{F}_p$; it is $q = p^{\lceil \frac{d+1}{p-1} \rceil}$, and the property of being a degree $\leq d$ polynomial is q-locally characterized. These works also showed that the property is also testable. Again, the test is simply to pick a random constraint and check if it is violated.

Indeed, for any q-locally characterized property $\mathcal{P}$ defined by constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$, one can design the following q-query test: choose a constraint $\mathcal{C}_i$ uniformly at random and reject only if the input function violates $\mathcal{C}_i$. Clearly, if the input function f is in $\mathcal{P}$, the test always accepts. The main challenge is in showing that if f is ε-far from $\mathcal{P}$, then this test rejects f with some probability $\delta = \delta(\varepsilon) > 0$. Equivalently put, f is ε-far from $\mathcal{P}$ then at least δm of the local constraints are violated. This was conjectured in [53] and proved in a sequence of works [12, 69, 15, 14, 13].

Theorem 13.2.3 ([13])**.** *Every q-locally characterized affine-invariant property is proximity obliviously testable with q queries.*

The proof of Theorem 13.2.3 is presented in Chapter 14.

13.3 Locality of Affine Invariant Properties via Linear Forms

In the context of linear-invariant and affine-invariant properties, we can define the notion of local characterization in a more algebraic way. As most of the research has focused on affine invariant properties we also limit our discussion to such properties. Many of the observations below naturally extend to linear invariant properties.

Consider an affine-invariant property $\mathcal{P} \subset \{\mathbb{F}^n \to [R]\}$ that is locally characterized by a collection $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ of q-local constraints. As $\mathcal{P}$ is affine-invariant, one can assume that $\mathcal{C}$ is also invariant under affine transformations. In other words, if $(a_1, a_2, \ldots, a_q; \sigma) \in \mathcal{C}$, then also $(Aa_1, Aa_2, \ldots, Aa_q; \sigma) \in \mathcal{C}$ for every affine transformation $A : \mathbb{F}^n \to \mathbb{F}^n$.

This can be rephrased using linear forms. Assume that $a_1, \ldots, a_q \in \mathbb{F}^n$ span a linear space of dimension $r \leq q$, and let $b_1, \ldots, b_r \in \mathbb{F}^n$ be a basis for this subspace. Let $\{\lambda_{i,j} \in \mathbb{F} : i \in [q], j \in [r]\}$ be such that $a_i = \sum_{j=1}^{r} \lambda_{i,j} b_j$. The assumption that $\mathcal{P}$ is affine invariant means that for every linear map $L : \mathbb{F}^n \to \mathbb{F}^n$ and ever $c \in \mathbb{F}^n$, we also have that $(La_1 + c, \ldots, La_q + c; \sigma) \in \mathcal{C}$. Let $L_1, \ldots, L_q \in \mathbb{F}^{r+1}$ be linear forms given by $L_i = (1, \lambda_{i,1}, \ldots, \lambda_{i,r})$. Then the following are also q-local constraint for $\mathcal{P}$:

$$(L_1(x_0, x_1, \ldots, x_r), \ldots, L_q(x_0, \ldots, x_r); \sigma) \qquad \forall x_0, \ldots, x_r \in \mathbb{F}^n.$$

The linear forms $L_1, \ldots, L_q$ are *affine forms*, where we recall the definition that a linear form is affine if its first coordinate is 1.

Next, we formalize this notion of affine local constraints.

Definition 13.3.1 (Affine local constraints)**.** *An* affine constraint of size q on k variables *is a tuple* $A = (L_1, \ldots, L_q; \sigma)$ *of* q *affine forms* $L_1, \ldots, L_q \in \mathbb{F}^k$ *and* $\sigma \in [R]^q$*.*

An affine constraint can be used to define affine invariant properties in the following way.

Definition 13.3.2 (Properties defined by affine constraints)**.** *Consider an affine constraint* (A, σ) *of size* q *on* k *variables. A function* $f : \mathbb{F}^n \to [R]$ *is said to be* (A, σ)-free *if there exist no* $x_1, \ldots, x_k \in \mathbb{F}^n$ *such that*

$$(f(L_1(x_1, \ldots, x_k)), \ldots, f(L_q(x_1, \ldots, x_k))) = \sigma.$$

On the other hand, if such $x_1, \ldots, x_\ell$ *exist, we say that* f induces (A, σ) at $x_1, \ldots, x_k$*.*

Given a (possibly infinite) collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\}$ *of affine constraints, a function* $f : \mathbb{F}^n \to [R]$ *is said to be* $\mathcal{A}$-free *if it is* (A^i, σ^i)*-free for every* $i \geq 1$*.*

The above discussion shows that an affine-invariant property is q-locally characterized if and only if it can be described using a bounded number of affine constraints of size q.

13.4 Subspace Hereditary Properties

Just as a necessary condition for proximity oblivious testability with one-sided error is local characterization, one can formulate a natural condition that is (almost) necessary for testability in general. In the context of affine-invariant properties, the condition can be succinctly stated as follows. We will assume throughout that $\mathcal{P}$ is a property of functions $f : \mathbb{F}^n \to [R]$ for all $n \geq 1$.

Definition 13.4.1 (Subspace hereditary properties)**.** *An affine-invariant property* $\mathcal{P}$ *is said to be* affine-subspace hereditary *if for any* $f : \mathbb{F}^n \to [R]$ *satisfying* $\mathcal{P}$*, the restriction of* f *to any affine subspace of* $\mathbb{F}^n$ *also satisfies* $\mathcal{P}$*.*

In [15], it is shown that every affine-invariant property testable by a "natural" tester is very "close" to a subspace hereditary property[1]. Thus, if we gloss over some technicalities, subspace hereditariness is a necessary condition for testability. In the opposite direction, [15] conjectures the following:

[1]We omit the technical definitions of "natural" and "close", since they are unimportant here. Informally, the behavior of a "natural" tester is independent of the size of the domain and "close" means that the property deviates from an actual affine subspace hereditary property on functions over a finite domain. See [15] for details, or [3] for the analogous definitions in a graph-theoretic context.

Conjecture 13.4.2 ([15])**.** *Every affine-subspace hereditary property is testable with one-sided error.*

Resolving Theorem 13.4.2 would yield a combinatorial *characterization* of the (natural) one-sided testable affine-invariant properties, similar to the characterization for testable dense graph properties [3].

Affine subspace hereditariness and affine constraints are related through the following simple observation.

Observation 13.4.3. *An affine-invariant property $\mathcal{P}$ is subspace hereditary if and only if it is equivalent to the property of $\mathcal{A}$-freeness for some fixed collection $\mathcal{A}$ of affine constraints (possibly infinite).*

Proof. Given an affine invariant property $\mathcal{P}$, a simple (though inefficient) way to obtain the set $\mathcal{A}$ is to let it be the following. For every $m \geq 1$ and every function $g : \mathbb{F}^m \to [R]$ that is not in $\mathcal{P}$, include the constraint that the restriction of $f : \mathbb{F}^n \to [R]$ to an m-dimensional subspace is not equal to g.

This can be expressed as an affine constraint. For every $a \in \mathbb{F}^m$ let $L_a \in \mathbb{F}^{m+1}$ be an affine form given by $L_a = (1, a_1, \ldots, a_m)$. The affine constraint we add is is

$$\mathcal{C}_g := ((L_a : a \in \mathbb{F}^n); (g(a) : a \in \mathbb{F}^n)) .$$

In particular, g is not $\mathcal{C}_g$-free, as g induces $\mathcal{C}_g$ at $x_1 = e_1, \ldots, x_m = e_m$, where e_i is the i-th unit vector. Hence the property defined by $\mathcal{A}$ is contained in $\mathcal{P}$. The containment in the other direction follows from $\mathcal{P}$ being affine-invariant and hereditary.

The other direction of the observation is trivial. □

Theorem 13.4.2 is not yet confirmed or refuted, however [13] proves testability under an additional assumption of "bounded complexity". Define the *Cauchy-Schwarz complexity* (See Theorem 11.1.1) of an affine constraint (A, σ) to be simply the Cauchy-Schwarz complexity of A. Let the Cauchy-Schwarz complexity of a collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\}$ of affine constraints to be the maximum of Cauchy-Schwarz complexity of the constrains (A^i, σ^i).

Theorem 13.4.4 ([13])**.** *Every subspace hereditary property of bounded Cauchy-Schwarz complexity is testable with one-sided error.*

All natural affine-invariant properties that we know of have bounded complexity (in fact, most are locally characterized). However, testing the subspace hereditary properties not covered by Theorem 13.4.4 are of theoretical interest.

13.5 Locality Dimension

The notion of q-local characterization uses the number of queries to measure the complexity of the characterization. In the algebraic setting of affine-invariant properties it is as natural to consider the dimension of the affine subspace containing those queries.

Definition 13.5.1 (Locality Dimension)**.** *The* locality dimension *of an affine-invariant property $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$ is the smallest K such that the following holds. There exists a collection $\mathcal{F} \subset \{\mathbb{F}^K \to [R]\}$ such that $f \in \mathcal{P}$ if and only if $f|_V \notin \mathcal{F}$ for all K-dimensional affine subspaces $V \subseteq \mathbb{F}^n$.*

As discussed above, every q-locally characterized $\mathcal{P}$ is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a collection of affine constraints with each constraint on at most $q + 1$ variables. Obviously one can extend this to include all the affine linear forms on these variables, or equivalent query f on the q-dimensional affine subspace containing these $q + 1$ variables. Thus every q-locally characterized property has locality dimension of at most q. In the other hand direction, if $\mathcal{P}$ is of locality dimension K, then it is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a finite collection of affine constraints, with each constraint of size $|\mathbb{F}|^K$ on $K + 1$ variables.

The notion of locality dimension is even more natural for affine-subspace hereditary properties.

Observation 13.5.2 (Locality Dimension for hereditary properties)**.** *The locality dimension of a hereditary affine-invariant property $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$ is the smallest K such that $f \in \mathcal{P}$ if and only if $f|_V \in \mathcal{P}$ for all K-dimensional affine subspaces $V \subseteq \mathbb{F}^n$.*

Chapter 14

One-sided algebraic property testing

In this Chapter we establish Theorem 13.2.3 and Theorem 13.4.4. That is, we show that every q-locally characterized affine-invariant property is proximity obliviously testable with one-sided error using q queries, and more generally we show that every affine invariant subspace hereditary property of bounded Cauchy-Schwarz complexity is testable with one-sided error.

14.1 Proof overview

Before delving into the technical details let us first present an overview of the proof. For simplicity, assume for now that $\mathcal{A}$ consists only of a single affine constraint (A, σ) where A is the tuple of affine linear forms $(L_1, \ldots, L_m)$, each over ℓ variables, and $\sigma \in [R]^m$. Our goal is to show that, when f is ε-far from being (A, σ)-free, then f contains many instances of violation of (A, σ)-freeness. The decomposition theorems presented in Chapter 9 allow us to find a polynomial factor partition of the domain $\mathbb{F}^n$ such that, for the purpose of counting the number of violations of (A, σ) freeness, f looks uniform in most of the parts. If the notion of ε-farness would have allowed to remove points from the domain of f, then the proof would have become much simpler.

Let us first consider this simpler case. Suppose that if remove any given ε-fraction of points from $\mathbb{F}^n$, there are still copies of (A, σ) in f. Now we look at the polynomial factor decomposition of f and remove all the points that are not in the uniform cells. We also remove all the points that have unpopular values in the uniform cells. That is for every cell, if there is a value that appears on very few points in that cell, then we remove all the points from that cell with that value. Now since the number of non-uniform cells is small, and also the number of unpopular points is small (by their definition), we have removed only few points from the domain. Hence according the modified notion of ε-farness, there are still copies of (A, σ) remaining in f. Now every point in any such copy must belong to a uniform cell and its value is not unpopular in that cell. It follows from these two properties that there are many such copies of (A, σ) in f.

