ABSTRACT

Higher-order Fourier analysis is a powerful tool in the study of problems in additive and extremal combinatorics, for instance the study of arithmetic progressions in primes, where the traditional Fourier analysis comes short. In recent years, higher-order Fourier analysis has found multiple applications in computer science in fields such as property testing and coding theory. In this thesis, we develop new tools within this theory with several new applications such as a characterization theorem in algebraic property testing. One of our main contributions is a strong near-equidistribution result for regular collections of polynomials.

The densities of small linear structures in subsets of Abelian groups can be expressed as certain analytic averages involving linear forms. Higher-order Fourier analysis examines such averages by approximating the indicator function of a subset by a function of bounded number of polynomials. Then, to approximate the average, it suffices to know the joint distribution of the polynomials applied to the linear forms. We prove a near-equidistribution theorem that describes these distributions for the group $\mathbb{F}_p^n$ when $p$ is a fixed prime. This fundamental fact was previously known only under various extra assumptions about the linear forms or the field size. We use this near-equidistribution theorem to settle a conjecture of Gowers and Wolf on the true complexity of systems of linear forms.

Our next application is towards a characterization of testable algebraic properties. We prove that every locally characterized affine-invariant property of functions $f : \mathbb{F}_p^n \rightarrow R$ with $n \in \mathbb{N}$, is testable. In fact, we prove that any such property $P$ is proximity-obliviously testable. More generally, we show that any affine-invariant property that is closed under subspace restrictions and has “bounded complexity” is testable. We also prove that any property that can be described as the property of decomposing into a known structure of low-degree polynomials is locally characterized and is, hence, testable.

We discuss several notions of regularity which allow us to deduce algorithmic versions of various regularity lemmas for polynomials by Green and Tao and by Kaufman and Lovett.
We show that our algorithmic regularity lemmas for polynomials imply algorithmic versions of several results relying on regularity, such as decoding Reed-Muller codes beyond the list decoding radius (for certain structured errors), and prescribed polynomial decompositions.

Finally, motivated by the definition of Gowers norms, we investigate norms defined by different systems of linear forms. We give necessary conditions on the structure of systems of linear forms that define norms. We prove that such norms can be one of only two types, and assuming that $|\mathbb{F}_p|$ is sufficiently large, they essentially are equivalent to either a Gowers norm or $L_p$ norms.
To my family
ACKNOWLEDGMENTS

First and foremost, I would like to express many thanks to my advisor Alexander Razborov. It has been a great honor for me to be a student of Sasha. He is a great mentor, and he has always been there for me whenever I have needed advice, and has been of great help whenever I needed his supports. I would like to thank my second adviser and friend Madhur Tulsiani, for his friendship, support and his great ideas and interesting problems, and for always being available for sharing his insights and intuition about a problem or a paper. I am very thankful to Shachar Lovett who has been a great friend, collaborator, a mentor, and for his never-ending collection of interesting open problems and ideas. It was a great experience when he hosted my visits to UCSD for two summers, where I learned a lot from him while also enjoying the beautiful San Diego summers. I am also thankful to Arnab Bhattacharyya for many interesting and fruitful discussions and collaborations.

Many thanks to Laci Babai and Janos Simons for many invaluable discussions on theory, and sharing their stories on the theory community and history of the mathematical problems. I would like to thank Denis Pankratov, for being a great office-mate and friend during the past few years at University of Chicago. I want to thank Raghav Kulkarni, Sushant Sachdeva, Abhishek Bhowmick, Ankan Saha, Shubhendu Trivedi, Yuan Li, Joshua Grochow, Kaave Hosseini and Joshua Brody for many interesting conversations. A special thanks to my friends Rasool Zandvakil, Gregor Jarosch, Laura Pilossoph, Rodrigo Dufeo and James Edwards for making my time in Chicago most enjoyable. I want to thank Sarah, for her amazing patience and support during the process of me writing this thesis.

I would like to thank my brother, Hamed, for without him I would not be where I am. I am very grateful to him for introducing me to theoretical computer science years ago when we were both much younger, and for keeping the habit of sharing with me the new exciting theory that he explores. I dedicate this dissertation to my loving family, my mother Sedigheh, my father Ibrahim, and my brothers Hamed and Amir.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2  HIGHER-ORDER FOURIER ANALYSIS</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Polynomials over Finite Fields</td>
<td>10</td>
</tr>
<tr>
<td>2.1.1 Non-classical Polynomials</td>
<td>11</td>
</tr>
<tr>
<td>2.1.2 The Derivative Polynomial</td>
<td>13</td>
</tr>
<tr>
<td>2.1.3 Gowers Norms</td>
<td>14</td>
</tr>
<tr>
<td>2.2 Rank and Regularity</td>
<td>15</td>
</tr>
<tr>
<td>2.2.1 Polynomial Factors</td>
<td>16</td>
</tr>
<tr>
<td>2.2.2 Analytic Measures of Uniformity</td>
<td>19</td>
</tr>
<tr>
<td>2.2.3 Equidistribution and Near-Orthogonality of Regular Factors</td>
<td>21</td>
</tr>
<tr>
<td>2.2.4 Strong Near-Orthogonality and Equidistribution</td>
<td>22</td>
</tr>
<tr>
<td>2.2.5 Regularization of Factors</td>
<td>26</td>
</tr>
<tr>
<td>2.3 Decomposition Theorems</td>
<td>30</td>
</tr>
<tr>
<td>2.3.1 Strong Decomposition Theorems</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2 Higher-order Fourier Expansion</td>
<td>35</td>
</tr>
<tr>
<td>2.4 Systems of Linear forms</td>
<td>36</td>
</tr>
<tr>
<td>2.4.1 Cauchy-Schwarz Complexity</td>
<td>36</td>
</tr>
<tr>
<td>2.4.2 The True Complexity</td>
<td>37</td>
</tr>
<tr>
<td>3  HOMOGENEOUS NON-CLASSICAL POLYNOMIALS</td>
<td>39</td>
</tr>
<tr>
<td>3.1 Properties of the Derivative Polynomial</td>
<td>41</td>
</tr>
<tr>
<td>3.2 A Homogeneous Basis, the Univariate Case</td>
<td>42</td>
</tr>
<tr>
<td>3.3 A Homogeneous Basis, the Multivariate Case</td>
<td>44</td>
</tr>
<tr>
<td>4  DISTRIBUTION OF POLYNOMIAL FACTORS OVER LINEAR FORMS</td>
<td>46</td>
</tr>
<tr>
<td>4.1 Near-equidistribution over Linear Forms</td>
<td>48</td>
</tr>
<tr>
<td>4.2 Strong near-orthogonality: Proof of Theorem 2.2.23</td>
<td>50</td>
</tr>
<tr>
<td>5  TRUE COMPLEXITY, ON A THEOREM OF GOWERS AND WOLF</td>
<td>57</td>
</tr>
<tr>
<td>5.1 The Gowers-Wolf Conjecture: Proof of Theorem 5.0.7</td>
<td>58</td>
</tr>
<tr>
<td>6  TESTABLE AFFINE-INARIANT PROPERTIES</td>
<td>65</td>
</tr>
<tr>
<td>6.1 Background</td>
<td>66</td>
</tr>
<tr>
<td>6.1.1 Algebraic Property Testing, Affine Invariance</td>
<td>66</td>
</tr>
<tr>
<td>6.1.2 Locally Characterized Properties</td>
<td>66</td>
</tr>
<tr>
<td>6.2 Locality for Affine-Invariant Properties</td>
<td>68</td>
</tr>
<tr>
<td>6.3 Subspace Hereditary Properties</td>
<td>69</td>
</tr>
</tbody>
</table>
8.6.1 Theorem 8.5.4, the Direct Part ......................... 167
8.6.2 Theorem 8.5.4, the Inverse Part ....................... 168

9 CONCLUSION ......................................................... 171
CHAPTER 1
INTRODUCTION

Traditional Fourier analysis has been shown to be a powerful tool in the study of problems in additive and extremal combinatorics. However, in the recent years there has been several cases, such as the study of arithmetic progressions, where traditional Fourier analysis comes short. An extension of the traditional Fourier analysis, called higher-order Fourier analysis, was initiated by Gowers’ seminal work [35], where he introduced a sequence of norms for functions on Abelian groups and used them to prove quantitative bounds for Szemerédi’s theorem on arithmetic progressions [73]. Higher-order Fourier analysis has been shown to be very useful in analysing the density of linear patterns in mathematical objects, and has been extensively studied in several works [44, 45, 11, 78, 77, 68, 36, 38, 37, 39, 72, 71]. We refer the interested reader to the monograph by Tao [76].

In this dissertation we are interested in the case where the underlying group is $F^n$, where $F = F_p$ for a fixed prime $p$. The traditional Fourier analysis over $F^n$ allows one to express a given function as a linear combination of the characters of $F^n$. Each character of $F^n$ is a linear phase corresponding to an $\alpha \in F^n$, defined as $\chi_\alpha(x) := e_p(\sum_{i=1}^n \alpha_i x_i)$, where $e_p(m) := e^{2\pi m/p}$. In higher-order Fourier analysis, the linear polynomials are replaced by higher-degree polynomials, and one would like to express a function $f : F^n \to \mathbb{C}$ as a linear combination of such functions. Such a higher-order expansion can be achieved in an approximate sense via the so-called “decomposition theorems” which are proved by an application of the “inverse theorems” for uniformity norms, such as Gowers norms [36, 75, 43]. This allows one to express a given function $f$ in the form $f = f_1 + \sum_{i=1}^k g_i$ where $f_1$ is a linear combination of higher-degree phases and $g_i$’s are sufficiently small in an appropriate norm allowing us to ignore them as negligible error [36]. One can then use such a decomposition of $f$ in order to reduce the problem of estimating the density of a linear pattern within $f$ to the problem of understanding the distribution of a collection of polynomials over a given
linear pattern.

One of the main useful properties of Fourier analysis is that the set of characters form an orthonormal basis. However, we lose this orthogonality when we include higher degree polynomials; for example a nonzero linear combination of two distinct quadratic polynomials will always have a nonzero bias. Green and Tao [43], and Kaufman and Lovett [54] resolve this issue by introducing a notion of regularity for polynomials, where a regular polynomial is one that is “distributionally well-behaved”. These set of works provide an approximate orthogonality which is sufficient for understanding quantities of the form \( \mathbb{E}_{x \in \mathbb{F}^n}[f_1(x) \cdots f_m(x)] \). This is not completely satisfactory towards answering questions about the density of arbitrary linear patterns, namely \( \mathbb{E}_{X \in (\mathbb{F}^n)^k}[f_1(L_1(X)) \cdots f_m(L_m(X))] \), where \( L_i \)'s are arbitrary linear forms over \( k \) variables. A main contribution of this dissertation which resolves this issue, is a new near-orthogonality for regular polynomials over arbitrary collections of linear forms. This improves on previous more restrictive results such as [15] for affine systems of linear forms, and [51] for fields of sufficiently large characteristic. This strong near-orthogonality result allows us to settle a conjecture of Gowers and Wolf regarding the complexity of a system of linear forms.

Gowers uniformity norms are commonly used to control averages of linear patterns such as arithmetic progressions within subsets of a group. Since higher-order Gowers norms represent stronger notions of uniformity, a question arises naturally: Given a linear pattern, i.e. a collection of linear forms, what is the smallest \( k \) for which the Gowers \( U^k \) norm controls the density of such patterns? This question was asked and studied by Gowers and Wolf [37] who conjectured a simple characterization for this value, and in [38] verified it for the case of functions over \( \mathbb{F}^n \) in the case when \( |\mathbb{F}| \) is sufficiently large. We use our strong near-orthogonality result to settle the Gowers-Wolf conjecture in full generality over \( \mathbb{F}^n \).

The decomposition theorems discussed above can be thought of as arithmetic “regularity lemmas”, where we think of regularity as a notion of uniformity that allows one to decompose
a given object into a collection of simpler objects which appear random according to certain statistics. Szemerédi’s celebrated regularity lemma [74, 61], which has been a fundamental tool in combinatorics [73, 59], can be thought of as a decomposition theorem for bivariate functions, where a given function, corresponding to the edges of a given graph is decomposed into a structured part, a small error, and a uniform or pseudorandom part. In the past decade such decomposition theorems have been developed in much more general settings and under various stronger or weaker notions and parameters of uniformity [36, 75, 40, 27]. Decomposition theorems can be used to reduce questions regarding large and complex mathematical objects to that of bounded complexity objects while only suffering a small error in parameters. This general phenomenon has been used in proving several major results in mathematics [42, 36, 83], and the reduction to a lower complexity object under a controllable error has also been shown to be useful in computer science [3, 16, 19]. The development and application of decomposition theorems requires answers to several interesting questions regarding the notion of uniformity that is to be used, the kinds of error that can be tolerated, the relations between error, uniformity and structure parameters that are required, and whether such a decomposition theorem exists or can be obtained efficiently. Several such questions are discussed in this dissertation.

The sequence of Gowers uniformity norms have been shown to have numerous applications such as the study of linear structures such as arithmetic progressions in subsets of a group, or as a test for correlation with low-degree phases. The $d+1$’th Gowers norm of a function $f : \mathbb{F}^n \rightarrow \mathbb{C}$ is defined by picking a random $d$ dimensional parallelepiped and considering the average product of the function over all its points, more formally

$$\|f\|_{U^{d+1}} := \left[ \mathbb{E}_{X,Y_1,\ldots,Y_d \in \mathbb{F}^n} \prod_{S \subseteq [d]} C^{|S|} f(X + \sum_{i \in S} Y_i) \right]^{1/2^d},$$

where $C$ is complex conjugation. Motivated by the definition of Gowers norms one may
be interested in introducing similar norms defined by different systems of linear forms. We study such norms and give nice necessary conditions on the structure of systems of linear forms which define a norm. In particular, we prove that such norms can be one of two types. We then show that norms of the first type, for which the complexity of the system of the linear forms is smaller than $|\mathbb{F}|$, are essentially equivalent to a Gowers norm. We further show that norms of the second type are equivalent to $L_p$ norms. Our results are inspired by similar necessary conditions in the context of graph normings by Hamed Hatami [47, 48].

Motivated by turning the numerous applications of decomposition theorems and regularity lemmas into algorithms, it is natural to ask whether it is possible to achieve them algorithmically. In the case of Szemerédi’s regularity lemma, the original proof was non-algorithmic and the question of finding an algorithm for computing the regularity partition was first considered by Alon et.al. [1]. The Szemerédi regularity lemma is often used to prove existence of certain substructures, such as triangles [67], in large graphs, and with an algorithmic version of the lemma one can efficiently “find” these structures in a given graph. Since then, there have been numerous improvements and extensions to the algorithmic version of the Szemerédi regularity lemma, of which the works [5, 28, 58, 25] constitute a partial list (see [25] for a detailed discussion). The arithmetic regularity lemmas of Green and Tao [43] and Kaufman and Lovett [54] show that one can modify a given collection of polynomials into a new “pseudorandom” collection of polynomials. These lemmas have various applications in computer science, such as (special cases of) Reed-Muller testing and worst-case to average-case reductions for polynomials. However, the proofs of [43, 54], which involve a transformation from the first collection to the pseudorandom collection, are not algorithmic. In this dissertation we consider analytic notions of regularity for polynomials and show that they allow for efficient algorithms.

One main application of our algorithmic regularity lemma is in decoding of Reed-Muller codes. In the case when the characteristic of $\mathbb{F}$ is large, the Gowers norm gives an approximate
test for checking if for a given polynomial $P$, there exists a $Q$ of degree at most $k$ within Hamming distance $1 - \frac{1}{|F|} - \varepsilon$ [43]. This is remarkable because the list-decoding radius of Reed-Muller codes of order $k$ codes is less than $1 - \frac{k}{|F|}$ for $k < |F|$ [33], and the test works even beyond that. Moreover, Tao and Ziegler [11] later showed that this test works not only when $P$ is a polynomial of degree $d < |F|$ but also when $P$ is an arbitrary function. Given this, it is natural to ask for the decoding analogue:

Given $P$ of degree $d$ over $\mathbb{F}^n$, if there exists $Q$ of degree $k$ such that $\text{dist}(P, Q) \leq 1 - \frac{1}{|F|} - \varepsilon$, can one find a $Q'$ (in time polynomial in $n$) of degree $k$ such that $\text{dist}(P, Q') \leq 1 - \frac{1}{|F|} - \eta$ for some $\eta$ depending on $\varepsilon$?

We use our algorithmic regularity lemma and solve the above decoding question for Reed-Muller codes beyond their list-decoding radius: This special case can be interpreted as the scenario when the received word $P$ is obtained from some codeword $Q$ of degree $k$ by adding a “noise” $P - Q$ of degree $d$. Thus, when the noise is “structured”, we can decode in the discussed sense even beyond the list-decoding radius. This shows that although there can be exponentially many such $Q$ due to the fact that we are in the regime beyond the list-decoding radius (see [55, 19]), the question is still tractable if we ask only for one such $Q$ and if we allow a loss from $\varepsilon$ to $\eta$. Such a decoding question was solved for Reed-Muller codes of order 2 over $\mathbb{F}_2$, for any given function $f$ (instead of a polynomial $P$ of bounded degree) by [79, 10].

Lastly a key contribution of this dissertation is an application of higher-order Fourier analysis to give a characterization of testable algebraic properties. The field of property testing, as initiated by [20, 8], is the study of algorithms that query their input a very small number of times and with high probability decide correctly whether the input satisfies a given property or is “far” from satisfying that property. A property is called 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 a large number of natural properties satisfy this strong requirement; see e.g. the surveys [24, 65, 64, 70] for
a general overview. An example of such phenomenon is the famous result of Blum, Luby
and Rubinfeld [20] who showed that linearity of a function $f : \mathbb{F}^n \rightarrow \mathbb{F}$ is testable by a test
which makes only 3 random queries. Some of the earliest results in the field of property
testing are about the properties of functions on $\mathbb{F}^n$, where $p$ is a fixed prime. However, other
settings such as graph and hypergraph properties have witnessed more progress in the past,
where several characterizations for testable properties have been proved [3, 7, 6, 26, 22].
This is mainly due to the fact that the concepts of uniformity and related approximations
are somewhat simpler in those settings. In this thesis we use the recently developed tools
in higher-order Fourier analysis in order to extend some of the results from the graph and
hypergraphs setting to the setting of functions on $\mathbb{F}^n$.

Specifically, we give a complete characterization of the so called “proximity oblivious”
testable properties that are invariant under affine transformations. We consider properties
of functions $f : \mathbb{F}^n \rightarrow \{1, \ldots, R\}$. Our main result shows that any such property that
is invariant with respect to affine transformations on $\mathbb{F}^n$ and that is locally characterized
is testable. Furthermore, we show that a large class of natural algebraic properties whose
query complexity had not been previously studied are locally characterized affine-invariant
properties and are, hence, testable. Our results are in the one-sided regime, where a tester
is required to accept any function that belongs to the studied property.

Recently, there has been several other developments in the study of affine-invariant prop-
erties; see [12] for a summary. Hatami and Lovett [52] proved that higher-order Fourier
analysis combined with ideas from [26] can be used to show that the distance to every
testable affine invariant property is constant-query estimable. This extends the ideas of [26]
to the algebraic property testing setting. In the two-sided error regime, Yoshida [81] uses
the decomposition theorems in order to give a full characterization of two-sided testable
affine invariant properties, analogous to the work of Alon et.al. [3] in the context of graphs.
In a more recent work, Yoshida [82] uses non-standard analysis in order to define limits of
functions sequences, which allows for a simpler characterization of two-sided testable affine-invariant properties. An approach to defining limit objects for Boolean functions over $\mathbb{F}^n$ using finitary higher-order Fourier analysis can be seen in [49].
In Chapter 2 we introduce higher-order Fourier analysis in detail and develop some of the essential tools that we need throughout this thesis. Notions of regularity and their associated norms and decomposition theorems as well as distributional properties of collections of polynomials over linear forms are discussed. In Chapter 3 we prove that there exists a basis for “non-classical” polynomials consisting only of “homogeneous non-classical polynomials”. Later we use this theorem in Chapter 4 to prove our main theorem which is a near-equidistribution theorem for collections of polynomials over arbitrary collections of linear forms. We use our near-equidistribution theorem to prove a conjecture of Gowers and Wolf regarding a complexity measure of systems of linear forms in Chapter 5. Chapter 6 is concerned with the field of algebraic property testing, where we prove a characterization theorem for affine invariant testable properties. Finally, in Chapter 8 we investigate norms that are defined by averaging over a system of linear forms, where we give necessary conditions on the structure of such systems of linear forms and use it to relate to $L_p$ or Gowers norms.
Fix a prime field $\mathbb{F} = \mathbb{F}_p$ for a prime $p \geq 2$. Define $|\cdot|$ to be the standard map from $\mathbb{F}$ to $\{0,1,\ldots,p-1\} \subset \mathbb{Z}$. Let $\mathbb{D}$ denote the complex unit disk $\{ z \in \mathbb{C} : |z| \leq 1 \}$, and let $\mathbb{T}$ denote the circle group $\mathbb{R}/\mathbb{Z}$, which is an Abelian group with group operation denoted $+$. Let $e : \mathbb{T} \to \mathbb{C}$ denote the character $e(x) = e^{2\pi i x}$, by abuse of notation, we may use $e : \mathbb{F} \to \mathbb{C}$ to denote $e(x) = e^{2\pi i x}/|\mathbb{F}|$ as well.

For integers $a, b$, we let $[a]$ denote the set $\{1, 2, \ldots, a\}$ and $[a, b]$ denote the set $\{a, a + 1, \ldots, b\}$. For real numbers, $\alpha, \sigma, \varepsilon$, we use the shorthand $\sigma = \alpha \pm \varepsilon$ to denote $\alpha - \varepsilon \leq \sigma \leq \alpha + \varepsilon$. The zero element in $\mathbb{F}^n$ is denoted by $0$. We will denote by lower case letters, e.g. $x, y$, elements of $\mathbb{F}^n$. We use capital letters, e.g. $X = (x_1, \ldots, x_k) \in (\mathbb{F}^n)^k$, to denote tuples of variables.
CHAPTER 2
HIGHER-ORDER FOURIER ANALYSIS

In this chapter we introduce higher-order Fourier analysis in detail, and prove several technical tools that will be required through this dissertation. We discuss several properties of polynomials over finite fields, and extensions of polynomials that are required when the field size is small. Notions of regularity for collections of polynomials and norms that capture these notions of regularity. Algebraic decomposition theorems associated with uniformity norms are presented, which allow to decompose a given function to a structured part given by a collection of polynomials and a pseudorandom part. We then show how one can use such a decomposition to write a higher-order expansion of the given function. We then discuss the distribution polynomials over collections of linear forms. This is useful because the density of a linear pattern in an arbitrary function can be related to the distribution of a regular collection of polynomials over a collection of linear forms via a decomposition theorem.

2.1 Polynomials over Finite Fields

Let $d \geq 0$ be an integer. A polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree $< d$ can be defined in one of two equivalent ways:

- **(Global Definition)** $P$ is a polynomial of degree $< d$ if and only if it can be written in the form

  $$P(x_1, ..., x_n) = \sum_{0 \leq i_1, ..., i_n; \atop i_1 + \cdots + i_n < d} c_{i_1, ..., i_n} x_1^{i_1} \cdots x_n^{i_n},$$

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

- **(Local Definition)** $P$ is a polynomial of degree $< d$ if it is $d$ times differentiable and its $d$-th derivative vanishes.
It is not difficult to check that the above two definitions are equivalent, using the Taylor series expansion to go from the local to the global definition. In finite characteristic, i.e. when $P : \mathbb{F}^n \to G$, the local definition of a polynomial used the notion of additive derivatives.

**Definition 2.1.1** (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 $P$ at $x$ and direction $y$. The degree of $P$ is the smallest $d$ for which the above holds.

It follows simply from the definition that for any direction $y \in \mathbb{F}^n$, $\text{deg}(D_y P) < \text{deg}(P)$. 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 way to suspect this is the fact that we are not able to divide by $d!$ in some cases, in order to be able to make use of Taylor expansions. For example when the range of $P$ is $\mathbb{R}/\mathbb{Z}$, it turns out that the global definition must be refined to the “non-classical polynomials”, which may have monomials that are different from the classical case. This phenomenon was noted by Tao and Ziegler [78] in the study of Gowers norms.

### 2.1.1 Non-classical Polynomials

**Definition 2.1.2** (Non-classical Polynomials (local definition)). For an integer $d \geq 0$, a function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a non-classical 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

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

(2.1)
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 non-classical polynomial of degree $\leq d$ whose image is contained in $\frac{1}{p} \mathbb{Z}/\mathbb{Z}$.

The following lemma of Tao and Ziegler [78] 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 standard map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. This lemma also characterizes the structure of non-classical polynomials.

**Lemma 2.1.3** (Lemma 1.7 in [78]). 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

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

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 $\alpha$ 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 $\mathbb{U}_{k+1} \overset{\text{def}}{=} \frac{1}{p^{k+1}} \mathbb{Z}/\mathbb{Z}$. Classical polynomials correspond to polynomials with 0 shift and 0 depth.

Note that Lemma 2.1.3 immediately implies the following important observation\(^1\):

**Remark 2.1.4.** If $Q : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $d$ and depth $k$, then $pQ$ is a polynomial of 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. Also, if $\lambda \in [1, p-1]$ is an integer, then $\deg(\lambda Q) = d$ and $\depth(\lambda Q) = k$.

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 thesis. Hence, all polynomials of depth $k$ take values in $\mathbb{U}_{k+1}$.

---

1. Recall that $\mathbb{T}$ is an additive group. If $n \in \mathbb{Z}$ and $x \in \mathbb{T}$, then $nx$ is shorthand for $x + \cdots + x$ if $n \geq 0$ and $-x - \cdots - x$ otherwise, where there are $|n|$ terms in both expressions.
2.1.2 The Derivative Polynomial

Given a non-classical polynomial $P$ of exact degree $d$, it is often useful to consider the properties of its $d$-th derivative. Motivated by this, we give the following definition.

**Definition 2.1.5** (Derivative Polynomial). Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-$d$ polynomial, possibly non-classical. Define the derivative polynomial $\partial P : (\mathbb{F}^n)^d \to \mathbb{T}$ by the following formula

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

where $h_1, \ldots, h_d \in \mathbb{F}^n$. Moreover for $k < d$ define

$$\partial_k P(x, h_1, \ldots, h_k) \overset{\text{def}}{=} D_{h_1} \cdots D_{h_k} P(x).$$

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

**Lemma 2.1.6.** Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-$d$ (non-classical) polynomial. Then the polynomial $\partial P(h_1, \ldots, h_d)$ is

(i) multilinear: $\partial P$ is additive in each $h_i$.

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

(iii) a classical nonzero polynomial of degree $d$.

(iv) homogeneous: all its monomials are of degree $d$.

Notice that by multilinear we mean additive in each direction $h_i$, which is not the usual use of the term “multilinear”.

**Proof.** The proof follows by the properties of the additive derivative $D_h$. Multilinearity of $\partial P$ follows from linearity of the additive derivative, namely for every function $Q$ and directions

---

2. Notice since $P$ is a degree $d$ polynomial, $D_{h_1} \cdots D_{h_d} P(x)$ does not depend on $x$ and thus we have the identity $\partial P(h_1, \ldots, h_d) = D_{h_1} \cdots D_{h_d} P(x)$ for any choice of $x \in \mathbb{F}^n$. 

13
$h_1, h_2$ we have the identity $D_{h_1 + h_2} Q(x) = D_{h_1} Q(x) + D_{h_2} Q(x + h_1)$. The invariance under permutations of $h_1, \ldots, h_d$ is a result of commutativity of the additive derivatives. Since $P$ is a degree-$d$ (non-classical) polynomial, $\partial P$ is nonzero by definition. Notice that since $D_0 Q \equiv 0$ for any function $Q$, we have $\partial P(h_1, \ldots, h_d) = 0$ if any of $h_i$ is equal to zero. Hence every monomial of $\partial P$ must depend on all $h_i$’s. The properties (iii) and (iv) now follow from this and the fact that $\deg(\partial P) \leq d$ and thus each monomial has exactly one variable from each $h_i$. □

2.1.3 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. For example, it is known through an inverse theorem for Gowers norms that a Gowers norm of a function can be used to test whether it is close to a polynomial of a given degree bound. Let us first define the multiplicative derivative of a complex valued function.

**Definition 2.1.7 (Multiplicative Derivative).** Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an element $h \in \mathbb{F}^n$, define the multiplicative derivative in direction $h$ of $f$ to be the function $\Delta_h f : \mathbb{F}^n \to \mathbb{C}$ satisfying $\Delta_h f(x) = f(x + h) f(x)$ for all $x \in \mathbb{F}^n$.

The *Gowers norm* of order $d$ for a function $f : \mathbb{F}^n \to \mathbb{C}$ is the expected multiplicative derivative of $f$ in $d$ random directions at a random point.

**Definition 2.1.8 (Gowers norm).** Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an integer $d \geq 1$, the Gowers norm of order $d$ for $f$ is given by

$$
\|f\|_{U^d} = \left| E_{y_1, \ldots, y_d, x} \left[ (\Delta_{y_1} \Delta_{y_2} \cdots \Delta_{y_d} f)(x) \right] \right|^{1/2^d}.
$$

Note that as $\|f\|_{U^1} = |E[f]|$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it is not difficult to show that $\| \cdot \|_{U^d}$ is indeed a norm, using the following lemma.
which was first proved in [35] as a key property of Gowers norms.

**Lemma 2.1.9** (Gowers Cauchy-Schwarz). Consider a family of functions \( f_S : \mathbb{F}^n \to \mathbb{C} \), where \( S \subseteq [d] \). Then

\[
\left| \mathbb{E}_{x,y_1,\ldots,y_d \in \mathbb{F}^n} \prod_{S \subseteq [d]} \mathcal{O}^{d-|S|} f_S(x + \sum_{i \in S} y_i) \right| \leq \prod_{S \subseteq [d]} \| f_S \|_{U^d}, \tag{2.2}
\]

It is a direct consequence of Definition 2.1.2 that a function \( f : \mathbb{F}^n \to \mathbb{C} \) with \( \| f \|_{\infty} \leq 1 \) satisfies \( \| f \|_{U^{d+1}} = 1 \) if and only if \( f = e(P) \) for a (non-classical) polynomial \( P : \mathbb{F}^n \to \mathbb{T} \) of degree \( \leq d \). The inverse theorem for Gowers norms is a robust version of this statement. We denote by \( \text{Poly}(\mathbb{F}^n \to \mathbb{T}) \) and \( \text{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T}) \), respectively, the set of all non-classical polynomials, and the ones of degree at most \( d \).

**Theorem 2.1.10** (Theorem 1.11 of [78]). Suppose \( \delta > 0 \), and \( d \geq 1 \) is an integer. There exists an \( \varepsilon = \varepsilon_{2.1.10}(\delta, d, \mathbb{F}) \) 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 \( P \in \text{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T}) \) of degree \( \leq d \) that is \( \varepsilon \)-correlated with \( f \), meaning

\[
\left| \mathbb{E}_{x \in \mathbb{F}^n} f(x)e(-P(x)) \right| \geq \varepsilon.
\]

### 2.2 Rank and Regularity

The rank of a polynomial is a notion of its complexity according to lower degree polynomials.

**Definition 2.2.1** (Rank of a polynomial). Given a polynomial \( P : \mathbb{F}^n \to \mathbb{T} \) and an integer \( d > 1 \), the \( d \)-rank of \( P \), denoted \( \text{rank}_d(P) \), 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 \( \infty \) 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 \( \text{rank}(P) \geq r \).

Note that for an integer \( \lambda \in [1, p-1] \), \( \text{rank}(P) = \text{rank}(\lambda P) \). 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 2.2.2.** 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 making an affine transformation, we can assume without loss of generality that \( A \) is the hyperplane \( \{ x_1 = 0 \} \). Let \( \pi : \mathbb{F}^n \to \mathbb{F}^{n-1} \) be the projection to \( A \). Let \( P'' = P - P' \circ \pi \). Clearly, \( P'' \) is zero on \( A \). For \( x \in \mathbb{F} \setminus \{0\} \), let \( h_x = (x, 0, \ldots, 0) \in \mathbb{F}^n \).

Note that \( D_{h_x} P'' \) is of degree \( \leq d - 1 \) and that \( (D_{h_x} P'')(y) = P''(y + h_x) \) for all \( y \in A \). Hence, for every \( x \in \mathbb{F} \setminus \{0\} \), \( P'' \) on \( h_x + A \) agrees with a polynomial \( Q_x \) of degree \( \leq d - 1 \).

So, for a function \( \Gamma : \mathbb{T}^{p+1} \to \mathbb{T} \), we can write \( P = \Gamma(\iota(x_1), P', Q_1, Q_2, \ldots, Q_{p-1}) \), where \( \iota(x_1), Q_1, \ldots, Q_{p-1} \) 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. \( \Box \)

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.

### 2.2.1 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 2.2.3** (Factors). If $X$ is a finite set then by a factor $\mathcal{B}$ we mean simply a partition of $X$ into finitely many pieces called atoms.

A function $f : X \to \mathbb{C}$ is called $\mathcal{B}$-measurable if it is constant on atoms of $\mathcal{B}$. For any function $f : X \to \mathbb{C}$, we may define the conditional expectation

$$\mathbb{E}[f|\mathcal{B}](x) = \mathbb{E}_{y \in \mathcal{B}(x)} [f(y)],$$

where $\mathcal{B}(x)$ is the unique atom in $\mathcal{B}$ that contains $x$. Note that $\mathbb{E}[f|\mathcal{B}]$ is $\mathcal{B}$-measurable.

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 \{ $x : (\phi_1(x), \ldots, \phi_C(x)) = (y_1, \ldots, y_C)$\} 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 2.2.4** (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 $\prod_{i=1}^C p^{k_i+1}$.

**Definition 2.2.5** (Conditional Expectation over Factors). Let $\mathcal{B}$ be a polynomial factor defined by $P_1, \ldots, P_C$. 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

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

Namely $\mathbb{E}[f|\mathcal{B}]$ is a constant on every atom of $\mathcal{B}$.
Following is a simple observation stating that $E[f|\mathcal{B}]$ has same correlation with any other function as $f$ does.

**Remark 2.2.6.** Let $f : \mathbb{F}^n \to \mathbb{C}$. Let $\mathcal{B}$ be a polynomial factor defined by polynomials $P_1, \ldots, P_C$. Let $g : \mathbb{F}^n \to \mathbb{C}$ be any $\mathcal{B}$-measurable function. Then

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

Finally, we define the rank of a factor similarly to that of a single polynomial.

**Definition 2.2.7** (Rank of a Factor). 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 have rank $r$ if $r$ is the least integer for which there exists $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \mod p^{k_1+1}, \ldots, \lambda_C \mod p^{k_C+1}) \neq 0^C$, such that $\operatorname{rank}_d(\sum_{i=1}^C \lambda_i P_i) \leq r$, where $d = \max_i \deg(\lambda_i P_i)$.

Given a polynomial factor $\mathcal{B}$ and a function $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, we say that $\mathcal{B}$ is $r$-regular if $\mathcal{B}$ is of rank larger than $r(|\mathcal{B}|)$.

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

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

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

Regular factors indeed do behave like a generic collection of polynomials, as we shall establish in a precise sense in Section 2.2.3. Thus, given any factor $\mathcal{B}$ that is not regular, it will often be useful to regularize $\mathcal{B}$, that is, find a refinement $\mathcal{B}'$ of $\mathcal{B}$ that is regular up to our desires. We distinguish between two kinds of refinements:

Next, we define the notion of conditional expectation with respect to a given factor.

**Definition 2.2.8** (Expectation over polynomial factor). Given a factor $\mathcal{B}$ and a function $f : \mathbb{F}^n \to \{0, 1\}$, the expectation of $f$ over an atom $y \in \mathbb{T}^{\mathcal{B}}$ is the average $E_{x : \mathcal{B}(x) = y}[f(x)]$,
which we denote by $E[f|y]$. The conditional expectation of $f$ over $B$, is the real-valued function over $\mathbb{F}^n$ given by $E[f|B](x) = E[f|B(x)]$. In particular, it is constant on every atom of the polynomial factor.

### 2.2.2 Analytic Measures of Uniformity

Regularity defined by the notion of rank is an algebraic/combinatorial notion of pseudorandomness, however 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 for every nonzero linear combination of polynomials in the polynomial factor to be unbiased, where the bias of a function is defined as follows.

**Definition 2.2.9 (Bias).** The bias of a function $f : \mathbb{F}^n \to \mathbb{T}$ is defined to be

$$\text{bias}(f) := \mathbb{E}_{x \in \mathbb{F}^n} [e_T(f(x))].$$

Green and Tao [43], and Kaufman and Lovett [54] proved the following relation between bias and rank of a polynomial.

**Theorem 2.2.10** ($d < |\mathbb{F}|$ [43], arbitrary $\mathbb{F}$ [54]). For any $\varepsilon > 0$ and integer $d > 0$, there exists $r = r_{2.2.10}(d, \varepsilon)$ 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 2.2.10 for classical polynomials. But their proof also works for non-classical ones without modification.

Note that $r_{2.2.10}(\delta, d)$, does not depend on the dimension $n$ of $\mathbb{F}^n$. Motivated by Theorem 2.2.10 we define unbiasedness for polynomial factors.

**Definition 2.2.11 ($\gamma$-unbiased factors).** Let $\mathcal{F}$ be a factor of degree $d$, and let $\gamma : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function. We say that $\mathcal{F}$ is $\gamma$-unbiased if for every collection of coefficients
\{(c_{i,j} \in \mathbb{F})_{1 \leq i \leq d, 1 \leq j \leq M_i}\}, we have
\[
\text{bias}\left(\sum_{i,j} c_{i,j} P_{i,j}\right) \leq \gamma(\dim(\mathcal{F})).
\]

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

**Definition 2.2.12 (Uniformity).** Let \(\varepsilon > 0\) be a real. A degree-\(d\) polynomial \(P : \mathbb{F}^n \to \mathbb{T}\) is said to be \(\varepsilon\)-uniform if
\[
\|e(P)\|_{U^d} < \varepsilon.
\]

Tao and Ziegler used Theorem 2.2.10 to show that high rank polynomials have small Gowers norm.

**Theorem 2.2.13 (Theorem 1.20 of [78]).** For any \(\varepsilon > 0\) and integer \(d > 0\), there exists an integer \(r(d, \varepsilon)\) such that the following is true. For any (non-classical) polynomial \(P : \mathbb{F}^n \to \mathbb{T}\) of degree \(\leq d\), if \(\|e(P)\|_{U^d} \geq \varepsilon\), then \(\text{rank}_d(P) \leq r\).

This immediately implies that a regular polynomial is also uniform.

**Corollary 2.2.14.** Let \(\varepsilon, d,\) and \(r(d, \varepsilon)\) be as in Theorem 2.2.13. Every \(r\)-regular polynomial \(P\) of degree \(d\) is also \(\varepsilon\)-uniform.

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

**Claim 2.2.15.** For every \(r, d \in \mathbb{N}\), there is \(\varepsilon(r, d)\) such that every \(\varepsilon\)-uniform polynomial of degree \(d\) has \(\text{rank}(P) > r\).