Unfortunately, the natural notion of ε-farness does not allow removing points from the domain of f, and instead one is only allowed to change the value of f on at most ε-faction of the points. This makes proving the theorem considerably more difficult. Now one has to appeal to the more technical decomposition theorems such as Theorem 9.4.3. This decomposition theorem, roughly speaking, shows the existence of a polynomial factor partition $\mathcal{B}$ of the domain such that for every atom c of $\mathcal{B}$, there is a sub-atom c' that comes from a finer partition $\mathcal{B}'$ such that most sub-atoms provide a good statistical samples of the demography of the atom that contains them. Now the most popular values inside sub-atoms will play an important role, and instead of removing points from the domain we will change the value of f to such values. In other words, unpopular values and mis-represented values inside every atom are changed to the most popular value inside the corresponding sub-atom.

Again, the number of changes will be small (less than ε-fraction of points). So the new function will still contain at least one copy of (A, σ). Note that the values in this copy are abundant in the corresponding

sub-atoms as we changed the unpopular values to the most popular values in the sub-atoms. Then, a key property of sub-atoms (i.e. the fact that s is fixed in Theorem 9.4.3) together with the uniformity of sub-atoms allows one to find many copies of this (A, σ) inside the corresponding sub-atoms.

We now proceed to give the detailed proof.

14.2 Big Picture Functions

Suppose we have a function $f : \mathbb{F}^n \to [R]$, and we want to find out whether it induces a particular affine constraint (A, σ), where $A = (L_1, \dots, L_m)$ is a sequence of affine forms on ℓ variables and $\sigma \in [R]^m$. Now, suppose $\mathbb{F}^n$ is partitioned by a polynomial factor $\mathcal{B}$ defined by nonclassical polynomials $P_1, \dots, P_C$ of degrees $d_1, \dots, d_C$ and depths $k_1, \dots, k_C$, respectively. Then, observe that if $b_1, \dots, b_m \in \mathbb{T}^C$ denote the atoms of $\mathcal{B}$ containing $L_1(x_1, \dots, x_\ell), \dots, L_m(x_1, \dots, x_\ell)$ respectively, it must be the case that $b_1, \dots, b_m$ are $\mathcal{B}$-consistent with A (as defined in Theorem 7.7.4). Thus, to locate where f might induce (A, σ), we should restrict our search to sequences of atoms consistent with A.

It will be convenient to "blur" the given function f so as to retain only atom-level information about it. That is, for every atom c of $\mathcal{B}$, we will define $f_{\mathcal{B}}(c) \subseteq [R]$ to be the set of all values that f takes within c. We denote by $\mathcal{P}([R])$ the power set of $[R]$.

Definition 14.2.1. *Given a function $f : \mathbb{F}^n \to [R]$ and a polynomial factor $\mathcal{B}$, the* big picture function *of f is the function $f_{\mathcal{B}} : \mathbb{T}^{|\mathcal{B}|} \to \mathcal{P}([R])$, defined by $f_{\mathcal{B}}(c) = \{f(x) : \mathcal{B}(x) = c\}$.*

On the other hand, given any function $g : \mathbb{T}^C \to \mathcal{P}([R])$, and a vector of degrees $\mathbf{d} = (d_1, \dots, d_C)$ and depths $\mathbf{k} = (k_1, \dots, k_C)$ (which we think of as corresponding to the degrees and depths of some polynomial factor of complexity C), we will define what it means for such a function to "induce" a copy of a given constraint.

Definition 14.2.2 (Partially induce)**.** *Suppose we are given vectors $\mathbf{d} = (d_1, \dots, d_C) \in \mathbb{Z}^C_{>0}$ and $\mathbf{k} = (k_1, \dots, k_C) \in \mathbb{Z}^C_{\geq 0}$, a function $g : \prod_{i \in [C]} \mathbb{U}_{k_i+1} \to \mathcal{P}([R])$, and an affine constraint (A, σ) of size m. We say that g* partially $(\mathbf{d}, \mathbf{k})$-induces *(A, σ) if there exist a sequence $b_1, \dots, b_m \in \mathbb{T}^C$ such that*

(i) $b_1, \dots, b_m$ is $(\mathbf{d}, \mathbf{k})$-consistent with A (see Theorem 7.7.4).

(ii) $\sigma_j \in g(b_j)$ for each $j \in [m]$.

Theorem 14.2.2 is justified by the following trivial observation.

Observation 14.2.3. *If $f : \mathbb{F}^n \to [R]$ induces a constraint (A, σ), then for a factor $\mathcal{B}$ defined by polynomials of respective degrees $(d_1, \dots, d_{|\mathcal{B}|}) = \mathbf{d}$ and respective depths $(k_1, \dots, k_{|\mathcal{B}|}) = \mathbf{k}$, the big picture function $f_{\mathcal{B}}$ partially $(\mathbf{d}, \mathbf{k})$-induces (A, σ).*

To handle a possibly infinite collection $\mathcal{A}$ of affine constraints (in order to prove Theorem 13.4.4), we will employ a compactness argument, analogous to one used in [4] in the context of graph properties, to bound the size of the constraint partially induced by the big picture function. Let us make the following definition:

Definition 14.2.4 (The compactness function)**.** *Suppose we are given positive integers C and d, and a possibly infinite collection of affine constraints $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \dots\}$, where (A^i, σ^i) is of size m_i. For fixed $\mathbf{d} = (d_1, \dots, d_C) \in [d]^C$ and $\mathbf{k} = (k_1, \dots, k_C) \in \left[0, \left\lfloor \frac{d-1}{p-1} \right\rfloor\right]^C$, denote by $\mathcal{G}(\mathbf{d}, \mathbf{k})$ the set of functions $g : \prod_{i=1}^C \mathbb{U}_{k_i+1} \to \mathcal{P}([R])$ that partially $(\mathbf{d}, \mathbf{k})$-induce some $(A^i, \sigma^i) \in \mathcal{A}$. The compactness function is defined as*

$$\Psi_{\mathcal{A}}(C, d) = \max_{\mathbf{d}, \mathbf{k}} \max_{g \in \mathcal{G}(\mathbf{d}, \mathbf{k})} \min_{\substack{(A^i, \sigma^i) \text{ partially} \\ (\mathbf{d}, \mathbf{k})\text{-induced by } g}} m_i$$

where the outer max *is over vectors $\mathbf{d} = (d_1, \dots, d_C) \in [d]^C$ and $\mathbf{k} = (k_1, \dots, k_C) \in \left[0, \left\lfloor \frac{d-1}{p-1} \right\rfloor\right]^C$. Whenever $\mathcal{G}(\mathbf{d}, \mathbf{k})$ is empty, we set the corresponding maximum to 0.*

Note that $\Psi_{\mathcal{A}}(C,d)$ is indeed finite, as the number of possible degree and depth sequences are bounded by d^{2C}, and the size of $\mathcal{G}(d_1,\dots,d_C)$ is bounded by $2^{Rp^{dC}}$.

Remark 14.2.5. *Note that if a function $g:\mathbb{T}^C\to\mathcal{P}([R])$ partially $(\mathbf{d},\mathbf{k})$-induces some constraint from $\mathcal{A}$ where $\mathbf{d}\in[d]^C$, then g must belong to $\mathcal{G}(\mathbf{d},\mathbf{k})$, and consequently it will necessarily partially induce some $(A^i,\sigma^i)\in\mathcal{A}$ whose size is at most $\Psi_{\mathcal{A}}(C,d)$.*

14.3 Proof of Testability

In this section we establish our main result Theorem 13.4.4 which in turn implies Theorem 13.2.3. It suffices to prove the following.

Theorem 14.3.1. *Let $d>0$ be an integer. Suppose we are given a possibly infinite collection of affine constraints $\mathcal{A}=\{(A^1,\sigma^1),(A^2,\sigma^2),\dots\}$ where each (A^i,σ^i) is an affine constraint of Cauchy-Schwarz complexity $\le d$, and of size m_i on ℓ_i variables. Then, there are functions $\ell_{\mathcal{A}}:(0,1)\to\mathbb{Z}_{>0}$ and $\delta_{\mathcal{A}}:(0,1)\to(0,1)$ such that the following is true. For every $\varepsilon\in(0,1)$, if a function $f:\mathbb{F}^n\to[R]$ is ε-far from being $\mathcal{A}$-free, then f induces at least $\delta_{\mathcal{A}}(\varepsilon)|\mathbb{F}|^{n\ell_i}$ many copies of some (A^i,σ^i) with $\ell_i<\ell_{\mathcal{A}}(\varepsilon)$.*

Moreover, if $\mathcal{A}$ is q-locally characterized, then $\ell_A(\varepsilon)\le q$ and hence is bounded by a constant independent of ε.

Theorem 13.4.4 follow immediately. Given $\mathcal{A}$ as in the statement of the theorems, let $f:\mathbb{F}^n\to[R]$ which is ε-far from $\mathcal{A}$. Let $\mathcal{A}_0\subset\mathcal{A}$ be the subset of all $(A^i,\sigma^i)\in\mathcal{A}$ with $\ell_i\le\ell_A(\varepsilon)$. Observe that $\mathcal{A}_0$ is a finite set, as we may assume that $m_i\le p^{\ell_i}$.

Consider the following test. Pick a uniform $(A^i,\sigma^i)\in\mathcal{A}_0$, pick uniformly $X=(x_1,\dots,x_{\ell_i})\in\mathbb{F}^{n\ell_i}$ and test if $(f(L_1(X)),\dots,f(L_{m_i}(X)))=\sigma^i$. ByTheorem 14.3.1, if f is ε-far from $\mathcal{A}$-freeness, this test rejects with probability at least $|\mathcal{A}_0|^{-1}\delta_{\mathcal{A}}(\varepsilon)$. For Theorem 13.2.3, note that if $\mathcal{A}$ is q-locally characterized, then the test is a one-sided proximity oblivious q-local test. Below, we prove Theorem 14.3.1.

Fix a function $f:\mathbb{F}^n\to[R]$ that is ε-far from being $\mathcal{A}$-free. For $i\in[R]$, define $f^{(i)}:\mathbb{F}^n\to\{0,1\}$ so that $f^{(i)}(x)$ equals 1 when $f(x)=i$ and equals 0 otherwise. Additionally, set the following parameters, where $\Psi_{\mathcal{A}}$ is the compactness function from Theorem 14.2.4:

$$\begin{array}{lllllll}\zeta & = & \frac{\varepsilon}{8R}, & & & & \\ \alpha(C) & = & p^{-2dC\Psi_{\mathcal{A}}(C,d)}, & & r(C) & = & r_{7.2.6}(\mathbb{F},d,\alpha(C)) \\ \Delta(C) & = & \frac{1}{16}\zeta^{\Psi_{\mathcal{A}}(C,d)}, & & \eta(C) & = & \frac{1}{8p^{dC\Psi_{\mathcal{A}}(C,d)}}\left(\frac{\varepsilon}{24R}\right)^{\Psi_{\mathcal{A}}(C,d)}.\end{array}$$

Intuitively, think of ζ as a small enough constant (depending only on ε,R), on $r(\cdot)$ as a large enough rank, and on $\alpha(\cdot),\Delta(\cdot),\eta(\cdot)$ as taking small enough positive values.

Decomposing by regular factors. As a first step, we apply (the multifunction generalization) of Theorem 9.4.3 to the functions $f^{(1)},f^{(2)},\dots,f^{(R)}$. By the theorem, there exists a polynomial factor $\mathcal{B}$ of degree d, and a refinement of it $\mathcal{B}'$ of degree d, both r-regular of complexity at most $C_{9.4.3}(\mathbb{F},d,r,\eta,\Delta,\zeta)$, such that the following holds. We can decompose each $f^{(i)}$ as

$$f^{(i)}=f_1^{(i)}+f_2^{(i)}+f_3^{(i)}$$

where

(i) $f_1^{(i)}=\mathbb{E}[f|\mathcal{B}']$.

(ii) $\|f_2^{(i)}\|_{U^{d+1}}<\eta(|\mathcal{B}'|)$.

(iii) $\|f_3^{(i)}\|_2\le\Delta(|\mathcal{B}|)$.

(iv) f_1 and f_1+f_3 take values in $[0,1]$; and f_2,f_3 take values in $[-1,1]$.

(v) $\mathcal{B}'$ ζ-represents $\mathcal{B}$ with respect to f.