The following is an analytical notion of uniformity for a factor along the same lines as Definition 2.2.12.
Definition 2.2.16 (Uniform Factor). Let $\gamma : \mathbb{N} \to \mathbb{R}^+$ 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 $\gamma$-uniform if for every collection $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \mod p^{k_1+1}, \ldots, \lambda_C \mod p^{k_C+1}) \neq 0^C$

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

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

Remark 2.2.17 (Equivalence between rank and uniformity). Similar to Corollary 2.2.14 it also follows from Theorem 2.2.13 that an $r$-regular degree-$d$ factor $\mathcal{B}$ is also $\varepsilon$-uniform when $r = r_{2.2.13}(d, \varepsilon)$ is as in Theorem 2.2.13. The converse of this also holds, since by Claim 2.2.15, and there is an approximate equivalence between regularity and uniformity.

2.2.3 Equidistribution and Near-Orthogonality of Regular Factors

In this section, we make precise the intuition that a high-rank collection of polynomials often behaves like 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 2.2.10).

Using a standard observation that relates the bias of a function to its distribution on its range, Theorem 2.2.10 implies the following.

Lemma 2.2.18 (Size of atoms). Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity $C$, and rank $r_{2.2.10}(d, \varepsilon)$, 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 $\varepsilon < \frac{1}{\|B\|}$, $B(x)$ attains every possible value in its range and thus has $\|B\|$ atoms.

Proof.

$$\Pr_{x}[B(x) = b] = \mathbb{E}_{x} \left[ \prod_{i} \frac{1}{p^{k_i + 1}} \sum_{\lambda_i = 0}^{p^{k_i + 1} - 1} e(\lambda_i(P_i(x) - b_i)) \right]$$

$$= \prod_{i} p^{-(k_i + 1)} \cdot \sum_{(\lambda_1, \ldots, \lambda_C) \in \prod_{i}[0, p^{k_i + 1} - 1]} \mathbb{E}_{x} \left[ e \left( \sum_{i} \lambda_i(P_i(x) - b_i) \right) \right]$$

$$= \prod_{i} p^{-(k_i + 1)} \cdot \left( 1 \pm \varepsilon \prod_{i} p^{k_i + 1} \right) = \frac{1}{\|B\|} \pm \varepsilon.$$

The first equality uses the fact that $P_i(x) - b_i$ is in $U_{k_i + 1}$ and that for any nonzero $x \in U_{k_i + 1}$, $\sum_{\lambda=0}^{p^{k_i + 1} - 1} e(\lambda x) = 0$. The third equality uses Theorem 2.2.10 and the fact that unless every $\lambda_i = 0$, the polynomial $\sum_i \lambda_i(P_i(x) - b_i)$ has rank at least $r_{2.2.10}(d, \varepsilon)$. □

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

**Lemma 2.2.19** (Equidistribution for unbiased factors). Suppose that $\gamma : \mathbb{N} \to \mathbb{R}^+$ is a decreasing function. Let $\mathcal{F}$ be a $\gamma$-unbiased factor of degree $d > 0$, 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 $b = (b_1, \ldots, b_C) \in U_{k_1 + 1} \times \cdots \times U_{k_C + 1}$. Then

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

### 2.2.4 Strong Near-Orthogonality and Equidistribution

One of the important and useful properties of the classical Fourier characters is that they form an orthonormal basis. Theorem 2.2.10, Lemma 2.2.18 and Lemma 2.2.19 provide an
approximate version of this phenomenon which is useful for several applications. However, this is not completely satisfactory, as we will see later that in order to study averages of the form
\[ E_{X \in (\mathbb{F}^n)^k} [f(L_1(X)) \cdots f(L_m(X))], \]
for a bounded function \( f : \mathbb{F}^n \to \mathbb{R} \), one needs to understand the distribution of the more sophisticated random variable
\[
A_{\mathcal{P}, \mathcal{L}}(X) := \begin{pmatrix}
P_1(L_1(X)) & P_2(L_1(X)) & \ldots & P_C(L_1(X)) \\
P_1(L_2(X)) & P_2(L_2(X)) & \ldots & P_C(L_2(X)) \\
\vdots & \vdots & & \vdots \\
P_1(L_m(X)) & P_2(L_m(X)) & \ldots & P_C(L_m(X))
\end{pmatrix},
\]
where \( X \) is the uniform random variable taking values in \((\mathbb{F}^n)^k\), \( \mathcal{P} = \{P_1, \ldots, P_C\} \) is a family of non-classical polynomials and \( \mathcal{L} = \{L_1, \ldots, L_m\} \subset \mathbb{F}^k \) is an arbitrary system of linear forms.

Since (non-classical) polynomials of a given degree satisfy various linear identities (e.g. every degree one polynomial \( P \) satisfies \( P(x + y + z) = P(x + y) + P(x + z) - P(x) \)), it is no longer possible to choose the polynomials in a way that this random matrix is almost uniformly distributed over all \( m \times C \) matrices \( A \in \prod_{i=1}^m \prod_{j=1}^C U_{k_j+1} \). Therefore, in this case one would like to obtain an almost uniform distribution on the points of the configuration space that are consistent with these linear identities. Note that [43] and [54] only address the case where the system of linear forms consists only of a single linear form; in other words they show that the entries in each row of this matrix are nearly independent. The stronger near-equidistribution would in particular imply that the columns of this matrix are nearly independent, with each column uniform modulo the required linear dependencies.

Hatami and Lovett [51] established this strong near-equidistribution in the case where the characteristic of the field \( \mathbb{F} \) is greater than the degree of the (classical) polynomial factor.
Lemma 2.2.20 ([51]). Let $\mathbb{F}$ be a prime field and $d < |\mathbb{F}|$. Let $\{L_1, \ldots, L_m\}$ be a system of linear forms. Let $\mathcal{F} = \{P_1, \ldots, P_k\}$ be a collection of homogeneous classical polynomials of degree at most $d$, such that $\text{rank}(\mathcal{F}) > \tau_{p,d}(\varepsilon)$. For every set of coefficients $\Lambda = \{\lambda_{i,j} \in \mathbb{F} : i \in [k], j \in [m]\}$, and

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

one of the following two cases holds:

$$P_\Lambda \equiv 0 \quad \text{or} \quad \mathbf{E}[e_\mathbb{F}(P_\Lambda)] < \varepsilon.$$

The proof technique used in [51] breaks down once $|\mathbb{F}|$ is allowed to be small. Bhat- tacharyya, et al. [15] later extended the result of [51] to the general characteristic case, but under the 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.

Theorem 2.2.21 (Near orthogonality [15]). Given $\varepsilon > 0$, suppose $\mathcal{B} = (P_1, \ldots, P_C)$ is a polynomial factor of degree $d > 0$ and rank $> r_{2,2,13}(d,\varepsilon)$, $A = (L_1, \ldots, L_m)$ is an affine constraint on $\ell$ variables, and $\Lambda$ is a tuple of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$. Define

$$P_{A,\mathcal{B},\Lambda}(x_1, \ldots, x_\ell) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)).$$

Then, one of the two statements below is true.

- For every $i \in [C]$, it holds that $\sum_{j \in [m]} \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ for all polynomials $Q_i : \mathbb{F}^n \to \mathbb{T}$ with the same degree and depth as $P_i$. Clearly, $P_{A,\mathcal{B},\Lambda} \equiv 0$ in this case.

- $P_{A,\mathcal{B},\Lambda}$ is non-constant. Moreover, $|\mathbf{E}_{x_1, \ldots, x_\ell}[e(P_{A,\mathcal{B},\Lambda}(x_1, \ldots, x_\ell))]| < \varepsilon$.

One can use the derivative technique used in [15] in a straightforward manner to prove the following near-orthogonality result for when the linear forms are not necessarily affine.
but they are allowed only to take 0 and 1 entries (See [18] for a proof).

Lemma 2.2.22 (Near orthogonality). For every decreasing function \( \varepsilon : \mathbb{N} \to \mathbb{R}^+ \) and parameter \( d \geq 1 \), there is a decreasing function \( \gamma : \mathbb{N} \to \mathbb{R}^+ \) such that the following holds. Suppose that \( F = \{ P_{i,j} \}_{1 \leq i \leq d, 1 \leq j \leq M_i} \) is a \( \gamma \)-uniform factor of degree \( d \). Let \( x \in \mathbb{F}^n \) be a fixed vector. For every integer \( k \) and set of not all zero coefficients \( \Lambda = \{ \lambda_{i,j,\omega} \}_{1 \leq i \leq d, 1 \leq j \leq M_i, \omega \in \{0,1\}^k} \) define

\[
P_{\Lambda}(y_1, \ldots, y_k) \overset{\text{def}}{=} \sum_{i \in [d], j \in [M_i]} \sum_{\omega \in \{0,1\}^k} \lambda_{i,j,\omega} \cdot P_{i,j}(x + \omega \cdot y),
\]

where \( y = (y_1, \ldots, y_k) \in (\mathbb{F}^n)^k \). Then either we have that \( P_{\Lambda} \) is a constant, i.e. it does not depend on \( y \), or

\[
\text{bias}(P_{\Lambda}) < \varepsilon (\dim(F))
\]

Finally, in the present dissertation we fully characterize the joint distribution of the matrix \( A_{P,\mathcal{L}} \) without any extra assumptions on the linear forms.

Theorem 2.2.23 (Near Orthogonality over Linear Forms). Let \( L_1, \ldots, L_m \) be linear forms on \( \ell \) variables and let \( \mathcal{B} = (P_1, \ldots, P_C) \) be an \( \varepsilon \)-uniform polynomial factor for some \( \varepsilon \in (0,1] \) defined only by homogeneous polynomials. For every tuple \( \Lambda \) of integers \( \{\lambda_{i,j}\}_{i \in [C], j \in [m]} \), define \( P_{\Lambda} : (\mathbb{F}^n)^\ell \to \mathbb{R} \) as

\[
P_{\Lambda}(X) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(X)).
\]

Then one of the following two statements holds:

- \( P_{\Lambda} \equiv 0 \).
- \( P_{\Lambda} \) is non-constant and \( \left| \mathbb{E}_{X \in (\mathbb{F}^n)^\ell} (e(P_{\Lambda})) \right| < \varepsilon \).

Furthermore \( P_{\Lambda} \equiv 0 \) if and only if for every \( i \in [C] \), we have \( \{\lambda_{i,j}\}_{j \in [m]} \in \Phi_{d_i,k_i}(\mathcal{L})^\perp \) where \( d_i, k_i \) are the degree and depth of \( P_i \), respectively.
Homogeneous non-classical polynomials are introduced in Chapter 3, where we also prove several useful facts about them. Theorem 2.2.23 is discussed in Chapter 4 and proved in Section 4.2.

2.2.5 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 [78]. We will first formally define what we mean by refining a polynomial factor.

**Definition 2.2.24** (Refinement). A factor $B'$ is called a refinement of $B$, and denoted $B' \succeq B$, if the induced partition by $B'$ is a combinatorial refinement of the partition induced by $B$. In other words, if for every $x, y \in \mathbb{F}^n$, $B'(x) = B'(y)$ implies $B(x) = B(y)$.

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

**Definition 2.2.25** (Semantic and syntactic refinements). $B'$ is called a syntactic refinement of $B$, and denoted $B' \succeq_{syn} B$, if the sequence of polynomials defining $B'$ extends that of $B$. It is called a semantic refinement, and denoted $B' \succeq_{sem} B$ if the induced partition is a combinatorial refinement of the partition induced by $B$. In other words, if for every $x, y \in \mathbb{F}^n$, $B'(x) = B'(y)$ implies $B(x) = B(y)$.

**Remark 2.2.26.** Clearly, being a syntactic refinement is stronger than being a semantic refinement. But observe that if $B'$ is a semantic refinement of $B$, then there exists a syntactic refinement $B''$ of $B$ that induces the same partition of $\mathbb{F}^n$, and for which $|B''| \leq |B'| + |B|$, because we can define $B''$ by just adding the defining polynomials of $B$ to those of $B'$.

The following, by now well-known, lemma states that given a classical polynomial factor one can refine it to a regular one.
Lemma 2.2.27 (Regularity Lemma for Polynomials). Let $d \geq 1$, $F : \mathbb{N} \to \mathbb{N}$ be a growth function, and let $\mathcal{F}$ be a classical polynomial factor of degree $d$. Then there exists a semantic $F$-regular refinement $\mathcal{F}' \succeq_{\text{sem}} \mathcal{F}$ of classical polynomials of same degree $d$, satisfying
\[
\dim(\mathcal{F}') = O_{F,d,\dim(\mathcal{F})}(1).
\]

Proof. We shall induct on the dimension vectors $(M_1, \ldots, M_d)$. Notice that the dimension vectors of a factor of degree $d$ takes values in $\mathbb{N}^d$, and we will use the reverse lexicographical ordering on this space for our induction. The proof for $d = 1$ is obvious, because non-zero linear functions have infinite rank$_0$, unless there is a linear dependency between the polynomials in $\mathcal{F}$, which we can simply discard by removing the polynomials that can be written as a linear combination of the others.

If $\mathcal{F}$ is already $F$-regular, then we are done. Otherwise, there is a set of coefficients $\{c_{i,j}\}_{1 \leq i \leq d, 1 \leq j \leq M_i}$, not all zero, such that
\[
\text{rank}_{k-1} \left( \sum_{i=1}^{d} \sum_{j=1}^{M_i} c_{i,j} P_{i,j} \right) < F(\dim(\mathcal{F})).
\]

Without loss of generality assume that $c_{k,M_k} \neq 0$. The above inequality means that $P_{k,M_k}$ can be written as a function of the rest of the polynomials in the factor and a set of at most $F(\dim(\mathcal{F}))$ polynomials of degree at most $k-1$. We will replace $P_{k,M_k}$ with these new polynomials. The new factor will have the following dimension vector
\[
(M_1, \ldots, M_{k-1} + F(\dim(\mathcal{F})), M_k - 1, M_{k+1}, \ldots, M_d).
\]

Now by the induction hypothesis the above can be regularized. \hfill \square

The same proof idea can be used to extend this lemma to non-classical polynomial factors.
Lemma 2.2.28 (Lemma 9.6 of [78]). Let \( r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) be a non-decreasing function and \( d > 0 \) be an integer. Then, there are functions \( \bar{C}^{(r,d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) and \( I^{(d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) such that the following is true. Suppose \( \mathcal{B} \) is an extended polynomial factor defined by polynomials \( P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T} \) of degree \( \leq d \). Then, there is a subspace \( V \subseteq \mathbb{F}^n \) and an \( r \)-regular extended factor \( \mathcal{B} \) consisting of polynomials \( Q_1, \ldots, Q_{\bar{C}} : V \to \mathbb{T} \) such that \( 2 \leq \deg(Q_i) \leq d \) for each \( i \), \( \mathcal{B} \) semantically refines the factor defined by \( P_1|_V, \ldots, P_C|_V \), \( \bar{C} \leq \bar{C}^{(r,d)}(C) \), and \( \dim(V) \geq n - I^{(d)}(\bar{C}) \).

The nature of the proof in regularity lemmas of the above type results in Ackermann-like functions even for “reasonable” growth functions.

The following lemma is the workhorse that allows us to construct regular refinements.

Lemma 2.2.29 (Polynomial Regularity Lemma). Let \( r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) be a non-decreasing function and \( d > 0 \) be an integer. Then, there is a function \( C^{(r,d)}_{2.2.29} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) such that the following is true. Suppose \( \mathcal{B} \) is a factor defined by polynomials \( P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T} \) of degree at most \( d \). Then, there is an \( r \)-regular factor \( \mathcal{B}' \) consisting of polynomials \( Q_1, \ldots, Q_{C'} : \mathbb{F}^n \to \mathbb{T} \) of degree \( \leq d \) such that \( \mathcal{B}' \succeq_{sem} \mathcal{B} \) and \( C' \leq C^{(r,d)}_{2.2.29}(C) \).

Moreover, if \( \mathcal{B} \) is itself a refinement of some \( \hat{\mathcal{B}} \) that has rank \( > (r(C') + C') \) and consists of polynomials, then additionally \( \mathcal{B}' \) will be a syntactic refinement of \( \hat{\mathcal{B}} \).

Proof. We can prove this lemma starting from Lemma 9.6 of [78]. To explain, let us define the notion of an extended factor. We say a polynomial factor \( \mathcal{B} \) is extended if for any polynomial \( Q \in \mathcal{B} \) that is not classical, \( pQ \in \mathcal{B} \) also. Note that an extended factor defined by polynomials \( P_1, \ldots, P_C \) is of high rank if for all tuples \( (\lambda_1, \ldots, \lambda_C) \in [0, p - 1]^C \), unless all the \( \lambda_i \)'s are zero, \( \sum_i \lambda_i P_i \) is of high \( (\max_i \deg(\lambda_i P_i)) \)-rank.

Let \( \mathcal{B}_1 \) be the extended factor defined by \( \left\{ p^k P_i \mid 0 \leq k \leq \text{depth}(P_i), i \in [C] \right\} \). Apply Lemma 2.2.28 to \( \mathcal{B}_1 \) in order to obtain a bounded index subspace \( V_1 \) and an extended \( R_1 \)-regular factor \( \bar{\mathcal{B}}_1 \) defined by polynomials \( Q_1, \ldots, Q_{\bar{C}} : V_1 \to \mathbb{T} \), where \( R_1 \) is a growth function (growing even faster than \( r \)) we specify later on in the proof and \( \bar{C} \leq \bar{C}^{(R_1,d)}(|\mathcal{B}_1|) \).
For \( a \in \mathbb{F}^n/V_1 \) and \( P \in \mathcal{B}_1 \), define \( P^a : V_1 \to \mathbb{F}^n \) to be \( P^a(x) = D_a P(x) = P(a + x) - P(x) \). Each \( P^a \) is of degree \( \leq d - 1 \). Also, since \( V_1 \) is the intersection of \( I \leq I^{(d)}(\bar{C}) \) hyperplanes, we can decide which coset in \( \mathbb{F}^n/V_1 \) an element \( x \in \mathbb{F}^n \) belongs to as a function of \( I \leq I^{(d)}(\bar{C}) \) (classical) linear functions \( \pi_1, \ldots, \pi_I \). Let \( \mathcal{B}_2 \) be the extended factor obtained by adding to \( \bar{B}_1 \) all the polynomials \( \{P^a \mid P \in \mathcal{B}_1, a \in \mathbb{F}^n/V_1 \} \) and \( \pi_1, \ldots, \pi_I \). Consider \( x \in \mathbb{F}^n \) and let \( x = a + y \) where \( y \in V_1 \), and \( a \in \mathbb{F}^n/V_1 \). Since \( P(x) = P(y) - P^{-a}(x) \), each polynomial in \( \mathcal{B}_1 \) is a function of the polynomials in \( \mathcal{B}_2 \) over all of \( \mathbb{F}^n \), and so \( \mathcal{B}_2 \) is a semantic refinement of \( \mathcal{B}_1 \) (and a syntactic refinement of \( \mathcal{B}_1 \)). Note that \( |\mathcal{B}_2| \leq C + dCI^{(d)}(C) + I^{(d)}(C) < \bar{C} + 2d\bar{C}I^{(d)}(C) \).

Now, suppose we repeat the steps in the previous paragraph with \( \mathcal{B}_2 \) taking the place of \( \mathcal{B}_1 \) and a different function \( R_2 \) taking the place of \( R_1 \). We specify \( R_2 \) later, but we will choose it so that it grows faster than \( r \). The new application of Lemma 2.2.28 to \( \mathcal{B}_2 \) produces an extended factor \( \bar{B}_2 \) that is \( R_2 \)-regular and a bounded index subspace \( V_2 \) such that the polynomials in \( \mathcal{B}_2 \) restricted to \( V_2 \) are measurable with respect to \( \bar{B}_2 \). We argue that \( \bar{B}_2 \) differs from \( \mathcal{B}_2 \) only by polynomials of degree \( \leq d - 1 \). Suppose \( \mathcal{B}_2 \) is not \( R_2 \)-regular to start off with. The function \( R_1 \) is chosen so that \( \bar{B}_1 \)'s rank, \( R_1(|\bar{B}_1|) > R_2(|\mathcal{B}_2|) + |\mathcal{B}_2| \). This means that if a linear combination of polynomials in \( \mathcal{B}_2 \), \( \sum_{S \in \mathcal{B}_2} \lambda S S \), has rank \( \leq R_2(|\mathcal{B}_2|) \) and \( d' = \max_{S : \lambda_S \neq 0} \deg(S) \), then there must be an \( S \not\in \bar{B}_1 \) with \( \lambda_S \neq 0 \) and degree \( d' \), since otherwise the rank condition of \( \mathcal{B}_1 \) would be violated. Since all the polynomials in \( \mathcal{B}_2 \) which are not in \( \bar{B}_1 \) have degree \( \leq d - 1 \), we conclude that \( d' \leq d - 1 \). Inspecting the proof of Lemma 2.2.28 in [78] shows that this means \( \bar{B}_2 \) consists of the polynomials of \( \bar{B}_1 \) along with other polynomials of degree \( \leq d - 1 \). In the same way as in the previous paragraph, we obtain an extended factor \( \mathcal{B}_3 \supseteq_{\text{syn}} \bar{B}_2 \), so that \( \mathcal{B}_3 \) is a semantic refinement of \( \mathcal{B}_2 \) over all of \( \mathbb{F}^n \). Note that since all the polynomials of \( \bar{B}_1 \) are already in \( \bar{B}_2 \), we only need to add \( \{P^a \mid P \in \mathcal{B}_2 \setminus \bar{B}_1, a \in \mathbb{F}^n/V_2 \} \), together with some linear functions. All these polynomials have degree at most \( \leq d - 2 \).
We keep repeating this process to obtain a sequence of extended factors $B_1, B_2, B_3, \ldots$ and $\bar{B}_1, \bar{B}_2, \bar{B}_3, \ldots$. Each $B_{i+1}$ semantically refines $B_i$ and syntactically refines $\bar{B}_i$. The process stops at step $i$ if $B_i$ becomes $R_i$-regular, where the sequence of growth functions $R_i$ satisfies $R_i(m) > R_{i+1}(m + 2dmI(d)(m)) + m + 2dmI(d)(m)$ and $R_d(m) = r(m)$. The functions $R_i$ are chosen so that $R_i(|\bar{B}_i|) > R_{i+1}(|B_{i+1}|) + |B_{i+1}|$, and therefore, by the above argument, $B_{i+1}$ differs from $B_i$ by polynomials of degree $\leq d - i$. So, we must stop after obtaining $B_d$ in the sequence. Also, since each $R_i$ grows faster than $r$, note that $R_i$-regularity for any $i \in [d]$ implies $r$-regularity. So, it must be that some $B_i$ for $i \leq d$ already becomes $r$-regular.

Given an extended factor $B''$ of rank $> r$, we can get a (standard) factor $B'$ of rank $> r$ by letting $B'$ be defined by the smallest subset of polynomials $S$ such that $\{p^i P \mid P \in S, i \in \mathbb{Z}_{\geq 0}\} \supseteq B''$. The last statement of the lemma follows from the same considerations as used above to argue that $\bar{B}_i$ syntactically refines $B_i$. □

2.3 Decomposition Theorems

“Decomposition theorems” [36, 75, 43] are important applications of inverse theorems. These theorems allow one to express a given function $f$ with certain properties as a sum $f_1 + \sum_{i=1}^k g_i$, where $f_1$ is “structured” and each $g_i$ has certain desired quasirandomness properties. The rough idea is that the structure of $f_1$ is strong enough for us to be able to analyze it reasonably explicitly, while the quasirandomness of $g_i$s is strong enough for many properties of $f_1$ to be unaffected if we “perturb” it to $f = f_1 + \sum_{i=1}^k g_i$. We refer the interested reader to [36] and [40] for a detailed discussion. The following decomposition theorem is a consequence of an inverse theorem for Gowers norms ([78, Theorem 1.11]).

**Theorem 2.3.1** (Strong Decomposition Theorem for Multiple Functions). Let $m, d \geq 1$ be integers, $\delta > 0$ a parameter, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth function. Given any
functions $f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{D}$, there exists a decomposition

$$f_i = g_i + h_i,$$

such that for every $1 \leq i \leq m$,

1. $g_i = \mathbb{E}[f_i | \mathcal{B}]$, where $\mathcal{B}$ is an $r$-regular polynomial factor of degree at most $d$ and complexity $C \leq C_{\text{max}}(p, m, d, \delta, r(\cdot))$,

2. $\|h_i\|_{U^{d+1}} \leq \delta$.

**Remark 2.3.2.** In the case of high-characteristic, i.e. $d < |\mathbb{F}|$, $\mathcal{B}$ will consist only of classical polynomials, and it is easily seen that we can further assume that the polynomials in factor $\mathcal{B}$ are homogeneous, which means that all their multinomials are of same degree. However, the case of non-classical polynomials is tricky and unclear, as for example it is not a priori clear what a good candidate definition for homogeneous non-classical polynomial is and whether such homogeneous non-classical polynomials span all polynomials. Chapter 3 addresses this issue, and proves that homogeneous non-classical polynomials span the space of all non-classical polynomials.

### 2.3.1 Strong Decomposition Theorems

Often, in order to obtain stronger statements about the structure and the qasirandomness in a decomposition theorem, one allows also a small $L^2$-error: that is, one writes $f$ as $f_1 + f_2 + f_3$ with $f_1$ structured, $f_2$ quasirandom, and $f_3$ small in $L^2$.

The strong decomposition theorem below shows that any Boolean function can be decomposed into the sum of a conditional expectation over a high rank factor, a function with small Gowers norm, and a function with small $L^2$-norm.

**Theorem 2.3.3** (Strong Decomposition Theorem; Theorem 4.4 of [16]). Suppose $\delta > 0$ and $d \geq 1$ are integers. Let $\eta : \mathbb{N} \to \mathbb{R}^+$ be an arbitrary non-increasing function and
Given $f : \mathbb{F}^n \to \{0, 1\}$ where $n > N$, there exist three functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B}$ of degree at most $d$ and complexity at most $C$ such that the following conditions hold:

(i) $f = f_1 + f_2 + f_3$.

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

(iii) $\|f_2\|_{U^{d+1}} \leq 1/\eta(|\mathcal{B}|)$.

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

(v) $f_1$ and $f_1 + f_3$ have range $[0, 1]$; $f_2$ and $f_3$ have range $[-1, 1]$.

(vi) $\mathcal{B}$ is $r$-regular.

It turns out though that this strong decomposition theorem is not quite sufficient for some applications. For example, in the study of algebraic property testing we require an even stronger decomposition theorem. The issue is that the bound on $f_3$ above is a constant $\delta$. Ideally, we would want $\delta$ to decrease as a function of the complexity of the polynomial factor, but such a decomposition theorem is simply not possible. However, analogous to what it is shown in [2] in the context graphs, here one can find two polynomial factors $\mathcal{B}' \succeq_{\text{syn}} \mathcal{B}$ such that, the structured part $f_1$ equals to $\mathbb{E}[f | \mathcal{B}']$, but now the $L^2$-norm of $f_3$ can be made arbitrarily small in terms of the complexity of the coarser factor $\mathcal{B}$. Furthermore for most atoms $c$ of $\mathcal{B}$, the function $f : \mathbb{F}^n \to \{0, 1\}$ have roughly the same density on $c$ and most of its subatoms in $\mathcal{B}'$. To make this precise, we make the following definition.

**Definition 2.3.4** (Polynomial factor represents another factor). Given a function $f : \mathbb{F}^n \to \{0, 1\}$, a polynomial factor $\mathcal{B}'$ that refines another factor $\mathcal{B}$ and a real $\zeta \in (0, 1)$, we say
\(B' \) \(\zeta\)-represents \(B\) with respect to \(f\) if for at most \(\zeta\) fraction of atoms \(c\) of \(B\), more than \(\zeta\) fraction of the atoms \(c'\) lying inside \(c\) satisfy \(|E[f|c] - E[f|c']| > \zeta\).

We can now state the following “Super Decomposition Theorem” from [16].

**Theorem 2.3.5** (Super Decomposition Theorem; Theorem 4.9 of [16]). Suppose \(\zeta > 0\) is a real and \(d, C_0 \geq 1\) are integers. Let \(\eta : \mathbb{N} \to \mathbb{R}^+\) and \(\delta : \mathbb{N} \to \mathbb{R}^+\) be arbitrary non-increasing functions, and \(r : \mathbb{N} \to \mathbb{N}\) be an arbitrary non-decreasing function. Then there exist \(N = N_{2.3.5}(\delta, \eta, r, d, \zeta)\) and \(C = C_{2.3.5}(\delta, \eta, r, d, \zeta)\) such that the following holds.

Given \(f : \mathbb{F}^n \to \{0, 1\}\) where \(n > N\), there exist functions \(f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}\), and polynomial factors \(B' \succeq \text{syn} B\) of degree at most \(d\) and of complexity at most \(C\), such that the following conditions hold:

1. \(f = f_1 + f_2 + f_3\).
2. \(f_1 = E[f|B']\).
3. \(\|f_2\|_{U^{d+1}} \leq \eta(|B'|)\).
4. \(\|f_3\|_2 \leq \delta(|B|)\).
5. \(f_1\) and \(f_1 + f_3\) have range \([0, 1]\); \(f_2\) and \(f_3\) have range \([-1, 1]\).
6. \(B\) and \(B'\) are both \(r\)-regular.
7. \(B' \) \(\zeta\)-represents \(B\) with respect to \(f\).

Although the above Super Decomposition Theorem may be useful by itself for other applications, we will need a particular variant. The factor \(B'\) is a syntactic refinement of \(B\), and thus is defined by adding new polynomials \(Q_1, \ldots, Q_{|B'|-|B|}\) to the polynomials defining \(B\). Then for each atom \(c\) in the coarser factor \(B\) we will select one atom \(c'\) of \(B'\) such that the following hold:
• There is a fixed \( s \in T^{|\mathcal{B}'|-|\mathcal{B}|} \) such that for every atom \( c \) in \( \mathcal{B} \) its corresponding atom \( c' \) is obtained by requiring \((Q_1, \ldots, Q_{|\mathcal{B}'|-|\mathcal{B}|})\) to be equal to \( s \).

• The \( L^2 \)-norm of \( f_3 \) conditioned inside every such atom (i.e., \( E_{x \in c'}[|f_3(x)|^2] \)) is small.

• Most subatoms \( c' \) will “well-represent” (in the sense of Definition 2.3.4) their corresponding atoms \( c \) from \( \mathcal{B} \).

Before stating this formally, let us also take this opportunity to remark that it is possible to adapt the proofs of the above decomposition theorems to decompose several functions \( f^{(1)}, \ldots, f^{(R)} : \mathbb{F}^n \to \{0,1\} \) simultaneously. Alternatively, this could be thought of as decomposing a single vector-valued function \( f : \mathbb{F}^n \to \{0,1\}^R \). Now we finally state the decomposition theorem that we will use in the proof of our main result.

**Theorem 2.3.6** (Subatom Selection; Theorem 4.12 of [16]). Suppose \( \zeta > 0 \) is a real and \( d, R \geq 1 \) are integers. Let \( \eta, \delta : \mathbb{N} \to \mathbb{R}^+ \) be arbitrary non-increasing functions, and let \( r : \mathbb{N} \to \mathbb{N} \) be an arbitrary non-decreasing function. Then, there exist \( C = C_{2.3.6}(\delta, \eta, r, \zeta, R) \) such that the following holds.

Given \( f^{(1)}, \ldots, f^{(R)} : \mathbb{F}^n \to \{0,1\} \), there exist functions \( f_1^{(i)} : f_2^{(i)} : f_3^{(i)} : \mathbb{F}^n \to \mathbb{R} \) for all \( i \in [R] \), a polynomial factor \( \mathcal{B} \) of degree \( d \) with atoms denoted by elements of \( T^{\mathcal{B}} \), a syntactic refinement \( \mathcal{B}' \succeq_{\text{sym}} \mathcal{B} \) of degree \( d \) with complexity at most \( C \) and atoms denoted by elements of \( T^{\mathcal{B}} \times T^{\mathcal{B}'-|\mathcal{B}|} \), and an element \( s \in T^{\mathcal{B}'-|\mathcal{B}|} \) such that the following is true:

(i) \( f^{(i)} = f_1^{(i)} + f_2^{(i)} + f_3^{(i)} \) for every \( i \in [R] \).

(ii) \( f_1^{(i)} = E[f^{(i)}|\mathcal{B}'] \) for every \( i \in [R] \).

(iii) \( \|f_2^{(i)}\|_{U^{d+1}} < \eta(|\mathcal{B}'|) \) for every \( i \in [R] \).

(iv) For every \( i \in [R] \), \( f_1^{(i)} \) and \( f_1^{(i)} + f_3^{(i)} \) have range \([0,1]\), and \( f_2^{(i)} \) and \( f_3^{(i)} \) have range \([-1,1]\).
(v) $\mathcal{B}$ and $\mathcal{B}'$ are both $r$-regular.

(vi) For every atom $c \in \mathbb{T}^{\mathcal{B}}$ of $\mathcal{B}$, the subatom $c' = (c, s) \in \mathbb{T}^{\mathcal{B}'}$ satisfy

$$
\mathbb{E}\left[ |f_3^{(i)}|^2 \left| (c, s) \right| \right] < \delta(|\mathcal{B}|)^2
$$

for every $i \in [R]$.

(vii) If $c$ is an atom of $\mathcal{B}$ chosen uniformly at random, then

$$
\Pr_{c} \left[ \max_{i \in [R]} \left( \mathbb{E}\left[ f^{(i)} \big| c \right] - \mathbb{E}\left[ f^{(i)} \big| (c, s) \right] \right) > \zeta \right] < \zeta.
$$

2.3.2 Higher-order Fourier Expansion

In the previous sections we saw that decomposition theorems can be used to express a given function $f$ that satisfies certain reasonable properties as $f = f_1 + \sum_{i=1}^{k} g_i$, where $f_1 = \Gamma(P_1, ..., P_C)$ for a set of (non-classical) polynomials, for some function $\Gamma : \mathbb{T}^C \to \mathbb{D}$, and $g_i$’s are small in norms that allow us to regard them as negligible error terms. Let $k_i$ denote the depth of the polynomial $P_i$ so that by Lemma 2.1.3, each $P_i$ takes values in $\mathbb{U}_{k_i+1} = \frac{1}{p^{k_i+1}} \mathbb{Z}/\mathbb{Z}$. Moreover, let $\Sigma := \mathbb{Z}_{p^{k_1+1}} \times \cdots \times \mathbb{Z}_{p^{k_C+1}}$. Using the Fourier expansion of $\Gamma_i$ we have

$$
f_1(x) = \sum_{\Lambda=(\lambda_1, ..., \lambda_C) \in \Sigma} \hat{\Gamma}_i(\Lambda) \cdot e \left( \sum_{j=1}^{C} \lambda_j P_j(x) \right),
$$

(2.3)

where $\hat{\Gamma}_i(\Lambda)$ is the Fourier coefficient of $\Gamma_i$ corresponding to $\Lambda$. Equation (2.3) implies an approximate higher-order Fourier expansion of $f$, a generalization of the classical expansion.

$$
f(x) \simeq \sum_{\Lambda=(\lambda_1, ..., \lambda_C) \in \Sigma} \hat{\Gamma}_i(\Lambda) \cdot e \left( \sum_{j=1}^{C} \lambda_j P_j(x) \right).
$$

(2.4)
2.4 Systems of Linear forms

Definition 2.4.1. A linear form on $k$ variables is a vector $L = (\ell_1, \ldots, \ell_k) \in \mathbb{F}^k$ and it maps $X = (x_1, \ldots, x_k) \in (\mathbb{F}^n)^k$ to $L(X) = \sum_{i=1}^k \ell_i x_i \in \mathbb{F}^n$.

For a linear form $L = (\ell_1, \ldots, \ell_k) \in \mathbb{F}^k$ we define $|L| \overset{\text{def}}{=} \sum_{i=1}^k |\ell_i|$.

Gowers norms are known to control the density of linear patterns in subsets of an Abelian group. Next, we present a notion of complexity associated with a collection of linear forms that determines an upper-bound on the order of the required Gowers norm to do so.

2.4.1 Cauchy-Schwarz Complexity

Let $A$ be a subset of $\mathbb{F}^n$ with the indicator function $1_A : \mathbb{F}^n \to \{0, 1\}$. Then the analytical average

$$E_{X \in (\mathbb{F}^n)^k} [1_A(L_1(X)) \cdots 1_A(L_m(X))]$$

represents the probability that $L_1(X), \ldots, L_m(X)$ all fall in $A$, where $X \in (\mathbb{F}^n)^k$ is chosen uniformly at random. Then roughly speaking, we will say that $A \subseteq \mathbb{F}^n$ is pseudorandom with regards to $L$ if

$$E_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^m 1_A(L_i(X)) \right] \approx \left( \frac{|A|}{p^n} \right)^m ;$$

That is, 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 $\alpha := |A|/p^n$ be the density of $A$, and define $f := 1_A - \alpha$. We have

$$E_X \left[ \prod_{i=1}^m 1_A(L_i(X)) \right]$$

$$= E_X \left[ \prod_{i=1}^m (\alpha + f(L_i(X))) \right] = \alpha^m + \sum_{S \subseteq [m], S \neq \emptyset} \alpha^{m-|S|} E_X \left[ \prod_{i \in S} f(L_i(X)) \right] .$$

36
Therefore, a sufficient condition for $A$ to be pseudorandom with regards to $\mathcal{L}$ is that
\[ \mathbb{E}_X \left[ \prod_{i \in S} f(L_i(X)) \right] \text{ is negligible for all nonempty subsets } S \subseteq [m]. \] Green and Tao [46] 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.

**Definition 2.4.2** (Cauchy-Schwarz complexity [46]). Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. The Cauchy-Schwarz complexity of $\mathcal{L}$ is the minimal $s$ such that the following holds. For every $1 \leq i \leq m$, 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.

Note that the Cauchy-Schwarz complexity of any system of $m$ linear forms in which any 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.

The reason for the term Cauchy-Schwarz complexity is the following lemma due to Green and Tao [46] the proof of which is based on a clever iterative application of the Cauchy-Schwarz inequality.

**Lemma 2.4.3** ([46], See also [37, Theorem 2.3]). Let $f_1, \ldots, f_m : \mathbb{F} \to \mathbb{D}$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of $m$ linear forms in $\ell$ variables of Cauchy-Schwarz complexity $s$. Then
\[ \left| \mathbb{E}_{X \in (\mathbb{F}^n)^\ell} \left[ \prod_{i=1}^m f_i(L_i(X)) \right] \right| \leq \min_{1 \leq i \leq m} \|f_i\|_{U^{s+1}}. \]

### 2.4.2 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 $s$ for which this condition holds for all $f : \mathbb{F}^n \to \mathbb{D}$. 37
Definition 2.4.4 (True complexity [37]). Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}$. The true complexity of $\mathcal{L}$ is the smallest $d \in \mathbb{N}$ with the following property. For every $\varepsilon > 0$, there exists $\delta > 0$ such that if $f : \mathbb{F}^n \to \mathbb{D}$ is any function with $\|f\|_{U^{d+1}} \leq \delta$, then

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^m f(L_i(X)) \right] \right| \leq \varepsilon.$$ 

An obvious bound on the true complexity is the Cauchy-Schwarz complexity of the system.
CHAPTER 3
HOMOGENEOUS NON-CLASSICAL POLYNOMIALS

Recall that a classical polynomial is called homogeneous if all of its monomials are of the same degree. Trivially a homogeneous classical polynomial \( P(x) \) satisfies \( P(cx) = |c|^d P(x) \) for every \( c \in \mathbb{F} \). We will use this property to define the class of nonclassical homogeneous polynomials.