Moreover, assume that the atoms of $\mathcal{B}'$ are indexed by (c,s) with $c \in \mathbb{T}^{|\mathcal{B}|}, s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$. Then exists a choice of $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$ for which the following is true:

(vi) For every atom c of $\mathcal{B}$, the sub-atom (c,s) of $\mathcal{B}'$ satisfies that

$$\mathbb{E}\left[|f_3^{(i)}(x)|^2 \mid \mathcal{B}'(x) = (c,s)\right] \leq \Delta(|\mathcal{B}|)^2.$$

(vii) For at most a ζ fraction of atoms c in $\mathcal{B}$ it holds that

$$\left|\mathbb{E}[f^{(i)}|c] - \mathbb{E}[f^{(i)}|(c,s)]\right| > \zeta.$$

We denote the sequence of polynomials defining $\mathcal{B}'$ by $P_1, \ldots, P_{|\mathcal{B}'|}$. Denote the degree of P_i by d_i and the depth of P_i by k_i. Since $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, we may assume $\mathcal{B}$ is generated by the polynomials $P_1, \ldots, P_{|\mathcal{B}|}$. We denote $C = |\mathcal{B}|$ and $C' = |\mathcal{B}'|$.

Note that $\|\mathcal{B}\| < p^{(k_{\max}+1)C} \leq p^{dC}$, where $k_{\max} \leq \lfloor (d-1)/(p-1) \rfloor$ is the maximum depth of a polynomial in $\mathcal{B}$.

Cleanup. Based on $\mathcal{B}'$ and $\mathcal{B}$, we construct a function $F : \mathbb{F}^n \to [R]$ that is $\frac{\varepsilon}{2}$-close to f and hence, still violates $\mathcal{A}$-freeness. The "cleaner" structure of F will help us locate the induced constraint violated by f.

The function F is the same as f except for the following. For every atom c of $\mathcal{B}$, let $t_c = \arg\max_{j \in [R]} \Pr[f(x) = j \mid \mathcal{B}'(x) = (c,s)]$ be the most popular value inside the corresponding subatom (c,s).

- **Poorly-represented atoms:** If there exists $i \in [R]$ such that $|\Pr[f(x) = i \mid \mathcal{B}(x) = c] - \Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)]| > \zeta$, then set $F(z) = t_c$ for every z in the atom c.
- **Unpopular values:** For any $i \in [R]$ such that $\Pr_x[f(x) = i \mid \mathcal{B}'(x) = (c,s)] < \zeta$, if z in the atom c satisfies $f(z) = i$, then set $F(z) = t_c$.
- For all other z in the atom c, set $F(z) = f(z)$.

A key property of the cleanup function F is that it supports a value inside an atom c of $\mathcal{B}$ only if the original function f acquires the value on at least an ζ fraction of the subatom (c,s). Furthermore as the following lemma shows it is $\varepsilon/2$-close to f, and therefore as f is ε-far from $\mathcal{A}$-freeness, we have that F is not $\mathcal{A}$-free.

Lemma 14.3.2. *The cleanup function F is $\varepsilon/2$-close to f, and therefore, it is not $\mathcal{A}$-free.*

Proof. The first step (fixing poorly represented atoms) applies to at most $\zeta R\|\mathcal{B}\|$ atoms, since $\mathcal{B}'$ ζ-represents $\mathcal{B}$ with respect to each $f^{(1)}, \ldots, f^{(R)}$. By Theorem 7.4.1, each atom of $\mathcal{B}$ occupies at most $\frac{1}{\|\mathcal{B}\|} + \alpha(C)$ fraction of the entire domain. So, the fraction of points whose values are set in the first step is at most $\zeta R\|\mathcal{B}\|(\frac{1}{\|\mathcal{B}\|} + \alpha(C)) < 2\zeta R = \varepsilon/4$.

In the second step (changing unpopular values), if $\Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)] < \zeta$, then $\Pr[f(x) = i \mid \mathcal{B}(x) = c] < \Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)] + \zeta < 2\zeta$. Hence, the fraction of the points z whose values are set in the second step is at most $2\zeta R = \varepsilon/4$.

Thus, the distance of F from f is bounded by $\varepsilon/2$. □

Locating a violated constraint. We now want to use F to "locate" a popular affine constraint induced in f. Setting $\mathbf{d} = (d_1, \ldots, d_C)$ and $\mathbf{k} = (k_1, \ldots, k_C)$, we have by Theorem 14.2.3 that the big picture function $F_{\mathcal{B}}$ of F will partially $(\mathbf{d}, \mathbf{k})$-induce some constraint from $\mathcal{A}$, and hence by Theorem 14.2.5, it will partially $(\mathbf{d}, \mathbf{k})$-induce some $(A, \sigma) \in \mathcal{A}$ of size $m \leq \Psi_{\mathcal{A}}(C, d)$ on ℓ variables. We will show that the original function f violates many instances of this constraint.

Denote the affine forms in A by $(L_1, \ldots, L_m)$ and the vector σ by $(\sigma_1, \ldots, \sigma_m)$. Since we can assume $\ell \leq m$ (without loss of generality by making a change of variables), we can now define

$$\ell_{\mathcal{A}}(\varepsilon) = \Psi_{\mathcal{A}}(C_{9.4.3}(\mathbb{F}, d, r, \eta, \Delta, \zeta), d). \tag{14.1}$$

Let $b_1, \ldots, b_m \in \prod_{i=1}^{C} \mathbb{U}_{k_i+1}$ correspond to the atoms of $\mathcal{B}$ where (A, σ) is partially $(\mathbf{d}, \mathbf{k})$-induced by $F_{\mathcal{B}}$. That is, $b_1, \ldots, b_m$ are consistent with A, and $\sigma_i \in F_{\mathcal{B}}(b_i)$ for every $i \in [m]$. Also, let $b'_1, \ldots, b'_m \in \prod_{i=1}^{C'} \mathbb{U}_{k_i+1}$ index the associated subatoms in $\mathcal{B}'$, obtained by letting $b'_j = (b_j, s)$ for every $j \in [m]$.

Lemma 14.3.3. *The subatoms $b'_1, \ldots, b'_m$ are consistent with A.*

Proof. Since $b_1, \ldots, b_m$ are already consistent with A, we only need to show that for every $i \in [C+1, C']$, the sequence $(b'_{1,i}, \ldots, b'_{m,i}) = (s_{i-C}, s_{i-C}, \ldots, s_{i-C})$ is (d_i, k_i)-consistent. That is, we need to show that there exists a homogeneous nonclassical polynomial P of degree d and depth k, and a choice of $x_1, \ldots, x_\ell \in \mathbb{F}^n$, such that $P(L_j(x_1, \ldots, x_\ell)) = s_{i-C}$ for all $j \in [m]$. This is where we use the assumption that the linear forms are affine. Take $x_2 = \ldots = x_\ell = 0$. Then $L_j(x_1, 0, \ldots, 0) = x_1$. So we only need to show the existence of P as above and of $x_1 \in \mathbb{F}^n$ for which $P(x_1) = s_{i-C}$. But this clearly holds, say by taking $P = P_i$ and x_1 any value in the (c, s) subatom. □

The main analysis. Let $X = (x_1, \ldots, x_\ell)$ where $x_1, \ldots, x_\ell \in \mathbb{F}^n$ are independently and uniformly chosen. Our goal is to prove a lower bound on

$$\Pr_X [f(L_1(X)) = \sigma_1 \wedge \cdots \wedge f(L_m(X)) = \sigma_m] = \mathbb{E}_X \left[f^{(\sigma_1)}(L_1(X)) \cdots f^{(\sigma_m)}(L_m(X)) \right]. \tag{14.2}$$

Theorem 14.3.1 follows if the above expectation is larger than the respective $\delta_{\mathcal{A}}(\varepsilon)$. We rewrite the expectation as

$$(14.2) = \mathbb{E}_X \left[(f_1^{(\sigma_1)} + f_2^{(\sigma_1)} + f_3^{(\sigma_1)})(L_1(X)) \cdots (f_1^{(\sigma_m)} + f_2^{(\sigma_m)} + f_3^{(\sigma_m)})(L_m(X)) \right]. \tag{14.3}$$

We can expand the expression inside the expectation as a sum of 3^m terms. The expectation of any term involving $f_2^{(\sigma_j)}$ for any $j \in [m]$ is bounded in magnitude by $\|f_2^{(\sigma_j)}\|_{U^{d+1}} \leq \eta(|\mathcal{B}'|)$, by Theorem 11.1.2 and the assumption that the Cauchy-Schwarz complexity of A is bounded by d. Hence, the expression (14.3) is at least

$$(14.3) \geq \mathbb{E}_X \left[(f_1^{(\sigma_1)} + f_3^{(\sigma_1)})(L_1(X)) \cdots (f_1^{(\sigma_m)} + f_3^{(\sigma_m)})(L_m(X)) \right] - 3^m \eta(|\mathcal{B}'|).$$

Next, because of the non-negativity of $f_1^{(\sigma_j)} + f_3^{(\sigma_j)}$ for every $j \in [m]$, we may further require that any event holds. We will require that $\mathcal{B}'(L_j(X)) = b'_j$ for all j. So

$$(14.3) \geq \mathbb{E}_X \left[\left(f_1^{(\sigma_1)} + f_3^{(\sigma_1)} \right) (L_1(X)) \cdots \left(f_1^{(\sigma_m)} + f_3^{(\sigma_m)} \right) (L_m(X)) \prod_{j \in [m]} 1_{[\mathcal{B}'(L_j(X)) = b'_j]} \right] - 3^m \eta(|\mathcal{B}'|), \tag{14.4}$$

where $1_{[\mathcal{B}'(L_j(X)) = b'_j]}$ is the indicator function of the event $\mathcal{B}'(L_j(X)) = b'_j$. In other words, now we are only counting patterns that arise from the selected subatoms $b'_1, \ldots, b'_m$.

Next, expand the product inside the expectation into 2^m terms. We will show that the contribution from each of the $2^m - 1$ terms involving $f_3^{(\sigma_k)}$ for any $k \in [m]$ is small. Each such term is trivially bounded from above by

$$\mathbb{E}_X \left[\left| f_3^{(\sigma_k)}(L_k(X)) \right| \prod_{j \in [m]} 1_{[\mathcal{B}'(L_j(X)) = b'_j]} \right]. \tag{14.5}$$

Here, we used fact that $|f_1^{(i)}(x)| \leq 1$ for all $i \in [R]$, which follows from (i) in our construction. The next lemma bounds the expression in (14.5). Let Λ_i denote the (d_i, k_i)-dependency set of $L_1, \ldots, L_m$ for $i \in [C']$.

Lemma 14.3.4. $\mathbb{E}_X \left[\left| f_3^{(\sigma_k)}(L_k(X)) \right| \prod_{j\in[m]} 1_{[\mathcal{B}'(L_j(X))=b'_j]} \right] \leq 2\Delta(C) \frac{\prod_{i\in[C']} |\Lambda_i|}{\|\mathcal{B}'\|^m}$.

Proof. In order to simplify the proof, apply a change of basis on $L_1, \ldots, L_m$ so that $L_k = e_1$. Namely, $L_k(X) = x_1$. To simplify notation let us also denote $x = x_1$ and $\mathbf{x} = (x_2, \ldots, x_\ell)$.