**Definition 3.0.5** (Homogeneity). A (nonclassical) polynomial \( P : \mathbb{F}^n \to \mathbb{T} \) is called homogeneous if for every \( c \in \mathbb{F} \) there exists a \( \sigma_c \in \mathbb{Z} \) such that \( P(cx) = \sigma_c P(x) \mod 1 \) for all \( x \in \mathbb{F}^n \).

**Remark 3.0.6.** It is not difficult to see that \( P(cx) = \sigma_c P(x) \mod 1 \) implies that \( \sigma_c = |c|^{\deg(P)} \mod p \), a property that we will use later. Indeed for \( d = \deg(P) \), we have \( 0 = \partial_d(P(cx) - \sigma_c P(x)) = (|c|^d - \sigma_c)\partial_d P(x) \mod 1 \). This, since \( \partial_d P(x) \) is a nonzero degree-\( d \) classical polynomial, implies \( \sigma_c = |c|^d \mod p \).

Notice that for a polynomial \( P \) to be homogeneous it suffices that there exists \( \sigma \in \mathbb{Z} \) for which \( P(\zeta x) = \sigma P(x) \mod 1 \), where \( \zeta \) is a generator of \( \mathbb{F}^* \). Throughout this chapter, we fix \( \zeta \in \mathbb{F}^* \) a generator of \( \mathbb{F}^* \). If \( P \) has depth \( k \), then we can assume that \( \sigma \in \mathbb{Z}_{p^k+1} \), as \( p^{k+1} P \equiv 0 \). The following lemma shows that \( \sigma \) is uniquely determined for all homogeneous polynomials of degree \( d \) and depth \( k \). Henceforth, we will denote this unique value by \( \sigma(d,k) \).

**Lemma 3.0.7.** For every \( d \) and \( k \), there is a unique \( \sigma = \sigma(d,k) \in \mathbb{Z}_{p^k+1} \), such that for every homogeneous polynomial \( P \) of degree \( d \) and depth \( k \), \( P(\zeta x) = |\sigma| P(x) \mod 1 \), where \( |\cdot| \) is the natural map from \( \mathbb{Z}_{p^k+1} \) to \( \{0,1,\ldots,p^{k+1}-1\} \subset \mathbb{Z} \).

**Proof.** Let \( P \) be a homogeneous polynomial of degree \( d \) and depth \( k \), and let \( \sigma \in \mathbb{Z}_{p^k+1} \) be such that \( P(\zeta x) = |\sigma| P(x) \mod 1 \). By Remark 3.0.6 we know that \( |\sigma| = |\zeta|^d \mod p \).

We also observe that \( P(x) = P(\zeta^{p-1} x) = |\sigma|^{p-1} P(x) \mod 1 \) from which it follows that \( \sigma^{p-1} = 1 \). We claim that \( \sigma \in \mathbb{Z}_{p^k+1} \) is uniquely determined by the two properties
i. $|\sigma| = |c|^d \mod p$, and

ii. $\sigma^{p-1} = 1$.

Suppose to the contrary that there are 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 $t = 1 \mod p$ 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 $x = 1 \mod p$ in $\mathbb{Z}_{p^{k+1}}$. Note that $ta_1, \ldots, ta_{p^k}$ is just a permutation of the first sequence and thus

$$t^{p^k} \prod a_i = \prod a_i.$$ 

Consequently $t^{p^k} = 1$, which combined with $t^0 = t$ implies $t = 1$. □

Lemma 2.1.3 allows us to express every (nonclassical) polynomial as a linear span of monomials of the form $|x_1|^{d_1} \cdots |x_n|^{d_n}$. 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 multivariate (nonclassical) polynomials linearly span the space of multivariate (nonclassical) polynomials. We will present the proof of this theorem in Section 3.3.

**Theorem 3.0.8.** There is a basis for $\text{Poly}(\mathbb{F}^n \to \mathbb{T})$ consisting only of homogeneous multivariate polynomials.

Theorem 3.0.8 allows us to make the extra assumption in the strong decomposition theorem (Theorem 2.3.1) that the resulting polynomial factor $\mathcal{B}$ consists only of homogeneous polynomials.

**Corollary 3.0.9** (Homogeneous Strong Decomposition). Let $d \geq 1$ be an integer, $\delta > 0$ a parameter, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth function. Given any functions
$f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{D}$, there exists a decomposition

$$f_i = g_i + h_i,$$

such that for every $1 \leq i \leq m$,

1. $g_i = \mathbf{E}[f_i|\mathcal{B}]$, where $\mathcal{B}$ is an $r$-regular polynomial factor of degree at most $d$ and complexity $C \leq C_{\text{max}}(p, d, \delta, r(\cdot))$, moreover $\mathcal{B}$ only consists of homogeneous polynomials.

2. $\|h_i\|_{U^{d+1}} \leq \delta$.

To simplify the notation, through the rest of this section we will omit writing “mod 1” in the description of the defined nonclassical polynomials.

### 3.1 Properties of the Derivative Polynomial

We start by proving the following simple observation.

**Claim 3.1.1.** Let $P : \mathbb{F} \to \mathbb{T}$ be a univariate polynomial of degree $d$. Then for every $c \in \mathbb{F}\setminus\{0\}$,

$$\deg \left( P(cx) - |c|^d P(x) \right) < d.$$

**Proof.** By Lemma 2.1.3 it suffices to prove the claim for a monomial $q(x) := \frac{|x|^s}{p^k+1}$ with $k(p-1)+s = d$. Note that $q(cx) - |c|^d q(x)$ takes values in $\frac{1}{p^k} \mathbb{Z}/\mathbb{Z}$ as $|c|^s - |c|^d$ is divisible by $p$. Hence

$$\deg \left( q(cx) - |c|^d q(x) \right) \leq (p-1)(k-1) < d. \quad (3.1)$$

\[ \square \]

It is not difficult to show that the above claim holds for any multivariate polynomial $P : \mathbb{F}^n \to \mathbb{T}$. We provide a proof for the multivariate case since it is a useful observation, although the univariate case suffices for our purposes in this section.
Claim 3.1.2. Let $P : \mathbb{F}^n \to \mathbb{T}$ be a univariate polynomial of degree $d$. Then for every $c \in \mathbb{F}\{0\}$,

$$\deg \left( P(cx) - |c|^d P(x) \right) < d.$$ 

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 Definition 2.1.5 and (Lemma 2.1.6) 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

$$|c|^{-d} \partial Q(y_1, \ldots, y_d) = \partial Q(c^{-1}y_1, \ldots, c^{-1}y_d) = D_{c^{-1}y_1} D_{c^{-1}y_2} \cdots D_{c^{-1}y_d} Q(x)$$

$$= \sum_{S \subseteq [d]} (-1)^{|S|} Q \left( c^{-1} \sum_{i \in S} y_i \right) = \sum_{S \subseteq [d]} (-1)^{|S|} P \left( \sum_{i \in S} y_i \right)$$

$$= \partial P(y_1, \ldots, y_d),$$

This implies that $\partial_d(Q - |c|^d P) \equiv 0$ and thus $\deg(Q - |c|^d P) < d$. \hfill \qedsymbol

### 3.2 A Homogeneous Basis, the Univariate Case

First we prove Theorem 3.0.8 for univariate polynomials.

**Lemma 3.2.1.** There is a basis of homogeneous univariate polynomials for $\text{Poly}(\mathbb{F} \to \mathbb{T})$.

Proof. We will prove by induction on $d$ that there is a basis $\{h_1, \ldots, h_d\}$ of homogeneous univariate polynomials for $\text{Poly}_{\leq d}(\mathbb{F} \to \mathbb{T})$ for every $d$. Let $\zeta$ be a fixed generator of $\mathbb{F}^*$. 42
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 $\sigma_d$. The base case of $d \leq p - 1$ is trivial as $\text{Poly}_{\leq p-1}(\mathbb{F} \to \mathbb{T})$ 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

$$f(x) := \frac{|\zeta x|^s}{p^k} - \frac{|\zeta|^s |x|^s}{p^k}.$$  

Claim 3.1.1 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 $e \neq s \mod (p-1)$:

$$f(x) = \sum_{\ell=1}^{k-1} a_\ell \frac{|x|^s}{p^\ell} + \sum_{e<d, e \neq s \mod (p-1)} b_e h_e(x).$$

Set $A := |\zeta|^s + \sum_{\ell=1}^{k-1} a_\ell p^{k-\ell}$, so that

$$\frac{|\zeta x|^s}{p^k} - A \frac{|x|^s}{p^k} = \sum_{e<d, e \neq d \mod (p-1)} b_e h_e(x). \quad (3.2)$$

By the induction hypothesis, for $e < d$, $h_e(\zeta x) = \sigma_e h(x)$ where $\sigma_e = |\zeta|^e \mod p$, and thus as $A = |\zeta|^s \mod p$, we have $\sigma_e \neq A \mod p$ when $e \neq s \mod (p-1)$. Consequently,

$$\sum_{e<d, e \neq d \mod (p-1)} b_e h_e(x) = \sum_{e<d, e \neq d \mod (p-1)} \frac{b_e}{\sigma_e - A} (\sigma_e - A) h_e(x)$$

$$= \sum_{e<d, e \neq d \mod (p-1)} \frac{b_e}{\sigma_e - A} (h_e(\zeta x) - Ah_e(x)).$$

43
Combing this with (3.2) we conclude that

\[
h_d(x) := |x|^s - \frac{1}{p^k} \sum_{e<d, e \neq d \text{ mod } (p-1)} \frac{b_e}{\sigma_e - A} h_e(x),
\]

satisfies

\[
h_d(\zeta x) = Ah_d(x).
\]

\[\square\]

3.3 A Homogeneous Basis, the Multivariate Case

We are now ready to prove the main result of this section.

**Theorem 3.0.8 (restated).** There is a basis for \( \text{Poly}(\mathbb{F}^n \to \mathbb{T}) \) consisting only of homogeneous multivariate polynomials.

**Proof.** We will show by induction on the degree \( d \), that every degree \( d \) monomial can be written as a linear combination of homogeneous polynomials. The base case of \( d < p \) is trivial as such monomials are classical and thus homogeneous themselves. Consider a (nonclassical) monomial \( M(x_1, \ldots, x_n) = |x_1|^{s_1} \cdots |x_n|^{s_n} \) of degree \( d = s_1 + \cdots + s_n + (p-1)(k-1) \). For every \( i \in [n] \) let \( g_i(x_i) := h_{s_i+(p-1)(k-1)}(x_i) \) where \( h_{s_i+(p-1)(k-1)}(\cdot) \) is the homogeneous univariate polynomial from Lemma 3.2.1. Every \( g_i \) takes values in \( \frac{1}{p^k} \mathbb{Z}/\mathbb{Z} \), and thus corresponds to a polynomial \( G_i : \mathbb{F} \to \mathbb{Z}_{p^k} \). Define \( F : \mathbb{F}^n \to \mathbb{Z}_{p^k} \) as

\[
F(x_1, \ldots, x_n) := G_1(x_1) \cdots G_n(x_n),
\]

and \( f : \mathbb{F}^n \to \mathbb{T} \) as

\[
f(x_1, \ldots, x_n) := \frac{F(x_1, \ldots, x_n)}{p^k}.
\]

It is simple to verify that \( \deg(f) = s_1 + \cdots + s_n + (p-1)(k-1) = d \), it has only one
monomial of \( \deg(f) \), which is \( \frac{|x_1|^{s_1} \cdots |x_n|^{s_n}}{p^k} = M(x_1, \ldots, x_n) \), and it is homogeneous. Thus \( M(x_1, \ldots, x_n) - f(x_1, \ldots, x_n) \) is of degree less than \( d \) and by the induction hypothesis can be written as a linear combination of homogeneous polynomials. \( \square \)
CHAPTER 4
DISTRIBUTION OF POLYNOMIAL FACTORS OVER LINEAR FORMS

A main contribution of this dissertation is a near-orthogonality result for (non-classical) polynomial factors of high rank \(^1\). Such a statement was proved in \([43, 78]\) for systems of linear forms corresponding to repeated derivatives or equivalently Gowers norms, and in \([51]\) for the case when the field is of high characteristic but with arbitrary system of linear forms and in \([15]\) for (non-classical) polynomials but only in the case of systems of affine linear forms. In Theorem 2.2.23 we establish the near-orthogonality over any arbitrary system of linear forms. Before stating this theorem we need to introduce the notion of consistency.

**Definition 4.0.1 (Consistency).** Let \(L = \{L_1, \ldots, L_m\} \) be a system of linear forms. A vector \((\beta_1, \ldots, \beta_m) \in \mathbb{T}^m\) is said to be \((d, k)\)-consistent with \(L\) if there exists a homogeneous polynomial \(P\) of degree \(d\) and depth \(k\) and a point \(X\) such that \(P(L_i(X)) = \beta_i\) for every \(i \in [m]\). Let \(\Phi_{d,k}(L)\) denote the set of all such vectors.

It is immediate from the definition that \(\Phi_{d,k}(L) \subseteq \mathbb{U}_{k+1}^m\) is a subgroup of \(\mathbb{T}^m\), or more specifically, a subgroup of \(\mathbb{U}_{k+1}^m\). Let

\[
\Phi_{d,k}(L)^\perp := \left\{ (\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m : \forall (\beta_1, \ldots, \beta_m) \in \Phi_{d,k}(L), \sum \lambda_i \beta_i = 0 \right\}.
\]

Equivalently \(\Phi_{d,k}(L)^\perp\) is the set of all \((\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m\) such that \(\sum_{i=1}^m \lambda_i P(L_i(X)) \equiv 0\) for every homogeneous polynomial \(P\) of degree \(d\) and depth \(k\).

**Theorem 2.2.23 (rephrased).** Let \(L_1, \ldots, L_m\) be linear forms on \(\ell\) variables and let \(B = (P_1, \ldots, P_C)\) be an \(\varepsilon\)-uniform polynomial factor for some \(\varepsilon \in (0, 1]\) defined only by homogeneous polynomials. For every tuple \(\Lambda\) of integers \((\lambda_{i,j})_{i \in [C], j \in [m]}\), define \(P_\Lambda : (\mathbb{F}^n)^\ell \rightarrow \mathbb{T}\)

---

\(^1\) The work in this chapter is based on the work \([50]\)
as

\[ P_\Lambda(X) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(X)). \]

Then one of the following two statements holds:

- \( P_\Lambda \equiv 0 \).
- \( P_\Lambda \) is non-constant and \( \left| E_{X \in (\mathbb{F}^n)^t} \left[ e(P_\Lambda) \right] \right| < \varepsilon \).

Furthermore \( P_\Lambda \equiv 0 \) if and only if for every \( i \in [C] \), we have \( (\lambda_{i,j})_{j \in [m]} \in \Phi_{d_i,k_i}(\mathcal{L})^\perp \) where \( d_i, k_i \) are the degree and depth of \( P_i \), respectively. We will present the proof of Theorem 2.2.23 in Section 4.2.

**Remark 4.0.2.** By Corollary 2.2.14 the assumption of \( \varepsilon \)-uniformity in Theorem 2.2.23 is satisfied for every factor \( B \) of rank at least \( r_{2.13}(d, \varepsilon) \). However, we would like to point out that in Theorem 2.2.23 by using the assumption of \( \varepsilon \)-uniformity instead of the assumption of high rank, we are able to achieve the quantitative bound of \( \varepsilon \) on the bias of \( P_\Lambda \).

**Remark 4.0.3.** In Theorem 2.2.23 in the second case where \( P_\Lambda \) is non-constant, it is possible to deduce a more general statement that \( \|e(P_\Lambda)\|_{U_t}^{2t} < \varepsilon \) for every \( t \leq \deg(P_\Lambda) \). Indeed assume that \( P_\Lambda \) is non-constant, and consider the derivative

\[ D_{Y_1} \ldots D_{Y_t} P_\Lambda(X) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} \sum_{S \subseteq [t]} (-1)^{|S|} P_i(L_j(X + \sum_{r \in S} Y_r)), \quad (4.1) \]

where \( Y_i := (y_{i,1}, \ldots, y_{i,t}) \in (\mathbb{F}^n)^t \). Notice that for every choice of \( j \in [m] \) and \( S \subseteq [t] \), \( L_j(X + \sum_{r \in S} Y_r) \) is an application of a linear form on the vector

\[ (x_1, \ldots, x_t, y_1, \ldots, y_1, \ldots, y_t, 1, \ldots, y_t, \ell) \in (\mathbb{F}^n)^{t+1}, \]
and since by Lemma 2.1.6 the polynomial $\partial P_\Lambda$ is nonzero, Theorem 2.2.23 implies

$$\|e(P_\Lambda)\|_{U^t}^{\partial t} = \mathbb{E}_{h_1, \ldots, h_t}[e(\partial_t P_\Lambda(h_1, \ldots, h_t))] \leq \varepsilon.$$  

### 4.1 Near-equidistribution over Linear Forms

It is well-known that statements similar to that of Theorem 2.2.23 imply “near-equidistributions” of the joint distribution of the polynomials applied to linear forms. Consider a highly uniform polynomial factor of degree $d > 0$, defined by a tuple of homogeneous polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective degrees $d_1, \ldots, d_C$ and depths $k_1, \ldots, k_C$, and let $\mathcal{L} = (L_1, \ldots, L_m)$ be a collection of linear forms on $\ell$ variables. As we mentioned earlier, we are interested in the distribution of the random matrix

$$
\begin{pmatrix}
P_1(L_1(X)) & P_2(L_1(X)) & \ldots & P_C(L_1(X)) \\
P_1(L_2(X)) & P_2(L_2(X)) & \ldots & P_C(L_2(X)) \\
\vdots & \vdots & \ddots & \vdots \\
P_1(L_\ell(X)) & P_2(L_\ell(X)) & \ldots & P_C(L_\ell(X))
\end{pmatrix},
$$

(4.2)

where $X$ is the uniform random variable taking values in $\mathbb{F}^n$. Note that by the definition of consistency, for every $1 \leq i \leq C$, the $i$-th column of this matrix must belong to $\Phi_{d_i, k_i}(\mathcal{L})$. Theorem 4.1.1 below says that Equation (4.2) is “almost” uniformly distributed over the set of all matrices satisfying this condition. The proof of Theorem 4.1.1 is standard and is identical to the proof of [15, Theorem 3.10] with the only difference that it uses Theorem 2.2.23 instead of the weaker near-orthogonality theorem of [15].

**Theorem 4.1.1 (Near-equidistribution over Linear Forms).** Given $\varepsilon > 0$, let $\mathcal{B}$ be an $\varepsilon$-uniform polynomial factor of degree $d > 0$ and complexity $C$, that is defined by a tuple of homogeneous polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ having respective degrees $d_1, \ldots, d_C$ and
depths $k_1, \ldots, k_C$. Let $\mathcal{L} = (L_1, \ldots, L_m)$ be a collection of linear forms on $\ell$ variables.

Suppose $(\beta_{i,j})_{i \in [C], j \in [m]} \in \mathbb{T}^{C \times m}$ is such that $(\beta_{i,1}, \ldots, \beta_{i,m}) \in \Phi_{d_i,k_i}(\mathcal{L})$ for every $i \in [C]$. Then

$$\Pr_{X \in (\mathbb{F}_p^n)\ell} [P_i(L_j(X)) = \beta_{i,j} \forall i \in [C], j \in [m]] = \frac{1}{K} \pm \varepsilon,$$

where $K = \prod_{i=1}^{C} |\Phi_{d_i,k_i}(\mathcal{L})|.$

Proof. We have

$$\Pr[P_i(L_j(X)) = \beta_{i,j} \forall i \in [C], \forall j \in [m]]$$

$$= \mathbb{E} \left[ \prod_{i,j} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j} = 0}^{p^{k_i+1} - 1} e(\lambda_{i,j}(P_i(L_j(X)) - \beta_{i,j})) \right]$$

$$= \left( \prod_{i \in [C]} p^{-(k_i+1)} \right)^m \sum_{(\lambda_{i,j})} \mathbb{E} \left[ - \sum_{i,j} \lambda_{i,j} \beta_{i,j} \right] \mathbb{E} \left[ \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(X)) \right],$$

where the outer sum is over $(\lambda_{i,j})_{i \in [C], j \in [m]}$ with $\lambda_{i,j} \in [0, p^{k_i+1} - 1]$. Let $\Lambda_i = \Phi_{d_i,k_i}(\mathcal{L}) \cap [0, p^{k_i+1} - 1]^m$, and note that $|\Lambda_i||\Phi_{d_i,k_i}| = p^{m(k_i+1)}$. Since $(\beta_{i,1}, \ldots, \beta_{i,m}) \in \Phi_{d_i,k_i}(\mathcal{L})$ for every $i \in [C]$, it follows that $\sum_{i,j} \lambda_{i,j} \beta_{i,j} = 0$ if $(\lambda_{i,1}, \ldots, \lambda_{i,m}) \in \Lambda_i$ for all $i \in [C]$. If the latter holds, then the expected value in the above expression is 0, and otherwise by Theorem 2.2.23, it is bounded by $\varepsilon$. Hence the above expression can be approximated by

$$p^{-m \sum_{i=1}^{C} (k_i+1)} \left( \prod_{i=1}^{C} |\Lambda_i| \right) \pm \varepsilon p^{m \sum_{i=1}^{C} (k_i+1)} = \frac{1}{K} \pm \varepsilon.$$
4.2 Strong near-orthogonality: Proof of Theorem 2.2.23

We prove Theorem 2.2.23 in this section. Our proof uses similar derivative techniques as used in [15], but in order to handle the general setting we will need a few technical claims which we present first. Recall that $|L| = \sum_{i=1}^{\ell} |\lambda_i|$ for a linear form $L = (\lambda_1, \ldots, \lambda_{\ell})$.

**Claim 4.2.1.** Let $d > 0$ be an integer, and $L = (\lambda_1, \ldots, \lambda_{\ell}) \in \mathbb{F}^\ell$ be a linear form on $\ell$ variables. There exists linear forms $L_i = (\lambda_{i,1}, \ldots, \lambda_{i,\ell}) \in \mathbb{F}^\ell$ for $i = 1, \ldots, m$, and 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 $d$.

In the case $|L| \leq d$ there is nothing to prove. Assume otherwise that $|L| > d$. We will use the fact that for every choice of $y_1, \ldots, y_{|L|} \in \mathbb{F}^n$,

$$\sum_{S \subseteq [|L|]} (-1)^{|S|} P \left( \sum_{i \in S} y_i \right) \equiv 0. \quad (4.3)$$

Let $X = (x_1, \ldots, x_{\ell}) \in (\mathbb{F}^n)^\ell$. Setting $|\lambda_i|$ of the vectors $y_1, \ldots, y_{|L|}$ to $x_i$ for every $i \in [\ell]$, Equation (4.3) implies

$$P(L(X)) = \sum_{i} \alpha_i P(M_i(X)),$$

where $M_i = (\tau_{i,1}, \ldots, \tau_{i,\ell})$, $|M_i| \leq |L| - 1$ and for every $j \in [\ell]$, $|\tau_{i,j}| \leq |\lambda_j|$. Repeatedly applying the same process to every $M_i$ with $|M_i| > d$ we arrive at the desired expansion. \qed
The next claim shows that we can further simplify the expression given in Claim 4.2.1. 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 4.2.2.** For any linear form $L \in \mathbb{F}^\ell$ and integer $d > 0$, there is a collection of coefficients $\{a_{M,c} \in \mathbb{Z}\}_{M \in \mathcal{L}_d, c \in \mathbb{F}^*}$ such that for every degree-$d$ polynomial $P : \mathbb{F}^n \to \mathbb{T}$,

$$P(L(X)) = \sum_{M \in \mathcal{L}_d, c \in \mathbb{F}^*} a_{M,c} P(cM(X)). \quad (4.4)$$

**Proof.** Similar to the proof of Claim 4.2.1 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 Claim 4.2.1 for the degree-$d$ polynomial $R(x) := P(c^{-1}x)$ and the linear form $L'$ we can write

$$P(L(X)) = P(c^{-1}L'(X)) = \sum_i \beta_i P(c^{-1}M_i(X)), \quad (4.5)$$

where for every $i$, $M_i = (\lambda_{i,1}, \ldots, \lambda_{i,\ell})$ satisfies $|M_i| \leq d$, and for every $j \in [\ell]$, $\lambda_{i,j} \leq \lambda_j$. Let $\mathcal{I}$ denote the set of indices $i$ such that the leading nonzero entry of $M_i$ is one. Then

$$P(L(x)) = \sum_{i \in \mathcal{I}} \alpha_i P(c^{-1}M_i(X)) + \sum_{j \notin \mathcal{I}} \alpha_j P(c^{-1}M_j(X)). \quad (4.6)$$

Notice that since the leading coefficient of $L'$ is 1, for every $j \notin \mathcal{I}$, $M_j$ has smaller support than $L$ and thus applying the induction hypothesis to the linear forms $c^{-1}M_j$ with $j \notin \mathcal{I}$ concludes the claim. □
Claim 4.2.2 applies to all polynomials of degree $d$. If we also specify the depth then we can obtain a stronger statement.

**Claim 4.2.3.** For every $d,k$, every system of linear forms $\{L_1, \ldots, L_m\}$, and constants $\{\lambda_i \in \mathbb{Z}\}_{i \in [m]}$, there exists $\{a_M \in \mathbb{Z}\}_{M \in \mathcal{L}_d}$ such that the following is true for every $(d,k)$-homogeneous polynomial $P : \mathbb{F}^n \to \mathbb{T}$:

- $\sum_{i=1}^m \lambda_i P(L_i(X)) \equiv \sum_{M \in \mathcal{L}_d} a_M P(M(X))$;

- For every $M$ with $a_M \neq 0$, we have $|M| \leq \deg(a_M P)$.

**Proof.** The proof is similar to that of Claim 4.2.2, except that now we repeatedly apply Claim 4.2.2 to every term of the form $\lambda P(L(X))$ to express it as a linear combination of $P(cM(X))$ for $M \in \mathcal{L}_{\deg(\lambda P)}$ and $c \in \mathbb{F}^*$. Then we use homogeneity 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 $\sigma_c = \sigma(d,k)^i$. By repeating this procedure we arrive at the desired expansion. □

We are now ready for the proof of our main theorem. For a linear form $L = (\lambda_1, \ldots, \lambda_\ell)$, let $lc(L)$ denote the index of its first nonzero entry, namely $lc(L) \overset{\text{def}}{=} \min_{i: \lambda_i \neq 0} i$.

**Proof of Theorem 2.2.23:** Let $d'$ be the degree of the factor. For every $i \in [C]$, by Claim 4.2.3 we have

$$\sum_{j=1}^m \lambda_{i,j} P_i(L_j(X)) = \sum_{M \in \mathcal{L}_{d'}} \lambda'_{i,M} P_i(M(X))$$

for some integers $\lambda'_{i,M}$ such that $|M| \leq \deg(\lambda'_{i,M} P_i)$ if $\lambda'_{i,M} \neq 0$. The simplifications of Claim 4.2.3 depend only on the degrees and depths of the polynomials. Hence if $\lambda'_{i,M} = 0$ for all $M \in \mathcal{L}_{d'}$, then $(\lambda_{i,1}, \ldots, \lambda_{i,m}) \in \Phi(d_i,k_i)$. So to prove the theorem, it suffices to show that $P_\Lambda$ has small bias if $\lambda'_{i,M} P_i \neq 0$ for some $M \in \mathcal{L}_{d'}$. 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 $\lambda'_{i,M} P_i \neq 0$. Choose $i^* \in [C]$ and $M^* \in \mathcal{M}$ in the following manner.

- First, let $M^* \in \mathcal{M}$ be such that $\text{lc}(M^*) = \min_{M \in \mathcal{M}} \text{lc}(M)$, 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 $i^* = 1$, $\text{lc}(M^*) = 1$, and let $d := \deg(\lambda'_{1,M^*} P_1)$. 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 $P_\Lambda$ has small bias. We prove this by deriving $P_\Lambda$ in specific directions in a manner that all the terms but $\lambda'_{1,M^*} P_1(M^*(X))$ vanish.

Given a vector $\alpha \in \mathbb{F}^{\ell}$, an element $y \in \mathbb{F}^n$, and a function $P : (\mathbb{F}^n)^\ell \to \mathbb{T}$, define the derivative of $P$ according to the pair $(\alpha, y)$ as

$$D_{\alpha,y} P(x_1, \ldots, x_\ell) \overset{\text{def}}{=} P(x_1 + \alpha_1 y, \ldots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell). \quad (4.8)$$

Note that for every $M \in \mathcal{M}$,

$$D_{\alpha,y} (P_i \circ M)(x_1, \ldots, x_\ell) = P_i(M(x_1, \ldots, x_\ell) + M(\alpha)y) - P_i(M(x_1, \ldots, x_\ell))$$

$$= (D_{\langle M,\alpha \rangle,y} P_i)(M(x_1, \ldots, x_\ell)).$$

Thus if $\alpha$ is chosen such that $\langle M, \alpha \rangle = 0$ then $D_{\alpha,y} (P_i \circ M) \equiv 0$.

Assume that $M^* = (w_1, \ldots, w_\ell)$, where $w_1 = 1$. Let $t := |M^*|$, $\alpha_1 := e_1 = (1, 0, 0, \ldots, 0) \in \mathbb{F}^{\ell}$, and let $\alpha_2, \ldots, \alpha_\ell$ be the set of all vectors of the form
In addition, pick \( \alpha_{t+1} = \cdots = \alpha_d = e_1 \).

Claim 4.2.4.

\[
D_{\alpha_1, y_1} \cdots D_{\alpha_d, y_d} P_\Lambda(X) = \left( D_{\langle M^*, \alpha_1 \rangle, y_1} \cdots D_{\langle M^*, \alpha_d \rangle, y_d} \sum_{i \in [C]: \deg(\lambda_i', M^* P_i) = d} \lambda_i' P_i \right) (M^*(X)).
\]

(4.9)

Proof. Deriving according to \((\alpha_1, y_1), \ldots, (\alpha_t, y_t)\) gives

\[
D_{\alpha_1, y_1} \cdots D_{\alpha_t, y_t} P_\Lambda(X) = \left( D_{\langle M^*, \alpha_1 \rangle, y_1} \cdots D_{\langle M^*, \alpha_t \rangle, y_t} \left( \sum_{i=1}^C \lambda_i' M^* P_i \right) \right) (M^*(X)).
\]

(4.10)

This is because for every \(M = (w_1', \ldots, w_\ell') \in \mathcal{M} \setminus \{M^*\}\), either \(w_1' = 0\) in which case \(\langle M, \alpha_1 \rangle = 0\), or otherwise \(w_1 = 1\) and \(|M| \leq t\), thus there must be an index \(\xi \in [\ell]\) such that \(w'_\xi < w_\xi\); By our choice of \(\alpha_2, \ldots, \alpha_t\) there is \(e, 2 \leq e \leq t\), such that \(\alpha_e = (-w'_\xi, 0, \ldots, 0, 1, 0, \ldots, 0)\) where 1 is in the \(\xi\)-th coordinate and thus \(\langle M_j, \alpha_e \rangle = 0\). Now, the claim follows after additionally deriving according to \((\alpha_{t+1}, y_{t+1}), \ldots, (\alpha_d, y_d)\). \(\square\)

Claim 4.2.4 implies that

\[
\mathbf{E}_{y_1, \ldots, y_d, x_1, \ldots, x_\ell} \left[ e \left( D_{\alpha_1, y_1} \cdots D_{\alpha_d, y_d} P_\Lambda(x_1, \ldots, x_\ell) \right) \right] = \left\| e \left( \sum_{i \in [C]: \deg(\lambda_i', M^* P_i) = d} \lambda_i' M^* P_i \right) \right\|_{U^d}^{2d} \leq \varepsilon^{2d},
\]

where the last inequality holds by the \(\varepsilon\)-uniformity of the polynomial factor. Now the theorem follows from the next claim from [15] which is a repeated application of the Cauchy-Schwarz inequality. We include a proof for self-containment.
Claim 4.2.5 ([15, Claim 3.4]). For any $\alpha_1, \ldots, \alpha_d \in \mathbb{F}^\ell \setminus \{0\},$

$$\mathbb{E}_{y, x_1, \ldots, x_\ell} \left[ e \left( (D_{\alpha_1, y_1} \cdots D_{\alpha_d, y_d} P_\Lambda)(x_1, \ldots, x_\ell) \right) \right] \geq \left( \mathbb{E}_{x_1, \ldots, x_\ell} e \left( P_\Lambda(x_1, \ldots, x_\ell) \right) \right)^{2^d}.$$  

Proof. It suffices to show that for any function $P(x_1, \ldots, x_\ell)$ and nonzero $\alpha \in \mathbb{F}^\ell,$

$$\left| \mathbb{E}_{y, x_1, \ldots, x_\ell} e \left( (D_{\alpha, y} P)(x_1, \ldots, x_\ell) \right) \right| \geq \left| \mathbb{E}_{x_1, \ldots, x_\ell} e \left( P(x_1, \ldots, x_\ell) \right) \right|^2.$$  

Recall that $(D_{\alpha, y} P)(x_1, \ldots, x_\ell) = P(x_1 + \alpha_1 y, \ldots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell).$ Without loss of generality, suppose $\alpha_1 \neq 0.$ We make a change of coordinates so that $\alpha$ can be assumed to be $(1, 0, \ldots, 0).$ More precisely, define $P' : (\mathbb{F}^n)^\ell \to \mathbb{T}$ as

$$P'(x_1, \ldots, x_\ell) = P \left( x_1, \frac{x_2 + \alpha_2 x_1}{\alpha_1}, \frac{x_3 + \alpha_3 x_1}{\alpha_1}, \ldots, \frac{x_\ell + \alpha_\ell x_1}{\alpha_1} \right),$$

so that $P(x_1, \ldots, x_\ell) = P'(x_1, \alpha_1 x_2 - \alpha_2 x_1, \alpha_1 x_3 - \alpha_3 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1),$ and thus

$$(D_{\alpha, y} P)(x_1, \ldots, x_\ell)$$

$$= P'(x_1 + \alpha_1 y, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1) - P'(x_1, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1).$$

Therefore

$$\left| \mathbb{E}_{y, x_1, \ldots, x_\ell} e \left( (D_{\alpha, y} P)(x_1, \ldots, x_\ell) \right) \right|$$

$$= \left| \mathbb{E}_{y, x_1, \ldots, x_\ell} e \left( P'(x_1 + \alpha_1 y, x_2, \ldots, x_\ell) - P'(x_1, x_2, \ldots, x_\ell) \right) \right|$$

$$= \mathbb{E}_{x_1, \ldots, x_\ell} \left| \mathbb{E}_{x_1} e \left( P'(x_1, x_2, \ldots, x_\ell) \right) \right|^2 \geq \mathbb{E}_{x_1, \ldots, x_\ell} \left| \mathbb{E}_{x_1} e \left( P'(x_1, x_2, \ldots, x_\ell) \right) \right|^2$$

$$= \left| \mathbb{E}_{x_1, x_2, \ldots, x_\ell} e \left( P(x_1, x_2, \ldots, x_\ell) \right) \right|^2.$$
The above proof also implies the following proposition just by omitting the application of Claim 4.2.3.

**Proposition 4.2.6.** Let $L_1, \ldots, L_m$ be linear forms on $\ell$ variables and let $B = (P_1, \ldots, P_C)$ be an $\varepsilon$-uniform polynomial factor of degree $d > 0$ for some $\varepsilon \in (0, 1]$ which is defined by only homogeneous polynomials. For every tuple $\Lambda$ of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$, define

$$P_\Lambda(X) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(X)),$$

where $P_\Lambda : (\mathbb{F}^n)^\ell \to \mathbb{T}$. Moreover assume that for every $i \in [C], j \in [m]$, $L_j \in L_{\deg(\lambda_{i,j} P_i)}$. Then, $P_\Lambda$ is of degree $d = \max_{i,j} \deg(\lambda_{i,j} P_i)$ and $\|e(P_\Lambda)\|_{U^d} < \varepsilon$. 

CHAPTER 5
TRUE COMPLEXITY, ON A THEOREM OF GOWERS AND WOLF

The work in this chapter is based on a previous joint work [50]. As discussed in Section 2.4.2, Cauchy-Schwarz complexity is an upper-bound on the true complexity of a system of linear forms. 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. Namely, that it is equal to 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 $L$ is defined as

$$L^d = \left( \prod_{j=1}^{d} \lambda_{i_j} : i_1, \ldots, i_d \in [k] \right) \in \mathbb{F}^k.$$

Later in [38, Theorem 6.1] they 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. In this chapter we verify the Gowers-Wolf conjecture in full generality by proving the following stronger theorem.

**Theorem 5.0.7.** Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. Assume that $L_1^{d+1}$ is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$. 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 $\|f_1\|_{L^{d+1}} \leq \delta$, we have

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{m} f_i(L_i(X)) \right] \right| \leq \varepsilon.$$

Theorem 5.0.7 was conjectured in [38], and left open even in the case of large $|\mathbb{F}|$. In [51], a partial near-orthogonality result is proved and used to prove Theorem 5.0.7 in the case where $|\mathbb{F}|$ is greater or equal to the Cauchy-Schwarz complexity of the system of linear form. Theorem 5.0.7 establishes this in its full generality using our near-orthogonality result for
regular non-classical polynomials. The following corollary to Theorem 5.0.7 is very useful when combined with the decomposition theorems such as Theorem 2.3.1.

**Corollary 5.0.8.** Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. Assume that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent. 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

$$\left| \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ \prod_{i=1}^m f_i(L_i(X)) \right] - \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ \prod_{i=1}^m g_i(L_i(X)) \right] \right| \leq \varepsilon,$$

**Proof.** Choosing $\delta = \delta(\varepsilon')$ as in Theorem 5.0.7 for $\varepsilon' := \varepsilon/m$, we have

$$\left| \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ \prod_{i=1}^m f_i(L_i(X)) \right] - \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ \prod_{i=1}^m g_i(L_i(X)) \right] \right|$$

$$= \sum_{i=1}^m \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ (f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right]$$

$$\leq \sum_{i=1}^m \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ (f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^m f_j(L_j(X)) \right] \leq m \cdot \delta \leq \varepsilon,$$

where the second inequality follows from Theorem 5.0.7 since $\|f_i - g_i\|_{U^{d+1}} \leq \delta$ and $L_i^{d+1}$ is not in the linear span of $\{L_j^{d+1}\}_{j \in [m] \setminus \{i\}}$. \hfill $\square$

### 5.1 The Gowers-Wolf Conjecture: Proof of Theorem 5.0.7

**Theorem 5.0.7 (rephrased).** Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. Assume that $L_1^{d+1}$ is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$. 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

$$\left| \mathbb{E}_{X \in [\mathbb{F}^n]^k} \left[ \prod_{i=1}^m f_i(L_i(X)) \right] \right| \leq \varepsilon.$$

(5.1)
We prove Theorem 5.0.7 in this section. Note that since $L_1^{d+1}$ is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$ we have that $L_1$ is linearly independent from each $L_j$ for $j > 1$. We claim that we may assume without loss of generality that $L_2, \ldots, L_m$ are pairwise linearly independent as well, namely that $\{L_1, \ldots, L_m\}$ has bounded Cauchy-Schwarz complexity. Assume that there are $j, \ell \in [m] \setminus \{1\}$ and a nonzero $c \in \mathbb{F}$ such that $L_\ell = cL_j$. Then we may define a new function $f_j'(x) = f_j(x)f_\ell(cx)$ so that $f_j'(L_j(X)) = f_j(L_j(X))f_\ell(L_\ell(X))$ and remove the linear form $L_\ell$ and functions $f_j, f_\ell$ from the system. Now

$$
\mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^m f_i(L_i(X)) \right] = \mathbb{E}_{X \in (\mathbb{F}^n)^k} f_j'(L_j(X)) \cdot \left[ \prod_{i \in [m] \setminus \{j, \ell\}} f_i(L_i(X)) \right],
$$

and thus it suffices to bound the right hand side of the above identity. We may repeatedly apply the above procedure in order to achieve a new system of pairwise linearly independent linear forms along with their corresponding functions while keeping $L_1$ and $f_1$ untouched.

Thus we may assume that $\{L_1, \ldots, L_m\}$ is of finite Cauchy-Schwarz complexity $s$ for some $s \leq m < \infty$. The case when $d \geq s$ follows from Lemma 2.4.3, thus we will consider the case when $d < s$. We will use Corollary 3.0.9 to write $f_i = g_i + h_i$ with

1. $g_i = \mathbb{E}[f_i | \mathcal{B}]$, where $\mathcal{B}$ is an $r$-regular polynomial factor of degree at most $s$ and complexity $C \leq C_{\max}(p, s, \eta, \delta, r(\cdot))$ defined by only homogeneous polynomials, where $r$ is a sufficiently fast growing growth function (to be determined later);

2. $\|h_i\|_{U^{s+1}} \leq \eta$.

We first show that by choosing a sufficiently small $\eta$ we may replace $f_i$’s in Equation (5.1) with $g_i$’s.
Claim 5.1.1. Choosing $\eta \leq \frac{\varepsilon}{2m}$ we have

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{m} f_i(L_i(X)) \right] - \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{m} g_i(L_i(X)) \right] \right| \leq \frac{\varepsilon}{2}.$$

Proof. We have

$$\left| \mathbb{E}_{X} \left[ \prod_{i=1}^{m} f_i(L_i(X)) \right] - \mathbb{E}_{X} \left[ \prod_{i=1}^{m} g_i(L_i(X)) \right] \right| = \left| \sum_{i=1}^{m} \mathbb{E}_{X} \left[ h_i(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^{m} f_j(L_j(X)) \right] \right| \leq \sum_{i=1}^{m} \frac{\|h_i\|_{U_{s+1}^k}}{m \cdot \eta} \leq \frac{\varepsilon}{2},$$

where the second inequality follows from Lemma 2.4.3 since the Cauchy-Schwarz complexity of $\mathcal{L}$ is $s$. \hfill $\square$

Thus it is sufficient to bound $\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{m} g_i(L_i(X)) \right] \right|$ by $\varepsilon/2$. For each $i$, $g_i = \mathbb{E}[f_i|\mathcal{B}]$ and thus

$$g_i(x) = \Gamma_i(P_1(x), \ldots, P_C(x)),$$

where $P_1, \ldots, P_C$ are the (nonclassical) homogeneous polynomials of degree $\leq s$ defining $\mathcal{B}$ and $\Gamma_i : \mathbb{T}^C \rightarrow \mathbb{D}$ is a function. Let $k_i$ denote the depth of the polynomial $P_i$ so that by Lemma 2.1.3, each $P_i$ takes values in $U_{k_i+1} = \frac{1}{p_{k_i+1}} \mathbb{Z}/\mathbb{Z}$. Moreover let $\Sigma := \mathbb{Z}_{p_{k_1+1}} \times \cdots \times \mathbb{Z}_{p_{k_C+1}}$. Using the Fourier expansion of $\Gamma_i$ we have

$$g_i(x) = \sum_{\Lambda = (\lambda_1, \ldots, \lambda_C) \in \Sigma} \hat{\Gamma}_i(\Lambda) \cdot e \left( \sum_{j=1}^{C} \lambda_j P_j(x) \right), \quad (5.2)$$
where $\hat{\Gamma}_i(\Lambda)$ is the Fourier coefficient of $\Gamma_i$ corresponding to $\Lambda$. Let $P_{\Lambda} := \sum_{j=1}^C \lambda_j P_j(x)$ for the sake of brevity so that we may write

$$
\mathbb{E}_X \left[ \prod_{i=1}^m g_i(L_i(X)) \right] = \sum_{\Lambda_1, \ldots, \Lambda_m \in \Sigma} \left( \prod_{i=1}^m \hat{\Gamma}_i(\Lambda_i) \right) \cdot \mathbb{E}_X \left[ e \left( \sum_{i=1}^m P_{\Lambda_i}(L_i(X)) \right) \right].
$$

We will show that for a sufficiently fast growing choice of the regularity function $r(\cdot)$ we may bound each term in Equation (5.3) by $\sigma := \varepsilon \frac{\sigma}{|\Sigma|^m}$, thus concluding the proof by the triangle inequality. We will first show that the terms for which $\deg(P_{\Lambda_1}) \leq d$ can be made small.

**Claim 5.1.2.** Let $\Lambda \in \Sigma$ be such that $\deg(P_{\Lambda}) \leq d$. For a sufficiently fast growing choice of $r(\cdot)$ and choice of $\delta \leq \sigma/2$, 

$$
\left| \hat{\Gamma}_1(\Lambda) \right| < \sigma.
$$

**Proof.** It follows from Equation (5.2) that

$$
\hat{\Gamma}_1(\Lambda) = \mathbb{E}_x [g_1(x)e(-P_{\Lambda}(x))] - \sum_{\Lambda' \in \Sigma \setminus \{\Lambda\}} \hat{\Gamma}_1(\Lambda') \cdot \mathbb{E}_x [e(P_{\Lambda'}(x) - P_{\Lambda}(x))].
$$

Note that $\|f_1\|_{U^{d+1}} \leq \delta$ and thus

$$
\left| \mathbb{E}_x [g_1(x)e(-P_{\Lambda}(x))] \right| = \left| \mathbb{E}_x [f_1(x)e(-P_{\Lambda}(x))] \right| 
\leq \|f_1 e(-P_{\Lambda})\|_{U^{d+1}} = \|f_1\|_{U^{d+1}} \leq \delta \leq \sigma/2,
$$

where we used the fact that $g_1 = \mathbb{E}[f_1|\mathcal{B}]$ and the fact that Gowers norms are increasing in $d$. Finally the terms of the form $\hat{\Gamma}_1(\Lambda') \cdot \mathbb{E}_x [e(P_{\Lambda'}(x) - P_{\Lambda}(x))]$ with $\Gamma' \neq \Gamma$ can be made arbitrarily small by choosing a sufficiently fast growing $r(\cdot)$ due to Remark 2.2.17, since $P_{\Lambda'} - P_{\Lambda} = P_{\Lambda' - \Lambda}$ is a nonzero linear combination of the polynomials defining the $r$-regular factor $\mathcal{B}$. 

The above claim allows us to bound the terms from Equation (5.3) corresponding to
tuples \((\Lambda_1, \ldots, \Lambda_m) \in \Sigma^m\) with \(\deg(P_{\Lambda_1}) \leq d\). This is because for such terms \(|\widehat{\Gamma}_1(\Lambda_1)| < \sigma\) by the above claim, and \(|\widehat{\Gamma}_i(\Lambda_i)| \leq 1\) since \(f_i\)'s take values in \(D\). It remains to bound the terms for which \(\deg(P_{\Lambda_1}) > d\). We will need the following claim.

**Lemma 5.1.3.** Assume that \(L_1^{d+1}\) is not in the linear span of \(L_2^{d+1}, \ldots, L_m^{d+1}\), and let \((\Lambda_1, \ldots, \Lambda_m) \in \Sigma^m\) be such that \(\deg(P_{\Lambda_1}) \geq d + 1\). Then

\[
\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) \not\equiv 0.
\]

This combined with Theorem 2.2.23 implies that for a sufficiently fast growing choice of \(r(\cdot), E_X[e(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)))] < \sigma\) which completes the proof of Theorem 5.0.7. Thus, we are left with proving Lemma 5.1.3.

**Proof of Lemma 5.1.3:** Assume to the contrary that \(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) \equiv 0\). Denoting the coordinates of \(\Lambda_i\) by \((\lambda_{i,1}, \ldots, \lambda_{i,C}) \in \Sigma^C\) we have

\[
\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) = \sum_{i \in [m], j \in [C]} \lambda_{i,j} P_{j}(L_i(X)) \equiv 0.
\]

Since the polynomial factor defined by \(P_1, \ldots, P_C\) is \(r\)-regular with a sufficiently fast growing growth function \(r(\cdot)\), Theorem 2.2.23 implies that for every \(j \in [C]\) we must have that

\[
\sum_{i=1}^{m} \lambda_{i,j} P_{j}(L_i(X)) \equiv 0. \quad (5.4)
\]

Since \(\deg(P_{\Lambda_1}) > d\), there must exist \(j \in [C]\) such that \(\lambda_{1,j} \neq 0\) and \(\deg(\lambda_{1,j} P_{j}) > d\). Let \(j^* \in [C]\) be such that \(\deg(\lambda_{1,j^*} P_{j^*})\) is maximized, and let \(d^* := \deg(P_{j^*})\) (note that \(\deg(\lambda_{1,j^*} P_{j^*}) \leq d^*\)). We will first prove that replacing \(P_{j^*}\) with a classical homogeneous polynomial \(Q\) of the same degree, Equation (5.4) for \(j = j^*\) would still hold.

**Claim 5.1.4.** Let \(Q: \mathbb{F}^n \to \mathbb{T}\) be a classical homogeneous polynomial with \(\deg(Q) = d^*\).
Then
\[ \sum_{i=1}^{m} \lambda_{i,j}^* Q(L_i(X)) \equiv 0. \]

**Proof.** Assume to the contrary that \( \sum_{i=1}^{m} \lambda_{i,j}^* Q(L_i(X)) \neq 0 \). By Claim 4.2.2 for degree \( d^* \) and linear forms \( L_1, \ldots, L_m \), we can find a set of coefficients \( \{a_{i,t} \in \mathbb{Z}\}_{i \in [\ell], t \in [m']} \), \( \{c_{i,t} \in \mathbb{F}\{0\}\}_{i \in [\ell], t \in [m']} \) and linear forms \( \{M_t\}_{t \in [m']} \) with \( |M_t| \leq d^* \) such that the first nonzero entry of every \( M_t \) is equal to 1, such that

\[ \sum_{i=1}^{m} \lambda_{i,j}^* Q(L_i(X)) = \sum_{t \in [\ell], t \in [m']} a_{i,t} \lambda_{i,j}^* Q(c_{i,t} M_t(X)) = \sum_{t=1}^{m'} \alpha_t Q(M_t(X)) \neq 0, \quad (5.5) \]

where \( \alpha_t = \sum_{i=1}^{\ell} a_{i,t} \lambda_{i,j}^* |c_{i,t}|^{d^*} \). Here the second equality follows from \( Q \) being classical and homogeneous. Furthermore,

\[ \sum_{i=1}^{m} \lambda_{i,j}^* P_j^*(L_i(X)) = \sum_{i \in [\ell], t \in [m']} a_{i,t} \lambda_{i,j}^* P_j^*(c_{i,t} M_t(X)) = \sum_{t=1}^{m'} \beta_t P_j^*(M_t(X)), \quad (5.6) \]

where \( \beta_t = \sum_{i=1}^{\ell} a_{i,t} \lambda_{i,j}^* \sigma_{i,t} \), \( \{\sigma_{i,t}\}_{i \in [\ell], t \in [m']} \) are integers whose existence follows from the homogeneity of \( P_j^* \). Moreover, by Remark 3.0.6 we know that \( \sigma_{i,t} \equiv |c_{i,t}|^{d^*} \mod p \), and hence

\[ \alpha_t \equiv \beta_t \mod p, \quad \forall t \in [m']. \quad (5.7) \]

Now notice that Equation (5.5) implies that there exists some \( t \in [m'] \) for which \( \alpha_t Q \neq 0 \), which, since \( Q \) is classical, is equivalent to \( \alpha_t \not\equiv 0 \mod p \). Hence also \( \beta_t \not\equiv 0 \mod p \). Let \( T = \{t \in [m'] : \beta_t \not\equiv 0 \mod p\} \), which we just verified is nonempty. Then for \( t \in T \),

\[ \deg(\beta_t P_j^*) = \deg(P_j^*) = d^*; \text{ and for } t \in [m] \setminus T, \deg(\beta_t P_j^*) \leq d^* - (p - 1) < d^*. \]

We can now decompose

\[ \sum_{i=1}^{m} \lambda_{i,j}^* P_j^*(L_i(X)) = \sum_{t \in T} \beta_t P_j^*(M_t(X)) + \sum_{t \in [m] \setminus T} \beta_t P_j^*(M_t(X)). \quad (5.8) \]
By Proposition 4.2.6, the first sum in Equation (5.8) is a nonzero polynomial of degree $d^*$, and by our previous argument, the sum for $t \notin T$ is a polynomial of degree less than $d^*$. Hence, we get that $\sum_{i=1}^{m} \lambda_{i,j^*} P_{j^*}(L_i(X))$ is a nonzero polynomial of degree $d^*$, which is a contradiction to our assumption. \[\square\]

We have proved that for every choice of a degree-$d^*$ classical homogeneous polynomial $Q$, $\sum_{i=1}^{m} \lambda_{i,j^*} Q(L_i(X)) \equiv 0$. Now choosing the polynomial $Q(x(1), \ldots, x(n)) = x(1) \cdots x(d^*)$ and looking at the coefficients of the monomials of degree $d^*$, we have

$$\sum_{i=1}^{m} \lambda_{i,j^*} L_i^{d^*} \equiv 0.$$ 

Recalling that $\lambda_{1,j^*} \neq 0$ and that $d^* > d$, this means that $L_{1}^{d+1}$ can be written as a linear combination of $L_2^{d+1}, \ldots, L_m^{d+1}$, a contradiction. \[\square\]
CHAPTER 6
TESTABLE AFFINE-INVARIANT PROPERTIES

The results in this chapter are based on previous work [15]. The field of property testing, as initiated by [20, 8] and defined formally by [66, 29], 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. 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 a large number of natural properties satisfy this strong requirement; see e.g. the surveys [24, 65, 64, 70] for a general overview.

Let \([R]\) denote the set \(\{1, \ldots, R\}\). Given a property \(P\) of functions in \(\{\mathbb{F}^n \to [R] \mid n \in \mathbb{Z}_{\geq 0}\}\), we say that \(f : \mathbb{F}^n \to [R]\) is \(\varepsilon\)-far from \(P\) if

\[
\min_{g \in P} \Pr_{x \in \mathbb{F}^n} [f(x) \neq g(x)] > \varepsilon,
\]

and we say that it is \(\varepsilon\)-close otherwise.

**Definition 6.0.5 (Testability).** A property \(P\) is said to be testable (with one-sided error) if there are functions \(q : (0, 1) \to \mathbb{Z}_{\geq 0}\), \(\delta : (0, 1) \to (0, 1)\), 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 \(f \in P\) and rejects with probability at least \(\delta(\varepsilon)\) if \(f\) is \(\varepsilon\)-far from \(P\). If, furthermore, \(q\) is a constant function, then \(P\) is said to be proximity-obliviously testable (PO testable).

The term proximity-oblivious testing is coined by Goldreich and Ron in [32]. As an example of a testable (in fact, PO testable) property, let us recall the famous result by

Blum, Luby and Rubinfeld [20] which initiated this line of research. They showed that linearity of a function \( f : \mathbb{F}^n \rightarrow \mathbb{F} \) is testable by a test which makes 3 queries. This test accepts if \( f \) is linear and rejects with probability \( \Omega(\varepsilon) \) if \( f \) is \( \varepsilon \)-far from linear.

### 6.1 Background

#### 6.1.1 Algebraic Property Testing, Affine Invariance

We say that a property \( \mathcal{P} \subseteq \{ \mathbb{F}^n \rightarrow [R] \} \) is linear-invariant if it is the case that for any \( f \in \mathcal{P} \) and for any linear transformation \( L : \mathbb{F}^n \rightarrow \mathbb{F}^n \), it holds that \( f \circ L \in \mathcal{P} \). Similarly, an affine-invariant property is closed under composition with affine transformations \( A : \mathbb{F}^n \rightarrow \mathbb{F}^n \) (an affine transformation \( A \) is of the form \( L + c \) where \( L \) is linear and \( c \) is a constant). The property of a function \( f : \mathbb{F}^n \rightarrow \mathbb{F} \) being affine is testable by a simple reduction to [20], 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) [8, 9, 23, 66, 4] 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 [57] made explicit note of this phenomenon and initiated a general study of the testability of affine-invariant properties (see also [30]). In particular, they asked for necessary and sufficient conditions for the testability of affine-invariant properties.

#### 6.1.2 Localy Characterized Properties

“Local characterization” is a necessary condition for PO testability. For a PO testable property \( \mathcal{P} \), if a function \( f \) does not satisfy \( \mathcal{P} \), then by Definition 6.0.5, the tester rejects \( f \) with positive probability. Since the test always accepts functions with the property, there must be \( q \) points \( x_1, \ldots, x_q \in \mathbb{F}^n \) that form a witness for non-membership in \( \mathcal{P} \). These are the queries that cause the tester to reject. Thus, denoting \( \sigma = (f(x_1), \ldots, f(x_q)) \in [R]^q \), we
say that $C = (x_1, x_2, \ldots, x_q; \sigma)$ forms a $q$-local constraint for $\mathcal{P}$. This means that whenever the constraint is violated by a function $g$, i.e., $(g(x_1), \ldots, g(x_q)) = \sigma$, we know that $g$ is not in $\mathcal{P}$. A property $\mathcal{P}$ is $q$-locally characterized if there exists a collection of $q$-local constraints $C_1, \ldots, C_m$ such that $g \in \mathcal{P}$ if and only if none of the constraints $C_1, \ldots, C_m$ are violated. It follows from the above discussion that if $\mathcal{P}$ is PO testable with $q$ queries, then $\mathcal{P}$ is $q$-locally characterized. We say $\mathcal{P}$ is locally characterized if it is $q$-locally characterized for some constant $q$.

We now give some examples of locally characterized affine-invariant properties. Consider the property of being affine. 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 automatically 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 $x, y, z$. More generally, consider the property of being a polynomial of degree at most $d$, for some fixed integer $d > 0$. The property is known to be PO testable due to independent work of [56, 53], and their test is based upon a $p \left\lceil \frac{d+1}{p-1} \right\rceil$-local characterization. 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 $C_1, \ldots, C_m$, one can design the following $q$-query test: choose a constraint $C_i$ uniformly at random and reject only if the input function violates $C_i$. Clearly, if the input function $f$ is in $\mathcal{P}$, the test always accepts. The question is the probability with which a function $\varepsilon$-far from $\mathcal{P}$ is rejected. We show that for affine-invariant properties, this test always rejects with probability bounded away from zero for every constant $\varepsilon > 0$.

**Theorem 6.1.1.** Every $q$-locally characterized affine-invariant property is proximity obliviously testable with $q$ queries.
6.2 Locality for Affine-Invariant Properties

In the context of affine-invariant properties, we can define the notion of local characterization in a more algebraic way than we did in the introduction. Recall that a hyperplane is an affine subspace of codimension 1.

Definition 6.2.1 (Locally characterized properties). An affine-invariant property $\mathcal{P} \subset \{ \mathbb{F}^n \to [R] : n \geq 0 \}$ is said to be locally characterized if both of the following hold:

- For every function $f : \mathbb{F}^n \to [R]$ in $\mathcal{P}$ and every hyperplane $A \subseteq \mathbb{F}^n$, $f|_A \in \mathcal{P}$.
- There exists a constant $K \geq 1$ such that if $f : \mathbb{F}^n \to [R]$ does not belong to $\mathcal{P}$ and $n > K$, then there exists a hyperplane $B \subseteq \mathbb{F}^n$ such that $f|_B \notin \mathcal{P}$.

The constant $K$ is said to be the locality of $\mathcal{P}$.

The following observation shows that an affine-invariant property is locally characterized if and only if it can be described using a bounded number of induced affine constraints from the previous section, and hence, is locally characterized in the sense of the introduction.

Lemma 6.2.2. If $\mathcal{P} \subseteq \{ \mathbb{F}^n \to [R] : n \geq 0 \}$ is a locally characterized affine-invariant property with locality $K$, then $\mathcal{P}$ is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a finite collection of induced affine constraints, with each constraint of size $p^K$ on $K + 1$ variables. On the other hand, if $\mathcal{P}$ is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a collection of induced affine constraints with each constraint on $\leq K + 1$ variables, then $\mathcal{P}$ has locality at most $K$.

Finally, we also make formal note of the observation in the introduction that if a property is testable, then it must be locally characterized.

Remark 6.2.3. If $K$ is a fixed integer and $\mathcal{P} \subset \{ \mathbb{F}^n \to [R] \}$ is an affine-invariant property that is testable with $K$ queries, then $\mathcal{P}$ is a locally characterized property with locality $K$.

So, we can view our main result as a converse statement.
6.3 Subspace Hereditary Properties

Just as a necessary condition for PO testability 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:

**Definition 6.3.1** (Subspace hereditary properties). An affine-invariant property $\mathcal{P}$ is said to be (affine) subspace hereditary if for any $f : \mathbb{F}^n \rightarrow \mathbb{R}$ satisfying $\mathcal{P}$, the restriction of $f$ to any affine subspace of $\mathbb{F}^n$ also satisfies $\mathcal{P}$.

In [17], it is shown that every affine-invariant property testable by a “natural” tester is very “close” to a subspace hereditary property\(^2\). Thus, if we gloss over some technicalities, subspace hereditariness is a necessary condition for testability. In the opposite direction, [17] conjectures the following:

**Conjecture 6.3.2** ([17]). Every subspace hereditary property is testable.

Resolving Conjecture 6.3.2 would yield a combinatorial characterization of the (natural) one-sided testable affine-invariant properties, similar to the characterization for testable dense graph properties [6]. Although we are not yet able to confirm or refute the full Conjecture 6.3.2, we can show testability if we make an additional assumption of “bounded complexity”, defined formally in Section 6.4.2.

**Theorem 6.3.3** (Informal). Every subspace hereditary property of “bounded complexity” is testable.

We will formally define the notion of complexity later on in Section 6.4.2, but for now, it suffices to know that it is an integer that we will associate with each property (independent of

---

2. 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 [17] for details, or [6] for the analogous definitions in a graph-theoretic context.
n). Also, $q$-locally characterized properties are of complexity at most $q$. All natural affine-invariant properties that we know of have bounded complexity (in fact, most are locally characterized). So, the subspace hereditary properties not covered by Theorem 6.3.3 seem to be mainly of theoretical interest.

6.4 Main Result

In this section, we describe our main result, Theorem 6.3.3, rigorously. Theorem 6.1.1 follows as a corollary. We first need to set up some notions. Just as a locally characterized property can be described by a list of constraints, subspace hereditary properties can also be described similarly, but here, the size of the list can be infinite. For affine-invariant properties, we can represent the constraints in a very special form, as “induced affine constraints”. We first describe these, then define the notion of complexity, and finally state the theorem.

6.4.1 Affine constraints

A linear form $L = (w_1, w_2, \ldots, w_k)$ is said to be affine if $w_1 = 1$. In this section, linear forms will always be assumed to be affine.

We specify a partial order $\preceq$ among affine forms. We say $(w_1, w_2, \ldots, w_k) \preceq (w'_1, w'_2, \ldots, w'_k)$ if $|w_i| \leq |w'_i|$ for all $i \in [k]$, where $|\cdot|$ is the obvious map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. An affine constraint is a collection of affine forms, with the added technical restriction of being downward-closed with respect to $\preceq$. For future references we state this as the following definition.

Definition 6.4.1 (Affine constraints). An affine constraint of size $m$ on $k$ variables is a tuple $A = (L_1, \ldots, L_m)$ of $m$ affine forms $L_1, \ldots, L_m$ over $\mathbb{F}$ on $k$ variables, where:

(i) $L_1(x_1, \ldots, x_k) = x_1$;

(ii) If $L$ belongs to $A$ and $L' \preceq L$, then $L'$ also belongs to $A$. 

70
Any subspace hereditary property can be described using affine constraints and forbidden patterns, in the following way.

**Definition 6.4.2 (Properties defined by induced affine constraints).**

- An induced affine constraint of size \( m \) on \( \ell \) variables is a pair \((A, \sigma)\) where \( A \) is an affine constraint of size \( m \) on \( \ell \) variables and \( \sigma \in [R]^m \).

- Given such an induced affine constraint \((A, \sigma)\), a function \( f : \mathbb{F}^n \to [R] \) is said to be \((A, \sigma)\)-free if there exist no \( x_1, \ldots, x_\ell \in \mathbb{F}^n \) such that

\[
(f(L_1(x_1, \ldots, x_\ell)), \ldots, f(L_m(x_1, \ldots, x_\ell))) = \sigma.
\]

On the other hand, if such \( x_1, \ldots, x_\ell \) exist, we say that \( f \) induces \((A, \sigma)\) at \( x_1, \ldots, x_\ell \).

- Given a (possibly infinite) collection \( A = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\} \) of induced affine constraints, a function \( f : \mathbb{F}^n \to [R] \) is said to be \( A \)-free if it is \((A^i, \sigma^i)\)-free for every \( i \geq 1 \).

As an example consider the property of having degree at most 1 as a polynomial, for function \( F : \mathbb{F}^n \to \mathbb{F} \). It is easy to see that \( F \) satisfies this property if and only if \( F(x_1) - F(x_1 + x_2) - F(x_1 + x_3) + F(x_1 + x_2 + x_3) = 0 \) for all \( x_1, x_2, x_3 \in \mathbb{F} \). Consequently the property can be defined by the set of induced affine constraints that forbid any values for \( F(x_1), F(x_1 + x_2), F(x_1 + x_3), F(x_1 + x_2 + x_3) \) that do not satisfy the identity \( F(x_1) - F(x_1 + x_2) - F(x_1 + x_3) + F(x_1 + x_2 + x_3) = 0 \).

The connection between affine subspace hereditariness and affine constraints is given by the following simple observation.

**Observation 6.4.3.** An affine-invariant property \( \mathcal{P} \) is subspace hereditary if and only if it is equivalent to the property of \( A \)-freeness for some fixed collection \( A \) of induced affine constraints.
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 \( n \) and a function \( f : \mathbb{F}^n \to \mathbb{R} \) that is not in \( \mathcal{P} \), we include in \( \mathcal{A} \) the constraint \((Af, \sigma_f)\), where \( Af \) is indexed by members of \( \mathbb{F}^n \) and contains \( \{L_z(X_1, \ldots, X_{n+1}) = X_1 + \sum_{i=1}^{n} z_i X_{i+1} : z = (z_1, \ldots, z_n) \in \mathbb{F}^n\} \), and \( \sigma_f \) is just set to \( f \).

Setting \( X_1 = 0 \) and \( X_{i+1} \) to the \( i \)th standard vector \( e_i \) for every \( i \in [n] \) shows that \( f \) is not \((Af, \sigma_f)\)-free. 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. \( \square \)

6.4.2 Complexity of Induced Affine Constraints

Recall from Definition 2.4.2, that if \( \mathcal{L} = \{L_1, \ldots, L_m\} \in \mathbb{F}^k \) contains two linear forms that are multiples of each other (that is \( L_i = \lambda L_j \) for \( i \neq j \) and \( \lambda \in \mathbb{F} \)), then the Cauchy-Schwarz complexity of \( \mathcal{L} \) is infinity. Otherwise, its Cauchy-Schwarz complexity is at most \(|\mathcal{L}| - 2\).

Note, also that sets of affine linear forms are always of finite complexity.

Definition 6.4.4. Given a collection \( \mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i)\} \) of induced affine constraints, we say that \( \mathcal{A} \) is of complexity \( \leq d \) if for each \( i \), the collection of affine forms \( A^i \) is of Cauchy-Schwarz complexity \( \leq d \) according to Definition 2.4.2.

6.4.3 Statement of the main result

Theorem 6.4.5 (Main theorem). For any integer \( d > 0 \) and (possibly infinite) fixed collection \( \mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\} \) of induced affine constraints, each of complexity \( \leq d \), there are functions \( q_{A^i} : (0, 1) \to \mathbb{Z}^+ \), \( \delta_{A^i} : (0, 1) \to (0, 1) \) and a tester \( T \) which, for every \( \varepsilon > 0 \), makes \( q_{A^i}(\varepsilon) \) queries, accepts \( \mathcal{A} \)-free functions and rejects functions \( \varepsilon \)-far from \( \mathcal{A} \)-free with probability at least \( \delta_{A^i}(\varepsilon) \). Moreover, \( q_{A^i} \) is a constant if \( \mathcal{A} \) is of finite size.
We do not have any explicit bounds on $\delta_A$ because the analysis depends on previous work based on ergodic theory. It would of course be interesting to have explicit bounds for some of the properties described in Section 6.1.2.

Let us finally note that Theorem 6.4.5 is quite nontrivial even if $A$ consists only of a single induced affine constraint of complexity greater than 1. Indeed, previously it was not known how to show testability in this case. A more detailed account of previous work is given in Section 6.4.4.

### 6.4.4 Comparison with Previous Work

Our testing result is part of, and culminates a sequence of works investigating the relationship between affine-invariance and testability. As described, Kaufman and Sudan [57] initiated the program. Subsequently, Bhattacharyya, Chen, Sudan, and Xie [14] investigated monotone linear-invariant properties of functions $f : \mathbb{F}_2^n \to \{0, 1\}$, where a property $\mathcal{P}$ is monotone if it satisfies the condition that for any function $g \in \mathcal{P}$, modifying $g$ by changing some outputs from 1 to 0 does not make it violate $\mathcal{P}$. Král, Serra and Vena [60] and, independently, Shapira [69] showed testability for any monotone linear-invariant property characterized by a finite number of linear constraints (of arbitrary complexity). For general non-monotone properties, Bhattacharyya, Grigorescu, and Shapira proved in [17] that affine-invariant properties of functions in $\{\mathbb{F}_2^n \to \{0, 1\}\}$ are testable if the complexity of the property is 1. Earlier this year, Bhattacharyya, Fischer and Lovett in [16] generalized [17] to show that affine-invariant properties of complexity $< p$ are testable. In this paper, we only have to restrict the complexity to be bounded, but the bound can be independent of $p$.

In terms of techniques, the general framework of the proof for testability here is very much the same as in [17] or [16]. However, the main difference here is that we work with collections of non-classical polynomials, rather than classical ones. Because the degrees of non-classical polynomials can change when multiplied by constants, the notions of rank and
regularity are much more subtle. We need to establish a new version of a “polynomial
regularity lemma” which allows us to decompose a given polynomial collection into a high
rank collection of non-classical polynomials. Also, as discussed earlier, we establish a new
equidistribution theorem for non-classical polynomials. We expect that these results will be
of independent interest.

6.5 Locality

In the context of affine-invariant properties, we can define the notion of local characterization
in a more algebraic way than we did in the introduction. Recall that a hyperplane is an affine
subspace of codimension 1.

Definition 6.5.1 (Locally characterized properties). An affine-invariant property \( \mathcal{P} \subset \{ \mathbb{F}^n \to [R] : n \geq 0 \} \) is said to be locally characterized if both of the following hold:

- For every function \( f : \mathbb{F}^n \to [R] \) in \( \mathcal{P} \) and every hyperplane \( A \leq \mathbb{F}^n \), \( f|_A \in \mathcal{P} \).

- There exists a constant \( K \geq 1 \) such that if \( f : \mathbb{F}^n \to [R] \) does not belong to \( \mathcal{P} \) and \( n > K \), then there exists a hyperplane \( B \leq \mathbb{F}^n \) such that \( f|_B \notin \mathcal{P} \).

The constant \( K \) is said to be the locality of \( \mathcal{P} \).

The following observation shows that an affine-invariant property is locally characterized
if and only if it can be described using a bounded number of induced affine constraints from
the previous section, and hence, is locally characterized in the sense of the introduction.

Lemma 6.5.2. If \( \mathcal{P} \subset \{ \mathbb{F}^n \to [R] : n \geq 0 \} \) is a locally characterized affine-invariant
property with locality \( K \), then \( \mathcal{P} \) is equivalent to \( \mathcal{A} \)-freeness, where \( \mathcal{A} \) is a finite collection
of induced affine constraints, with each constraint of size \( p^K \) on \( K + 1 \) variables. On the other
hand, if \( \mathcal{P} \) is equivalent to \( \mathcal{A} \)-freeness, where \( \mathcal{A} \) is a collection of induced affine constraints
with each constraint on \( \leq K + 1 \) variables, then \( \mathcal{P} \) has locality at most \( K \).
Finally, we also make formal note of the observation in the introduction that if a property is testable, then it must be locally characterized.

**Remark 6.5.3.** If $K$ is a fixed integer and $\mathcal{P} \subset \{ \mathbb{F}^n \to \mathbb{R} \}$ is an affine-invariant property that is testable with $K$ queries, then $\mathcal{P}$ is a locally characterized property with locality $K$.

So, we can view our main result as a converse statement.

### 6.6 Strong Near-Orthogonality

We will use the strong near-orthogonality theorem of [15] for affine systems of linear forms, Theorem 2.2.21. Recall that this is a weaker version of Theorem 2.2.23, however we use the former for its slightly simpler statement, and the fact that it does not require using homogeneous non-classical polynomials.

**Remark 6.6.1.** The proof of Theorem 2.2.21 also shows the following. Suppose, in the setting of Theorem 2.2.21, that for every $P_i \in \mathcal{B}$ and $L_j \in A$, either $|L_j| \leq \deg(\lambda_{i,j}P_i)$ or $\lambda_{i,j} = 0$. Then, unless every $\lambda_{i,j} = 0 \pmod{p^{k_{i,j}+1}}$, we have that $P_{A,B,\Lambda}$ is non-constant and $|E[e(P_{A,B,\Lambda}(x_1,\ldots,x_\ell))]| < \varepsilon$. The only modification needed to the above proof is that the transformation from $\Lambda$ to $\Lambda'$ can be omitted.

To show equidistribution of $(P_i(L_j(x_1,\ldots,x_\ell)))$, we can use Theorem 2.2.21 in the same manner we used Theorem 2.2.10 to show the equidistribution of $(P_i(x))$ in Lemma 2.2.18. Before we do so, however, let us give a name to those $\Lambda$ for which the first case of Theorem 2.2.21 holds.

**Definition 6.6.2.** Given an affine constraint $A = (L_1,\ldots,L_m)$ on $\ell$ variables and integers $d,k > 0$ such that $d > k(p - 1)$, the $(d,k)$-dependency set of $A$ is the set of tuples $(\lambda_1,\ldots,\lambda_m) \in [0,p^{k+1} - 1]$ such that $\sum_{i=1}^{m} \lambda_i P(L_i(x_1,\ldots,x_\ell)) \equiv 0$ for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $d$ and depth $k$. 75
Theorem 2.2.21 says that if $B$ is a regular factor, $P_{A,B,A} \equiv 0$ exactly when the first condition holds. In other words:

**Corollary 6.6.3.** Fix an integer $C > 0$, tuples $(d_1, \ldots, d_C) \in \mathbb{Z}_0^C$ and $(k_1, \ldots, k_C) \in \mathbb{Z}_0^C$, and an affine constraint $(L_1, \ldots, L_m)$ on $\ell$ variables. For $i \in [C]$, let $\Lambda_i$ be the $(d_i, k_i)$-dependency set of $A$.

Then, for any polynomial factor $B = (P_1, \ldots, P_C)$, where each $P_i$ has degree $d_i$ and depth $k_i$, and $B$ has rank $> r_{2.2.13} \left( \max_i d_i, \frac{1}{2} \right)$, 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 $i \in [C]$, $(\lambda_{i,1} \mod p^{k_i+1}), \ldots, (\lambda_{i,m} \mod p^{k_i+1}) \in \Lambda_i$.

**Proof.** The “if” direction is obvious. For the “only if” direction, we use Theorem 2.2.21 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 any polynomial $Q_i$ with degree $d_i$ and depth $k_i$. This is equivalent to saying $(\lambda_{i,1} \mod p^{k_i+1}), \ldots, (\lambda_{i,m} \mod p^{k_i+1}) \in \Lambda_i$.

Similar to Theorem 4.1.1 we define consistency as following.

**Definition 6.6.4 (Consistency).** Let $A$ be an affine constraint of size $m$. A sequence of elements $b_1, \ldots, b_m \in \mathbb{T}$ are said to be $(d, k)$-consistent with $A$ 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 $A$, it holds that $\sum_{i=1}^m \lambda_i b_i = 0$.

Given vectors $d = (d_1, \ldots, d_C) \in \mathbb{Z}_0^C$ and $k = (k_1, \ldots, k_C) \in \mathbb{Z}_0^C$, a sequence of vectors $b_1, \ldots, b_m \in \mathbb{T}^C$ are said to be $(d, k)$-consistent with $A$ if for every $i \in [C]$, the elements $b_{1,i}, \ldots, b_{m,i}$ are $(d_i, k_i)$-consistent with $A$.

If $B$ is a polynomial factor, the term $B$-consistent with $A$ is a synonym for $(d, k)$-consistent with $A$ where $d = (d_1, \ldots, d_C)$ and $k = (k_1, \ldots, k_C)$ are respectively the degree and depth sequences of polynomials defining $B$. 
Now, the following is proved identically to Theorem 4.1.1.

**Theorem 6.6.5.** Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity $C$, and rank $r_{2.2.10}(d, \varepsilon)$, that is defined by a tuple of 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 $A = (L_1, \ldots, L_m)$ be an affine constraint on $\ell$ variables.

Suppose $b_1, \ldots, b_m \in \mathbb{T}^C$ are atoms of $\mathcal{B}$ that are $\mathcal{B}$-consistent with $A$. Then

$$\Pr_{x_1, \ldots, x_\ell} [\mathcal{B}(L_j(x_1, \ldots, x_\ell)) = b_j \forall j \in [m]] = \frac{\prod_{i=1}^C |\Lambda_i|}{\|\mathcal{B}\|^m} \pm \varepsilon$$

where $\Lambda_i$ is the $(d_i, k_i)$-dependency set of $A$.

### 6.7 Degree-structural Properties

The conditions required in Theorem 6.1.1 and Theorem 6.3.3 are very general, and so, we expect that they are satisfied by many 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, testable by Theorem 6.1.1. We give the definition below in Definition 6.7.1. First let us list some examples of degree-structural properties. Let $d$ be a fixed positive integer. Each of the following definitions defines a degree-structural property.