As a first step, we apply the Cauchy-Schwartz inequality and obtain

$$\begin{aligned}
&\left(\mathbb{E}_{X\in(\mathbb{F}^n)^\ell} \left[\left| f_3^{(\sigma_k)}(x_1) \right| \prod_{j\in[m]} 1_{[\mathcal{B}'(L_j(X))=b'_j]} \right] \right)^2 \\
&= \left(\mathbb{E}_{x\in\mathbb{F}^n, \mathbf{x}\in(\mathbb{F}^n)^{\ell-1}} \left[\left| f_3^{(\sigma_k)}(x) \right| 1_{[\mathcal{B}'(x)=b'_k]} \prod_{j\in[m]} 1_{[\mathcal{B}'(L_j(x,\mathbf{x}))=b'_j]} \right] \right)^2 \\
&\leq \mathbb{E}_x \left[|f_3^{(\sigma_k)}(x)|^2 1_{[\mathcal{B}'(x)=b'_k]} \right] \mathbb{E}_x \left(\mathbb{E}_{\mathbf{x}} \prod_{j\in[m]} 1_{[\mathcal{B}'(L_j(x,\mathbf{x}))=b'_j]} \right)^2 . \qquad (14.6)
\end{aligned}$$

We first bound the first term in the right hand side. By (vi) in our construction and Theorem 7.4.1, we have

$$\begin{aligned}
&\mathbb{E}_x \left[|f_3^{(\sigma_k)}(x)|^2 1_{[\mathcal{B}'(x)=b'_k]} \right] \\
&= \mathbb{E}_x \left[|f_3^{(\sigma_k)}(x)|^2 \mid \mathcal{B}'(x) = b'_k \right] \Pr_x[\mathcal{B}'(x) = b'_k] \\
&\leq \Delta^2(C) \Pr_x[\mathcal{B}'(x) = b'_k] \leq \Delta^2(C) \left(\frac{1}{\|\mathcal{B}'\|} + \alpha(C') \right) \leq \frac{2\Delta^2(C)}{\|\mathcal{B}'\|}. \qquad (14.7)
\end{aligned}$$

The second term in the right hand side of (14.6) is equal to

$$\begin{aligned}
&\mathbb{E}_x \left(\mathbb{E}_{\mathbf{x}} \prod_{j\in[m]} 1_{[\mathcal{B}'(L_j(x,\mathbf{x}))=b'_j]} \right)^2 = \\
&\frac{1}{\|\mathcal{B}'\|^{2m}} \mathbb{E}_x \left[\left(\mathbb{E}_{\mathbf{x}} \prod_{\substack{i\in[C'] \\ j\in[m]}} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} e\left(\lambda_{i,j} \cdot (P_i(L_j(x,\mathbf{x})) - b'_{i,j}) \right) \right)^2 \right] \\
&= \frac{1}{\|\mathcal{B}'\|^{2m}} \mathbb{E}_x \left[\left(\sum_{\substack{(\lambda_{i,j})\in \\ \prod_{i,j}[0,p^{k_i+1}-1]}} e\left(- \sum_{\substack{i\in[C'] \\ j\in[m]}} \lambda_{i,j} b'_{i,j} \right) \mathbb{E}_{\mathbf{x}} e\left(\sum_{\substack{i\in[C'] \\ j\in[m]}} \lambda_{i,j} P_i(L_j(x,\mathbf{x})) \right) \right)^2 \right] \\
&\leq \frac{1}{\|\mathcal{B}'\|^{2m}} \sum_{\substack{(\lambda_{i,j}),(\tau_{i,j})\in \\ \prod_{i,j}[0,p^{k_i+1}-1]}} \left| \mathbb{E}_{x,\mathbf{x},\mathbf{y}} \left[e\left(\sum_{\substack{i\in[C'] \\ j\in[m]}} \lambda_{i,j} P_i(L_j(x,\mathbf{x})) \right) e\left(- \sum_{\substack{i\in[C'] \\ j\in[m]}} \tau_{i,j} P_i(L_j(x,\mathbf{y})) \right) \right] \right|. \qquad (14.8)
\end{aligned}$$

Here, $\mathbf{y} = (y_2, \ldots, y_\ell)$ where $y_2, \ldots, y_\ell \in \mathbb{F}^n$ are new independent uniform random variables.

We can bound the above using Theorem 7.6.1. Let A' denote the set of $2m$ linear forms: $\{L_j(x_1, x_2, \ldots, x_\ell) \mid j \in [m]\} \cup \{L_j(x_1, y_2, \ldots, y_\ell) \mid j \in [m]\}$ in variables $x_1, \ldots, x_\ell, y_2, \ldots, y_\ell$. Let Λ'_i denote the (d_i, k_i)-dependency set of A'.

Applying Theorem 7.6.1 (just as in the proof of Theorem 7.7.5), we get that

$$(14.8) \leq \frac{\prod_{i\in[C']} |\Lambda'_i|}{\|\mathcal{B}'\|^{2m}} + \alpha(C') \leq \frac{2\prod_{i\in[C']} |\Lambda'_i|}{\|\mathcal{B}'\|^{2m}}. \qquad (14.9)$$

Thus, the next step is to compute $|\Lambda_i'|$.

Claim 14.3.5. *For each $i \in [C']$ it holds that $|\Lambda_i'| = |\Lambda_i|^2 \cdot p^{k_i+1}$.*

Proof. Recall that by our initial change of basis, $L_k(x, \mathbf{x}) = L_k(x, \mathbf{y}) = x$. For any $\lambda, \tau \in \Lambda_i$ and any $\alpha \in \mathbb{Z}_{p^{k_i+1}}$, note that $(\lambda_1 + \alpha, \lambda_2, \ldots, \lambda_m, \tau_1 - \alpha, \tau_2, \ldots, \tau_m) \in \Lambda_i'$. Hence, $|\Lambda_i'| \geq |\Lambda_i|^2 \cdot p^{k_i+1}$. To show $|\Lambda_i'| \leq |\Lambda_i|^2 \cdot p^{k_i+1}$, we give a map from Λ_i' to $\Lambda_i \times \Lambda_i$ that is p^{k_i+1}-to-1.

Suppose that $\sum_{j=1}^m \lambda_j Q(L_j(x_1, x_2, \ldots, x_\ell)) + \sum_{j=1}^m \tau_j Q(L_j(x_1, y_2, \ldots, y_\ell)) \equiv 0$ for every nonclassical homogeneous polynomial Q of degree d_i and depth k_i. Setting $x_2 = \ldots = x_\ell = 0$ shows that

$$\sum_{j=1}^m \tau_j Q(L_j(x_1, y_2, \ldots, y_\ell)) = -\left(\sum_{j=1}^m \lambda_j\right) Q(x_1),$$

which implies that

$$\left(-\sum_{j=2}^m \lambda_j, \lambda_2, \ldots, \lambda_m\right) \in \Lambda_i.$$

Similarly, setting $y_2 = \ldots = y_\ell = 0$ shows that

$$\sum_{j=1}^m \lambda_j Q(L_j(x_1, x_2, \ldots, x_\ell)) = -\left(\sum_{j=1}^m \tau_j\right) Q(x_1),$$

which implies that

$$\left(-\sum_{j=2}^m \tau_j, \tau_2, \ldots, \tau_m\right) \in \Lambda_i.$$

Consequently,

$$(\lambda, \tau) \mapsto \left(\left(-\sum_{j=2}^m \lambda_j, \lambda_2, \ldots, \lambda_m\right), \left(-\sum_{j=2}^m \tau_j, \tau_2, \ldots, \tau_m\right)\right)$$

is a map from Λ_i' to $\Lambda_i \times \Lambda_i$. To see that it is p^{k_i+1}-to-1, note that if $(\lambda_1, \ldots, \lambda_m, \tau_1, \ldots, \tau_m) \in \Lambda_i'$ then also

$$(\lambda_1 + \gamma, \lambda_2, \ldots, \lambda_m, \tau_1 - \gamma, \tau_2, \ldots, \tau_m) \in \Lambda_i'$$

for every $\gamma \in \mathbb{Z}_{p^{k_i+1}}$, and that these p^{k_i+1} elements are all mapped to the same element in $\Lambda_i \times \Lambda_i$ by our map. □

Thus

$$(14.9) \leq \frac{2\prod_{i\in[C']} |\Lambda_i|^2 p^{k_i+1}}{\|\mathcal{B}'\|^{2m}} = \frac{2\prod_{i\in[C']} |\Lambda_i|^2}{\|\mathcal{B}'\|^{2m-1}}.$$

Combining this with Equation (14.7) and Equation (14.6), we obtain

$$(14.6) \leq 4\Delta^2(C)\frac{\prod_{i\in[C']} |\Lambda_i|^2}{\|\mathcal{B}'\|^{2m}}. \tag{14.10}$$

This concludes the proof of the lemma.

□

Finally, we turn to the main term in the expansion of Equation (14.4). We know from Theorem 14.3.3 that the subatoms $b'_1, \ldots, b'_m$ are consistent with A. Thus

$$\begin{aligned}
&\mathbb{E}_X\left[f_1^{(\sigma_1)}(L_1(X))\cdots f_1^{(\sigma_m)}(L_m(X))\cdot\prod_{j\in[m]}1_{[\mathcal{B}'(L_j(X))=b'_j]}\right]\\
&=\Pr_X[\mathcal{B}'(L_1(X))=b'_1\wedge\cdots\wedge\mathcal{B}'(L_m(X))=b'_m]\cdot\\
&\qquad\mathbb{E}_X\left[f_1^{(\sigma_1)}(L_1(X))\cdots f_1^{(\sigma_m)}(L_m(X))\mid\forall j\in[m],\ \mathcal{B}'(L_j(X))=b'_j\right]\\
&\geq\left(\frac{\prod_{i=1}^{C'}|\Lambda_i|}{\|\mathcal{B}'\|^m}-\alpha(C')\right)\zeta^m. \qquad (14.11)
\end{aligned}$$

Let us justify the last line. The first term is due to the lower bound on the probability from Theorem 7.7.5. The second term in (14.11) follows since each $f_1^{(\sigma_j)}$ is constant on the atoms of $\mathcal{B}'$, and because by construction, the big picture function $F_{\mathcal{B}}$ of the cleanup function F, on which (A,σ) was partially induced, supports a value inside an atom b of $\mathcal{B}$ only if the original function f acquires the value on at least an ζ fraction of the subatom (c,s). By our choice of $\alpha(C')$, we can further deduce that

$$\mathbb{E}_X\left[f_1^{(\sigma_1)}(L_1(X))\cdots f_1^{(\sigma_m)}(L_m(X))\cdot\prod_{j\in[m]}1_{[\mathcal{B}'(L_j(X))=b'_j]}\right]\geq\frac{\prod_{i=1}^{C'}|\Lambda_i|}{2\|\mathcal{B}'\|^m}. \qquad (14.12)$$

Setting $\beta=\frac{\prod_{i=1}^{C'}|\Lambda_i|}{\|\mathcal{B}'\|^m}$ and combining the bounds from (14.4), Theorem 14.3.4 and (14.12), we conclude

$$(14.2)\geq\beta\left(\frac{1}{2}\left(\frac{\varepsilon}{8R}\right)^m-2^{m+1}\Delta(C)\right)-3^m\cdot\eta(C').$$

To complete the proof, we need to choose parameters. We have $\|\mathcal{B}'\|\leq p^{dC'}$ and $m\leq\Psi_{\mathcal{A}}(C,d)$. Thus $\beta\geq\|\mathcal{B}'\|^{-\Psi_{\mathcal{A}}(C,d)}$. We choose $\Delta(C)=\frac{1}{16}(\frac{\varepsilon}{8R})^{\Psi_{\mathcal{A}}(d,C)}$, $\eta(C')<\frac{1}{8\|\mathcal{B}'\|^{\Psi_{\mathcal{A}}(C,d)}}\left(\frac{\varepsilon}{24R}\right)^{\Psi_{\mathcal{A}}(C,d)}$, and both C and C' are upper-bounded by $C_{9.4.3}(\mathbb{F},d,r,\eta,\Delta,\zeta)$, we can now define

$$\delta_{\mathcal{A}}(\varepsilon)=\frac{1}{4}p^{-d\Psi_A(C_{9.4.3}(\Delta,\eta,\rho,\zeta,R))C_{9.4.3}(\Delta,\eta,\rho,\zeta,R)}\cdot\left(\frac{\varepsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C_{9.4.3}(\Delta,\eta,\rho,\zeta,R),d)} \qquad (14.13)$$

to conclude the proof.

Chapter 15

Degree structural properties

Theorem 13.2.3 shows that every locally characterizable property is proximity-oblivious testable with one-sided error. The condition of being locally characterizable is quite general, and as a result, we expect that there are many such interesting algebraic properties. This, in fact, turns out to be the case. We show that a class of properties that we call *degree-structural* are all locally characterized and are, hence, proximity-obliviously testable by Theorem 13.2.3.

Before giving the formal definition, let us first list some examples of degree-structural properties. Let d be a fixed positive integer, $\mathbb{F} = \mathbb{F}_p$ a fixed prime finite field. Each of the following definitions defines a degree-structural property.