- **Degree $\leq d$:** The degree of the function $F : \mathbb{F}^n \to \mathbb{F}$ as a polynomial is at most $d$;

- **Splitting:** A function $F : \mathbb{F}^n \to \mathbb{F}$ *splits* if it can be written as a product of at most $d$ linear functions;

- **Factorization:** A function $F : \mathbb{F}^n \to \mathbb{F}$ *factors* if $F = GH$ for polynomials $G, H : \mathbb{F}^n \to \mathbb{F}$ such that $\deg(G) \leq d - 1$ and $\deg(H) \leq d - 1$;

- **Sum of two products:** A function $F : \mathbb{F}^n \to \mathbb{F}$ is a *sum of two products* if there
are polynomials $G_1, G_2, G_3, G_4$ such that $F = G_1G_2 + G_3G_4$ and $\deg(G_i) \leq d - 1$ for $i \in \{1, 2, 3, 4\}$;

- **Having square root:** A function $F : \mathbb{F}^n \to \mathbb{F}$ has a square root if $F = G^2$ for a polynomial $G$ with $\deg(G) \leq d/2$;

- **Low $d$-rank:** for a fixed integer $r > 0$, a function $F : \mathbb{F}^n \to \mathbb{F}$ 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 $\Gamma : \mathbb{F}^r \to \mathbb{F}$ such that $F = \Gamma(G_1, \ldots, G_r)$.

In fact, roughly speaking, any property that can be described as the property of decomposing into a known structure of low-degree polynomials is degree-structural.

**Definition 6.7.1** (Degree-structural property). Given an integer $c > 0$, a vector of non-negative integers $d = (d_1, \ldots, d_c) \in \mathbb{Z}_{\geq 0}^c$, and a function $\Gamma : \mathbb{F}^c \to \mathbb{F}$, define the $(c, d, \Gamma)$-structured property to be the collection of functions $F : \mathbb{F}^n \to \mathbb{F}$ for which there exist polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ satisfying $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ for all $x \in \mathbb{F}^n$ and $\deg(P_i) \leq d_i$ for all $i \in [c]$.

We say a property $\mathcal{P}$ is degree-structural if there exist integers $\sigma, \Delta > 0$ and a set of tuples $S \subset \{(c, d, \Gamma) \mid c \in [\sigma], d \in [0, \Delta]^c, \Gamma : \mathbb{F}^c \to \mathbb{F}\}$, such that a function $F : \mathbb{F}^n \to \mathbb{F}$ satisfies $\mathcal{P}$ if and only if $F$ is $(c, d, \Gamma)$-structured for some $(c, d, \Gamma) \in S$. We call $\sigma$ the scope and $\Delta$ the max-degree of the degree-structural property $\mathcal{P}$.

It is straightforward to see that the examples above satisfy this definition.

### 6.7.1 Every Degree Structural Property is Testable

In this section, we prove our main result for degree-structural properties stating that if $\mathcal{P}$ is degree-structural (recall Definition 6.7.1), then $\mathcal{P}$ is locally characterized. The proof uses many of the tools established in Section 2.2.3.
Theorem 6.7.2. Every degree-structural property with bounded scope and max-degree is a locally characterized affine-invariant property.

Combining Theorem 6.7.2 with Theorem 6.1.1 implies PO testability for all degree-structural properties.

Proof. Let \( P \) be a degree-structural property with scope \( \sigma \) and max-degree \( \Delta \). Denote by \( S \) the set of tuples \((c, d, \Gamma)\) such that \( c \leq \sigma \) and \( P \) is the union over all \((c, d, \Gamma) \in S\) of \((c, d, \Gamma)\)-structured functions. It is clear that \( P \) is affine-invariant, as having degree bounded by a constant is an affine-invariant property. It is also immediate that \( P \) is closed under taking restrictions to subspaces, since if \( F \) is \((c, d, \Gamma)\)-structured, then \( F \) restricted to any hyperplane is also \((c, d, \Gamma)\)-structured. The non-trivial part of the theorem is to show that the locality is bounded. In other words we need to show that there is a constant \( K \) such that for \( n \geq K \), if \( F : \mathbb{F}^n \to \mathbb{T} \) is a function with \( F|_A \in P \) for every hyperplane \( A \leq \mathbb{F}^n \), then \( F \in P \).

First, let us bound the degree of \( F \). We know that \( F|_A \in P \) for every hyperplane \( A \). Therefore, \( \deg(F|_A) \leq p\sigma\Delta \) for every \( A \), as \( F|_A \) is a function of at most \( \sigma \) polynomials each of degree at most \( \Delta \) over a field of characteristic \( p \). It follows that \( F \) itself is of degree \( \leq p\sigma\Delta \).

Let \( r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) be a function to be fixed later. Define \( r_2 : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \) so that \( r_2(m) > r(C_{2.2.29}^{(r,p\sigma\Delta)}(m + \sigma)) + C_{2.2.29}^{(r,p\sigma\Delta)}(m + \sigma) + p \).

We apply Lemma 2.2.29 to \( \{F\} \) to find an \( r_2 \)-regular polynomial factor \( B \) of degree \( \leq p\sigma\Delta \), defined by polynomials \( R_1, \ldots, R_C : \mathbb{F}^n \to \mathbb{T} \), where \( C \leq C_{2.2.29}^{r_2d} \). Since \( F \) is measurable with respect to \( B \), there exists a function \( \Sigma : \mathbb{T}^C \to \mathbb{F} \), 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 \( \depth(R_i) \). By taking \( K \) to be sufficiently large, we can guarantee the existence of an \( i_0 \in [n] \) such that \( x_{i_0} \) is not involved in any of these
monomials. Consequently \( \deg(R'_i) = \deg(R_i) \) and \( \text{depth}(R'_i) = \text{depth}(R_i) \) for all \( i \in [C] \), where \( R'_1, \ldots, R'_C \) are the restrictions of \( R_1, \ldots, R_C \), respectively, to the hyperplane \( \{x_{i_0} = 0\} \). Also by Lemma 2.2.2, \( R'_1, \ldots, R'_C \) have rank \( > r_2(C) - p \). Since \( F|_{x_{i_0}=0} \in \mathcal{P} \), by definition of \( \mathcal{P} \), there must exist \((c, d, \Gamma) \in S \) with \( c \leq \sigma \) such that

\[
\Sigma(R'_1, \ldots, R'_C) = \Gamma(P_1, \ldots, P_c),
\]

where \( \deg(P_i) \leq d_i \) for all \( i \in [c] \).

Now, apply Lemma 2.2.29 to find an \( r \)-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 Lemma 2.2.29, we obtain a syntactic refinement of \( \{R'_1, \ldots, R'_C\} \). That is, we obtain a tuple \( \mathcal{B}' \) of polynomials \( R'_1, \ldots, R'_C, S_1, \ldots, S_D : \mathbb{F}^n \to \mathbb{T} \) such that it has degree \( \leq p\sigma \Delta \), its rank \( > r(C+D) \), and \( C+D \leq C^{(r)}_{2.2.29}(C+\sigma) \), and for each \( i \in [c] \), \( P_i = \Gamma_i(R'_1, \ldots, R'_C, S_1, \ldots, S_D) \) for some function \( \Gamma_i : \mathbb{T}^{C+D} \to \mathbb{T} \). So for all \( x \in \mathbb{F}^n \),

\[
\Sigma(R'_1(x), \ldots, R'_C(x)) = \Gamma(\Gamma_1(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_D(x)), \ldots, \Gamma_c(R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_D(x))).
\]

Applying Lemma 2.2.18, we see that if the rank of \( \mathcal{B}' \) is \( > r_{2.2.18}(p\sigma \Delta, \varepsilon) \) where \( \varepsilon > 0 \) is sufficiently small (say \( \varepsilon = \|\mathcal{B}'\|/2 \)), then \((R'_1(x), \ldots, R'_C(x), S_1(x), \ldots, S_D(x)) \) acquires every value in its range. Thus, we have the identity

\[
\Sigma(a_1, \ldots, a_c) = \Gamma(\Gamma_1(a_1, \ldots, a_C, b_1, \ldots, b_D), \ldots, \Gamma_c(a_1, \ldots, a_C, b_1, \ldots, b_D)),
\]

for every \( a_i \in \cup_{\text{depth}(R'_i)+1} \) and \( b_i \in \cup_{\text{depth}(S_i)+1} \). Thus, we can substitute \( R_i \) for \( R'_i \) and 0
for $S_i$ in the above equation and still retain the identity

$$F(x) = \Sigma(R_1(x), \ldots, R_C(x))$$

$$= \Gamma(\Gamma_1(R_1(x), \ldots, R_C(x), 0, \ldots, 0), \ldots, \Gamma_c(R_1(x), \ldots, R_C(x), 0, \ldots, 0))$$

$$= \Gamma(Q_1(x), \ldots, Q_c(x))$$

where $Q_i : \mathbb{F}^n \rightarrow \mathbb{T}$ 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 $\depth(R_i) = \depth(R'_i)$, we can apply Theorem 6.7.3 below to conclude that $\deg(Q_i) \leq \deg(P_i) \leq d_i$ for every $i \in [c]$, as long as the rank of $B'$ is $> r_{6.7.3}(p\sigma\Delta)$. Finally, we show that $Q_1, \ldots, Q_c$ map to $U_1 = \iota(F)$ and, so, are classical. Indeed, since $P_1, \ldots, P_c$ are classical, $\Gamma_1, \ldots, \Gamma_c$ must map to $\iota(F)$ on all of $\prod_{i=1}^C \cup_{\depth(R'_i)+1} \times \prod_{i=1}^D \cup_{\depth(S_i)+1} \geq \prod_{i=1}^C \cup_{\depth(R_i)+1} \times \{0\}^D$. Hence, $F \in \mathcal{P}$. □

The following theorem, used in the proof above, shows that a function of a high rank collection of polynomials has the degree one would expect. Thus, it displays yet another way in which high-rank polynomials behave “generically”. The proof is via another application of the near-orthogonality result in Theorem 2.2.21.

**Theorem 6.7.3.** For an integer $d > 0$, let $P_1, \ldots, P_C : \mathbb{F}^n \rightarrow \mathbb{T}$ be polynomials of degree $\leq d$ and rank $> r_{6.7.3}(d, C)$, and let $\Gamma : \mathbb{T}^C \rightarrow \mathbb{T}$ be an arbitrary function. Define the polynomial $F : \mathbb{F}^n \rightarrow \mathbb{T}$ by $F(x) = \Gamma(P_1(x), \ldots, P_C(x))$. Then, for every collection of polynomials $Q_1, \ldots Q_C : \mathbb{F}^n \rightarrow \mathbb{T}$ with $\deg(Q_i) \leq \deg(P_i)$ and $\depth(Q_i) \leq \depth(P_i)$ for all $i \in [C]$, if $G : \mathbb{F}^n \rightarrow \mathbb{T}$ is the polynomial $G(x) = \Gamma(Q_1(x), \ldots, Q_C(x))$, it holds that $\deg(G) \leq \deg(F)$.

**Proof.** Let $f(x) = e(F(x))$ and $\gamma(x_1, \ldots, x_C) = e(\Gamma(x_1, \ldots, x_C))$. Let $D = \deg(F)$. Then, for every $x, y_1, \ldots, y_{D+1} \in \mathbb{F}^n$,

$$\Delta_{y_{D+1}} \cdots \Delta_{y_1} f(x) = 1.$$
We need to show that \( g(x) = e(G(x)) \) also satisfies \( \Delta g_{D+1} \cdots \Delta g_1 g(x) = 1. \)

Let \( k_1, \ldots, k_C \) be the depths of \( P_1, \ldots, P_C \), respectively. Then, each \( P_i \) takes values in \( \mathbb{U}_{k_i+1} \). Let \( \Sigma \) denote the group \( \mathbb{Z}_{p^{k_1+1}} \times \cdots \times \mathbb{Z}_{p^{k_C+1}} \). Considering the Fourier transform of \( \gamma \), we have

\[
f(x) = \gamma(P_1(x), \ldots, P_C(x)) = \sum_{\beta \in \Sigma} \hat{\gamma}(\beta) e\left( \sum_{i=1}^C \beta_i P_i(x) \right).
\]

Next, we look at the derivative.

\[
\Delta g_{D+1} \cdots \Delta g_1 f(x) = \Delta g_{D+1} \cdots \Delta g_1 \left( \sum_{\beta \in \Sigma} \hat{\gamma}(\beta) e\left( \sum_{i=1}^C \beta_i P_i(x) \right) \right)
= \sum_{\alpha \in \Sigma} \prod_{J \subseteq [D+1]} \hat{\gamma}(\alpha_J) e\left( \sum_{i=1}^C \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right) \right) \cdot (-1)^{|J|+1} \sum_{i=1}^C \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right).
\]

Denoting \( \delta(\alpha) = \prod_{J \subseteq [D+1]} \hat{\gamma}(\alpha_J) \) for \( \alpha = (\alpha_J)_{J \subseteq [D+1]} \in \Sigma^{P([D+1])} \), we have

\[
\Delta g_{D+1} \cdots \Delta g_1 f(x) = \sum_{\alpha \in \Sigma^{P([D+1])}} \delta(\alpha) e\left( \sum_{i=1}^C \sum_{J \subseteq [D+1]} (-1)^{|J|+1} \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right) \right). \tag{6.1}
\]

For any \( i \), if there is a \( J \) such that \( |J| > \deg(\alpha_{J,i} P_i) \), we can use Eq. (2.1) to rewrite \( \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right) \) as a linear combination (over \( \mathbb{Z} \)) of \( \left\{ P_i \left( x + \sum_{j \in J'} y_j \right) : |J'| < |J| \right\} \).

We repeat this process until for every \( i \) and \( J \), either \( \alpha_{J,i} = 0 \) or \( |J| \leq \deg(\alpha_{J,i} P_i) \). Denoting by \( \mathcal{A} \) the set of \( \alpha \in \Sigma^{P([D+1])} \) that satisfy this condition, we have obtained a new set of coefficients \( \delta'(\alpha) \) such that

\[
\Delta g_{D+1} \cdots \Delta g_1 f(x) = \sum_{\alpha \in \mathcal{A}} \delta'(\alpha) e\left( \sum_{i=1}^C \sum_{J \subseteq [D+1]} \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right) \right).
\]

Now, the crucial observation is that if instead of \( P_1, \ldots, P_C \), we had \( Q_1, \ldots, Q_C \), the same
decomposition applies.

\[
\Delta y_{D+1} \cdots \Delta y_1 g(x) = \sum_{\alpha \in A} \delta'(A) e \left( \sum_{i=1}^{C} \sum_{J \subseteq [D+1]} \alpha_{J,i} Q_i \left( x + \sum_{j \in J} y_j \right) \right).
\] (6.2)

The reason is that Eq. (6.1) remains valid as is if \( f \) is replaced by \( g \) and the \( P_i \)'s are replaced by \( Q_i \)'s, and furthermore since \( \deg(P_i) \geq \deg(Q_i) \) and \( \depth(P_i) \geq \depth(Q_i) \), the applications of Eq. (2.1) remain valid also. Therefore, Eq. (6.2) is also valid.

But now, we argue that \( \delta'(\cdot) \) are uniquely determined. Let \( k = \max_i k_i \leq d/(p-1) \).

**Claim 6.7.4.** If \( P_1, \ldots, P_C \) are of rank \( > r_{2.2.13}(d, \frac{1}{|A|}) + 1 \), the functions

\[
\left\{ e \left( \sum_{i=1}^{C} \sum_{J \subseteq [D+1]} \alpha_{J,i} P_i \left( x + \sum_{j \in J} y_j \right) \right) : \alpha \in A \right\}
\]

are linearly independent over \( \mathbb{C} \).

**Proof.** Note that all these functions have \( L^2 \)-norm equal to 1. Hence it suffices to show that their pairwise inner products are all bounded in absolute value by \( 1/|A| \). To prove this consider \( \alpha, \beta \in A \), and note that by Theorem 2.2.21 and, in particular, Remark 6.6.1, unless all the \( \alpha_{J,i} - \beta_{J,i} \) are zero,

\[
\left| \mathbb{E} \left[ e \left( \sum_{i=1}^{C} \sum_{J \subseteq [D+1]} (\alpha_{J,i} - \beta_{J,i}) P_i \left( x + \sum_{j \in J} y_j \right) \right) \right] \right| < \frac{1}{|A|}.
\]

Therefore, since \( \Delta y_{D+1} \cdots \Delta y_1 f(x) = 1 \), we must have \( \delta'(\alpha) = 1 \) when \( \alpha \) is the all-zero tuple, and \( \delta'(\alpha) = 0 \) for every non-zero \( \alpha \). Plugging this into Eq. (6.2), we get

\[
\Delta y_{D+1} \cdots \Delta y_1 g(x) = 1.
\]
6.8 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 \((A,\sigma)\), where \( A = (L_1, \ldots, L_m) \) is a sequence of affine forms on \( \ell \) variables and \( \sigma \in [R]^m \). Now, suppose \( \mathbb{F}^n \) is partitioned by a polynomial factor \( \mathcal{B} \) defined by polynomials \( P_1, \ldots, P_C \) of degrees \( d_1, \ldots, d_C \) and depths \( k_1, \ldots, k_C \). Then, observe that if \( b_1, \ldots, b_m \in \mathbb{T}^C \) denote the atoms of \( \mathcal{B} \) containing \( L_1(x_1, \ldots, x_\ell), \ldots, L_m(x_1, \ldots, x_\ell) \) respectively, it must be the case that \( b_1, \ldots, b_m \) are \( \mathcal{B} \)-consistent with \( A \) (as defined in Definition 6.6.4). Thus, to locate where \( f \) might induce \((A,\sigma)\), 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 \( c \).

**Definition 6.8.1.** 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 \( \mathbf{d} = (d_1, \ldots, d_C) \) and depths \( \mathbf{k} = (k_1, \ldots, k_C) \) (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 6.8.2 (Partially induce).** Suppose we are given vectors \( \mathbf{d} = (d_1, \ldots, d_C) \in \mathbb{Z}_{>0}^C \) and \( \mathbf{k} = (k_1, \ldots, k_C) \in \mathbb{Z}_{\geq 0}^C \), a function \( g : \prod_{i \in [C]} \bigcup_{k_i+1} \to \mathcal{P}([R]) \), and an induced affine constraint \((A,\sigma)\) of size \( m \). We say that \( g \) partially \((\mathbf{d},\mathbf{k})\)-induces \((A,\sigma)\) if there exist a sequence \( b_1, \ldots, b_m \in \mathbb{T}^C \) that is \((\mathbf{d},\mathbf{k})\)-consistent with \( A \), and \( \sigma_j \in g(b_j) \) for each \( j \in [m] \).

Definition 6.8.2 is justified by the following trivial observation.
Remark 6.8.3. If \( f : \mathbb{F}^n \to [R] \) induces a constraint \((A, \sigma)\), then for a factor \( B \) defined by polynomials of respective degrees \((d_1, \ldots, d_{|B|}) = d\) and respective depths \((k_1, \ldots, k_{|B|}) = k\), the big picture function \( f_B \) partially \((d, k)\)-induces \((A, \sigma)\).

To handle a possibly infinite collection \( \mathcal{A} \) of affine constraints, we will employ a compactness argument, analogous to one used in [7] to bound the size of the constraint partially induced by the big picture function. Let us make the following definition:

**Definition 6.8.4 (The compactness function).** Suppose we are given positive integers \( C \) and \( d \), and a possibly infinite collection of induced affine constraints \( \mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\} \), where \((A^i, \sigma^i)\) is of size \( m_i \). For fixed \( d = (d_1, \ldots, d_C) \in [d]^C \) and \( k = (k_1, \ldots, k_C) \in [0, \left\lfloor \frac{d-1}{p-1} \right\rfloor]^C \), denote by \( G(d, k) \) the set of functions \( g : \prod_{i=1}^C \bigcup_{k_i+1} \to \mathcal{P}([R]) \) that partially \((d, k)\)-induce some \((A^i, \sigma^i) \in \mathcal{A} \). The compactness function is defined as

\[
\Psi_{\mathcal{A}}(C, d) = \max_{d, k} \max_{g \in G(d, k)} \min_{(A^i, \sigma^i) \text{ partially } (d, k)\text{-induced by } g} m_i
\]

where the outer max is over vectors \( d = (d_1, \ldots, d_C) \in [d]^C \) and \( k = (k_1, \ldots, k_C) \in [0, \left\lfloor \frac{d-1}{p-1} \right\rfloor]^C \). Whenever \( G(d, k) \) 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 \( G(d_1, \ldots, d_C) \) is bounded by \( 2^{Rp^dC} \).

**Remark 6.8.5.** Note that if a function \( g : \mathbb{T}^C \to \mathcal{P}([R]) \) partially \((d, k)\)-induces some constraint from \( \mathcal{A} \) where \( d \in [d]^C \), then \( g \) must belong to \( G(d, k) \), and consequently it will necessarily partially induce some \((A^i, \sigma^i) \in \mathcal{A} \) whose size is at most \( \Psi_{\mathcal{A}}(C, d) \).

### 6.9 Proof of Testability

We prove the main result, Theorem 6.4.5, in this section. In fact, we will show the following.
Theorem 6.9.1. Let $d > 0$ be an integer. Suppose we are given a possibly infinite collection of affine constraints $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\}$ where each $(A^i, \sigma^i)$ is an affine constraint of complexity $\leq d$, and of size $m_i$ on $\ell_i$ variables. Then, there are functions $\ell_A : (0, 1) \to \mathbb{Z}_{>0}$ and $\delta_A : (0, 1) \to (0, 1)$ such that the following is true for any $\varepsilon \in (0, 1)$. If a function $f : \mathbb{F}^n \to [R]$ is $\varepsilon$-far from being $\mathcal{A}$-free, then $f$ induces at least $\delta_A(\varepsilon) p^{\ell_A(\varepsilon)}$ many copies of some $(A^i, \sigma^i)$ with $\ell_i < \ell_A(\varepsilon)$.

Moreover, if $\mathcal{A}$ is locally characterized, then $\ell_A(\varepsilon)$ is a constant independent of $\varepsilon$.

Theorem 6.4.5 immediately follows. Consider the following test: choose uniformly at random $x_1, \ldots, x_{\ell_A(\varepsilon)} \in \mathbb{F}^n$, let $H$ denote the affine space $\left\{x_1 + \sum_{j=2}^{\ell_A(\varepsilon)} \lambda_j x_j : \lambda_j \in \mathbb{F}\right\}$, and check whether $f$ restricted to $H$ is $\mathcal{A}$-free or not, thus making $\leq p^{\ell_A(\varepsilon)}$ queries. By Theorem 6.9.1, if $f$ is $\varepsilon$-far from $\mathcal{A}$-freeness, this test rejects with probability at least $\delta_A(\varepsilon)$.

Proof Theorem 6.9.1:

Preliminaries. Fix a function $f : \mathbb{F}^n \to [R]$ that is $\varepsilon$-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 $\Psi_A$ is the compactness function from Definition 6.8.4:

\[
\begin{align*}
\alpha(C) &= p^{-2dC \Psi_A(C,d)}, \\
\rho(C) &= r_{2.2.13}(d, \alpha(C)), \\
\zeta &= \frac{\varepsilon}{8R}, \\
\Delta(C) &= \frac{1}{16} P \Psi_A(C,d), \\
\eta(C) &= \frac{1}{8p^{dC \Psi_A(C,d)}} \left(\frac{\varepsilon}{24R}\right)^{\Psi_A(C,d)}.
\end{align*}
\]

Decomposing by regular factors. Next, apply Theorem 2.3.6 to the functions $f^{(1)}, f^{(2)}, \ldots, f^{(R)}$ in order to get polynomial factors $\mathcal{B} \succeq_{\text{syn}} \mathcal{B}$ of complexity at most $C_{2.3.6}(\Delta, d, \rho, \zeta, \eta)$, an element $s \in \mathbb{T|\mathcal{B}|\mathcal{B}|}$, and functions $f^{(i)}_1, f^{(i)}_2, f^{(i)}_3 : \mathbb{F}^n \to \mathbb{R}$ for each $86$
\(i \in [R]\) with the desired properties. The sequence of polynomials generating \(B'\) will be denoted by \(P_1, \ldots, P_{|B'|}\). Since \(B'\) is a syntactic refinement, we can assume \(B\) is generated by the polynomials \(P_1, \ldots, P_{|B|}\). Let \(C = |B|\) and \(C' = |B'|\). Note that \(|B| < p^{(k_{\text{max}}+1)C} \leq p^{dC}\), where \(k_{\text{max}} \leq \lfloor (d-1)/(p-1) \rfloor\) is the maximum depth of a polynomial in \(B\). Denote the degree of \(P_i\) by \(d_i\) and the depth of \(P_i\) by \(k_i\).

**Cleanup.** Based on \(B'\) and \(B\), we construct a function \(F : \mathbb{F}^R \rightarrow [R]\) that is \(\frac{\varepsilon}{2}\)-close to \(f\) and hence, still violates \(A\)-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 \(B\), let \(t_c = \arg \max_{j \in [R]} \Pr[f(x) = j | B'(x) = (c, s)]\) be the most popular value inside the corresponding subatom \((c, s)\).

- **Poorly-represented atoms:** If there exists \(i \in [R]\) such that \(|\Pr[f(x) = i | B(x) = c] - \Pr[f(x) = i | B'(x) = (c, s)]| > \zeta\), then set \(F(z) = t_c\) for every \(z\) in the atom \(c\).

- **Unpopular values:** Otherwise, for any \(z\) in the atom \(c\) with \(0 < \Pr_x[f(x) = f(z) | B'(x) = (c, s)] < \zeta\), set \(F(z) = t_c\).

A key property of the cleanup function \(F\) is that it supports a value inside an atom \(c\) of \(B\) only if the original function \(f\) acquires the value on at least an \(\zeta\) fraction of the subatom \((c, s)\). Furthermore as the following lemma shows it is \(\varepsilon/2\)-close to \(f\), and therefore, it is not \(A\)-free.

**Lemma 6.9.2.** The cleanup function \(F\) is \(\varepsilon/2\)-close to \(f\), and therefore, it is not \(A\)-free.

**Proof.** The first step applies to at most \(\zeta|B|\) atoms, since \(B'\) \(\zeta\)-represents \(B\) with respect to each \(f^{(1)}, \ldots, f^{(R)}\). By Lemma 2.2.18, each atom occupies at most \(1/|B| + \alpha(C)\) fraction of the entire domain. So, the fraction of points whose values are set in the first step is at most \(\zeta|B|(1/|B| + \alpha(C)) < 2\zeta\).
In the second step, if \( \Pr[f(x) = f(z) \mid \mathcal{B}'(x) = (c,s)] < \zeta \), then \( \Pr[f(x) = f(z) \mid \mathcal{B}(x) = c] < \Pr[f(x) = f(z) \mid \mathcal{B}'(x) = (c,s)] + \zeta < 2\zeta \). Hence, the fraction of the points 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 \( 2\zeta + \varepsilon/4 < \varepsilon/2 \). \( \square \)

**Locating a violated constraint.** We now want to use \( F \) to “find” the affine constraint induced in \( f \). Setting \( d = (d_1, \ldots, d_C) \) and \( k = (k_1, \ldots, k_C) \), we have by Remark 6.8.3 that the big picture function \( F_B \) of \( F \) will partially \((d,k)\)-induce some constraint from \( \mathcal{A} \), and hence by Remark 6.8.5, it will partially \((d,k)\)-induce some \((A,\sigma) \in \mathcal{A} \) of size \( m \leq \Psi_A(C,d) \) on \( \ell \) 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 \( \sigma \) 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

\[
\ell_A(\varepsilon) = \Psi_A(C,3.6(\Delta,\eta,\rho,\zeta,R),d). \tag{6.3}
\]

Let \( b_1, \ldots, b_m \in \prod_{i=1}^{C+1} \bigcup_{k_i+1} \) correspond to the atoms of \( \mathcal{B} \) where \((A,\sigma) \) is partially \((d,k)\)-induced by \( F_B \). That is, \( b_1, \ldots, b_m \) are consistent with \( A \), and \( \sigma_i \in F_B(b_i) \) for every \( i \in [m] \). Also, let \( b'_1, \ldots, b'_m \in \prod_{i=1}^{C'} \bigcup_{k_{i+1}} \) index the associated subatoms in \( \mathcal{B}' \), obtained by letting \( b'_j = (b_j, s) \) for every \( j \in [m] \).

**Lemma 6.9.3.** 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. This holds because a constant function is of degree \( \leq d_i \). \( \square \)
The main analysis. Let $\mathbf{x} = (x_1, \ldots, x_\ell)$ where $x_1, \ldots, x_\ell$ are independent random variables taking values in $\mathbb{F}^n$ uniformly. Our goal is to prove a lower bound on

$$
\mathbb{P}_\mathbf{x}[f(L_1(\mathbf{x})) = \sigma_1 \land \cdots \land f(L_m(\mathbf{x})) = \sigma_m] = \mathbb{E}_\mathbf{x}[f^{(\sigma_1)}(L_1(\mathbf{x})) \cdots f^{(\sigma_m)}(L_m(\mathbf{x}))].
$$

(6.4)

The theorem obviously follows if the above expectation is larger than the respective $\delta_A(\varepsilon)$. We rewrite the expectation as

$$
\mathbb{E}_\mathbf{x}

\left[
(f^{(\sigma_1)}_1 + f^{(\sigma_1)}_2 + f^{(\sigma_1)}_3)(L_1(\mathbf{x})) \cdots (f^{(\sigma_m)}_1 + f^{(\sigma_m)}_2 + f^{(\sigma_m)}_3)(L_m(\mathbf{x}))
\right].

(6.5)

We can expand the expression inside the expectation as a sum of $3^m$ terms. The expectation of any term involving $f^{(\sigma_j)}_2$ for any $j \in [m]$ is bounded in magnitude by $\|f^{(\sigma_j)}_2\|_{U^{d+1}} \leq \eta(|\mathcal{B}'|)$, by Lemma 2.4.3 and the fact that the complexity of $A$ is bounded by $d$. Hence, the expression (6.5) is at least

$$
\mathbb{E}_\mathbf{x}

\left[
(f^{(\sigma_1)}_1 + f^{(\sigma_1)}_3)(L_1(\mathbf{x})) \cdots (f^{(\sigma_m)}_1 + f^{(\sigma_m)}_3)(L_m(\mathbf{x}))
\right] - 3^m \eta(|\mathcal{B}'|).

\text{(6.6)}

Now, because of the non-negativity of $f^{(\sigma_j)}_1 + f^{(\sigma_j)}_3$ for every $j \in [m]$, this is at least

$$
\mathbb{E}_\mathbf{x}

\left[
(f^{(\sigma_1)}_1 + f^{(\sigma_1)}_3)(L_1(\mathbf{x})) \cdots (f^{(\sigma_m)}_1 + f^{(\sigma_m)}_3)(L_m(\mathbf{x})) \prod_{j \in [m]} 1_{[\mathcal{B}'(L_j) = b'_j]}
\right] - 3^m \eta(|\mathcal{B}'|),

\text{(6.6)}

where $1_{[\mathcal{B}'(L_j) = b'_j]}$ is the indicator function of the event $\mathcal{B}'(L_j(\mathbf{x})) = b'_j$. In other words, now we are only counting patterns that arise from the selected subatoms $b'_1, \ldots, b'_m$. We 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 $f^{(\sigma_k)}_3$ for any $k \in [m]$ is small. Such a term is trivially
bounded from above by

\[
\mathbb{E} \left[ f_3^{(\sigma_k)}(L_k(x)) \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right].
\] (6.7)

Without loss of generality, we assume that \( k = 1 \). This is convenient as by Definition 6.4.1 (i) we have \( L_1(x) = x_1 \). (For other values of \( k \), we can do a change of variables, replacing \( x_1 \) with \( L_k(x) \), so that we can assume \( L_k(x) = x_1 \).) With the assumption \( k = 1 \), the square of (6.7) is equal to the following.

\[
\left( \mathbb{E} \left[ f_3^{(\sigma_k)}(x_1) \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right] \right)^2 
\leq \mathbb{E} \left[ f_3^{(\sigma_k)}(x_1)^2 \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right] \mathbb{E} \left[ x_1 \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right]^2. \tag{6.8}
\]

By Theorem 2.3.6 (vi) and Lemma 2.2.18, we have

\[
\mathbb{E} \left[ f_3^{(\sigma_k)}(x_1)^2 \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right] \leq \Delta^2(C) \mathbb{Pr}_{x_1}[B'(x_1) = b'_k] 
\leq \Delta^2(C) \left( \frac{1}{\|B'\|} + \alpha(C') \right) \leq \frac{2 \Delta^2(C)}{\|B'\|}. \tag{6.9}
\]

Let \( y = (y_2, \ldots, y_\ell) \) where \( y_2, \ldots, y_\ell \) are independent random variables taking values in \( \mathbb{F}^n \) uniformly. The second term in the right hand side of (6.8) is equal to

\[
\frac{1}{\|B'\|^2} \mathbb{E} \left[ x_1 \prod_{i \in [C']} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j} = 0}^{p^{k_i+1}-1} e \left( \lambda_{i,j} \cdot (P_i(L'_j(x)) - b'_{i,j}) \right) \right]^2
\]
\[
\frac{1}{\|B\|^2m} \mathbb{E} \left[ \left( \sum_{(\lambda_{i,j}) \in \Pi_{i,j}[0, p^{k_i+1}-1]} e^{-\sum_{i \in [C^\prime]} \sum_{j \in [m]} \lambda_{i,j} b_{i,j}^\prime} \right) \mathbb{E} \left( \sum_{i \in [C^\prime]} \sum_{j \in [m]} \lambda_{i,j} P_i(L_j(x)) \right)^2 \right] \\
\leq \frac{1}{\|B\|^2m} \sum_{(\lambda_{i,j}),(\tau_{i,j}) \in \Pi_{i,j}[0, p^{k_i+1}-1]} \mathbb{E}_{x,y} \left[ e^{\sum_{i \in [C^\prime]} \sum_{j \in [m]} \lambda_{i,j} P_i(L_j(x))} e^{-\sum_{i \in [C^\prime]} \sum_{j \in [m]} \tau_{i,j} P_i(L_j(x,y))} \right].
\]

(6.10)

We can bound the above using Theorem 2.2.21. 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 \( \Lambda_i \) and \( \Lambda_i' \) denote the \((d_i, k_i)\)-dependency set of \( A \) and \( A' \) respectively.

**Lemma 6.9.4.** For each \( i \), \( |\Lambda_i'| = |\Lambda_i|^2 \cdot p^{k_i+1} \)

**Proof.** Recall that \( L_1(x) = L_1(x_1, y) = x_1 \). For any \( \lambda, \tau \in \Lambda_i \) and any \( \alpha \in [0, p^{k_i+1}-1] \), note that \((\lambda_1 + \alpha \mod p^{k_i+1}), \lambda_2, \ldots, \lambda_m, \tau_1 - \alpha \mod p^{k_i+1}, \tau_2, \ldots, \tau_m) \in \Lambda_i' \). 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 \( \Lambda_i \) to \( \Lambda_i \times \Lambda_i \) that is \( p^{k_i+1} \)-to-1. Suppose \( \sum_{j=1}^{m} \lambda_j Q(L_j(x_1, x_2, \ldots, x_\ell)) + \sum_{j=1}^{m} \tau_j Q(L_j(x_1, y_2, \ldots, y_\ell)) \equiv 0 \) for every polynomial \( Q \) of degree \( d_i \) and depth \( k_i \). Setting \( x_2 = \ldots = x_\ell = 0 \) shows that

\[
\sum_{j=1}^{m} \tau_j Q(L_j(x_1, y_2, \ldots, y_\ell)) = -\left( \sum_{j=1}^{m} \lambda_j \right) Q(x_1),
\]

91
and similarly setting \( y_2 = \ldots = y_\ell = 0 \) shows

\[
\sum_{j=1}^{m} \lambda_j Q(L_j(x_1, x_2, \ldots, x_\ell)) = -\left( \sum_{j=1}^{m} \tau_j \right) Q(x_1).
\]

In particular \( \sum_{j=1}^{m} \lambda_j = -\sum_{j=1}^{m} \tau_j \). Consequently,

\[
(\lambda, \tau) \mapsto \left( \left( -\sum_{j=2}^{m} \lambda_j \pmod{p^{k_i+1}}, \lambda_2, \ldots, \lambda_m \right), \left( -\sum_{j=2}^{m} \tau_i \pmod{p^{k_i+1}}, \tau_2, \ldots, \tau_m \right) \right)
\]

is a map from \( \Lambda'_i \) to \( \Lambda_i \times \Lambda_i \). To see that it is \( p^{k_i+1} \)-to-1, note that

\[
(\lambda_1 + \tau_1 - \gamma \pmod{p^{k_i+1}}, \lambda_2, \ldots, \lambda_m, \gamma, \tau_2, \ldots, \tau_m) \in \Lambda'_i
\]

for every \( \gamma \in [0, p^{k_i+1} - 1] \), and these elements are all mapped to the same element in \( \Lambda_i \times \Lambda_i \).

Applying Theorem 2.2.21 (just as in the proof of Theorem 6.6.5), we get that

\[
(6.10) \leq \frac{1}{\|B'\|^{2m}} \left( \prod_{i=1}^{C'} |\Lambda_i|^{2p^{k_i+1}} + \|B'\|^{2m} \alpha(C') \right) \leq \frac{\prod_{i=1}^{C'} |\Lambda_i|^2}{\|B'\|^{2m}} + \alpha(C').
\]

Combining this with Eq. (6.9) and Eq. (6.8), we obtain

\[
(6.7) \leq 2\Delta(C) \sqrt{\frac{\prod_{i=1}^{C'} |\Lambda_i|^2}{\|B'\|^{2m}}} + \alpha(C').
\]

Finally, we turn to the main term in the expansion of Eq. (6.6). We know from
Lemma 6.9.3 that the subatoms $b'_1, \ldots, b'_m$ are consistent with $A$. Thus

$$E_x \left[ f_1^{(\sigma_1)}(L_1(x)) \cdots f_m^{(\sigma_m)}(L_m(x)) \cdot \prod_{j \in [m]} 1_{[B'(L_j) = b'_j]} \right]$$

$$= \Pr[B'(L_1(x)) = b'_1 \land \cdots \land B'(L_m(x)) = b'_m].$$

$$E_x \left[ f_1^{(\sigma_1)}(L_1(x)) \cdots f_m^{(\sigma_m)}(L_m(x)) | \forall j \in [m], B'(L_j(x)) = b'_j \right]$$

$$\geq \left( \frac{\prod_{i=1}^{C'} |A_i|}{\|B'\|^m} - \alpha(C') \right) \zeta^m. \quad (6.12)$$

Let us justify the last line. The first term is due to the lower bound on the probability from Theorem 6.6.5. The second term in (6.12) follows since each $f_1^{(\sigma_j)}$ is constant on the atoms of $B'$, and because by construction, the big picture function $F_B$ of the cleanup function $F$, on which $(A, \sigma)$ was partially induced, supports a value inside an atom $b$ of $B$ only if the original function $f$ acquires the value on at least an $\zeta$ fraction of the subatom $(c, s)$.

Setting $\beta = (\prod_{i=1}^{C'} |A_i|)^m$ and combining the bounds from (6.6), (6.11) and (6.12), we conclude

$$(6.4) \geq (\beta - \alpha(C'))) \cdot \left( \frac{\varepsilon}{8R} \right)^m - 2^m \Delta(C) \sqrt{\beta^2 + \alpha(C')} - 3^m \cdot \eta(C')$$

$$\geq \frac{\beta}{2} \cdot \left( \frac{\varepsilon}{8R} \right) \Psi_A(C,d) - 2\Psi_A(C,d) + 1 \beta \cdot \Delta(C) - 3\Psi_A(C,d) \cdot \eta(C')$$

Since $\|B'\| \leq p^{dC'}$,

- $\frac{1}{\|B'\| \Psi_A(C,d)} \leq \beta \leq 1$,
- $\Delta(C) = \frac{1}{16} (\frac{\varepsilon}{8R}) \Psi_A(d,C)$,
- $\eta(C') < \frac{1}{8\|B'\| \Psi_A(C,d)} \left( \frac{\varepsilon}{24R} \right)^\Psi_A(C,d)$.
and both $C$ and $C'$ are upper-bounded by $C_{2,3,6}(\Delta, \eta, \rho, \zeta, R)$, we can now define

$$
\delta_A(\varepsilon) = \frac{1}{4} p - d \Psi_A(C_{2,3,6}(\Delta, \eta, \rho, \zeta, R)) C_{2,3,6}(\Delta, \eta, \rho, \zeta, R) \cdot \left( \frac{\varepsilon}{8R} \right)^{\Psi_A(C_{2,3,6}(\Delta, \eta, \rho, \zeta, R), d)}
$$

(6.13)

to conclude the proof.
CHAPTER 7
ALGORITHMIC ASPECTS OF REGULARITY

The results in this chapter are based on [18]1. This chapter is concerned with the algorithmic versions of regularity lemmas for polynomials over finite fields. These regularity lemmas proved by Green and Tao [43] and by Kaufman and Lovett [54] as discussed in Section 2.2.5 show that one can modify a given collection of polynomials $\mathcal{F} = \{P_1, \ldots, P_m\}$ into a new collection $\mathcal{F'}$ so that the polynomials in $\mathcal{F'}$ are regular in the sense of Definition 2.2.7. These lemmas have various applications, such as (special cases of) Reed-Muller testing and worst-case to average-case reductions for polynomials. However, as mentioned in the introduction, the transformation from $\mathcal{F}$ to $\mathcal{F'}$ in these works were not algorithmic. In this chapter, we show that the analytic notions of regularity such as the ones defined in Section 2.2.2, while being qualitatively equivalent to the above, allow for efficient algorithms.

In particular, for high enough characteristic, in time polynomial in number of variables, we can refine $\mathcal{F}$ into $\mathcal{F'}$ where every nonzero linear combination of polynomials in $\mathcal{F'}$ has desirably small Gowers norm. In the case of small fields, we give algorithmic regularity lemmas for two different notions of regularity. First is the notion of strong regularity introduced by Kaufman and Lovett [54] that allows one to show that a polynomial approximable by a strongly regular factor is actually exactly computable by it. The second notion is rank of a polynomial factor, which Tao and Ziegler [78] showed is equivalent to having the analytic property of low Gowers norm.

Throughout this chapter a polynomial is always a classical polynomial unless it is otherwise specified to be a non-classical polynomial.

---

7.1 Applications

A regular factor is a basic ingredient of higher-order Fourier analysis, and we believe that our algorithmic regularity lemmas will therefore be helpful in a variety of future applications. In the following, we examine a few interesting applications of relevance to coding theory, complexity theory and algorithmic algebra.

7.1.1 Inverse Theorem and Reed-Muller Decoding in High Characteristic

One of our main motivations is finding an algorithmic version of the inverse Gowers Theorem for polynomials.

**Theorem 7.1.1** (Inverse Gowers Theorem for polynomials in high characteristic, [43]). If a polynomial \( P \) of degree \( d < |\mathbb{F}| \) satisfies \( \|eP\|_{U^{k+1}} \geq \varepsilon \), then there exists a polynomial \( Q \) of degree at most \( k \), such that

\[
|\langle e_{\mathbb{F}}P, e_{\mathbb{F}}Q \rangle| = \left| \mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}P(x) - Q(x) \right| \geq \eta,
\]

for some \( \eta \) depending only on \( \varepsilon \) and \( d \).

The Inverse Gowers Theorem has also been established over small fields by Tao and Ziegler [78] (Theorem 2.1.10), but here, one needs to extend the given theorem statement to non-classical polynomials, which are discussed in Section 2.1.1.

Theorem 7.1.1 has a direct interpretation in terms of Reed-Muller codes over \( \mathbb{F} \). The Reed-Muller code of order \( k \) over \( \mathbb{F}^n \) is simply the set of polynomials of degree at most \( k \) over \( \mathbb{F}^n \). If \( d < |\mathbb{F}| \) and for a given polynomial \( P \) of degree \( d \), there exists a degree-\( k \) polynomial \( Q \) such that \( \Pr \{ x \in \mathbb{F}^n \} P(x) = Q(x) \geq \frac{1}{|\mathbb{F}|} + \varepsilon \), which is the same as saying that \( \text{dist}(P, Q) \leq 1 - \frac{1}{|\mathbb{F}|} - \varepsilon \) (for \( \text{dist}(P, Q) \) denoting the normalized Hamming distance), then it follows from the definition of Gowers norms and the Gowers-Cauchy-Schwarz inequality.

96
that $\|e_F t P\|_{U^{k+1}} \geq \varepsilon$ for a nonzero $t \in \mathbb{F}$. Then, Theorem 7.1.1 gives that there exists a polynomial $\tilde{Q}$ of degree $k$ such that $|\langle e_F t P, e_F \tilde{Q} \rangle| \geq \eta$, or equivalently, $\text{dist}(P, Q') \leq 1 - \frac{1}{|\mathbb{F}|} - \eta'$ for some $\eta' > 0$ and $Q'$ of degree $k$.

Thus, when $d < |\mathbb{F}|$, the Gowers norm gives an approximate test for checking if for a given polynomial $P$, there exists a $Q$ of degree at most $k$ within Hamming distance $1 - \frac{1}{|\mathbb{F}|} - \varepsilon$. If there exists a $Q$, then the Gowers norm is large and if the Gowers norm is larger than $\varepsilon$, then there exists a $Q'$ within distance $1 - \frac{1}{|\mathbb{F}|} - \eta'$. This is remarkable because the list-decoding radius of Reed-Muller of order $k$ codes is only $1 - \frac{k}{|\mathbb{F}|}$ for $k < |\mathbb{F}|$ [33], and the test works even beyond that. In fact the Hamming distance of a random $P$ is $1 - \frac{1}{|\mathbb{F}|} - o(1)$ from all $Q$ of degree $k$, and the test works all the way up to that distance. Moreover, Tao and Ziegler [78] later showed that this test works not only when $P$ is a polynomial of degree $d < |\mathbb{F}|$ (as in the case of Theorem 7.1.1) but also when $P$ is an arbitrary function and $2^k < |\mathbb{F}|$.

Given the above, it is natural to consider the decoding analogue of the above question:

Given $P$ of degree $d$ over $\mathbb{F}^n$, if there exists $Q$ of degree $k$ such that $\text{dist}(P, Q) \leq 1 - \frac{1}{|\mathbb{F}|} - \varepsilon$, can one find a $Q'$ (in time polynomial in $n$) of degree $k$ such that $\text{dist}(P, Q') \leq 1 - \frac{1}{|\mathbb{F}|} - \eta$ for some $\eta$ depending on $\varepsilon$?

Note that $d$, $k$ and $|\mathbb{F}|$ are assumed to be constants and dependence on these is allowed, but not on $n$. Also, observe that there might be exponentially many such $Q$ since we are in the regime beyond the list-decoding radius (see [55]), but the question turns out to be tractable since we only ask for one such $Q$ and allow a loss from $\varepsilon$ to $\eta$. Such a decoding question was solved for Reed-Muller codes of order 2 over $\mathbb{F}_2^n$ for any given function $f$ (instead of a polynomial $P$ of bounded degree) by [79].

In Section 7.5.2, we show an algorithmic version of Theorem 7.1.1, meaning we explicitly find a $Q$ which has correlation at least $\eta$. In Section 7.5.3, we show how this algorithmic

---

2. For fields of low characteristic ($|\mathbb{F}| \leq k$), the only testing results of this flavor (which works beyond the list-decoding radius for arbitrary input functions) were proved by Samorodnitsky [68] for Reed-Muller codes of order 2 over $\mathbb{F}_2^n$ and by Green and Tao [41] for Reed-Muller codes of order 2 over $\mathbb{F}_5^n$. 97
inverse theorem leads to solving the above decoding question for polynomials $P$ of degree $d$ and the Reed-Muller code of order $k$ for $k \leq d < |\mathbb{F}|$. This special case can be interpreted as follows: if $P$ is of degree-$d$ for some degree $k \leq d < |\mathbb{F}|$, then we can think of $P$ being obtained from some $Q$ of degree $k$ (which is a codeword) by adding the “noise” $P - Q$ of degree $d$. Thus, when the noise is “structured” i.e. given by a degree-$d$ polynomial, we can decode in the above sense (of finding some codeword within a given distance) even beyond the list-decoding radius. Note that $d$ can be much larger than $k$, so the noise set is still much larger than the code.

Both our algorithmic version of the regularity lemma and the decoding algorithm, are randomized algorithms that run in time $O(n^d)$ and output the desired objects with high probability. This is linear time for regularity since even writing a polynomial of degree $d$ takes time $\Omega(n^d)$. The algorithms can be derandomized in polynomial time (at the expense of a very large exponent) using pseudorandom generators for low-degree polynomials. For the question of finding a degree $k$ polynomial within a given distance of $P$, it is possible that one may be able to do this in time $O(n^k)$, but we do not achieve this.

### 7.1.2 Efficient Worst-Case to Average-Case Reductions for Polynomials

The central component in Green and Tao’s proof of Theorem 7.1.1 can be viewed as a worst-case to average-case reduction that is interesting by itself. Let $P : \mathbb{F}^n \to \mathbb{F}$ be a degree-$d$ polynomial which can be weakly approximated by a lower degree polynomial $Q : \mathbb{F}^n \to \mathbb{F}$, of degree $< d$. Here by weak approximation we mean $\langle e_{\mathbb{F}} P, e_{\mathbb{F}} Q \rangle \geq \delta$. Then Green and Tao show that if $d < |\mathbb{F}|$, $P$ can in fact be computed by lower degree polynomials, i.e. there exist $Q_1, \ldots, Q_m$ of degree $< d$ and a function $\Gamma : \mathbb{F}^M \to \mathbb{F}$, such that $P = \Gamma(Q_1, \ldots, Q_M)$. $M$ depends only on $d, \delta$ and $|\mathbb{F}|$. Our algorithmic regularity lemma for polynomials over high characteristic, when plugged into the arguments of Green and Tao, show that this result

---

3. We thank Shachar Lovett for this observation
can be made algorithmic, meaning that $Q_1, \ldots, Q_M$ and $\Gamma$ can be found in $\text{Poly}(n)$ time (assuming constant $d$, $\delta$ and $|\mathbb{F}|$).

Green and Tao (as well as Lovett, Meshulam and Samorodnitsky [63]) showed that their proof techniques break down in the general case when $|\mathbb{F}|$ may be small. In particular, Theorem 7.1.1 is false when $|\mathbb{F}| \leq d$. However, Kaufman and Lovett [54] showed that the above worst-case to average-case reduction is still true for small fields. In order to prove this result, they introduced a more involved notion of regularity called strong regularity. Informally, strong regularity for a factor ensures that not only is the joint distribution of the polynomials in the factor close to uniform, but in fact, the joint distribution of the derivatives of the polynomial is also close to uniform.

We prove that given a factor $\mathcal{F}$, we can find a strongly unbiased refinement $\mathcal{F}'$ in polynomial time. Inserting this algorithmic regularity lemma into Kaufman and Lovett's proof shows that given a polynomial $P$ that is known to be correlated with a lower degree polynomial, we can explicitly find polynomials $Q_1, \ldots, Q_M$ and a function $\Gamma$ such that $P = \Gamma(Q_1, \ldots, Q_M)$ and $M$ is bounded. A randomized algorithm for recovering $Q_1, \ldots, Q_M$ and $\Gamma$ with high probability runs in time $O(n^d)$, where $d = \deg(P)$.

### 7.1.3 Algorithmic Inverse Theorem for Polynomials: Fields of Small Characteristic

Over small fields, the correspondence between high rank and low Gowers norm, which drives our proof over large characteristic, breaks down and one needs to consider non-classical polynomials. (see Section 2.1.1 for formal definition).

The corresponding notion of regularity for factors $\mathcal{F}$ over a small field is that no nonzero linear combination $P$ of polynomials in $\mathcal{F}$ can be written as a function of a bounded number $M$ of non-classical polynomials of degree less than $\deg(P)$ (Definition 2.2.7). We prove an algorithmic regularity lemma that in polynomial time, finds a regular refinement when given
as input an arbitrary non-classical polynomial factor of bounded degree.

Using the notion of high-rank non-classical polynomial factors, we obtain an algorithmic version of the inverse theorem for bounded degree polynomials (for the case when the degree $d$ is greater than the order $p$ of the underlying prime field) by Tao and Ziegler [78]. We manage to somewhat simplify the proof of the inverse theorem for bounded degree polynomials (i.e. Theorem 2.1.10 when $f$ itself is a phase polynomial), by observing that one of its steps can be generalized to avoid a few other parts of the proof.

7.1.4 Polynomial decompositions

Given a positive integer $k$, a vector of positive integers $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and a function $\Gamma : F^k \to F$, we say that a function $P : F^n \to F$ is $(k, \Delta, \Gamma)$-structured if there exist polynomials $P_1, P_2, \ldots, P_k : F^n \to F$ with each $\deg(P_i) \leq \Delta_i$ such that for all $x \in F^n$,

$$P(x) = \Gamma(P_1(x), P_2(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ are said to form a $(k, \Delta, \Gamma)$-decomposition (or simply, a polynomial decomposition). For instance, an $n$-variate polynomial over the field $F$ of total degree $d$ factors nontrivially exactly when it is $(2, (d-1, d-1), \text{prod})$-structured where $\text{prod}(a, b) = a \cdot b$.

In [13], Bhattacharyya noted that for any fixed $k$, $\Delta$, $\Gamma$ and $F$, our algorithmic version of the Green-Tao regularity lemma over high characteristic allows one to decide $(k, \Delta, \Gamma)$-structure in time $\text{Poly}(n)$ for all input polynomials $P : F^n \to F$ whose degree is strictly less than $|F|$. However, [13] left open the question of solving the polynomial decomposition problem over small fields, such as $F_2$, and input polynomials of degree larger than $|F|$. Here, we resolve this issue.

At its core, the proof of correctness in [13] used our algorithmic version of the Green-Tao regularity lemma that applies only for large characteristic. The algorithmic regularity lemma
for non-classical factors can then be used in the proof of [13] to show that over any fixed prime order field $\mathbb{F}$ and for any fixed choice of positive integer $k$, vector of positive integers $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and function $\Gamma : \mathbb{F}^k \to \mathbb{F}$, the $(k, \Delta, \Gamma)$-decomposition problem is solvable in $\text{Poly}(n)$ time when the input is a bounded degree polynomial.

### 7.2 Algorithmic Bogdanov-Viola Lemma; Approximate Regularization

The Bogdanov-Viola Lemma [21] states that if a polynomial of degree $d$ is biased, then it can be approximated by a bounded set of polynomials of lower degree. Recall the definition of bias of a function $P : \mathbb{F}^n \to \mathbb{F}$:

$$\text{bias}(P) = \left| \text{E}_{x \in \mathbb{F}^n} [e_{\mathbb{F}}(P(x))] \right|$$

The following is an easy to observe algorithmic version of this lemma. We reproduce the proof by Green and Tao [43] while pointing out that all the steps can be performed efficiently.

**Lemma 7.2.1** (Algorithmic Bogdanov-Viola lemma). *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

$$\text{bias}(P) \geq \delta,$$

runs in time $O_{\delta, \beta, \sigma}(n^d)$, and with probability $1 - \beta$ returns functions $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ and $\Gamma : \mathbb{F}^C \to \mathbb{F}$, and a set of polynomials $P_1, \ldots, P_C$, where $C \leq \frac{|\mathbb{F}|^5}{\delta^2 \sigma \beta}$ and $\text{deg}(P_i) < d$ for all $i \in [C]$, for which

- $\Pr_x(P(x) \neq \tilde{P}(x)) \leq \sigma$, and
- $\tilde{P}(x) = \Gamma(P_1(x), \ldots, P_C(x)).$*
Proof. The proof will be an adaptation of the proof from [43]. Notice that we can compute the explicit description of \( P \) in \( O(n^d) \) queries. For every \( a \in \mathbb{F} \) define the measure \( \mu_a(t) \) \( \overset{\text{def}}{=} \Pr(P(x) = a + t) \). It is easy to see that if \( \text{bias}(P(x)) \geq \delta \) then, for every \( a \neq b \),

\[
\| \mu_a - \mu_b \| \geq \frac{4\delta}{|\mathbb{F}|}.
\] (7.1)

We will try to estimate each of these distributions. Let

\[
\tilde{\mu}_a(t) \overset{\text{def}}{=} \frac{1}{C} \sum_{1 \leq i \leq C} \mathbb{1}_{P(x_i) = a + t},
\]

where \( C > \frac{|\mathbb{F}|^5}{\delta \beta_1} \), and \( x_1, x_2, \ldots, x_C \in \mathbb{F}^n \) are chosen uniformly at random. Therefore by an application of Chebyshev’s inequality

\[
\Pr \left( |\tilde{\mu}_a(t) - \mu_a(t)| > \frac{\delta}{2|\mathbb{F}|^2} \right) < \beta_1,
\]

for all \( t \in \mathbb{F} \) and therefore

\[
\Pr \left( \| \tilde{\mu}_a - \mu_a \| > \frac{\delta}{2|\mathbb{F}|^2} \right) < \beta_1. \quad (7.2)
\]

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 \( h = (h_1, \ldots, h_C) \in (\mathbb{F}^n)^C \) be chosen
uniformly at random, where $C$ is a sufficiently large constant to be chosen later. Define

$$
\mu_{\text{obs}}(t) \overset{\text{def}}{=} \frac{1}{C} \sum_{1 \leq i \leq C} \mathbb{1}_{D_{h_i}P(x) = t},
$$

and let

$$
\bar{P}_h(x) \overset{\text{def}}{=} \arg \min_{r \in F} \| \tilde{\mu}_r - \mu_{\text{obs}}^{(x)} \|.
$$

Now choosing $C \geq |F|^5 / \delta^2 \sigma \beta^2$ follows

$$
\Pr_h(\bar{P}_h(x) \neq P(x)) \leq \Pr_h \left( \| \mu_{\text{obs}}^{(x)} - \mu P(x) \| \geq \frac{\delta}{|F|} \right) \leq \sigma \beta^2,
$$

(7.3)

where the first inequality follows from (7.1) and (7.2). Therefore

$$
\Pr_{x, h}(\bar{P}_h(x) \neq P(x)) = \mathbb{E}_{x, h} \mathbb{1}_{\bar{P}_h(x) \neq P(x)} \leq \sigma \beta^2,
$$

and thus

$$
\Pr_h \left[ \Pr_x (\bar{P}_h(x) \neq P(x)) \geq \sigma \right] \leq \beta.
$$

Let $P_i := D_{h_i}P$, so that $P_i$ is of degree $\leq d$ and $\bar{P}_h$ is a function of $P_1, \ldots, P_C$. Now setting $\beta_1 := \frac{\beta}{2|F|^2}$ and $\beta_2 := \frac{\beta}{2}$ finishes the proof. $\square$

**Remark 7.2.2** (Derandomization). We note that Lemma 7.2.1 can be derandomized in Poly$(n)$ steps using known pseudorandom generators for low degree polynomials over finite fields [21, 62, 80], as shown in [13], following a suggestion by Shachar Lovett. In particular, Viola [80] showed that for any integer $d \geq 1$ and $0 < \varepsilon < 1$, there is an explicit generator $g : \mathbb{F}^s \rightarrow \mathbb{F}^n$ such that $s = d \log |\mathbb{F}| n + O(d 2^d \log(1/\varepsilon))$ and for every polynomial $P : \mathbb{F}^n \rightarrow \mathbb{F}$,

$$
| \mathbb{E}_{Y \leftarrow g(U_s)} e(P(Y)) - \mathbb{E}_{X \leftarrow U_n} e(P(X)) | < \varepsilon
$$

103
where $U_s$ is the uniform distribution on $\mathbb{F}^s$ and $U_n$ the uniform distribution on $\mathbb{F}^n$. In order to get a deterministic algorithm, we run over all possible seeds $s$ to the generator $g$. As the discussion in [13] shows, one can use this technique to derandomize Lemma 7.2.1 in $\text{Poly}(n)$ time. (The exponent on $n$ becomes a large constant proportional to $d$ and $C$ in the proof above.)

### 7.2.1 Unbiased Almost-Refinement

Using Lemma 7.2.1, we can deduce the following which can be thought of as an algorithmic analogue of the regularity lemma, with the caveat that the refinement is approximate.

**Lemma 7.2.3** (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 $O_{\gamma, \rho, \sigma, \text{dim}(\mathcal{F})}(n^d)$ and with probability $1 - \rho$ returns a $\gamma$-unbiased factor $\mathcal{F}'$ with $\text{dim}(\mathcal{F}') = O_{\gamma, \rho, \sigma, \text{dim}(\mathcal{F})}(1)$, such that $\mathcal{F}'$ is $\sigma$-close to being a refinement of $\mathcal{F}$.

**Proof.** The proof idea is similar to that of Lemma 2.2.27 in the sense that we do the same type of induction. The difference is that at each step we will have to control the errors that we introduce through the use of Lemma 7.2.1 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 $O_{\gamma, \rho, \sigma, \text{dim}(\mathcal{F})}(n^d)$ 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. Assume that $\mathcal{F}$ is $\gamma$-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 \(3\gamma(\dim(\mathcal{F}))\). To do so, for each linear combination \(\sum_{i,j} c_{i,j} P_{i,j}\) independently select a set of vectors \(x_1, \ldots, x_C\) uniformly at random from \(\mathbb{F}^n\), and let \(\tilde{\text{bias}}(\sum_{i,j} c_{i,j} P_{i,j}) \overset{\text{def}}{=} \left| \frac{1}{C} \sum_{\ell \in [C]} c_{\ell} f(y_\ell) \right|\), with \(y_\ell = \sum_{i,j} c_{i,j} \cdot P_{i,j}(x_\ell)\). Choosing \(C = \mathcal{O}(\dim(\mathcal{F})^{1/4} \gamma^{2/3} \log(1/\rho))\), we can distinguish bias \(\gamma\) from bias \(\gamma/2\), 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 \(3\gamma(\dim(\mathcal{F}))\) 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 \(1 - \frac{\rho}{4}\), bias(\(\sum_{i,j} c_{i,j} P_{i,j}\)) \(\geq \gamma(\dim(\mathcal{F}))/2\), by Lemma 7.2.1 we can find, with probability \(1 - \frac{\rho}{4}\), a set of polynomials \(Q_1, \ldots, Q_r\) of degree \(k - 1\) such that

- \(\sum_{i,j} c_{i,j} P_{i,j}\) is \(\frac{\sigma}{2}\)-close to a function of \(Q_1, \ldots, Q_r\),
- \(r \leq \frac{16|\mathbb{F}|^5}{\gamma(\dim(\mathcal{F}))^2 \sigma \cdot \rho'}\).

We replace one polynomial of highest degree that appears in \(\sum_{i,j} c_{i,j} P_{i,j}\) with polynomials \(Q_1, \ldots, Q_r\).

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 \(\gamma\)-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 \(\tilde{\mathcal{F}}\) be the outcome of one step of our algorithm. With probability \(1 - \frac{\rho}{4}\), \(\tilde{\mathcal{F}}\) 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 \(\gamma\)-unbiased factor \(\mathcal{F}'\) which is \(\frac{\sigma}{2}\)-close to being a
refinement of $\tilde{F}$ and therefore, with probability at least $1 - (\rho + \rho + \rho + \rho') > 1 - \rho$, is $\sigma$-close to being a refinement of $F$. \hfill \Box

**Remark 7.2.4.** Note that Lemma 7.2.3 can be derandomized using PRGs for low-degree polynomials. Remark 7.2.2 discusses derandomization of Lemma 7.2.1. We can similarly also make the step for estimating the bias be deterministic, by using a different PRG for fooling each linear combination of polynomials and then running over all seeds to these PRGs. The resulting running time is $\text{Poly}(n)$ with a large exponent.

Let $P : \mathbb{F}^n \to \mathbb{F}$ be a function, and let $\mathcal{F}$ be a polynomial factor such that $P$ is measurable in $\mathcal{F}$. This means that there exists a function $\Gamma : \mathbb{F}^{\dim(\mathcal{F})} \to \mathbb{F}$ such that $P(x) = \Gamma(\mathcal{F}(x))$. Although we know that such a $\Gamma$ exists, but we do not have explicit description of $\Gamma$. It is a simple but useful observation that we can provide query access to $\Gamma$ in case $\mathcal{F}$ is unbiased. It is worth recording this as the following lemma.

**Lemma 7.2.5 (Query access).** Suppose that $\beta \in (0, 1]$ and let $\mathcal{F} = \{P_1, \ldots, P_m\}$ be a $\gamma$-unbiased factor of degree $d$, where $\gamma : \mathbb{F}^n \to \mathbb{F}$ is a decreasing function which decreases suitably fast. Suppose that we are given query access to a function $f : \mathbb{F}^n \to \mathbb{F}$, where $f$ is measurable in $\mathcal{F}$, namely there exists $\Gamma : \mathbb{F}^{\dim(\mathcal{F})} \to \mathbb{F}$ such that $f(x) = \Gamma(\mathcal{F}(x))$. There is a randomized algorithm that, takes as input a value $a = (a_1, \ldots, a_m) \in \mathbb{F}^m$, makes a single query to $f$, runs in $O(n^d)$, and returns a value $y \in \mathbb{F}$ such that $\Gamma(a) = y$ with probability greater than $1 - \beta$.

**Proof.** Choose $\gamma(x) := \frac{1}{2p^x}$. Applying Lemma 2.2.19 for the case of classical polynomials, we get

$$|\mathcal{F}^{-1}(a)| \geq \frac{1}{\|\mathcal{F}\|} - \frac{1}{2\|\mathcal{F}\|} = \frac{1}{2\|\mathcal{F}\|}.$$  

Let $x_1, \ldots, x_K$ be chosen uniformly at random from $\mathbb{F}^n$, where $K \geq 2\|\mathcal{F}\| \log \frac{1}{\beta}$. Then with probability at least $1 - \beta$, there is $i^* \in [K]$ such that $\mathcal{F}(x_{i^*}) = (a_1, \ldots, a_m)$. This means that $x_{i^*} \in \mathcal{F}^{-1}(a)$. Notice that we can find this $i^*$, since we have explicit description of
polynomials in $\mathcal{F}$ and we can evaluate them on each $x_i$. Our algorithm will query $f$ on input $x_i^*$ and return $f(x_i^*)$. 

\[ \Box \]

### 7.3 Algorithmic Uniform Refinement in high Characteristic

In this section we shall assume that the field $\mathbb{F}$ has large characteristic, namely $|\mathbb{F}|$ is greater than the degree of the polynomials and the degree of the factors that we study. We will prove the following lemma which states that we can efficiently refine a given factor to a uniform factor.

**Lemma 7.3.1** (Uniform Refinement). 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 $O_{\rho, \gamma, \dim(\mathcal{F})} (n^d)$, and with probability $1 - \rho$ outputs a $\gamma$-uniform factor $\tilde{\mathcal{F}}$, where $\tilde{\mathcal{F}}$ is a refinement of $\mathcal{F}$, is of same degree $d$, and $|\dim(\tilde{\mathcal{F}})| \ll_{\sigma, \gamma, \dim(\mathcal{F})} 1$.

**Note 7.3.2.** Notice that unlike Lemma 7.2.3, here $\mathcal{F}'$ is an exact refinement of $\mathcal{F}$. This will be achieved by the use of Proposition 7.3.4 which roughly states that approximation of a polynomial by a uniform factor implies its exact computation.

The following proposition of Green and Tao states that if a polynomial is approximated well enough by a regular factor, then it is computed by the factor.

**Proposition 7.3.3** ([43]). Suppose that $d \geq 1$ is an integer. There exists a constant $\sigma_d > 0$ such that the following holds. Suppose that $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree $d$ and that $\mathcal{F}$ is an $F$-regular factor of degree $d - 1$ for some increasing function $F : \mathbb{N} \to \mathbb{N}$ which increases suitably rapidly in terms of $d$. Suppose that $\bar{P} : \mathbb{F}^n \to \mathbb{F}$ is an $\mathcal{F}$-measurable function and that $\Pr(P(x) \neq \bar{P}(x)) \leq \sigma_d$. Then $P$ is itself $\mathcal{F}$-measurable.

Our proof of Lemma 7.3.1 will require the following variant of Proposition 7.3.3 which follows by Remark 2.2.17.
Proposition 7.3.4. Suppose that $d \geq 1$ is an integer. There exists $\sigma_{7.3.4}(d) > 0$ such that the following holds. Suppose that $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree $d$ and that $\mathcal{F}$ is a $\gamma_{7.3.4}$-uniform factor of degree $d - 1$ for some decreasing function $\gamma_{7.3.4}$ which decreases suitably rapidly in terms of $d$. Suppose that $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ is an $\mathcal{F}$-measurable function and that $\Pr(P(x) \neq \tilde{P}(x)) \leq \sigma_{7.3.4}(d)$. Then $P$ is itself $\mathcal{F}$-measurable.

We will prove Lemma 7.3.1 by induction on the dimension vector of $\mathcal{F}$. For the induction step, we check whether there is a linear combination of polynomials in $\mathcal{F}$ that has large Gowers norm. We will use 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 $Q_1, ..., Q_r$, then use the induction hypothesis to refine $\{Q_1, ..., Q_r\}$ to a uniform factor $\{\tilde{Q}_1, ..., \tilde{Q}_r\}$ and use Proposition 7.3.4 to conclude that $P$ is measurable in $\{\tilde{Q}_1, ..., \tilde{Q}_r\}$.

Proof of Lemma 7.3.1: Let $\mathcal{F}$ be a factor of degree $d$. Similar to the proof of Lemma 7.2.3 we will induct on dimension vector $(M_1, ..., M_d)$. At all steps of the proof we may assume that all polynomials in $\mathcal{F}$ are homogeneous, that is, all multinomials are of same degree. This is because otherwise, we may replace each polynomial $P_{i,j}$ with $i$ homogeneous polynomials that sum up to $P_{i,j}$. Moreover, as in the proof of Lemma 7.2.3, without loss of generality we may assume that the polynomials in $\mathcal{F}$ are linearly independent, since otherwise we can detect and remove linear dependencies in $O_{\dim(\mathcal{F})}(n^d)$ time.

Given the above assumptions, the base case for $d = 1$ is simple, a linearly independent homogeneous factor is 0-uniform. Let $d > 1$ and assume that $\mathcal{F}$ is not $\gamma$-uniform, 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

$$\|e_\mathcal{F}(\sum_{i,j} c_{i,j} \cdot P_{i,j})\|_{U^k} \geq \gamma(\dim(\mathcal{F})),$$

where $k = \deg(\sum_{i,j} c_{i,j} \cdot P_{i,j})$. We will use the following randomized algorithm to detect such a linear combination:
We will estimate $\|e_\mathbb{F}(\sum_{i,j} c_{i,j} \cdot P_{i,j})\|_{U^k}^{2k}$ for each of the $|\mathbb{F}|^{\dim(F)}$ linear combinations and check whether it is greater than $\frac{3\gamma(\dim(F))^{2k}}{4}$. To do this, as in the proof of Lemma 7.2.3, independently for each linear combination $\sum_{i,j} c_{i,j} \cdot P_{i,j}$ select $C$ sets of vectors $x^{(\ell)}, y_1^{(\ell)}, ..., y_k^{(\ell)}$ chosen uniformly at random from $\mathbb{F}^n$ for $1 \leq \ell \leq C$, and compute

$$\left| \sum_{\ell \in [C]} e_\mathbb{F}\left( \sum_{I \subseteq [k]} (-1)^{|I|} \sum_{i,j} c_{i,j} \cdot P_{i,j}(x^{(\ell)} + y_I^{(\ell)}) \right) \right|$$

Choosing $C = O_{\dim(F)} \left( \frac{1}{\gamma(\dim(F))^2 \log(\frac{1}{\rho})} \right)$, we can distinguish between Gowers norm $\leq \frac{\gamma(F)}{2^{1/2^k}}$ and Gowers norm $\geq \gamma(F)$, with $\rho' = \frac{\rho}{4|\mathbb{F}|^{\dim(F)}}$ probability of error.

Let $Q := \sum_{i,j} c_{i,j} \cdot P_{i,j}$, with $\deg(Q) = k$, the detected linear combination for which the estimated $\|e_\mathbb{F}(Q)\|_{U^k}^{2k}$ is greater than $\frac{3\gamma(\dim(F))^{2k}}{4}$. By a union bound on all the linear combinations, with probability at least $1 - \frac{\rho}{4}$, $\|e_\mathbb{F}(Q)\|_{U^k}^{2k} \geq \frac{\gamma(\dim(F))^{2k}}{2}$. Notice that

$$\text{bias}(DQ) = \|e(Q)\|_{U^k}^{2k} \geq \frac{\gamma(\dim(F))}{2},$$

where $DQ : (\mathbb{F}^n)^k \rightarrow \mathbb{F}$ is defined as

$$DQ(y_1, ..., y_k) \overset{\text{def}}{=} D_{y_1} D_{y_2} \cdots D_{y_k} Q(x).$$

Let $\tilde{\sigma}$ and $\tilde{\gamma}$ be as in Proposition 7.3.4. By Lemma 7.2.1 we can find, with probability $1 - \frac{\rho}{4}$, a set of polynomials $Q_1, ..., Q_r$ of degree $k - 1$ such that

- $DQ$ is $\tilde{\sigma}$-close to being measurable with respect to $Q_1, ..., Q_r$.
- $r \leq \frac{8|\mathbb{F}|^5}{\gamma(\dim(F))^{2k+1} \tilde{\sigma} \rho}$.

By the induction hypothesis, with probability $1 - \frac{\rho}{4}$, we can refine $\{Q_1, ..., Q_r\}$ to a new factor $\{\tilde{Q}_1, ..., \tilde{Q}_{r'}\}$, which is $\tilde{\gamma}$-uniform. It follows from Proposition 7.3.4 that $DQ$ is in fact
measurable in \{\tilde{Q}_1, \ldots, \tilde{Q}_{r'}\}. Since \(|\mathbb{F}| > k\) we can write the following Taylor expansion

\[
Q(x) = \frac{DQ(x, \ldots, x)}{k!} + R(x),
\]

(7.4)

where \(R(x)\) is a polynomial of degree \(\leq k - 1\) which can be computed from \(Q\). Thus we can replace a maximum degree polynomial \(P_{i,j}\), that appears in \(Q\), with polynomials \(\tilde{Q}_1, \ldots, \tilde{Q}_{r'}\), and \(R(x)\). The proof follows by using the induction hypothesis for parameters \(\ell_4\) and \(\gamma\).

**Remark 7.3.5.** The above proof can be derandomized in polynomial time. Lemma 7.2.1 can be made deterministic as discussed after its proof. Here, the additional randomness is in estimating the Gowers norms of each linear combination of polynomials. Again, using pseudorandom generators for degree-\(d\) polynomials lets us make this estimation be deterministic. For more details, see [13].

### 7.4 Algorithmic Regularity in Low Characteristic

As mentioned in previous sections, Gowers uniformity for polynomial factors fails to address “bias implies low rank” in the case when the field \(\mathbb{F}\) has small characteristic. Kaufman and Lovett [54] introduce a stronger notion of regularity in order to handle the general case. To define their notion of regularity, we first have to define the derivative space of a factor.

**Definition 7.4.1 (Derivative Space).** Let \(\mathcal{F} = \{P_1, \ldots, P_m\}\) be a polynomial factor over \(\mathbb{F}^n\). Recall that

\[
D_h P(x) = P(x + h) - P(x),
\]

is the additive derivative of \(P\) in direction \(h\), where \(h\) is a vector from \(\mathbb{F}^n\). We define

\[
\text{Der}(\mathcal{F}) \overset{\text{def}}{=} \{(D_h P_i)(x) : i \in [m], h \in \mathbb{F}^n\},
\]

as the derivative space of \(\mathcal{F}\).
**Definition 7.4.2 (Strong regularity of polynomials [54]).** Suppose that \( F : \mathbb{N} \to \mathbb{N} \) is an increasing function. Let \( \mathcal{F} = \{ P_1, \ldots, P_m \} \) be a polynomial factor of degree \( d \) and \( \Delta : \mathcal{F} \to \mathbb{N} \) assigns a natural number to each polynomial in \( \mathcal{F} \). We say that \( \mathcal{F} \) is strongly \( F \)-regular with respect to \( \Delta \) if

1. For every \( i \in [m] \), \( 1 \leq \Delta(P_i) \leq \deg(P_i) \).

2. For any \( i \in [m] \) and \( r > \Delta(P_i) \), there exist a function \( \Gamma_{i,r} \) such that

   \[
   P_i(x + y_J) = \Gamma_{i,r}(P_j(x + y_J) : j \in [m], J \subseteq [r], |J| \leq \Delta(P_j)),
   \]

   where \( x, y_1, \ldots, y_r \) are variables in \( \mathbb{F}^n \), and \( y_J = \sum_{k \in J} y_k \) for every \( J \subseteq [r] \).

3. For any \( r \geq 0 \), let \( \mathcal{C} = \{ c_{i,I} \in \mathbb{F} \}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} \) be a collection of field elements. Define

   \[
   Q_{\mathcal{C}}(x, y_1, \ldots, y_r) \overset{\text{def}}{=} \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} \cdot P_i(X + Y_I).
   \]

Let \( \mathcal{F}' \subseteq \mathcal{F} \) be the set of all \( P_i \)'s which appear in \( Q_{\mathcal{C}} \). There do not exist polynomials \( \widetilde{P}_1, \ldots, \widetilde{P}_\ell \in \text{Der}(\mathcal{F}') \), and \( I_1, \ldots, I_\ell \subseteq [r] \), with \( \ell \leq F(\dim(\mathcal{F})) \), for which \( Q_{\mathcal{C}} \) can be expressed as

   \[
   \Gamma \left( \widetilde{P}_1(x + y_{I_1}), \ldots, \widetilde{P}_\ell(x + y_{I_\ell}) \right),
   \]

for some function \( \Gamma : \mathbb{F}^\ell \to \mathbb{F} \).

The following lemma states that every polynomial factor can be refined to a strongly regular factor.