- **Degree $\leq d$:** All polynomials $F : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d$;
- **Product of linear forms:** Polynomials $F : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d$ which are the product of at most d linear functions;
- **Composite polynomials:** Polynomials $F : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d$ which factor as $F = GH$ where G, H are polynomials of degree $1 \leq \deg(G), \deg(H) \leq d - 1$;
- **Having square root:** Polynomials $F : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d$ which can be factored as $F = G^2$ for a polynomial G of degree $\leq d/2$;
- **Sum of two products:** Polynomials $F : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d$ which can be decomposed as $F = G_1G_2 + G_3G_4$, where $1 \leq \deg(G_i) \leq d - 1$ for all $i \in \{1, 2, 3, 4\}$;
- **Low d-rank:** For fixed integers $r \geq 1, R \geq 2$, a function $F : \mathbb{F}^n \to [R]$ *has d-rank at most r* if there exist polynomials $G_1, \ldots, G_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d - 1$ and a function $\Gamma : \mathbb{F}^r \to [R]$ such that $F = \Gamma(G_1, \ldots, G_r)$.

In fact, roughly speaking, any property that can be described as the property of being decomposable into a known structure of low-degree polynomials is degree-structural. The following definition formalizes this notion.

Definition 15.0.1 (Degree-structural property)**.** *Let $\mathbb{F} = \mathbb{F}_p$ be a prime finite field and let $R \geq 2$. Given an integer $c \geq 1$, a vector of non-negative integers $\mathbf{d} = (d_1, \ldots, d_c) \in \mathbb{Z}^c_{\geq 0}$, and a function $\Gamma : \mathbb{F}^c \to [R]$, define the* $(c, \mathbf{d}, \Gamma)$-structured property *to be the collection of functions $F : \mathbb{F}^n \to [R]$ for which there exist polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ of degrees $\deg(P_i) \leq d_i$ for all $i \in [c]$, such that $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ for all $x \in \mathbb{F}^n$.*

We say a property $\mathcal{P} \subseteq \{F : \mathbb{F}^n \to [R]\}$ is degree-structural *if there exist integers $c, d \geq 1$ and a set of tuples $S \subset \{(c, \mathbf{d}, \Gamma) \mid \mathbf{d} \in [0, d]^c, \Gamma : \mathbb{F}^c \to [R]\}$, such that a function $F \in \mathcal{P}$ if and only if F is $(c, \mathbf{d}, \Gamma)$-structured for some $(c, \mathbf{d}, \Gamma) \in S$. We call R the* range*, c the* scope *and d the* max-degree *of the degree-structural property $\mathcal{P}$.*

Remark 15.0.2. *We could allow for an even more general definition of degree-structural properties, by allowing the polynomials P_i above to be nonclassical with some bound on the depth. Everything below can be extended to this setting as well.*

It is straightforward to see that the examples described above all satisfy this definition. By definition, any degree structural property is affine invariant. In this section, we present a result from [13] that shows that any degree-structural property with bounded scope and max-degree has a local characterization.

Theorem 15.0.3. *Fix a prime finite field $\mathbb{F}$. Every degree-structural property $\mathcal{P}$ with range R, scope c and max-degree d is q-locally characterized for some $q = q(\mathbb{F}, R, c, d)$.*

An immediate corollary of Theorem 15.0.3 and Theorem 13.2.3 is that any degree-structural property is proximity-oblivious locally testable.

Corollary 15.0.4. *Fix a prime finite field $\mathbb{F}$. Every degree-structural property $\mathcal{P}$ with range R, scope c and max-degree d is q-locally testable for some $q = q(\mathbb{F}, R, c, d)$.*

The proof of Theorem 15.0.3 will rely on Theorem 7.6.1. In the original paper [13], the proof utilized a special version of Theorem 7.6.1 for affine constraints, as these are the systems of linear forms which arise in affine invariant property testing. Here, we present a slightly different version, which avoids the need to introduce this specialized form of Theorem 7.6.1.

15.1 Proof of Theorem 15.0.3

We proceed to give the proof of Theorem 15.0.3. The proof would utilize the following lemma due to [11], which shows that locally characterized properties cannot distinguish between polynomials of high rank, as long as they have the same degree and depth.

Lemma 15.1.1. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime finite field, and fix $d, c, q \geq 1$. Let $\mathcal{P}$ be a q-locally characterized property of functions $\mathbb{F}^n \to [R]$. Let $\mathcal{B}$ be a polynomial factor of rank $> r_{15.1.1}(\mathbb{F}, d, c, q)$ defined by homogeneous nonclassical polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ of degrees $\leq d$, and define $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ for some function $\Gamma : \mathbb{T}^c \to [R]$.*

Assume that $F \in \mathcal{P}$. Then, for any homogeneous polynomials $Q_1, \ldots, Q_c : \mathbb{F}^n \to \mathbb{T}$ for which $\deg(Q_i) = \deg(P_i)$ and $\mathrm{depth}(Q_i) = \mathrm{depth}(P_i)$, it holds that the function $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$ is also in $\mathcal{P}$.

Proof. Assume that $G \notin \mathcal{P}$. As $\mathcal{P}$ is q-locally characterized, there must be q points $x_1, \ldots, x_q \in \mathbb{F}^n$ such that the values of $G(x_1), \ldots, G(x_q)$ witness the fact that $G \notin \mathcal{P}$. We will show that there exists points $x'_1, \ldots, x'_q \in \mathbb{F}^n$ which witness the fact that $F \notin \mathcal{P}$, a contradiction.

Let e denote the dimension of the linear subspace spanned by $x_1, \ldots, x_q$, and assume without loss of generality that $x_1, \ldots, x_e$ are linearly independent. Let $L_1, \ldots, L_q : \mathbb{F}^e \to \mathbb{F}$ be linear forms such that $x_i = L_i(x_1, \ldots, x_e)$. For every Q_i, we have that $Q_i(L_1(x_1, \ldots, x_e)), \ldots, Q_i(L_q(x_1, \ldots, x_e))$ are (d_i, k_i)-consistent with $L_1, \ldots, L_m$, where $d_i = \deg(P_i) = \deg(Q_i)$ and $k_i = \mathrm{depth}(P_i) = \mathrm{depth}(Q_i)$ (see Theorem 7.7.4). Theorem 7.7.5 then implies that, as long as $\mathcal{B}$ has high enough rank (concretely, rank at least $r_{7.2.2}(\mathbb{F}, d, \varepsilon)$ for $\varepsilon = 1/2|\mathcal{B}|^m$) then $((P_i(L_j(X))) : X \in (\mathbb{F}^n)^e)$ attain all values in $\mathbb{T}^{mc}$ which are $(\mathbf{d}, \mathbf{k})$-consistent with $L_1, \ldots, L_m$, where $\mathbf{d} = (d_1, \ldots, d_c)$ and $\mathbf{k} = (k_1, \ldots, k_c)$. In particular, there are some $x'_1, \ldots, x'_e \in \mathbb{F}^n$ such that $P_i(L_j(x'_1, \ldots, x'_e)) = Q_i(L_j(x_1, \ldots, x_e))$ for all $i \in [c], j \in [m]$. This implies $F(x_i) = G(x_i)$ for all $i \in [q]$, which in turn shows $F \notin \mathcal{P}$. □

As a specific corollary, we obtain that under the definitions of Theorem 15.1.1, if $F : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree D then $\deg(G) \leq D$.

Corollary 15.1.2. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime finite field, and fix $c, d \geq 1$. Let $\mathcal{B}$ be a polynomial factor of rank $> r_{15.1.2}(\mathbb{F}, d, c)$ defined by homogeneous nonclassical polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ of degrees $\leq d$, and define $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ for some function $\Gamma : \mathbb{T}^c \to \mathbb{F}$.*

Assume that $\deg(F) = D$. *Then, for any homogeneous polynomials* $Q_1, \ldots, Q_c : \mathbb{F}^n \to \mathbb{T}$ *for which* $\deg(Q_i) = \deg(P_i)$ *and* $\mathrm{depth}(Q_i) = \mathrm{depth}(P_i)$, *it holds that the function* $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$ *has* $\deg(G) \leq D$.

Proof. The corollary follows as the property of being a polynomial of degree $\leq D$ is q-locally characterized for $q \leq 2^D$. Moreover, if $\deg(F) = D$ and F is a function of c nonclassical polynomials of degree $\leq d$, then $D \leq dcp^{\lceil d/(p-1) \rceil}$. □

Proof of Theorem 15.0.3. Let $\mathcal{P}$ be a degree-structural property with range R, scope c and max-degree d. Denote by S the set of tuples $(c, \mathbf{d}, \Gamma)$ such that $\mathcal{P}$ is the union over all $(c, \mathbf{d}, \Gamma) \in S$ of $(c, \mathbf{d}, \Gamma)$-structured functions. It is clear that $\mathcal{P}$ is affine-invariant, as having degree bounded by some d_i is an affine-invariant property. It is also immediate that $\mathcal{P}$ is closed under taking restrictions to subspaces, since if F is $(c, \mathbf{d}, \Gamma)$-structured, then F restricted to any hyperplane is also $(c, \mathbf{d}, \Gamma)$-structured. The non-trivial part of the theorem is to show that the locality dimension is bounded. In other words, we need to show that there is a constant $K = K(\mathbb{F}, R, c, d)$, such that for $n \geq K$, if $F : \mathbb{F}^n \to [R]$ is a function with $F|_H \in \mathcal{P}$ for every K-dimensional subspace $H \leq \mathbb{F}^n$, then this implies that $F \in \mathcal{P}$. To do that, it suffices (by induction) to prove a weaker statement: for $n \geq K$, if $F : \mathbb{F}^n \to [R]$ is a function with $F|_H \in \mathcal{P}$ for every hyperplane (namely, an $(n-1)$-dimensional subspace) $H \leq \mathbb{F}^n$, then this implies that $F \in \mathcal{P}$.

For each $t \in [R]$ let $F_t : \mathbb{F}^n \to \{0, 1\} \subset \mathbb{F}$ be given by $F_t(x) = 1_{F(x)=t}$. The first step will be bound the degrees of F_t. Recall that we assume that $F|_H \in \mathcal{P}$ for every hyperplane H. Thus

$$F|_H(x) = \Gamma_H(P_{H,1}(x), \ldots, P_{H,c}(x))$$

for some polynomials $P_{H,i}$ of degrees $\deg(P_{H,i}) \leq d_i \leq d$ and some $\Gamma_H : \mathbb{F}^c \to [R]$. This implies that $(F_t)|_H(x) = \Gamma_{H,t}(P_{H,1}(x), \ldots, P_{H_c}(x))$ where $\Gamma_{H,t}(z) = 1$ if $\Gamma_H(z) = t$ and $\Gamma_{H,t}(z) = 0$ otherwise. In particular, $\deg((F_t)|_H) \leq D := |\mathbb{F}|cd$ for all $t \in [R]$. As the property of being low degree is locally characterized, by making sure that $K > D$ we obtain that also $\deg(F_t) \leq D$.

Let $r_1 : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a function to be determined later. Define $r_2 : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ so that $r_2(\cdot) > r_1(C_{7.5.5}^{(\mathbb{F},r,D)}(\cdot + c)) + C_{7.5.5}^{(\mathbb{F},r,D)}(\cdot + c) + |\mathbb{F}|$. Apply Theorem 7.5.5 to the polynomial factor $\{F_1, \ldots, F_R\}$ to obtain an r_2-regular polynomial factor $\mathcal{B}$ of degree $\leq D$, defined by homogeneous nonclassical polynomials $R_1, \ldots, R_C : \mathbb{F}^n \to \mathbb{T}$, where $C \leq C_{7.5.5}^{\mathbb{F},r_2,d}(R)$. Since each F_t is measurable with respect to $\mathcal{B}$, so is F. So, there exists a function $\Sigma : \mathbb{T}^C \to [R]$ such that

$$F(x) = \Sigma(R_1(x), \ldots, R_C(x)).$$

From each R_i pick a monomial with degree equal to $\deg(R_i)$ and a monomial (possibly the same one) with depth equal to $\mathrm{depth}(R_i)$. By taking K to be sufficiently large, and as we assume that $n > K$, we can gaurantee the existence of an $i_0 \in [n]$ such that x_{i_0} is not involved in any of these monomials. Let $R'_1, \ldots, R'_C$ be the restrictions of $R_1, \ldots, R_C$, respectively, to the hyperplane $H = \{x_{i_0} = 0\}$. By the choice of i_0 we have that $\deg(R'_i) = \deg(R_i)$ and $\mathrm{depth}(R'_i) = \mathrm{depth}(R_i)$ for all $i \in [C]$, Also, by Theorem 7.0.2, $R'_1, \ldots, R'_C$ have rank $> r_2(C) - p$. Since $F|_{x_{i_0}=0} \in \mathcal{P}$ by our assumption, by definition of $\mathcal{P}$, there must exist $(c, \mathbf{d}, \Gamma) \in S$ such that

$$\Sigma(R'_1(x), \ldots, R'_C(x)) = \Gamma(P_1(x), \ldots, P_c(x)),$$

where $\deg(P_i) \leq d_i$ for all $i \in [c]$.