**Lemma 7.4.3 (Strong-Regularity Lemma, [54]).** Let \( F : \mathbb{N} \to \mathbb{N} \) be an increasing function. Let \( \mathcal{F} = \{ P_1, \ldots, P_m \} \) be a polynomial factor of degree \( d \). There exists a refinement \( \mathcal{F}' = \{ P_1, \ldots, P_r \} \) of same degree \( d \) along with a function \( \Delta : \mathcal{F} \to \mathbb{N} \) such that

- \( \mathcal{F}' \) is strongly \( F \)-regular with respect to \( \Delta \),
\[ \dim(F') = O_{F_\dim(F),d}(1). \]

Moreover they prove that if a polynomial is approximated by a strongly regular factor, then it is in fact measurable with respect to the factor.

**Lemma 7.4.4 ([54])**. For every \( d \) there exists a constant \( \varepsilon_d = 2^{-\Omega(d)} \) such that the following holds. Let \( P : \mathbb{F}^n \to \mathbb{F} \) be a polynomial of degree \( d \), \( P_1, ..., P_s \) be a strongly regular collection of polynomials of degree \( d - 1 \) with degree bound \( \Delta \) and \( \Gamma : \mathbb{F}^s \to \mathbb{F} \) be a function such that \( \Gamma(P_1, ..., P_s) \) is \( \varepsilon_d \)-close to \( P \). Then there exists \( \Gamma' : \mathbb{F}^s \to \mathbb{F} \) such that

\[ P(x) = \Gamma'(P_1(x), ..., P_s(x)). \]

Lemma 7.4.3 and Lemma 7.4.4 combined with Bogdanov-Viola Lemma immediately gives the following theorem.

**Theorem 7.4.5** (Bias implies low rank for general fields [54]). Let \( P : \mathbb{F}^n \to \mathbb{F} \) be a polynomial of degree \( d \) such that \( \text{bias}(P) \geq \delta > 0 \). Then \( \text{rank}_{d-1} \leq c(d, \delta, \mathbb{F}) \).

In Section 7.4.1 we will give an algorithmic version of this theorem. To do this, as in previous sections, we will have to use a notion of regularity which is well suited for algorithmic applications. Uniform factors (Definition 2.2.16) fail to address fields of low characteristic, for the reason that to refine a factor to a uniform factor we make use of division by \( d! \) which is not possible in fields with \( |\mathbb{F}| \leq d \). Strong regularity of [54] suggests how to extend the notion of unbiased factors to a stronger version in quite a straightforward fashion.

**Definition 7.4.6** (Strongly \( \gamma \)-unbiased factors). Suppose that \( \gamma : \mathbb{N} \to \mathbb{R}^+ \) is a decreasing function. Let \( \mathcal{F} = \{P_1, ..., P_m\} \) be a polynomial factor of degree \( d \) over \( \mathbb{F}^n \) and let \( \Delta : \mathcal{F} \to \mathbb{N} \) assign a natural number to each polynomial in the factor. We say that \( \mathcal{F} \) is strongly \( \gamma \)-unbiased with degree bound \( \Delta \) if

1. For every \( i \in [m] \), \( 1 \leq \Delta(P_i) \leq \deg(P_i) \).
2. For every \( i \in [m] \) and \( r > \Delta(P_i) \), there exists a function \( \Gamma_{i,r} \) such that

\[
P_i(x + y_{[r]}) = \Gamma_{i,r}(P_j(x + y_J) : j \in [m], J \subseteq [r], |J| \leq \Delta(P_j)).
\]

3. For any \( r \leq 0 \) and not all zero collection of field elements

\[
\mathcal{C} = \{c_{i,I} \in \mathbb{F} \}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)}
\]

we have

\[
\text{bias}\left( \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} \cdot P_i(x + y_I) \right) \leq \gamma(\dim(F)),
\]

for random variables \( x, y_1, \ldots, y_r \in \mathbb{F}^n \).

We will say \( F \) is \( \Delta \)-bounded if it only satisfies (1) and (2).

In Section 7.4.1 we will show how to refine a given factor to a strongly unbiased one. This will allow us prove an algorithmic analogue of Lemma 7.4.4, which will be presented in the next section.

### 7.4.1 Algorithmic approximate strongly unbiased refinement

In this section we will use strongly unbiased factors (Definition 7.4.6) to prove Proposition 7.3.3 and Proposition 7.3.4, respectively, without the assumption on \( |\mathbb{F}| \) being larger than \( d \). The following claim states that although in part (3) there is no upper bound on \( r \), choosing a suitably faster decreasing function \( \gamma' = \gamma^{2d} \), one can assume that \( r \leq d \).

**Claim 7.4.7.** Let \( \mathcal{F} = \{P_1, \ldots, P_m\} \) be a polynomial factor of degree \( d \), and \( \Delta : \mathcal{F} \rightarrow \mathbb{N} \) be a function such that \( \mathcal{F} \) is \( \Delta \)-bounded. Suppose that \( \gamma : \mathbb{N} \rightarrow \mathbb{R}^+ \) is a decreasing function such that for every \( r \in [d], x \in \mathbb{F}^n \) and not all zero collection of coefficients
where $y_1, \ldots, y_r$ are variables from $\mathbb{F}^n$. Then $\mathcal{F}$ is strongly $\gamma$-unbiased.

**Proof.** Let $r > d$ be an integer, $x \in \mathbb{F}^n$ be fixed and $\{c_{i,I}\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)}$ be a set of coefficients, not all of which are zero. We will show that

$$\text{bias} \left( \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} \cdot P_i(x + y_I) \right) \leq \gamma (\dim(\mathcal{F}))^{2d},$$

(7.5)

Define $Q_{\mathcal{C}}(y_1, \ldots, y_r) \overset{\text{def}}{=} \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} \cdot P_i(x + y_I)$, and let $I \subseteq [r]$ be maximal such that $c_{i,I}$ is nonzero for some $i \in [m]$. Without loss of generality assume that $I = \{1, 2, \ldots, |I|\}$. We will derive $Q_{\mathcal{C}}$ in directions $y_1, \ldots, y_{|I|}$, namely

$$DQ_{\mathcal{C}, y_1, \ldots, y_{|I|}}(h_1, \ldots, h_{|I|}) := D_{h_1} \cdots D_{h_{|I|}} Q_{\mathcal{C}}(y_1, \ldots, y_r),$$

where $h_i$ is only supported on $y_i$ coordinates. Notice that, since $I$ was maximal, $DQ_{\mathcal{C}}$ does not depend on any $y_i$ with $i \in [r] \setminus I$, namely, for every such $J \subseteq [r]$ with $J \not\subseteq I$, and every $i \in [m]$, $D_{h_1} \cdots D_{h_{|I|}} P_i(x + y_J) \equiv 0$. Moreover

$$D_{h_1} \cdots D_{h_{|I|}} Q_{\mathcal{C}}(y_1, \ldots, y_r) = \sum_{i \in [m], J \subseteq [r], |J| \leq \Delta(P_i)} c_{i,J} \cdot D_{h_1} \cdots D_{h_{|I|}} P_i(x + y_J)$$

$$= \sum_{i \in [m]} c_{i,I} \cdot D_{h_1} \cdots D_{h_{|I|}} P_i(x + y_I)$$

$$= \sum_{i \in [m], J \subseteq I} c_{i,I} \cdot (-1)^{|J|} \cdot P_i((x + y_I) + h_J).$$

Applying the claim below for the random variable $x' := x + y_I$ and taking the bias over
random variables $x', h_1, \ldots, h_{\mid I\mid}$ we have that

$$\gamma(\dim(F))^{2d} \geq \bias(DQ_{\mathcal{C},y_1,\ldots,y_r}(x', h_1, \ldots, h_{\mid I\mid})) \geq \bias(Q_{\mathcal{C}}(y_1, \ldots, y_r))^{2\mid I\mid}.$$  

Now the claim follows by the fact that $\mid I\mid \leq d$. \hfill $\square$

Following is a simple repeated application of the Cauchy-Schwarz inequality.

Claim 7.4.8.

$$\bias(P_{\Lambda})^{2^r} \leq \bias(DP_{\Lambda}).$$

Proof. It suffices to show that we have

$$\mid \mathbb{E}_{y_1,\ldots,y_k,h\in F^n} e_F(P_{\Lambda}(y_1 + h, y_2, \ldots, y_k) - P_{\Lambda}(y_1, y_2, \ldots, y_k)) \mid \geq \mathbb{E}_{y_1,\ldots,y_k} e_F(P_{\Lambda}(y_1, \ldots, y_k))^{2^r}.$$

This is a simple application of Cauchy Schwarz inequality

$$\mid \mathbb{E}_{y_1,\ldots,y_k,h} e_F(P_{\Lambda}(y_1 + h, y_2, \ldots, y_k) - P_{\Lambda}(y_1, \ldots, y_k)) \mid = \mathbb{E}_{y_2,\ldots,y_k} \left( \mathbb{E}_z e_F(P_{\Lambda}(z, y_2, \ldots, y_k)) \right)^{2^r} \geq \mathbb{E}_{z,y_2,\ldots,y_k} e_F(P_{\Lambda}(z, y_2, \ldots, y_k \in F^n))^{2^r}.\hfill \square$$

Claim 7.4.7 suggests a way of refining a factor to a strongly uniform one: We will estimate the bias of every possible linear combination of the polynomials over all points $x + y_I$ where $I \subseteq [r]$ and $r \leq d$, and once we have detected a biased combination, refine using Lemma 7.2.1. This is made possible by the fact that $r$ and $\dim(F)$ are both bounded.
Lemma 7.4.9 (Algorithmic strongly γ-unbiased refinement). Suppose that \( \sigma, \beta \in (0, 1] \) are parameters, and \( \gamma : \mathbb{N} \to \mathbb{R}^+ \) is a decreasing function. There is a randomized algorithm that given a factor \( F = \{P_1, ..., P_m\} \) of degree \( d \), runs in \( O(\gamma(n^d)) \), with probability \( 1 - \beta \), returns a strongly \( \gamma \)-unbiased factor \( F' \) with degree bound \( \Delta \) such that \( F' \) is \( \sigma \)-close to being a refinement of \( F \).

Proof. We will design a refinement process which has to stop in a finite number of steps which only depends on \( \dim(F) \), \( \sigma \), \( \beta \), and \( \gamma \). We will start with the initial factor being \( F \) and we will set \( \Delta(P_i) := \deg(P_i) \). Notice that this automatically satisfies the first two properties of Definition 7.4.6, namely \( \Delta \)-boundedness. This is because for every \( r > \deg(P_i) \) we can use the derivative property

\[
P_i(x + y[I]) = \sum_{I \subseteq [r]} (-1)^{r-|I|} \cdot P_i(x + y_I).
\]

to write \( P_i(x + y[I]) \) as a function of \( \{P_i(x + y_J)\}_{J \subseteq [r], |J| \leq d} \).

As explained in proof of Lemma 7.3.1, at each step, we may assume without loss of generality that all polynomials in the factor are homogeneous. Moreover, we will assume that for every \( r \leq d \), the set of polynomials \( \{P_i(x + y[I])\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} \) is linearly independent, where \( x, y_1, \ldots, y_r \in \mathbb{F}^n \) are variables. This is because at each step we have access to explicit description of the polynomials in the factor, therefore we can compute each of the \( |\mathbb{F}|^{B(r)} \) possible linear combinations, where \( B(r) = \sum_{i \in [m]} \sum_{j \in [\Delta(P_i)]} \binom{r}{j} \), and check whether it is equal to zero or not. Suppose that for a set of coefficients \( \mathcal{C} = \{c_{i,I} \in \mathbb{F}\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} \) we have \( Q_{\mathcal{C}} \overset{\text{def}}{=} \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} P_i(x + y_I) \equiv 0 \). Let \( P_i \) be a polynomial that appears in \( Q_{\mathcal{C}} \) and let \( I \) be maximal such that \( c_{i,I} \neq 0 \). We can let \( \Delta(P_i) := |I| - 1 \), and remove \( P_i \) from the factor if \( \Delta(P_i) \) becomes zero.

We will stop refining if \( F \) is of degree 1. Notice that a linearly independent factor of degree 1 is not biased at all, and therefore strongly 0-unbiased. Assume that \( F \) is not
strongly $\gamma$-unbiased with respect to the current $\Delta$, then there exists $r$ and a set of coefficients $C = \{c_{i,I} \in \mathbb{F}\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)}$ for which

$$
\text{bias}(Q_C) = \text{bias}\left(\sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} P_i(x + y_I)\right) > \gamma(\dim(F)).
$$

To detect this, we will use the following algorithm: Set $K := |\mathbb{F}|^B(r)$. We can estimate the bias of each of the $K$ possible linear combinations $Q_C$ and check whether our estimate is greater than $3\gamma(\dim(F))$. Letting $\beta' := \beta \frac{\gamma(\dim(F))}{4}$, we can distinguish $\text{bias} \geq \gamma(\dim(F))$ from $\text{bias} \leq \gamma(\dim(F))$ correctly with probability $1 - \beta'$ by computing the average of $e_{\mathbb{F}}(Q_C)$ on $O(\dim(F))\left(\frac{1}{\gamma(\dim(F))^2}\right)$ random sets of vectors $x, y_1, \ldots , y_r$. Let $C$ be such that the estimated bias for $Q_C$ was greater than $3\gamma(\dim(F))$, and let $k = \deg(Q_C)$. By a union bound with probability $1 - \beta'$, $\text{bias}(Q_C) \geq \frac{\gamma(\dim(F))}{2}$, and by Lemma 7.2.1 we can find, with probability $1 - \beta'$, a set of polynomials $Q_1, \ldots , Q_s : (\mathbb{F}^n)^{r+1} \rightarrow \mathbb{F}$ of degree $k - 1$ such that

- $Q_C$ is $\frac{\beta}{2}$-close to a function of $Q_1, \ldots , Q_s$,
- $s \leq \frac{16|\mathbb{F}|^5}{\gamma(\dim(F)) \cdot \sigma \cdot \beta}$.

Moreover we know from the proof of Lemma 7.2.1 that for each $j \in [s]$ we have

$$
Q_j(x, y_1, \ldots , y_r) = Q_C(x + h_x, y_1 + h_1, \ldots , y_r + h_r) - Q_C(x, y_1, \ldots , y_r)
= \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)} c_{i,I} \cdot D_{h_x + h_I} P_i(x + y_I).
$$

for some fixed vectors $h_x, h_1, \ldots , h_r \in \mathbb{F}^n$. Let $R_{i,I}^{(j)} := D_{h_x + h_I} P_i$ so that $\deg(R_{i,I}^{(j)}) < \deg(P_i)$ and $Q_j$ is measurable in $\{R_{i,I}^{(j)}(x + y_I)\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)}$.

Let $P_i$ be a polynomial of maximum degree that appears in $Q_C$, and let $I$ be maximal such that $c_{i,I} \neq 0$. We will add each $R_{i,I}^{(j)}$ to $\mathcal{F}$ with $\Delta(R_{i,I}^{(j)}) = \deg(R_{i,I}^{(j)})$ and let $\Delta(P_i) := |I| - 1$ and discard $P_i$ if $|I|$ becomes zero. Notice that the new factor is $\Delta$-bounded because $P_i(x + 117$
$y_1$) can be written as a function of $\{Q_j(x, y_1, \ldots, y_r)\}_{j \in [s]}$ and $\{P_j(x + y_j)\}_{J \subseteq I, |J| \leq \Delta(P_j)}$. Moreover each $Q_j$ is measurable in $\{R_{i, I}^{(j)}(x + y_I)\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(P_i)}$.

We can prove that the process stops after a constant number of steps by a strong induction on the dimension vector $(M_1, \ldots, M_d)$ of the factor $\mathcal{F}$, where $M_i$ is the number of polynomials of degree $i$ in the factor and $d$ is the degree of the factor. The induction step follows from the fact that at every step we discard $P_i$ or decrease the $\Delta(P_i)$ for some $i$ and add polynomials of lower degree. Let $\mathcal{F}'$ and $\Delta : \mathcal{F}' \rightarrow \mathbb{N}$ be the output of the algorithm. The claim follows by our choices for error distances and the probabilities.

Remark 7.4.10. The randomness involved in estimating the bias can be removed by using PRGs for polynomials over $\mathbb{F}^n$.

7.4.2 Algorithmic exact strongly unbiased refinement

It is not difficult to check that Kaufman and Lovett’s proof of Lemma 7.4.4, can be adapted to a proof of the following analogue for our notion of strongly unbiased factors.

Lemma 7.4.11 (Approximation by Strongly Unbiased Factor implies Computation). For every $d \geq 0$ there exists a constant $\sigma_d$ and a decreasing function $\gamma : \mathbb{N} \rightarrow \mathbb{R}^+$ such that the following holds. Let $P : \mathbb{F}^n \rightarrow \mathbb{F}$ be a polynomial of degree $d$, $\mathcal{F} = \{P_1, \ldots, P_m\}$ be a strongly $\gamma$-unbiased polynomial factor of degree $d - 1$ with degree bound $\Delta$, and let $\Gamma : \mathbb{F}^m \rightarrow \mathbb{F}$ be a function such that $P$ is $\sigma_d$-far from $\Gamma(\mathcal{F})$. Then $P$ is in fact measurable in $\mathcal{F}$.

Remark 7.4.12. (Exact refinement) One can use the above lemma to modify the refinement process of Lemma 7.4.9 to achieve exact refinement instead of approximate refinement. This can be done by using induction on the degree of the factor and at every round refining the new polynomials that are to be added to the factor to a strongly unbiased set of polynomials. Then Lemma 7.4.11 ensures that the removed polynomial is measurable in the new set of polynomials and therefore at every step we have exact refinement of the original factor.
7.5 Applications in High Characteristic

In this section we show some applications of the uniformity notion, and our algorithm for uniform refinement of a factor. Throughout the section we shall assume that the field \( F \) has large characteristic, namely \(|F|\) is greater than the degree of the polynomials and the degree of the factors that we study.

7.5.1 Computing Polynomials with Large Gowers Norm

Following is an immediate corollary of Lemma 7.2.1, Lemma 7.2.5, and Proposition 7.3.4.

Lemma 7.5.1. Suppose that an integer \( d \) satisfies \( 0 \leq d < |F| \). 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 \( \dim(\mathcal{F}) = O_{\delta, \beta, \sigma}(1) \) such that \( P \) is measurable in \( \mathcal{F} \), namely there is a function \( \Gamma \) such that

\[
P(x) = \Gamma\left((P_{i,j}(x))_{1 \leq i \leq d-1, 1 \leq j \leq M_i}\right).
\]

Moreover for every query to the value of \( \Gamma(\cdot) \) for a vector from \( \mathbb{F}^{\dim(\mathcal{F})} \) the algorithm returns the correct answer with probability \( 1 - \beta \).

Proof. Let \( \sigma_d \) and \( \gamma_1 \) be as in Proposition 7.3.4, and let \( \gamma_2 \) be as in Lemma 7.2.5. Set \( \gamma_d := \min\{\gamma_1, \gamma_2\} \). By Lemma 7.2.1 since \( \text{bias}(P) \geq \delta \), in time \( O_{\sigma_d, \beta}(n^d) \), with probability \( 1 - \frac{\beta}{2} \) we can find a polynomial factor \( \mathcal{F} \) of degree \( d - 1 \) such that

- \( P \) is \( \frac{\sigma_d}{2} \)-close to a function of \( \mathcal{F} \)
- \( \dim(\mathcal{F}) \leq \frac{4|F|^5}{\delta^2 \sigma_d \beta} \)

By Lemma 7.3.1, with probability \( 1 - \frac{\beta}{2} \), we can find a \( \gamma_d \)-uniform factor \( \mathcal{F}' \) such that \( \mathcal{F}' \) is \( \frac{\sigma_d}{2} \)-close to being a refinement of \( \mathcal{F} \).
Thus with probability greater than $1 - \beta$, $P$ is $\sigma_d$-close to a function of the strongly $\gamma_d$-unbiased factor $\mathcal{F}'$, and by Lemma 7.4.11 $P$ is measurable in $\mathcal{F}'$. The query access to $\Gamma$ can be granted by Lemma 7.2.5.

The following is an algorithmic version of Proposition 6.1 of [43], which states that given a polynomial $P$ such that $e_\mathcal{F}(P)$ has large Gowers norm, then we can compute $P$ by a few lower degree polynomials.

**Proposition 7.5.2** (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 $\|e_\mathcal{F}(P(x))\|_{U^d} \geq \delta$, runs in $O_{\delta, \beta}(n^d)$ and with probability $1 - \beta$, returns a polynomial factor $\mathcal{F}$ of degree $d - 1$ such that

- There is a function $\Gamma : \mathbb{F}^{\dim(\mathcal{F})} \to \mathbb{F}$ such that $P = \Gamma(\mathcal{F})$.
- $\dim(\mathcal{F}) = O_{d, \delta, \beta}(1)$.

Moreover, for every query to the value of $\Gamma(\cdot)$ for a vector from $\mathbb{F}^{\dim(\mathcal{F})}$, the algorithm returns the correct answer with probability $1 - \beta$.

*Proof.* Write $\partial^d P(h_1, \ldots, h_d) := D^{h_1} \cdots D^{h_d} P(x)$. Since $P$ has degree $d$, $\partial^d P$ does not depend on $x$. From the definition of the $U^d$ norm, we have

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

Applying Lemma 7.5.1 to $\partial^d P$, with probability $1 - \frac{\beta}{2}$, we can find a factor $\tilde{\mathcal{F}}$ of degree $d - 1$, such that $\dim(\tilde{\mathcal{F}}) = O_{\delta, \beta, d}(1)$ and $\partial^d P$ is measurable in $\tilde{\mathcal{F}}$.

It is easy to check that since $|\mathbb{F}| > d$, we have the following Taylor expansion

$$P(x) = \frac{1}{d!} \partial^d P(x, \ldots, x) + Q(x),$$

120
7.5.2 Algorithmic Inverse Theorem in High Characteristic

Proposition 7.5.2 allows us to prove the following algorithmic version of Theorem 2.1.10.

**Theorem 7.5.3.** 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 $O_{\delta, \beta}(n^d)$ and with probability $1 - \beta$, returns a polynomial $Q$ of degree $\leq k$ such that

$$|\langle e_{\mathbb{F}}(P), e_{\mathbb{F}}(Q) \rangle| \geq \eta.$$

**Proof.** By Proposition 7.5.2, with probability $1 - \frac{\beta}{4}$ we can find a polynomial factor $\tilde{F}$ of degree $d - 1$ such that $P$ is measurable in $\tilde{F}$. Let $\gamma : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function to be specified later. By Lemma 7.3.1, with probability $1 - \frac{\beta}{4}$, we can refine $\tilde{F}$ to a $\gamma$-uniform factor $F = \{P_1, ..., P_m\}$ of same degree $d - 1$, with $\dim(F) = O_{\gamma, \beta, \epsilon}(1)$. Since $P$ is measurable in $F$, there exists $\Gamma : \mathbb{F}^{\dim(F)} \to \mathbb{F}$ such that $\tilde{P} = \Gamma(F)$. Letting $f := e_{\mathbb{F}}(\tilde{P})$, and using the Fourier decomposition of $e_{\mathbb{F}}(\Gamma)$ we can write

$$e_{\mathbb{F}}(\tilde{P}(x)) = \sum_{i=1}^{L} c_i e_{\mathbb{F}}(\langle \alpha^{(i)}, F \rangle(x)), \quad (7.6)$$

where $L = |\mathbb{F}|^{\dim(F)} = O_{\gamma, \beta}(1)$, $\alpha^{(i)} \in \mathbb{F}^{\dim(F)}$, and

$$\langle \alpha^{(i)}, F \rangle(x) \overset{\text{def}}{=} \sum_{j=1}^{m} \alpha^{(i)}_j \cdot P_j(x).$$

Notice that the terms in (7.6), unlike Fourier characters, are not orthogonal. But since the
factor $\mathcal{F}$ is $\gamma$-uniform, Lemma 2.2.22 ensures approximate orthogonality. Let $Q_i := \langle \alpha^{(i)}, \mathcal{F} \rangle$.

Choose $\gamma(u) \leq \frac{\sigma}{|\mathcal{F}| u}$, so that $\gamma(\dim(\mathcal{F})) \leq \frac{\sigma}{L^2}$, where $\sigma := \frac{\varepsilon^{2k+1}}{4}$. It follows from the near orthogonality of the terms in (7.6) by Lemma 2.2.22 that

$$|c_i - \langle f, e_{\mathcal{F}}(Q_i) \rangle| \leq \frac{\sigma}{L},$$

(7.7)

and

$$\left| \|f\|_2^2 - \sum_{i=1}^L c_i^2 \right| \leq \sigma.$$  

(7.8)

Claim 7.5.4. There exists $\delta'(\varepsilon, |\mathcal{F}|) \in (0, 1]$ 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 $|\langle f, e_{\mathcal{F}}(Q_i) \rangle| \geq \delta'$.

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

$$\varepsilon^{2k+1} \leq \|f\|_U^{2k+1} \leq \|f\|_2^2 \|f\|_\infty^{-2k+1},$$

(7.9)

and (7.8) imply that there exists $i \in [L]$ such that $c_i^2 \geq \frac{\varepsilon^{2k+1}}{L} - \sigma = \frac{3\varepsilon^{2k+1}}{4L}$, which combined with (7.7) implies that $|\langle f, e_{\mathcal{F}}(Q_i) \rangle| \geq \frac{\varepsilon^{2k+1}}{2L}$.

Now for the induction step, assume that $d > k + 1$. We will decompose (7.6) 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 $g := \sum_{i \in S} c_i e_{\mathcal{F}}(Q_i)$ and $h := \sum_{i \in [L]\setminus S} c_i e_{\mathcal{F}}(Q_i)$. Notice that by the triangle inequality of Gowers norm, our choice of $\gamma$, and the fact that $\mathcal{F}$ is $\gamma$-uniform

$$\|h\|_U^{k+1} \leq \sum_{i \in [L]\setminus S} |c_i| \cdot \|e_{\mathcal{F}}(Q_i)\|_U^{k+1} \leq L \cdot \frac{\varepsilon^{2k+1}}{4L^2} = \frac{\varepsilon^{2k+1}}{4L},$$

122
and thus
\[ \|g\|_{U^k+1} \geq \frac{\varepsilon}{2}. \]

Now the claim follows by the base case.

Let \( \delta'(\varepsilon, |F|) \) be as in the above claim, and choose \( \eta := \frac{\delta'}{4} \). We will use the following theorem of Goldreich and Levin [31] which gives an algorithm to find all the large Fourier coefficients of \( e_F(\Gamma) \).

**Theorem 7.5.5** (Goldreich-Levin theorem [31]). Let \( \zeta, \rho \in (0, 1] \). There is a randomized algorithm, which given oracle access to a function \( \Gamma : \mathbb{F}^m \rightarrow \mathbb{F} \), runs in time \( O(m^2 \log m \cdot \text{poly}(\frac{1}{\zeta}, \log(\frac{1}{\rho}))) \) and outputs a decomposition

\[ \Gamma = \sum_{i=1}^\ell b_i \cdot e_F(\langle \eta_i, x \rangle) + \Gamma', \]

with the following guarantee:

- \( \ell = O(\frac{1}{\zeta^2}) \).

- \( \Pr[\exists i : |b_i - \hat{\Gamma}(\eta_i)| > \zeta/2] \leq \rho. \)

- \( \Pr[\forall \alpha \text{ such that } |\hat{f}(\alpha)| \geq \zeta, \exists i \eta_i = \alpha] \geq 1 - \rho. \)

We will use the above theorem with parameters \( \zeta := \frac{\delta'}{2} \) and \( \rho := \frac{\beta}{4} \), and use the randomized algorithm in Lemma 7.2.5 to provide answer to all its queries to \( \Gamma \). Choose \( \gamma \) suitably small, so that with probability at least \( 1 - \frac{\beta}{4} \) our answer to all queries to \( \Gamma \) are correct. By Claim 7.5.4 there is \( i \in [L] \) such that \( \hat{\Gamma}(\alpha) = c_i \geq \frac{3\delta'}{4} \). With probability \( 1 - \frac{\beta}{4} \) there is \( j \) such that \( \eta_j = \alpha_i \) and with probability at least \( 1 - \frac{\beta}{4} \), \( |b_j - c_i| \leq \frac{\zeta}{2} \leq \frac{\delta'}{4} \), 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 \( |\langle f, e_F(Q_i) \rangle| \geq \frac{\delta'}{4} = \eta. \)
In this section we discuss the task of decoding Reed-Muller codes when the noise is “structured”. Recall that the codewords in a Reed-Muller code of order $k$, correspond to evaluations of all degree $k$ polynomials in $n$ variables over $\mathbb{F}$. We will consider the task of decoding codewords when the noise is structured in the sense that the applied noise to the codeword itself is a polynomial of higher degree $d$. In other words, we are given a polynomial $P$ of degree $d > k$ with the promise that it is “close” to an unknown degree $k$ polynomial, and the task is to find a degree $k$ polynomial $Q$ that is somewhat close to $P$. The assumption of $d > k$ is made, because the case $d \leq k$ is trivial.

**Theorem 7.5.6** (Reed-Muller Decoding). Suppose that $d > k > 0$ are integers, and let $\varepsilon \in (0, 1]$ and $\beta \in (0, 1]$ be parameters. There is $\delta(\varepsilon, d, k) > 0$ and a randomized algorithm that, given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree $d$, with the promise that there exists a polynomial $Q$ of degree $k$ such that $\Pr(P(x) = Q(x)) \geq \frac{1}{|\mathbb{F}|} + \varepsilon$, runs in $O(n^d)$, and with probability $1 - \beta$ returns a polynomial $\tilde{Q}$ of degree $k$ such that

$$\Pr(P(x) = \tilde{Q}(x)) \geq \frac{1}{|\mathbb{F}|} + \delta.$$ 

Before presenting the proof of Theorem 7.5.6 we will look at the relations between Hamming distance of $P$ and $Q$, correlation between $e_\mathbb{F}(P)$ and $e_\mathbb{F}(Q)$, and Gowers norm of $e_\mathbb{F}(P)$. The following claim relates the Hamming distance of two polynomials $P$ and $Q$ to the correlation between $e_\mathbb{F}(P)$ and $e_\mathbb{F}(Q)$.

**Claim 7.5.7.** Suppose that $\varepsilon \in (0, 1]$, and let $P, Q : \mathbb{F}^n \to \mathbb{F}$ be two functions such that $\Pr_x(P(x) = Q(x)) \geq 1/|\mathbb{F}| + \varepsilon$. Then there exists a nonzero $t \in \mathbb{F}$ for which

$$\left| \left< e_\mathbb{F}(t \cdot P), e_\mathbb{F}(t \cdot Q) \right> \right| = \left| \mathbb{E}_x e_\mathbb{F}(t \cdot (P(x) - Q(x))) \right| \geq \varepsilon.$$ 

124
Proof. We have

\[
\frac{1}{|F|} + \varepsilon \leq \mathbb{P}_{x \in \mathbb{F}^n} (P(x) = Q(x)) = \mathbb{E}_{x \in \mathbb{F}^n} \left[ 1_{P(x) = Q(x)} \right] = \mathbb{E}_{x \in \mathbb{F}^n} \left[ \frac{1}{|F|} \sum_{t \in F} e_F(t \cdot (P(x) - Q(x))) \right] = \frac{1}{|F|} \sum_{t \in F} \langle e_F(t \cdot P), e_F(t \cdot Q) \rangle = \frac{1}{|F|} \left( 1 + \sum_{t \in F \setminus \{0\}} \langle e_F(t \cdot P), e_F(t \cdot Q) \rangle \right),
\]

and thus there is \( t \in F \setminus \{0\} \) for which \(| \langle e_F(t \cdot P), e_F(t \cdot Q) \rangle | \geq \varepsilon \). \( \square \)

Assume that \( P : \mathbb{F}^n \to \mathbb{F} \) is a polynomial of degree \( d \) and \( Q : \mathbb{F}^n \to \mathbb{F} \) is a polynomial of degree \( k < d \). Using Lemma 2.1.9 with \( f_\emptyset := e_F(P - Q) \) and \( f_S := 1 \) for \( S \neq \emptyset \) implies that

\[
| \langle e_F(P), e_F(Q) \rangle | = \left| \mathbb{E}_{x \in \mathbb{F}^n} e_F(P(x) - Q(x)) \right| \leq ||e_F(P(x) - Q(x))||_{U^d} = ||e_F(P(x))||_{U^d},
\]

where the last equality follows from the fact that \( \deg(Q(x)) = k < d \).

**Proof of Theorem 7.5.6:** Since there exists a polynomial \( Q \) of degree \( k \) with \( \mathbb{P}(P(x) \neq Q(x)) \leq \varepsilon \), by Claim 7.5.7 there exists a nonzero \( t \in F \) such that

\[
| \langle e_F(t \cdot P), e_F(t \cdot Q) \rangle | \geq \varepsilon.
\]

Notice that

\[
||e_F(t \cdot P)||_{U^d} \geq ||e_F(t \cdot P)||_{U^{k+1}} \geq \langle e_F(t \cdot P), e_F(t \cdot Q) \rangle \geq \varepsilon,
\]

where the second inequality follows from (7.10) since \( \deg(t \cdot Q) \leq k \). Notice that we may find such a constant \( t \in F \setminus \{0\} \) with high probability by estimating the \( U^{k+1} \) norm of \( tP \) for
all $|\mathbb{F}| - 1$ possible choices of $t$. Set $\tilde{P} := t \cdot P$.

Now by Theorem 7.5.3 there exists $\eta_{d,\beta,\varepsilon} \in (0,1]$ such that we can find in time $O(n^d)$, with probability $1 - \frac{\beta}{2}$, a polynomial $\tilde{Q}$ such that

$$\left| \langle e_\mathbb{F}(\tilde{P}), e_\mathbb{F}(\tilde{Q}) \rangle \right| \geq \eta.$$

The following claim shows that, there is a constant shift of $e_\mathbb{F}(\tilde{Q})$ that approximates $e_\mathbb{F}(\tilde{P})$.

**Claim 7.5.8.** Let $\tilde{P}$ and $\tilde{Q}$ be as above. There is a randomized algorithm that with probability $1 - \frac{\beta}{2}$ returns an $h \in \mathbb{F}$ for which

$$\Pr(\hat{P}(x) = \tilde{Q}(x) + h) \geq \frac{1}{|\mathbb{F}|} + \frac{\eta}{2|\mathbb{F}|^2}.$$

**Proof.** Since $\tilde{P}(x) - \tilde{Q}(x)$ takes values in $\mathbb{F}$, there must be a choice of $r \in \mathbb{F}$ such that

$$\Pr(\tilde{P}(x) - \tilde{Q}(x) = r) \geq \frac{1}{|\mathbb{F}|} + \frac{\eta}{2|\mathbb{F}|^2}.$$ Similar to the proof of Lemma 7.2.1 defining $\mu_0(r) = \Pr(\tilde{P}(x) - \tilde{Q}(x) = r)$, we can find the estimate measure $\mu_{\text{obs}}$ by $K = O(|\mathbb{F}|,\delta,\beta(1)$ random queries to $\tilde{P}(x) - \tilde{Q}(x)$ such that

$$\Pr\left( \exists r \in \mathbb{F} : |\mu_0(r) - \mu_{\text{obs}}(r)| \geq \frac{\eta}{2|\mathbb{F}|^2} \right) \leq \frac{\beta}{2}.$$ Choosing $h \in \mathbb{F}$ such that $\mu_{\text{obs}}(h)$ is maximized, we have

- $\Pr(\tilde{P}(x) = Q(x) + h) \geq \frac{1}{|\mathbb{F}|} + \frac{\eta}{2|\mathbb{F}|^2}$
- $\deg(Q(x) + h) = k$.

Since $\tilde{P} = t \cdot P$, the same also holds between $P$ and $t^{-1}(Q + h)$. The probability of correctness is now ensured by a union bound. □
Remark 7.5.9. The randomness involved in using the Goldreich-Levin algorithm can be removed. Note that at the expense of a greater running time, we can simply look at all the Fourier coefficients of $\Gamma$ since its domain is of constant size. Finally, Claim 7.5.7 can be derandomized in the same way as Lemma 7.2.1, by using PRGs with seed length $O(\log n)$ for fooling polynomials over $\mathbb{F}^n$.

7.6 Applications in Low Characteristic

7.6.1 Computing a biased Polynomial

Having Lemma 7.4.9 and Lemma 7.4.11 in hand we immediately have the following analogue of Theorem 7.4.5, which states that if a polynomial is biased then we can find a factor that computes it.

Theorem 7.6.1 (Computing a biased polynomial). Let $\beta \in (0, 1]$ be an error parameter. 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 $O_{\delta, \beta}(n^d)$ and with probability $1 - \beta$, returns a polynomial factor $\mathcal{F} = \{P_1, ..., P_{c(d, \delta)}\}$ of degree $d - 1$ such that $P$ is measurable in $\mathcal{F}$.

Proof. Let $\sigma_d$ and $\gamma_d$ be as in Lemma 7.4.11. By Lemma 7.2.1 since $\text{bias}(P) \geq \delta$, in time $O_{\sigma_d, \beta}(n^d)$, with probability $1 - \frac{\beta}{2}$ we can find a polynomial factor $\mathcal{F} = \{P_1, ..., P_{c(d, \delta)}\}$ of degree $d - 1$ such that

- $P$ is $\frac{\gamma_d^2}{2}$-close to a function of $\mathcal{F}$
- $\dim(\mathcal{F}) \leq \frac{4|\mathcal{F}|^5}{\delta^2 \sigma_d \beta}$

By Lemma 7.4.9, with probability $1 - \frac{\beta}{2}$, we can find a strongly $\gamma_d$-unbiased factor $\mathcal{F}'$ with degree bound $\Delta$ such that $\mathcal{F}'$ is $\frac{\gamma_d^2}{2}$-close to being a refinement of $\mathcal{F}$.

Thus with probability greater than $1 - \beta$, $P$ is $\sigma_d$-close to a function of the strongly $\gamma_d$-unbiased factor $\mathcal{F}'$, and it follows from Lemma 7.4.11 that $P$ is measurable in $\mathcal{F}'$. \qed
The algorithm can be made deterministic in time $\text{Poly}(n)$ (with a large exponent), as indicated in the previous sections.

### 7.6.2 Worst Case to Average Case Reduction

Here we will show how Theorem 7.6.1 implies an algorithmic version of worst case to average case reduction from [54]. To present the result, we first have to define what it means for a factor to approximate a polynomial.

**Definition 7.6.2 ($\delta$-approximation).** We say that a function $f : F^n \rightarrow F$ $\delta$-approximates a polynomial $P : F^n \rightarrow F$ if

$$\left| \mathbb{E}_{x \in F^n} [e^F(P(x) - f(x))] \right| \geq \delta.$$  

Kaufman and Lovett use Lemma 7.4.4 to show the following reduction.