Next, apply Theorem 7.5.5 again to find an r_1-regular refinement of the factor defined by the tuple of polynomials $(R'_1, \ldots, R'_C, P_1, \ldots, P_c)$. Because of our choice of r_2 and the last part of Theorem 7.5.5, we obtain a syntactic refinement of $\{R'_1, \ldots, R'_C\}$. That is, we obtain a polynomial factor $\mathcal{B}'$ defined by homogeneous nonclassical polynomials $R'_1, \ldots, R'_C, S_1, \ldots, S_E : \mathbb{F}^n \to \mathbb{T}$ such that it has degree $\leq D$, rank $> r_1(C + E)$ where $C + E \leq C_{7.5.5}^{\mathbb{F},d,r_1}(C + \sigma)$. In particular, for each $i \in [c]$ we have

$$P_i(x) = \Gamma_i(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_E(x))$$

for some function $\Gamma_i : \mathbb{T}^{C+E} \to \mathbb{T}$. So for all $x \in \mathbb{F}^n$,

$$\Sigma(R'_1(x), \ldots, R'_C(x)) = \\ \Gamma(\Gamma_1(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_E(x)), \ldots, \Gamma_c(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_E(x))).$$

Applying Theorem 7.4.1, we see that if the rank of $\mathcal{B}'$ is $> r_{7.4.1}(\mathbb{F}, D, \varepsilon)$ where $\varepsilon > 0$ is sufficiently small (say $\varepsilon = \frac{1}{2\|\mathcal{B}'\|}$), then $(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_E(x))$ acquires every value in its range. Thus, we have the identity

$$\Sigma(a_1, \ldots, a_C) = \Gamma(\Gamma_1(a_1, \ldots, a_C, b_1, \ldots, b_E), \ldots, \Gamma_c(a_1, \ldots, a_C, b_1, \ldots, b_E)),$$

for every $a_i \in \mathbb{U}_{\mathrm{depth}(R'_i)+1}$ and $b_i \in \mathbb{U}_{\mathrm{depth}(S_i)+1}$. Hence, we can substitute R_i for R'_i and 0 for S_i in the above equation and still retain the identity

$$\begin{aligned} F(x) &= \Sigma(R_1(x), \ldots, R_C(x)) \\ &= \Gamma(\Gamma_1(R_1(x), \ldots, R_C(x), 0, \ldots, 0), \ldots, \Gamma_c(R_1(x), \ldots, R_C(x), 0, \ldots, 0)) \\ &= \Gamma(Q_1(x), \ldots, Q_c(x)) \end{aligned}$$

where $Q_i : \mathbb{F}^n \to \mathbb{T}$ are defined as $Q_i(x) = \Gamma_i(R_1(x), \ldots, R_C(x), 0, \ldots, 0)$. Since for every i, $\deg(R_i) = \deg(R'_i)$ and $\mathrm{depth}(R_i) = \mathrm{depth}(R'_i)$, we apply Theorem 15.1.2 to conclude that $\deg(Q_i) \leq \deg(P_i) \leq d_i$ for every $i \in [c]$, as long as the rank of $\mathcal{B}'$ is $> r_{15.1.2}(\mathbb{F}, d, C)$.

Finally, we argue that $Q_1, \ldots, Q_c$ are classical polynomials. Indeed, since $P_1, \ldots, P_c$ are classical polynomials, $\Gamma_1, \ldots, \Gamma_c$ must map to $\mathbb{U}_1$ on all of $\prod_{i=1}^{C} \mathbb{U}_{\mathrm{depth}(R'_i)+1} \times \prod_{i=1}^{E} \mathbb{U}_{\mathrm{depth}(S_i)+1} \supseteq \prod_{i=1}^{C} \mathbb{U}_{\mathrm{depth}(R_i)+1} \times \{0\}^E$. Hence, $Q_1, \ldots, Q_c$ take values in $\mathbb{U}_1$, and hence are classical polynomials. We conclude that $F \in \mathcal{P}$. $\square$

Chapter 16

Estimating the distance from algebraic properties

In Chapter 14 and Chapter 15 we discussed one-sided error testable affine-invariant properties. In this section we discuss two different settings, testability with two-sided error and parameter estimation. First let us define the notion of testability with two-sided error.

Definition 16.0.1 (Testability with two-sided error)**.** *A property $\mathcal{P} \subset \{\mathbb{F}^n \to [R]\}$ is said to be* testable *with two-sided error if there is a function $q : (0,1) \to \mathbb{N}$ and a randomized algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{F}^n \to [R]$, makes at most $q(\varepsilon)$ queries to the oracle for f, accepts with probability at least $2/3$ if $f \in \mathcal{P}$ and rejects with probability at least $2/3$ if f is ε-far from $\mathcal{P}$.*

Parameter estimation. A *function parameter* is any map π which maps a function $f : \mathbb{F}^n \to [R]$ to $\pi(f) \in [0,1]$. A function parameter is affine invariant if for any $f : \mathbb{F}^n \to [R]$ and any invertible affine map $A : \mathbb{F}^n \to \mathbb{F}^n$ it holds that $\pi(f) = \pi(f \circ A)$. An example of a function parameter is the distance from an affine invariant property. That is, if $\mathcal{P} \subset \{f : \mathbb{F}^n \to [R]\}$ is an affine invariant property, then $\pi(f) = \mathrm{dist}(f, \mathcal{P})$ is an affine-invariant function parameter.

Definition 16.0.2 (Parameter estimation)**.** *A function parameter π of functions $f : \mathbb{F}^n \to [R]$ is said to be* estimable *if there is a function $q : (0,1) \to \mathbb{N}$ and a randomized algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{F}^n \to [R]$, makes at most $q(\varepsilon)$ queries to the oracle for f, and outputs a value which is within ε of $\pi(f)$ with probability at least $2/3$.*

Hatami and Lovett [47] showed that the distance to any constant-query testable (with two-sided error) hereditary affine-invariant property is constant-query estimable.

Theorem 16.0.3 ([47])**.** *For every two-sided error testable hereditary affine-invariant property $\mathcal{P}$ the parameter* $\mathrm{dist}(f, \mathcal{P})$ *is estimable.*

We will present a sketch of the proof of Theorem 16.0.3. We would like to prove that given an oracle access to a function $f : \mathbb{F}^n \to [R]$ and an error parameter $\varepsilon > 0$, one can query the function f on $q(\varepsilon)$ points, and output an estimate of $\mathrm{dist}(f, \mathcal{P})$ that is, with probability $\geq 2/3$ (say), within ε of the correct value. The test we study is very natural. We show that there exists a constant $m = m(\mathcal{P}, \varepsilon)$ such that for a random affine subspace H of dimension m it holds that

$$\Pr_H \left[|\Pr[\mathrm{dist}(f, \mathcal{P})] - \Pr[\mathrm{dist}(f|_H, \mathcal{P})]| \leq \varepsilon \right] \geq 2/3.$$

Note that crucially, m is independent of n.

The proof of Theorem 16.0.3 combines higher-order Fourier analysis with the framework of Fischer and Newman [28] which obtained similar results for graph properties. At a high level, the approach for the graph

case and the affine-invariant case are similar. One applies a regularization process, which allows to represent a graph (or a function) by a small structure. Then, one argues that a large enough random sample of the graph or function should have a similar small structure representing it. Hence, properties of the main object can be approximated by properties of a large enough sample of it. Fischer and Newman [28] implemented this idea in the graph case. Adapting this to the algebraic case inevitably introduces some new challenges.

To simplify the presentation, we focus on the case $R = 2$ from now on. That is, we assume that all functions are $f : \mathbb{F}^n \to \{0,1\}$. Also, we fix the required success probability at $2/3$ (any other constant strictly less than 1 would also work). The following theorem formalizes the discussion above.

Theorem 16.0.4. *Let $\mathcal{P} \subset \{f : \mathbb{F}^n \to \{0,1\} : n \in \mathbb{N}\}$ be an affine-invariant hereditary property which is two-sided testable. Then, for any $\varepsilon > 0$, there exists a constant $m = m(\mathcal{P}, \varepsilon)$ such that the following holds. Let $H \subset \mathbb{F}^n$ be a uniformly chosen m-dimensional affine subspace. Then*

$$\Pr_H \left[|\Pr[\mathrm{dist}(f, \mathcal{P})] - \Pr[\mathrm{dist}(f|_H, \mathcal{P})]| \leq \varepsilon \right] \geq 2/3.$$

16.1 Proof sketch of Theorem 16.0.4

Let $\mathcal{P} \subset \{f : \mathbb{F}^n \to \{0,1\} : n \in \mathbb{N}\}$ be an affine-invariant hereditary property which is two-sided testable. Let $f : \mathbb{F}^n \to \{0,1\}$ be a function, and let $\tilde{f}$ be the restriction of the function to a random m-dimensional affine subspace $H \subset \mathbb{F}^n$. We show that, if $m = m(\mathcal{P}, \varepsilon)$ is chosen large enough, then

- **Completeness:** If $\mathrm{dist}(f, \mathcal{P}) \leq \delta$, then with high probability, $\mathrm{dist}(\tilde{f}, \mathcal{P}) \leq \delta + \varepsilon$.
- **Soundness:** If $\mathrm{dist}(f, \mathcal{P}) \geq \delta + \varepsilon$, then with high probability, $\mathrm{dist}(\tilde{f}, \mathcal{P}) \geq \delta$.

Let us first fix notations. Let $A : \mathbb{F}^m \to \mathbb{F}^n$ be a random full rank affine transformation. Note that $Af : \mathbb{F}^m \to \{0,1\}$ defined as $Af(x) = f(Ax)$ is the restriction of f to the affine subspace which is the image of A. Thus in the above discussion $\tilde{f} = Af$.

The proof of the completeness is simple. If $\delta(f, \mathcal{P}) \leq \delta$ then there exists $g \in \mathcal{P}$ for which $\delta(f, g) \leq \delta$. With high probability over a random restriction, the distance of Af and Ag is at most $\delta + o_m(1)$. This is true because (i) $\mathbb{E}_A \mathrm{dist}(Af, Ag) = \mathrm{dist}(f, g)$, as each point in $\mathbb{F}^n$ has equal probability to be in the image of A, and (ii) a random affine subspace is pairwise independent with regards to whether an element is contained in it. Thus, by Chebychev's inequality,

$$\Pr_A[|\mathrm{dist}(Af, Ag) - \mathrm{dist}(f, g)| \geq \varepsilon] \leq \frac{1}{\varepsilon^2 |\mathbb{F}|^m}.$$

Clearly, choosing m large enough guarantees that the probability is bounded by $1/10$ (say).

The main challenge (as in nearly all works in property testing) is to establish soundness. That is, we wish to show that if a function f is far from $\mathcal{P}$ then, with high probability, a random restriction of it is also far from $\mathcal{P}$ as well. The main idea is to show that if for a typical restriction Af is δ-close to a function $h : \mathbb{F}^m \to \{0,1\}$ which is in $\mathcal{P}$, then h can be "pulled back" to a function $g : \mathbb{F}^n \to \{0,1\}$ which is both about δ-close to f and also very close to $\mathcal{P}$. This will contradict our initial assumption that f is $(\delta + \varepsilon)$-far from $\mathcal{P}$. In order to do so we apply the machinery of higher order Fourier analysis. Our description in this overview subsection will hide various "cheats" but will present the correct general outline. For the full details we refer to the original paper [47].

First, we apply two-sided testability deduce that a restriction to a low-dimensional subspace can distinguish between $f \in \mathcal{P}$ and f which is ε-far from $\mathcal{P}$.

Claim 16.1.1. *For any $\varepsilon > 0$, there is $k = k(\mathcal{P}, \varepsilon)$ and a partition of the set of functions $\{\mathbb{F}^k \to \{0,1\}\}$ to two disjoint sets F_+, F_- such that the following holds. Let $B : \mathbb{F}^k \to \mathbb{F}^n$ be a random full rank affine transformation. Then for any $f : \mathbb{F}^n \to \{0,1\}$:*

- *If $f \in \mathcal{P}$ then $\Pr[Bf \in F_+] \geq 0.9$.*

- *If* $\mathrm{dist}(f, \mathcal{P}) \geq \varepsilon$ *then* $\Pr[Bf \in F_-] \geq 0.9$.

Proof. The assumption that $\mathcal{P}$ is two-sided locally testable means that for some $q = q(\varepsilon)$, there is a local test which uses q queries, that can distinguish with probability 0.9 between $f \in \mathcal{P}$ and f for which $\mathrm{dist}(f, \mathcal{P}) \geq \varepsilon$. As the property if affine-invariant, we may assume that the queries come from affine forms in some $k \leq q$ variables. In particular, all the queries are contained in an affine subspace of dimension k. The sets F_+, F_- are determined by which restrictions make the tester accept and which ones make it reject. □

Next, we apply the decomposition theorems discussed in Chapter 9. These allow us to decompose f to "structured" parts which we will study, and "pseudo-random" parts which do not affect the distribution of restrictions to k-dimensional subspaces. In order to do so, for a function $f : \mathbb{F}^n \to \{0,1\}$ define by $\mu_{f,k}$ the distribution of its restriction to k-dimensional subspaces. That is, for any $v : \mathbb{F}^k \to \{0,1\}$ let

$$\mu_{f,k}[v] = \Pr_B[Bf = v],$$

where to recall $B : \mathbb{F}^k \to \mathbb{F}^n$ is a random full rank affine transformation.

We will need to slightly generalize this definition to randomized functions. In our context, these can be modeled as $f : \mathbb{F}^n \to [0,1]$. Such a function describes a distribution over (deterministic) functions $f' : \mathbb{F}^n \to \{0,1\}$ as follows: for each $x \in \mathbb{F}^n$ independently, sample $f'(x) \in \{0,1\}$ so that $\mathbb{E}[f'(x)] = f(x)$. We extend the definition of $\mu_{f,k}$ to functions $f : \mathbb{F}^n \to [0,1]$ by setting

$$\mu_{f,k}[v] = \mathbb{E}_{f'}\left[\mu_{f',k}[v]\right].$$

Our definition implies that if two functions $f, g : \mathbb{F}^n \to [0,1]$ have distributions $\mu_{f,k}$ and $\mu_{g,k}$ close in statistical distance, then random restrictions to k-dimensional affine subspaces cannot distinguish f from g. This will be useful in the analysis of the soundness.

We next decompose our function $f : \mathbb{F}^n \to \{0,1\}$ based on the above intuition. Let d to be determined later. Theorem 9.1.1 gives a decomposition

$$f = f_1 + f_2$$

where $f_1 = \Gamma(P_1(x), \ldots, P_C(x)) : \mathbb{F}^n \to [0,1]$ for a high-rank polynomial factor $\{P_1, \ldots, P_C\}$ and where $\|f_2\|_{U^d}$ is small enough. In the actual proof one has to use the strong decomposition theorem (Theorem 9.3.1) into three parts $f = f_1 + f_2 + f_3$. However for the sake of sketching the proof one can ignore this technicality. For an appropriately chosen d ($d > |\mathbb{F}|^k$ suffices) we can then replace f with f_1 for the purposes of analyzing its restrictions to k-dimensional subspaces. That is,

$$\mu_{f,k} \approx \mu_{f_1,k}$$

where $f_1 : \mathbb{F}^n \to [0,1]$ and where closeness is in statistical distance. Thus, from now on we restrict our attention to f_1, namely the "structured part" of f.

The next step is to show that the same type of decomposition can be applied to the restriction Af of f, recalling that $A : \mathbb{F}^m \to \mathbb{F}^n$ is a random full rank affine transformation. We will later choose $m \gg k$ for the proof to work. Clearly, $Af = Af_1 + Af_2$. We next analyze the typical behavior of Af_1 and Af_2.

First, note that $Af_1 = \Gamma(Q_1(x), \ldots, Q_C(x))$ where $Q_i = AP_i$ are the restrictions of $P_1, \ldots, P_C$. One can show (and we omit the details in this proof sketch) that as $P_1, \ldots, P_C$ is a high-rank polynomial factor, then $Q_1, \ldots, Q_C$ is also a high-rank polynomial factor. Thus f_1 and Af_1 have the same "high level" factorization Γ to high-rank polynomials.

Next, we analyze Af_2. We claim that with high probability over the choice of A, if m is chosen large enough, then $\|Af_2\|_{U^d} \approx \|f_2\|_{U^d}$. This holds since

$$\mathbb{E}_A \|Af_2\|_{U^d}^{2^d} = \|f_2\|_{U^d}^{2^d} \pm O(2^d |\mathbb{F}|^{-m}).$$

So, Af and Af_1 also have similar distribution of their restrictions to random k-dimensional subspaces. That is,

$$\mu_{Af,k} \approx \mu_{Af_1,k}.$$

Next, assume that $\text{dist}(Af, \mathcal{P}) \leq \varepsilon$. By definition, there exists $h : \mathbb{F}^m \to \{0,1\}$, where $h \in \mathcal{P}$, such that $\text{dist}(Af, h) \leq \varepsilon$. We apply the same decomposition process to h. Thus, we decompose

$$h = h_1 + h_2,$$

where h_1 is structured and $\|h_2\|_{U^d}$ is very small. Thus as before, $\mu_{h,k} \approx \mu_{h_1,k}$.

By choosing the exact parameters of "high rank" for h_1 to be lower than these for Af_1, but still high enough (for exact details, see the original paper [47]), we may assume that the polynomials that appear in the decomposition of h_1 extend $Q_1, \ldots, Q_C$ which appear in the decomposition of Af_1. That is, we have

$$h_1(x) = \Gamma'(Q_1(x), \ldots, Q_{C'}(x))$$

for some $C' > C$, such that the entire set of polynomials $Q_1, \ldots, Q_{C'}$ is of high rank. The important aspect here is that the polynomials $Q_1, \ldots, Q_C$ which compose Af_1 are part of the description of h_1 (however, it may be the case that Γ' ignores them; we will see soon that this is impossible). To summarize: both Af_1 and h_1 can be defined in terms of the same basic "building blocks", namely high rank polynomials $Q_1, \ldots, Q_{C'}$.

The next step is to "pull back" h to a function $\phi : \mathbb{F}^n \to \{0,1\}$, such that ϕ is very close to $\mathcal{P}$, and such that $\text{dist}(f, \phi) \approx \text{dist}(Af, h)$. This will show that

$$\text{dist}(f, \mathcal{P}) \leq \text{dist}(f, \phi) + o_m(1) = \text{dist}(Af, h) + o_m(1) \leq \delta + o_m(1).$$

Setting $m = m(k, \varepsilon)$ large enough would show that $\text{dist}(f, \mathcal{P}) \leq \delta + \varepsilon$, which is our goal.

The first step is to pull back h_1. To recall, $h_1(x) = \Gamma'(Q_1(x), \ldots, Q_{C'}(x))$, and moreover $Q_i = AP_i$ for $1 \leq i \leq C$. So, for $C < i \leq C'$ we need to define pullback polynomials $P_i : \mathbb{F}^n \to \{0,1\}$ such that (i) $Q_i = AP_i$; and (ii) $P_1, \ldots, P_{C'}$ are of high rank. This can be done for example by letting $P_i = DQ_i$ for any affine map $D : \mathbb{F}^n \to \mathbb{F}^m$ for which AD is the identity map on $\mathbb{F}^m$. So, define $\phi_1 : \mathbb{F}^n \to [0,1]$ given by

$$\phi_1(x) = \Gamma'(P_1(x), \ldots, P_{C'}(x)).$$

Note that $A\phi_1 = h_1$; that is, ϕ_1 is the pull-back of the "structured part" h_1 of h. However, it does not in general generate a function close to f. This makes sense, as we still have not used the finer "pseudo-random" structure of h_2.

However, we can already show something about ϕ_1: it is very close to $\mathcal{P}$. More concretely, as $P_1, \ldots, P_{C'}$ are high rank polynomials, and also $Q_1, \ldots, Q_{C'}$ are high rank polynomials, we have

$$\mu_{\phi_1,k} \approx \mu_{h_1,k}.$$

Recall that we have shown $\mu_{h_1,k} \approx \mu_{h,k}$, and that $h \in \mathcal{P}$. Thus, the tester which distinguishes functions in $\mathcal{P}$ from those ε-far from $\mathcal{P}$ cannot distinguish ϕ_1 from functions in $\mathcal{P}$. Hence, ϕ_1 must be ε-close to $\mathcal{P}$.

The final step is to define the more refined pull-back ϕ of h. Define an atom as a subset $\{x \in \mathbb{F}^n : P_1(x) = a_1, \ldots, P_{C'}(x) = a_{C'}\}$ for values $a_1, \ldots, a_{C'}$. Note that the functions f_1, h_1 are constant over atoms. We next define $\phi : \mathbb{F}^n \to [0,1]$ by redefining ϕ_1 inside each atom, so that the average over each atom of ϕ and ϕ_1 is the same, but such that ϕ is as close as possible to f given this constraint. Concretely, we can consider three cases:

- An atom where the average of f over the atom equals the value ϕ_1 assigns to this atom. In this atom, we simply set $\phi(x) = f(x)$ for all points x in this atom.

- An atom where the average of f over the atom is larger than the value ϕ_1 assigns to this atom. In this atom, for any x, if $f(x) = 0$ then we set $\phi(x) = 0$, and if $f(x) = 1$ then we set $\phi(x) = \alpha$ where the value α is chosen so that the average of ϕ and ϕ_1 over the atom is the same.

- An atom where the average of f over the atom is lower than the value ϕ_1 assigns to this atom. This is analogous to the previous case.

One can show that under this choice, ϕ is indeed a proper pullback of h, in the sense that

$$\operatorname{dist}(f, \phi) = \operatorname{dist}(f_1, \phi_1) \approx \operatorname{dist}(Af, h_1) \approx \operatorname{dist}(Af, h).$$

Moreover,

$$\operatorname{dist}(\phi, \mathcal{P}) \approx \operatorname{dist}(\phi_1, \mathcal{P}) = o_m(1).$$

We thus conclude that $\operatorname{dist}(f, \mathcal{P}) \leq \operatorname{dist}(Af, h) + o_m(1) \leq \delta + o_m(1)$.

Part IV

Open Problems

Chapter 17

Open problems

We conclude this survey with a quick overview of the main open problems in this area. Several of these have already been mentioned in the text, and we repeat them here for completeness.

17.1 Testability of hereditary properties

A beautiful and fundamental result of Alon and Shapira [3] says that every hereditary graph property is testable. In [15] the analogue of Alon and Shapira's result is conjectured for algebraic properties.

Theorem 13.4.2 (restated). *Every affine invariant subspace hereditary property is testable with one-sided error.*

Resolving Conjecture 13.4.2 would yield a combinatorial *characterization* of the one-sided testable affine-invariant properties, similar to the characterization for testable dense graph properties [3]. In Theorem 13.4.4, we made major progress towards resolving Conjecture 13.4.2 by proving the testability under an additional assumption of "bounded complexity". The authors of the survey believe that it is likely that Conjecture 13.4.2 is in fact false, and some assumption on the complexity of the property is necessary.

17.2 Testing correlation with classical polynomials

As we saw in Section 6.2, estimating the Gowers norm $\|f\|_{U_{d+1}}$ can be used as a test for whether a function f has significant or negligible correlation with degree d nonclassical polynomials. We also saw in Chapter 5 that $d = 3$ is the first case where the Gowers norm is not a test for correlation with classical polynomials. The $d = 3$ still remains a fascinating open problem.

Problem 17.2.1. *Does there exist a tester which queries a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ on a constant number of positions, and which can distinguish whether f has noticeable or negligible correlation with classical cubic polynomials?*

More precisely, we would like a universal test with a constant number of queries for which the following holds. For every $\varepsilon > 0$, there exists a constant $0 < \delta(\varepsilon) < \varepsilon$, such that with probability at least $2/3$ the test correctly distinguishes between functions that have correlation at most $\delta(\varepsilon)$ with all classical cubic polynomials, and the ones that have correlation at least ε with some classical cubic polynomial.

As we saw in Chapter 16, a tester for which the number of queries depends on the error parameter exists and was given in [47], however the above problem is left open. See [48] for more results and discussion on the subject.

17.3 Quantitative bounds for inverse theorems

Recall that the inverse theorem for Gowers norms, Theorem 6.2.3, shows existence of $\varepsilon(\delta, d, \mathbb{F})$ such that the following holds. For every function $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leq 1$ and $\|f\|_{U^{d+1}} \geq \delta$, there exists a polynomial $P \in \mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$ of degree $\leq d$ that is ε-correlated with f, meaning

$$|\mathbb{E}_{x \in \mathbb{F}^n} f(x) e(-P(x))| \geq \varepsilon.$$

The asymptotics of the required dependence of ε on δ is not well understood and in fact it is a major open problem whether this dependence can be made polynomial, even in the special case of $d = 3$ and $|\mathbb{F}| = 2$.

Conjecture 17.3.1 (Polynomial Inverse Gowers conjecture for U^3)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. If $\left\|(-1)^f\right\|_{U^3} \geq \varepsilon$, then there exists a quadratic classical polynomial q such that $\Pr[f(x) = q(x)] \geq \frac{1}{2} + \delta$, where $\delta = c\varepsilon^c$ for some absolute constant $c > 0$.*

An evidence for the importance of understanding quantitative bounds for the inverse theorem is the work of Green-Tao [42] and Lovett [55], where it is proved that the above conjecture is equivalent to a very important conjecture in additive combinatorics, namely the well-known polynomial Freiman-Ruzsa conjecture. The best known lower-bound for δ is quasi-polynomial in ε which follows from the work of Sanders on the Bogolyubov-Ruzsa conjecture [67].

17.4 Complexity of linear forms

In Theorem 11.2.4 we proved that if $\mathcal{L} = \{L_1, \ldots, L_m\}$ is a system of linear forms for which $L_1^{d+1}, L_2^{d+1}, \ldots, L_m^{d+1}$ are linearly independent, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any collection of functions $f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{D}$ with $\|f_1\|_{U^{d+1}} \leq \delta$, we have

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[\prod_{i=1}^m f_i(L_i(X)) \right] \right| \leq \varepsilon.$$

Our proof of this fact goes through inverse theorem for Gowers norms and regularity lemmas for polynomials and as a result the dependence of δ on ε is really bad. It is left open whether this theorem can be proved with reasonable bounds on δ. We state an open problem due to Gowers and Wolf [38] that focuses on the polynomial regime.

Problem 17.4.1 (Problem 7.6 in [38], reformulated)**.** *Does there exist an integer $d \geq 1$ and a system of linear forms $L_1, \ldots, L_m$ with $L_1^{d+1}, \ldots, L_m^{d+1}$ linearly independent, such that the following holds: For every positive real number r, there exists $\varepsilon > 0$ and functions $f_i : \mathbb{F}^n \to \mathbb{C}$ such that $\|f_i\|_{U^d} \leq \varepsilon^r$ for every i, and yet*

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^d} \prod_{i=1}^m f_i(L_i(X)) \right| > \varepsilon.$$

17.5 Norms defined by linear forms.

A possibly challenging open problem is to give a full characterization of collections of linear forms $L_1, \ldots, L_m$ for which

$$\left| \mathbb{E}_{X \in \mathbb{F}^k} \prod_{i=1}^m f(L_i(X)) \right|^{1/m},$$

defines a norm on the space $\{f : \mathbb{F}^n \to \mathbb{R}\}$. Note that both the Gowers norm as well as the L_p norms for even values of p are of this type. In fact, one can show a characterization when $\mathbb{F}$ has characteristic larger than the true complexity of the linear forms. It turns out that in this case, all such norms are essentially either equivalent to L_p norm for some $p > 0$ or equivalent to the Gowers norm of some order d. However, the techniques do not seem to extend to the general question which is the interesting regime.

Bibliography

[1] Noga Alon and Richard Beigel. Lower bounds for approximations by low degree polynomials over z m. In *Computational Complexity, 16th Annual IEEE Conference on, 2001.*, pages 184–187. IEEE, 2001.

[2] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. Inform. Theory*, 51(11):4032–4039, 2005.

[3] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. on Comput.*, 37(6):1703–1727, 2008.

[4] Noga Alon and Asaf Shapira. Every monotone graph property is testable. *SIAM J. on Comput.*, 38(2):505–522, 2008.

[5] L. Babai, L. Fortnow, and C. Lund. Addendum to: "Nondeterministic exponential time has two-prover interactive protocols" [Comput. Complexity **1** (1991), no. 1, 3–40; MR1113533 (92h:68031)]. *Comput. Complexity*, 2(4):374, 1992.

[6] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd Annual ACM Symposium on the Theory of Computing*, pages 21–32, New York, 1991. ACM Press.

[7] Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[8] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos Kiwi, and Madhu Sudan. Linearity testing over characteristic two. *IEEE Trans. Inform. Theory*, 42(6):1781–1795, November 1996.

[9] 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(6):1539–1596, 2010.

[10] Arnab Bhattacharyya. Polynomial decompositions in polynomial time. Technical report, Electronic Colloquium on Computational Complexity (ECCC), 2014. `http://eccc.hpi-web.de/report/2014/018`.

[11] Arnab Bhattacharyya, Abhishek Bhowmick, and Chetan Gupta. On higher-order fourier analysis over non-prime fields. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 60. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[12] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. *Theory Comput.*, 7:75–99, 2011.

[13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, STOC '13, pages 429–436, New York, NY, USA, 2013. ACM.

[14] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1337–1355, 2013.

[15] Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A unified framework for testing linear-invariant properties. In *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 478–487, 2010.

[16] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1870–1889, 2015.

[17] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science FOCS 2010*, pages 488–497. IEEE Computer Soc., Los Alamitos, CA, 2010.

[18] Abhishek Bhowmick and Shachar Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *arXiv preprint arXiv:1506.02047*, 2015.

[19] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990)*, volume 47 (3), pages 549–595, 1993.

[20] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 41–51, Washington, DC, USA, 2007. IEEE Computer Society.

[21] Henri Cohen. *Number theory: Volume I: Tools and diophantine equations*, volume 239. Springer Science & Business Media, 2008.

[22] David Conlon, Jacob Fox, and Benny Sudakov. Hypergraph ramsey numbers. *Journal of the American Mathematical Society*, 23(1):247–266, 2010.

[23] P. Erdös and R. Rado. Combinatorial theorems on classifications of subsets of a given set. *Proceedings of the London mathematical Society*, 3(1):417–439, 1952.

[24] P Erdös and G Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470, 1935.

[25] Paul Erdös and Paul Turán. On Some Sequences of Integers. *J. London Math. Soc.*, S1-11(4):261, 1936.

[26] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[27] Eldar Fischer. The art of uninformed decisions: A primer to property testing. In G. Paun, G. Rozenberg, and A. Salomaa, editors, *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, volume 1, pages 229–264. World Scientific Publishing, 2004.

[28] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM J. Comput.*, 37(2):482–501 (electronic), 2007.

[29] Harry Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[30] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[31] Oded Goldreich and Tali Kaufman. Proximity oblivious testing and the role of invariances. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 579–592. Springer, Heidelberg, 2011.

[32] Oded Goldreich and Leonid A Levin. A hard-core predicate for all one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 25–32. ACM, 1989.

[33] Oded Goldreich and Dana Ron. On proximity-oblivious testing. *SIAM J. Comput.*, 40(2):534–566, 2011.

[34] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *Proc. 36th Annual International Conference on Automata, Languages, and Programming*, pages 500–512, 2009.

[35] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[36] W. T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010.

[37] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010.

[38] W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, 21(1):36–69, 2011.

[39] Ben Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[40] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3-norm. *Proc. Edin. Math. Soc.*, 51:73–153, 2008.

[41] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009.

[42] Ben Green and Terence Tao. An equivalence between inverse sumset theorems and inverse conjectures for the u 3 norm. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 149, pages 1–19. Cambridge Univ Press, 2010.

[43] Benjamin Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[44] Hamed Hatami, Pooya Hatami, and James Hirst. Limits of Boolean functions on $\mathbb{F}_p^n$. *Electron. J. Combin.*, 21(4):Paper 4.2, 15, 2014.

[45] Hamed Hatami, Pooya Hatami, and Shachar Lovett. General systems of linear forms: equidistribution and true complexity. *arXiv preprint arXiv:1403.7703*, 2014.

[46] Hamed Hatami and Shachar Lovett. Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms. *Geom. Funct. Anal.*, 21(6):1331–1357, 2011.

[47] Hamed Hatami and Shachar Lovett. Estimating the distance from testable affine-invariant properties. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 237–242. IEEE, 2013.

[48] Hamed Hatami and Shachar Lovett. Correlation testing for affine invariant properties on $\mathbb{F}_p^n$ in the high error regime. *SIAM Journal on Computing*, 43(4):1417–1455, 2014.

[49] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[50] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 423–432, 2004.

[51] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. *Foundations of Computer Science, IEEE Annual Symposium on*, 0:166–175, 2008.

[52] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. on Comput.*, 36(3):779–802, 2006.

[53] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 403–412, 2008.

[54] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.

[55] Shachar Lovett. Equivalence of polynomial conjectures in additive combinatorics. *Combinatorica*, 32(5):607–618, 2012.

[56] Shachar Lovett. *An Exposition of Sanders' Quasi-Polynomial Freiman-Ruzsa Theorem.* Number 6 in Graduate Surveys. Theory of Computing Library, 2015.

[57] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. *Theory Comput.*, 7:131–145, 2011.

[58] Edouard Lucas. Théorie des fonctions numériques simplement périodiques. *American Journal of Mathematics*, 1(2):184–196, 1878.

[59] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes.* Elsevier, 1977.

[60] Ryan O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014.

[61] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 76–85. IEEE, 2008.

[62] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2009.

[63] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[64] Ronitt Rubinfeld. Sublinear time algorithms. In *Proceedings of International Congress of Mathematicians 2006*, volume 3, pages 1095–1110, 2006.

[65] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[66] Alex Samorodnitsky. Low-degree tests at large distances. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 506–515. ACM, New York, 2007.

[67] Tom Sanders. On the Bogolyubov-Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[68] Tom Sanders. The structure theory of set addition revisited. *Bulletin of the American Mathematical Society*, 50:93–127, 2013.

[69] Asaf Shapira. Green's conjecture and testing linear-invariant properties. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 159–166. ACM, New York, 2009.

[70] Madhu Sudan. Invariance in property testing. Technical Report 10-051, Electronic Colloquium in Computational Complexity, March 2010.

[71] Balazs Szegedy. On higher order fourier analysis. *arXiv preprint arXiv:1203.2260*, 2012.

[72] E. Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.*, 20:89–104, 1969.

[73] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriui Vladimirovivc Linnik.

[74] Terence Tao. Structure and randomness in combinatorics. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 3–15. IEEE, 2007.

[75] Terence Tao. Some notes on non-classical polynomials in finite characteristic. *Blog post available at* `https://terrytao.wordpress.com/2008/11/13/some-notes-on-non-classical-polynomials-in-finite-characteristic`, 2008.

[76] Terence Tao. *Higher order Fourier analysis*, volume 142. American Mathematical Soc., 2012.

[77] Terence Tao and Van H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Paperback edition [of MR2289012].

[78] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Anal. PDE*, 3(1):1–20, 2010.

[79] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012.

[80] Emanuele Viola. Selected results in additive combinatorics: An exposition. *Theory of Computing, Graduate Surveys*, 2:1–15, 2011.

[81] Emmanuele Viola. The sum of D small-bias generators fools polynomials of degree D. *Computational Complexity*, 18(2):209–217, 2009.

[82] Yuichi Yoshida. A characterization of locally testable affine-invariant properties via decomposition theorems. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 154–163. ACM, 2014.

[83] Yuichi Yoshida. Gowers norm, function limits, and parameter estimation. *arXiv preprint arXiv:1410.5053*, 2014.

[84] Tamar Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97 (electronic), 2007.