**Theorem 7.6.3** (Theorem 3 of [54]). Let $P(x)$ be a polynomial of degree $k$, $g_1, \ldots, g_c$ polynomials of degree $d$, and $\Lambda : F^c \rightarrow F$ a function such that composition $\Lambda(g_1(x), \ldots, g_c(x))$ $\delta$-approximates $P$. Then there exist $c'$ polynomials $h_1, \ldots, h_{c'}$ and a function $\Gamma : F^c' \rightarrow F$ such that

$$P(x) = \Gamma(h_1(x), \ldots, h_{c'}(x)).$$

Moreover, $c' = c'(d, c, k)$ and each $h_i$ is of the form $p(x + a) - p(x)$ or $g_j(x + a)$, where $a \in F^n$. In particular, if $k \leq k - 1$ then $\deg(h_i) \leq k - 1$ also.

Here we will design a randomized algorithm that given $g_1, \ldots, g_k$ can compute a set of $h_1, \ldots, h_{c'}$ efficiently.

**Theorem 7.6.4** (Worst-case to average case reduction). Let $\delta, \beta \in (0, 1]$ be parameters. There is a randomized algorithm that takes as input

- A polynomial $P : F^n \rightarrow F$ of degree $d$
• A polynomial factor $\mathcal{F} = \{P_1, \ldots, P_m\}$ of degree $d - 1$

• A function $\Lambda$ such that $\Lambda(\mathcal{F})$ $\delta$-approximates $P$

and with probability at least $1 - \beta$, returns a polynomial factor $\mathcal{F}' = \{R_1, \ldots, R_{m'}\}$ and a function $\Gamma : \mathbb{F}^{\dim(\mathcal{F}')} \rightarrow \mathbb{F}$ such that

$$P(x) = \Gamma(R_1(x), \ldots, R_{m'}(x)),$$

moreover $c' = O_{\dim(\mathcal{F}), \beta, \delta, d}(1)$.

Proof. Looking at the Fourier decomposition of $e_\mathbb{F}(\Lambda)(y_1, \ldots, y_m)$, since $\Lambda(P_1, \ldots, P_m)$ $\delta$-approximates $P$, there must exist $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{F}^c$ such that $Q_\alpha(x) = \sum_{i \in [m]} \alpha_i \cdot P_i(x)$, $\delta'$-approximates $P$, where $\delta' \geq \frac{\delta}{|\mathbb{F}|^m}$. We will estimate $|\text{bias}(P - Q_\alpha)|$ for every $\alpha \in \mathbb{F}^m$.

For each $\alpha$, we can distinguish $\text{bias}(P - Q_\alpha) \leq \frac{\delta'}{2}$ from $\text{bias}(P - Q_\alpha) \geq \delta'$, with probability $1 - \frac{3\beta}{|\mathbb{F}|^m}$, by evaluating $e_\mathbb{F}(P(x) - Q_\alpha(x))$ on $C = O_{\dim(\mathcal{F})} \left( \frac{1}{\delta^2} \log\left( \frac{1}{\beta} \right) \right)$ random inputs. Let $\alpha^* \in \mathbb{F}^m$ be such that our estimate for $\text{bias}(P - Q_\alpha)$ is greater than $\frac{3\delta'}{4}$.

With probability at least $1 - \frac{\beta}{3|\mathbb{F}|^m}$ we will find such $\alpha^*$, and by a union bound with probability at least $1 - \frac{\beta}{3}$, $\text{bias}(P - Q_{\alpha^*}) \geq \frac{\delta'}{2}$. Now applying Theorem 7.6.1 to $P - Q_{\alpha^*}$ with parameters $\frac{\beta}{3}$ and $\frac{\delta'}{2}$, we find a polynomial factor $\mathcal{F}' = \{R_1, \ldots, R_{m'}\}$ of degree $d - 1$, such that with probability $1 - \frac{\beta}{3}$, $P - Q_{\alpha^*}$ is measurable in $\mathcal{F}'$. Namely there exists $\Gamma'$ such that $P - Q_{\alpha^*}(x) = \Gamma'(\mathcal{F}'(x))$ and therefore

$$P(x) = \Gamma'(R_1(x), \ldots, R_{m'}(x)) + Q_{\alpha^*}(x).$$

□

This algorithm can be derandomized in time $\text{Poly}(n)$, using the remarks made in the previous sections.
7.7 Algorithmic Inverse Theorem for Polynomials

The proof of the inverse theorem for Gowers norms (Theorem 2.1.10 [78]) is based on ergodic theory and a pure combinatorial proof of the inverse theorem for general functions is not yet known, but [78] give an analytical proof for the case of bounded degree polynomials. In this section we will show how to conclude an algorithmic inverse theorem for polynomials of bounded degree from the proof presented in [78]. Observing that one can generalize one of the steps in the proof by [78] we manage to somewhat simplify the proof of the inverse theorem for bounded degree polynomials by skipping a few of the steps.

Theorem 7.7.1 (Algorithmic Inverse for Polynomials). Let $s > 0$ be an integer, and let $\varepsilon, \mu \in (0, 1]$ be parameters. There is a randomized algorithm that given a (non-classical) degree-$s$ polynomial $P : \mathbb{F}^n \to \mathbb{T}$ with $\|P\|_{U^s} \geq \varepsilon$, runs in $O_{\varepsilon,s,\beta}(n^d)$ and with probability $1 - \mu$ returns a polynomial $Q$ with $\deg(Q) \leq s - 1$ such that

$$|\langle e(P), e(Q) \rangle| \geq \delta(\varepsilon, s) > 0.$$ 

To prove this we first show how to compute such a polynomial by a lower degree polynomial factor.

Theorem 7.7.2 (Computing polynomials with large Gowers norm). Let $s > 0$ be an integer and let $\mu \in (0, 1]$ be a parameter. There is a randomized algorithm that given a (non-classical) polynomial $P : \mathbb{F}^n \to \mathbb{T}$ with $\|P\|_{U^s}^{2s} > \varepsilon$, runs in $O_{s,\beta}(n^s)$ and with probability $1 - \mu$ outputs the following

1. A polynomial $Q : \mathbb{F}^n \to \mathbb{T}$ with

   (a) $\deg(Q) = s$.

   (b) $P$ and $Q$ have the same $s$-th derivative. In particular $\deg(P - Q) \leq s - 1$. 

130
2. Polynomials (possibly non-classical) \( Q_1, \ldots, Q_C : \mathbb{F}^n \to \mathbb{T} \) for some \( C = O_{s, \beta, \varepsilon}(1) \) such that \( Q \) is measurable in \( \{Q_1, \ldots, Q_C\} \).

We will prove this theorem in Section 7.7.2. Being able to algorithmically compute functions of large Gowers norm allows us to algorithmically refine any polynomial factor to a uniform factor, in the sense of Definition 2.2.16 in low characteristic. Notice that the two obstacles in the proof of Lemma 7.3.1 were that we were not able to compute polynomials of high Gowers norm in the low characteristic case and the reason was that we used (7.4) which is not possible when \( |\mathbb{F}| \leq d \). Theorem 7.7.2 allows us to compute each linear combination of polynomials in the factor that has large Gowers norm, and thus the following is a corollary to Theorem 7.7.2.

**Corollary 7.7.3 (Uniform Refinement in Low Characteristic).** Suppose \( d \) 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 \( O_{\rho, \gamma, \dim(\mathcal{F})}(nd) \), and with probability \( 1 - \rho \) outputs a \( \gamma \)-uniform factor \( \widetilde{\mathcal{F}} \), where \( \widetilde{\mathcal{F}} \) is a refinement of \( \mathcal{F} \), is of same degree \( d \), and \( |\dim(\widetilde{\mathcal{F}})| \ll_{\sigma, \gamma, \dim(\mathcal{F})} 1 \).

Notice that both \( \mathcal{F} \) and \( \mathcal{F}' \) are now (non-classical) polynomial factors. Theorem 7.7.1 follows as a corollary by a proof identical to that of Theorem 7.5.6. Theorem 7.7.2 and Corollary 7.7.3 can be derandomized just as it was possible in the high characteristic case.

### 7.7.1 Derivative Polynomials as Symmetric Multilinear Polynomials

Let \( P : \mathbb{F}^n \to \mathbb{T} \) be a degree-d polynomial, possibly non-classical. Define the derivative polynomial \( DP : (\mathbb{F}^n)^d \to \mathbb{T} \) by the following formula

\[
DP(h_1, \ldots, h_d) \overset{\text{def}}{=} D_{h_1} \cdots D_{h_d} P(x),
\]
where \(h_1, \ldots, h_d \in \mathbb{F}^n\). Lemma 2.1.6 shows some useful properties of the derivative polynomial. Motivated by the properties of the derivative function, we define a class of polynomials.

**Definition 7.7.4 (Classical Symmetric Multilinear Polynomials).** For every \(d\), denote by \(\text{HCSM}_d(\mathbb{F}^n)\) the set of all polynomials \(Q : (\mathbb{F}^n)^s \to \mathbb{F}, s \leq d\), which satisfy properties (i), (ii), (iii) and (iv) from Lemma 2.1.6. In particular for every degree-\(d\) polynomial \(P\), \(DP \in \text{HCSM}_d(\mathbb{F}^n)\). The reason we use the name “HCSM” is that \(\text{CSM}\) which was defined in [78] is reserved for the subclass of classical symmetric multilinear polynomials which are the derivative polynomials of classical polynomials.

**Remark 7.7.5.** It is observed in [78] that if \(P : \mathbb{F}^n \to \mathbb{T}\) is a classical degree-\(d\) polynomial then \(DP(h_1, \ldots, h_d)\) vanishes whenever at least \(p\) of the \(h_1, \ldots, h_d\) are equal. Let \(\text{CSM}_d(\mathbb{F}^n)\) denote the set of polynomials \(P \in \text{HCSM}_d(\mathbb{F}^n)\) which satisfy this condition.

In what follows we shall study some explicit structural properties of \(\text{CSM}_d\), \(\text{HCSM}_d\), and derivative polynomials.

**Definition 7.7.6 (Concatenation).** Let \(P \in \text{HCSM}_k(\mathbb{F}^n)\) and \(Q \in \text{HCSM}_l(\mathbb{F}^n)\), for integers \(l, k \geq 1\). Define the concatenation \(P \ast Q \in \text{HCSM}_{k+l}(\mathbb{F}^n)\) as

\[
(P \ast Q)(y_1, \ldots, y_{k+l}) \overset{\text{def}}{=} \sum_{A \subseteq [k+l], \left|A\right|=k} P((y_i)_{i \in A}) \cdot Q((h_j)_{j \in [k+l] \setminus A}).
\]

The following lemma shows how the derivative polynomial acts on product of two polynomials.

**Lemma 7.7.7 ([78]).** Let \(P, Q : \mathbb{F}^n \to \mathbb{T}\) be degree-\(k\) and degree-\(l\) polynomials respectively. Then we have the following identity

\[
D(PQ) = DP \ast DQ.
\]
The following lemma by Tao and Ziegler shows that every polynomial in CSM\(_d(\mathbb{F}^n)\) has a \(d\)-th root which is a classical polynomial.

**Lemma 7.7.8** (Part of Lemma 4.5 in [78]). Let \(d \geq 1\) be an integer, then for every \(Q \in \text{CSM}_d(\mathbb{F}^n)\) there is a degree-\(d\) classical polynomial \(P : \mathbb{F}^n \to T\) for which \(DP = Q\).

It is not difficult to extend this to the more general case of HCSM\(_d(\mathbb{F}^n)\) and non-classical polynomials.

**Lemma 7.7.9.** Let \(d \geq 1\) be an integer, then for every \(Q \in \text{HCSM}_d(\mathbb{F}^n)\) there is a degree-\(d\) (non-classical) polynomial \(P : \mathbb{F}^n \to T\) for which \(DP = Q\).

**Proof.** We know that \(Q\) is a classical polynomial so it there is a unique representation of \(Q\) as following

\[
Q(h_1, \ldots, h_d) = \tau \left( \sum_{i_1, \ldots, i_d \in [n]} c_{i_1, \ldots, i_d} h_{i_1, i_1} \cdots h_{i_d, i_d} \right),
\]

where \(h_k = (h_{k,1}, \ldots, h_{k,n})\), and \(c_{i_1, \ldots, i_d} \in \mathbb{F}\) are coefficients. As a consequence of symmetry of \(Q\), the coefficients \(c_{i_1, \ldots, i_d}\) are symmetric with respect to permutations of \((i_1, \ldots, i_d)\) and thus \(Q\) must be an integer linear combination of expressions of the form

\[
\tau \left( \sum_{(i_1, \ldots, i_d) : \{i_1, \ldots, i_d\} = A} h_{i_1, i_1} \cdots h_{i_d, i_d} \right) \tag{7.11}
\]

where \(A\) is a multiset of cardinality \(d\) with elements from \([n]\). For \(j \in [n]\), let \(a_j\) denote the multiplicity of \(j\) in the set \(A\). Using this we can rewrite (7.11) as

\[
Q_1 * Q_2 * \cdots * Q_n,
\]

where \(Q_j(y_1, \ldots, y_{a_j}) = y_{1,j}y_{2,j} \cdots y_{a_j,j}\). Assume that \(a_j = r_j + (p - 1)t_j\) where \(1 \leq r_j \leq \ldots \)

133
Define
\[ P_j(x) = \frac{|x_j|^{r_j}}{p^{t_j}} \mod 1. \]

Notice that \( \deg(P_j) = r_j + (p - 1)t_j = a_j \) and thus by Lemma 2.1.6 \( DP_j \in \text{HCSM}_{a_j}(\mathbb{F}^n) \).

It is easy to see that \( DP_j(y_1, \ldots, y_{a_j}) \) only depends on variables \( y_{1,j}, \ldots, y_{a_j,j} \) and finally it follows by multilinearity of \( DP_j \) that there exists \( c \in \mathbb{F} \setminus \{0\} \) such that
\[ DP_j(y_1, \ldots, y_{a_j}) = c_{y_{1,j}} \cdots y_{a_j,j}. \]

Now letting \( P'_j := c^{-1}P_j \), it follows by Lemma 7.7.7 that
\[ D(\prod_{j \in [n]} P'_j) = DP'_1 \ast \cdots \ast DP'_n = Q_1 \ast Q_2 \ast \cdots \ast Q_n. \]

We shall define the symmetric power of a polynomial.

**Definition 7.7.10 (Symmetric Power).** Let \( Q \) be a polynomial in \( \text{HCSM}_k(\mathbb{F}^n) \) for \( k \geq 1 \). Define the symmetric power \( \text{Sym}^m(Q) \in \text{HCSM}_{mk}(\mathbb{F}^n) \) for \( m \geq 1 \) by
\[ \text{Sym}^m(Q)(y_1, \ldots, y_{mk}) \overset{\text{def}}{=} \sum_{\mathcal{A}} \prod_{A \in \mathcal{A}} Q((y_i)_{i \in A}), \]
where the sum ranges over all partitions \( \mathcal{A} \) of \( \{1, \ldots, mk\} \) into \( m \) subsets of size \( k \).

Lemma 6.2 in [78] shows that for every \( R \in \text{CSM}_d(\mathbb{F}^n) \) there is a classical degree \( md \) polynomial for which \( DQ = \text{Sym}^mR \). By slight modification of their proof we extend this to \( \text{HCSM}_d(\mathbb{F}^n) \).

**Lemma 7.7.11 (Symmetric Power Rule).** Let \( d, m \) be integers and let \( R \in \text{HCSM}_d(\mathbb{F}^n) \) be a classical symmetric multilinear polynomial. Then there exists a degree-\( md \) polynomial
Q : \mathbb{F}^n \rightarrow \mathbb{T} such that
\[ DQ = \text{Sym}^m(R). \]

Moreover if \( m > 1 \) then \( Q \) can be written as a function of a polynomial of degree \( \leq md - 1 \).

**Proof.** By Lemma 7.7.9 there is a (non-classical) degree-\( d \) polynomial \( P \) for which \( DP = R \).
Assume that \( P \) has depth \( k > 0 \).

Let \( M \geq 0 \) be such that \( p^M \leq m < p^{M+1} \). There exists a degree \( d + M(p-1) \) polynomial \( \tilde{P}_M : \mathbb{F}^n \rightarrow \mathbb{T} \) for which \( p^M \cdot \tilde{P}_M \equiv P \mod 1 \). Notice that \( \tilde{P}_M \) is not unique, as \( p^M T \equiv 0 \mod 1 \) for any polynomial \( T \) of depth at most \( M \). We can pull \( \tilde{P}_M \) to \( \mathbb{Z}/p^{k+M+1} \mathbb{Z} \) and obtain a degree-(\( d + M(p-1) \)) polynomial \( P_M : \mathbb{F}^n \rightarrow \mathbb{Z}/p^{k+M+1} \) for which \( P = \frac{P_M}{p^{k+M}} \mod 1 \).

Let \( Q := \frac{(P_M)_{m'}}{p^{k+1}} \mod 1 \). We claim that

(i) \( \deg(Q) \leq md \).

(ii) \( DQ = \text{Sym}^m(R) \).

Parts (i) and (ii) together imply that \( \deg(Q) = md \). The last part of the theorem follows from the fact that \( d + M(p-1) < md \) when \( m > 1 \) and that \( Q \) can be expressed as a function of \( \tilde{P}_M \) which is of degree \( d + M(p-1) \).

Part (i) will be a special case of the following claim.

**Claim 7.7.12.** Let \( j \geq 0 \) and \( m' \leq m \) be parameters. Then
\[ \deg \left( \frac{D_{h_1} \cdots D_{h_j} P_M}{p^{k+1}} \mod 1 \right) \leq d - j + (m' - 1) \cdot \max(d - j, 1). \]

**Proof.** We break the claim to two cases.

**Case 1, \((d - j + (m' - 1) \cdot \max(d - j, 1) < 0)\):** This in particular implies that \( m' < j - d \). We need to show that \( \frac{D_{h_1} \cdots D_{h_j} P_M}{p^{k+1}} \mod 1 \equiv 0 \) or equivalently that \( D_{h_1} \cdots D_{h_j} P_M \) is divisible by \( p^{k+1} \). Notice that \( \deg(P_M) = d + M(p-1) \) which implies that \( \deg(D_{h_1} \cdots D_{h_j} P_M) \leq d -
For any fixed choice of vectors $h_1, \ldots, h_j \in \mathbb{F}^n$. Thus $p^{M-a}D_{h_1} \cdots D_{h_j} P_M = 0$ for any choice of $a$ for which $d - j + a(p - 1) \leq 0$. This implies that $D_{h_1} \cdots D_{h_j} P_M$ is divisible by $p^{a+k+1}$. It is not difficult to check that choosing $a := \lfloor \frac{m' - 1}{p-1} \rfloor$ we have that $(D_{h_1} \cdots D_{h_j} P_M)$ is divisible by $p^{k+1}$ since $m' < p^{\lfloor \frac{m' - 1}{p-1} \rfloor} + 1 = p^{a+1}$.

**Case 2.** $(d - j + (m' - 1) \cdot \max(d - j, 1) \geq 0)$: We will prove this by downward induction on $j$ and upward induction on $m'$. It is sufficient to show that

$$\deg \left( D_{h_{j+1}} \left( \frac{D_{h_1} \cdots D_{h_j} P_M}{p^{k+1}} \right) \right) \mod 1 \leq d - j + (m' - 1) \cdot \max(d - j, 1) - 1.$$

We will make use of the following identity $(\frac{r+s}{m'}) = \sum_{i=0}^{m'} \binom{r}{i} \binom{s}{m' - i}$ and write

$$D_{h_{j+1}} \left( \frac{D_{h_1} \cdots D_{h_j} P_M}{p^{k+1}} \right) = \sum_{i=1}^{m'} \binom{D_{h_1} \cdots D_{h_{j+1}} P_M}{i} \frac{D_{h_1} \cdots D_{h_j} P_M}{m' - i} \mod 1,$$

where the equalities are modulo 1. Now by the two strong induction hypotheses the degree of each summand in the right-hand-side summation is at most

$$d - (j + 1) + (i - 1) \max(d - (j + 1), 1) + d - j + (m' - i) \max(d - j, 1) \leq d - j + (m' - 1) \max(d - j, 1) - 1,$$

which concludes the claim.

Now (i) follows by choosing $j = 0$ and $m' = m$. We move to (ii), namely we want to prove that $DQ = \text{Sym}^m DP = \text{Sym}^m R$. Notice that

$$D_{h} \left( \frac{P_M}{m} \right) = \sum_{i=1}^{m} \binom{D_{h} P_M}{i} \frac{P_M}{m-i} \mod 1,$$

and thus the terms with $i \geq 2$ have degree at most $mk - i \leq mk - 2$ and thus do not
contribute to \( DQ \). Therefore

\[
DQ(h_1, \ldots, h_{md}) = D \left( D_{h_{md}} \frac{P_M}{p^{k+1}} \right) (h_1, \ldots, h_{md-1})
\]

\[
= D \left( \frac{D_{h_{md}} P_M}{p^{k+1}} \right) \ast D \left( \frac{P_M}{p^{k+1}} \right)
\]

\[
= \sum_{1 \leq i_1 < \cdots < i_{d-1} < md} DP(h_{i_1}, \ldots, h_{i_{d-1}}, h_{md}) D \left( \frac{P_M}{m-1} \right) (h_{j_1}, \ldots, h_{j_{(m-1)d}}),
\]

where \( 1 \leq j_1 < \cdots < j_{(m-1)d} < md \) are such that \( \{j_1, \ldots, j_{(m-1)d}\} = \{1, \ldots, md - 1\} \setminus \{i_1, \ldots, i_{d-1}\} \) and the second equality follows by Lemma 7.7.7. Now the claim \( DQ = \Sym^m DP \) follows by induction on \( m \).

\[\square\]

### 7.7.2 Computing Polynomials with Large Gowers Norm

In this section we will prove Theorem 7.7.2. The next theorem is an algorithmic generalization of Theorem 6.6 in [78] from CSM polynomials to the more general class of HCSM polynomials. The theorem shows what kind of structure can be obtained from biased classical symmetric multilinear polynomials.

**Theorem 7.7.13.** Let \( s \geq 0 \) be an integer and \( \varepsilon, \beta \in (0, 1] \) be parameters. Suppose that \( P \in \text{HCSM}_s(\mathbb{F}^n) \) is such that \( \text{bias}(P) > \varepsilon \). Then there exists a bounded index subspace \( V \) of \( \mathbb{F}^n \) such that on \( V^s \), \( P \) is a linear combination (over \( \mathbb{F} \)) of a bounded number of expressions of the form

\[
\Sym^{m_1}(Q_1) \ast \cdots \ast \Sym^{m_r}(Q_r),
\]

(7.13)

for some \( m_1, \ldots, m_r \geq 1 \) and \( 2 \leq k_1, \ldots, k_r \leq s - 1 \) and \( Q_i \in \text{HCSM}_{k_i}(V) \) for \( i \in [r] \) with

\[m_1 k_1 + \cdots + m_r k_r = s.\]

Moreover, there is a randomized algorithm that given such \( P \), runs in \( O_{\beta, \varepsilon}(n^s) \) and with
probability 1 − \beta outputs \mathcal{P} as a linear combination of terms of form Eq. (7.13) explicitly.

We are given a polynomial \( P \in \text{HCSM}_s(\mathbb{F}^n) \) that is biased. We can use Theorem 7.6.1 to compute \( P \) with a few polynomials of degree \( \leq s - 1 \). We will show that this can be done using only classical symmetric multilinear polynomials. This requires going into the proofs of Lemma 7.2.1 and Lemma 7.4.9 and the following claim which states that a derivative of a classical symmetric multilinear polynomial can be written as linear combination of classical symmetric multilinear polynomials.

**Claim 7.7.14.** Let \( s > 0 \) be an integer, \( P \in \text{HCSM}_s(\mathbb{F}^n) \), and \( h = (h_1, \ldots, h_s) \in (\mathbb{F}^n)^s \). Then the additive derivative of \( P \) in direction \( h \), \( D_h P \), can be written as a linear combination of \( 2^s - 1 \) polynomials \( \{Q_S\}_{S \subset [s], S \neq \emptyset} \), where \( Q_S \in \text{HCSM}_{s-1}(\mathbb{F}^n) \).

**Proof.** We can write

\[
D_h P(x_1, \ldots, x_s) = P(x_1 + h_1, \ldots, x_s + h_s) - P(x_1, \ldots, x_s) = \sum_{S \subset [s], S \neq \emptyset} P((x_i)_{i \in S}, (h_i)_{i \in S})
\]

where the second equality follows from multilinearity of \( P \). Now letting \( Q_S := P((x_i)_{i \notin S}, (1^n)^{s-|S|}) \) we have that \( Q_S \in \text{HCSM}_{s-|S|}(\mathbb{F}^n) \subseteq \text{HCSM}_{s-1}(\mathbb{F}^n) \). \( \Box \)

Now we are able to show that a biased HCSM polynomial can be computed by an HCSM factor of lower degree.

**Definition 7.7.15 (HCSM-Factors).** Let \( d > 0 \) be an integer. A collection \( \mathcal{F} = \{P_{i,j}\}_{1 \leq i \leq d, 1 \leq j \leq m_i} \) for natural numbers \( m_1, \ldots, m_d \) with \( P_{i,j} \in \text{HCSM}_i(\mathbb{F}^n) \) is called an HCSM-factor of degree \( d \).

**Definition 7.7.16 (Measurability).** Let \( d \geq k > 0 \) and let \( \mathcal{F} = \{P_{i,j}\}_{1 \leq i \leq k, 1 \leq j \leq m_i} \) be an HCSM-factor of degree \( k \). A polynomial \( P \in \text{HCSM}_d(\mathbb{F}^n) \) is said to be measurable in \( \mathcal{F} \), if
there exists a collection

\[ A = \{ \alpha = (i_\alpha, j_\alpha, S_\alpha) : 1 \leq i_\alpha \leq k, j_\alpha < M_i_\alpha, S_\alpha \subseteq [d], |S_\alpha| = i_\alpha \}, \]

and a function \( \Gamma : \mathbb{F}^A \rightarrow \mathbb{F} \) such that

\[ P(h_1, \ldots, h_d) = \Gamma \left( P_{i_\alpha j_\alpha} \left( (h_\ell)_{\ell \in S_\alpha} \right) : \alpha \in A \right) \]

Claim 7.7.14 allows us to compute a biased HCSM polynomial by an HCSM-factor of lower degree.

**Corollary 7.7.17.** Let \( \beta \in (0, 1] \) be an error parameter. There is a randomized algorithm that given a polynomial \( P \in \text{HCSM}_s(\mathbb{F}^n) \) with \( \text{bias}(P) \geq \delta > 0 \), runs in \( O_{\delta, \beta, \gamma}(n^d) \) and with probability \( 1 - \beta \), returns an HCSM-factor \( \mathcal{F} = \{ P_{i,j} \}_{1 \leq i \leq d-1, 1 \leq j \leq m_i} \) of degree \( d-1 \) such that \( P \) is measurable in \( \mathcal{F} \).

**Proof.** We observe that Lemma 7.2.1 which is used in the proof of Theorem 7.6.1 and in every step of Lemma 7.4.9, approximates (and therefore computes) a polynomial by a set of its derivatives. The corollary then follows by using Claim 7.7.14 at every step. The \( \gamma \)-unbiased factor can be achieved using the similar modification of Lemma 7.2.3. \( \square \)

**Definition 7.7.18 (Strongly \( \gamma \)-unbiased HCSM-factors).** Let \( \gamma : \mathbb{N} \rightarrow \mathbb{R}^+ \) be a decreasing function, and let \( d \geq k > 0 \). Suppose that \( \mathcal{F} = \{ P_{i,j} \}_{1 \leq i \leq d, 1 \leq j \leq m_i} \) is an HCSM-factor of degree \( d \) and \( \Delta = \{ \delta_{i,j} \in [i] \}_{1 \leq i \leq d, 1 \leq j \leq m_i} \) is a collection of degree bounds. We say that \( \mathcal{F} \) is strongly \( \gamma \)-unbiased if for every \( i \in [d], \mathcal{F}_i = \{ P_{i,j}(h_1, \ldots, h_i) \}_{1 \leq j \leq M_d} \) is strongly \( \gamma \)-unbiased with degree bounds \( \Delta_i = \{ \delta_{i,j} \}_{1 \leq j \leq m_i} \) in the sense of Definition 7.4.6.

**Lemma 7.7.19 (Computation by strongly \( \gamma \)-unbiased factors).** Let \( \beta \in (0, 1] \) be an error parameter, and let \( \gamma : \mathbb{N} \rightarrow \mathbb{R}^+ \) be a decreasing function. There is a randomized algorithm that given a polynomial \( P \in \text{HCSM}_d(\mathbb{F}^n) \) with \( \text{bias}(P) \geq \delta > 0 \), runs in \( O_{\delta, \beta, \gamma}(n^d) \) and with
probability $1 - \beta$, returns a strongly $\gamma$-unbiased HCSM-factor $F = \{P_{i,j}\}_{1 \leq i \leq d-1, 1 \leq j \leq m_i}$ of degree $d - 1$ with degree bound $\Delta = \{\delta_{i,j} \in \delta\}_{1 \leq i \leq d-1, 1 \leq j \leq m_i}$ such that $P$ is measurable in $F$.

**Proof.** By Corollary 7.7.17 we can find an HCSM-factor $F = \{P_{i,j}\}_{1 \leq i \leq d-1, 1 \leq j \leq m_i}$ such that $P$ is measurable in $F$. Now one can use Lemma 7.4.9 and Claim 7.7.14 to repeatedly refine the factor until we obtain a strongly $\gamma$-unbiased HCSM-factor. \qed

Now we are ready to prove Theorem 7.7.13.

**Proof of Theorem 7.7.13:** We are given a polynomial $P \in \text{HCSM}_s(\mathbb{F}^n)$ that is biased. Let $\gamma : \mathbb{N} \to \mathbb{R}^+$ be a decreasing function to be specified later. By Lemma 7.7.19 with high probability we can find a strongly $\gamma$-unbiased HCSM-factor $F = \{P_{i,j}\}_{1 \leq i \leq s-1, 1 \leq j \leq m_i}$ such that $P$ is measurable in $F$. Namely there exists a collection $A = \{\alpha = (i_\alpha, j_\alpha, S_\alpha) : 1 \leq i_\alpha \leq s - 1, 1 \leq j_\alpha < M_{i_\alpha}, S_\alpha \subseteq [s], |S_\alpha| = i_\alpha\}$ and a function $\Gamma : \mathbb{F}^A \to \mathbb{F}$ such that

$$P(h_1, \ldots, h_s) = \Gamma \left( P_{i_\alpha, j_\alpha} \left( (h_\ell)_{\ell \in S_\alpha} \right) : \alpha \in A \right)$$

For technical reasons to be stated later linear polynomials are problematic for the argument and we will pass to a bounded index subspace $V$ of $\mathbb{F}^n$ where $F$ has no linear polynomials. Notice that $\Gamma$ must be invariant under permutations of $h_1, \ldots, h_s$. We will show that $\Gamma$ is moreover multilinear.

**Claim 7.7.20 (HCSM Analogue of Proposition 8.1 from [78]):** For a suitably fast decreasing choice of the function $\gamma : \mathbb{N} \to \mathbb{R}^+$, $\Gamma((x_\alpha)_{\alpha \in A})$ is a linear combination (over $\mathbb{F}$) of monomials

$$x_{\alpha_1} \cdots x_{\alpha_r},$$

where $\alpha_\ell = (i_{\alpha_\ell}, j_{\alpha_\ell}, S_{\alpha_\ell})$ for $\ell = 1, \ldots, r$ are elements of $A$ for which $S_{\alpha_1}, \ldots, S_{\alpha_r}$ partition $\{1, \ldots, s\}$. 

140
Proof. The proof will follow the lines of the proof from [78], except now the factor is an HCSM-factor instead of CSM-factor and the factor is strongly $\gamma$-unbiased instead of regular. We will deduce the equidistribution properties of the HCSM-factor from Lemma 2.2.19.

Split $A = A_\exists s \cup A_\not\exists s$, where $A_\exists s$ is the set of $\alpha = (i_\alpha, j_\alpha, S_\alpha) \in A$ for which $s \in S_\alpha$ and $A_\not\exists s = A \setminus A_\exists s$. We will split $\mathbb{F}^A = \mathbb{F}^{A_\exists s} \times \mathbb{F}^{A_\not\exists s}$ in the natural way, namely for every $x \in \mathbb{F}^A$ we write $x = (x_\exists s, x_\not\exists s)$ where $x_\exists s \in \mathbb{F}^{A_\exists s}$ and $x_\not\exists s \in \mathbb{F}^{A_\not\exists s}$. We will show the following linearity

$$\Gamma(x_\exists s + y_\exists s, x_\not\exists s) = \Gamma(x_\exists s, x_\not\exists s) + \Gamma(y_\exists s, x_\not\exists s),$$

(7.14)

for every $x_\exists s, y_\exists s \in \mathbb{F}^{A_\exists s}$ and $x_\not\exists s \in \mathbb{F}^{A_\not\exists s}$. To prove (7.14), by multilinearity of the polynomials $P$ and $P_{i,j}$ for $i \in [s - 1], j \in [m_i]$, it suffices to find $h_1, \ldots, h_s, h'_s \in \mathbb{F}^n$ such that

- $P_{i_\alpha, j_\alpha}((h_\ell)_{\ell \in S_\alpha}) = x_\exists s$, for every $\alpha \in A_\exists s$,

- $P_{i_\alpha, j_\alpha}((h_\ell)_{\ell \in S_\alpha}) = x_\not\exists s$, for every $\alpha \in A_\not\exists s$, and

- $P_{i_\alpha, j_\alpha}((h_\ell)_{\ell \in S_\alpha, i \not= s}, h'_s) = y_\exists s$, for every $\alpha \in A_\exists s$.

The existence of such vectors $h_1, \ldots, h_d, h'_d$ is implied by the claim below. Now we can write

$$\Gamma(x_\exists s, x_\not\exists s) + \Gamma(y_\exists s, x_\not\exists s) = P(h_1, \ldots, h_s) + P(h_1, \ldots, h_{s-1}, h'_s) = P(h_1, \ldots, h_s, h'_s)$$

$$= \Gamma \left( \left( P_{i_\alpha, j_\alpha}((h_\ell)_{\ell \in S_\alpha \setminus \{s\}}, h_s + h'_s) \right)_{\alpha \in A_\exists s} \cdot \left( P_{i_\alpha, j_\alpha}((h_\ell)_{\ell \in S_\alpha})_{\alpha \in A_\not\exists s} \right) \right)$$

$$= \Gamma(x_\exists s + y_\exists s, x_\not\exists s),$$

where the second equality is multilinearity of $P$ and the last equality is multilinearity of the polynomials $P_{i,j}$. The following claim shows that choosing a suitably fast decreasing $\gamma$ implies existence of vectors $h_1, \ldots, h_s, h'_s$. 

141
Claim 7.7.21. Let $s > 0$. There is a sufficiently fast decreasing choice of $\gamma : \mathbb{N} \to \mathbb{R}^+$ such that for every degree $(s - 1)$ $\gamma$-unbiased HCSM-factor $\mathcal{F} = \{P_{i,j}\}_{1 \leq i \leq s-1, 1 \leq j \leq m_i}$ the following holds. For every collection

$$A = \{\alpha = (i_\alpha, j_\alpha, S_\alpha) : i_\alpha \in [s - 1], j_\alpha \in [m_i], S_\alpha \subseteq [s + 1], |S_\alpha| = i_\alpha\},$$

and collection of field elements $(x_\alpha)_{\alpha \in A}$, there exists $h_1, \ldots, h_{s+1} \in \mathbb{F}^n$ such that for every $\alpha \in A$,

$$P_{i_\alpha,j_\alpha}((h_\ell)_{\ell \in S_\alpha}) = x_\alpha.$$

Proof. We will prove this by showing that the distribution of the vector

$$(P_{i_\alpha,j_\alpha}((h_\ell)_{\ell \in S_\alpha}))_{\alpha \in A}$$

for uniformly at random $h_1, \ldots, h_{s+1}$ can be made arbitrarily close to the uniform distribution over $\mathbb{F}^{|A|}$. To do this, it is sufficient to prove that for every collection of coefficients $\{\lambda_\alpha \in \mathbb{F}\}_{\alpha \in A}$ we have

$$\left| \mathbb{E}_{h_1, \ldots, h_{s+1} \in \mathbb{F}^n} e\left( \sum_{\alpha \in A} \lambda_\alpha P_{i_\alpha,j_\alpha}((h_\ell)_{\ell \in S_\alpha}) \right) \right| \leq \gamma(\dim(\mathcal{F}))^{2-s+1}.$$

This will be a corollary to Lemma 2.2.19. Let $\beta \in A$ be such that $|S_\beta| = i_\beta$ is maximized as $\lambda_\beta$ is nonzero. Without loss of generality assume that $S_\beta = \{1, \ldots, k\}$, where $i_\beta = k$. We will derive $P_A(h_1, \ldots, h_{s+1}) \overset{\text{def}}{=} \sum_{\alpha \in A} \lambda_\alpha P_{i_\alpha,j_\alpha}((h_\ell)_{\ell \in S_\alpha})$ in directions $h_1, \ldots, h_k$. Notice that for any $\alpha$ with $S_\alpha \neq S_\beta$, the summand corresponding to $\alpha$ does not depend on at least one vector among $h_1, \ldots, h_k$ and will therefore vanish. Thus we have

142
$$D_{Y_1}D_{Y_2} \cdots D_{Y_k} P_A(h_1, \ldots, h_{s+1}) = \sum_{\alpha \in A: S_{\alpha} = \{1, \ldots, k\}} \lambda_{\alpha} D_{y_1}D_{y_2} \cdots D_{y_k} P_{i_\alpha j_\alpha}(h_1, \ldots, h_k)$$

$$= \sum_{\alpha \in A: S_{\alpha} = \{1, \ldots, k\}} \lambda_{\alpha} P_{i_\alpha}(y_1, \ldots, y_k),$$

where $Y_1, \ldots, Y_k \in (\mathbb{F}^n)^{s+1}$ are vectors that are supported only on $h_1, \ldots, h_k$ respectively and $y_i$ is the restriction of $Y_i$ to the $h_i$ coordinates. The second equality follows from the fact that $P_{i,j}$ are multilinear and classical. Now similar to Claim 7.4.8 we have that

$$\left| \mathbb{E}_{h_1, \ldots, h_{s+1} \in \mathbb{F}^n} e \left( \sum_{\alpha \in A} \lambda_{\alpha} P_{i_\alpha j_\beta}(h_\ell)_{\ell \in S_{\alpha}} \right) \right|^{2^k} \leq \left| \mathbb{E}_{y_1, \ldots, y_k \in \mathbb{F}^n} e \left( \sum_{\alpha \in A: S_{\alpha} = S_{\beta}} \lambda_{\alpha} P_{i_\alpha j_\alpha}(y_1, \ldots, y_k) \right) \right| \leq \gamma(|\dim(\mathcal{F})|),$$

where the first inequality is repeated Cauchy-Schwarz inequality and the second inequality follows by $\mathcal{F}$ being $\gamma$-unbiased. Now the claim follows by the fact that $k = i_\beta \leq s - 1$. □

By symmetry (7.14) implies that

$$\Gamma(x_{\exists j} + y_{\exists j}, x_{\not\exists j}) = \Gamma(x_{\exists j}, x_{\not\exists j}) + \Gamma(y_{\exists j}, x_{\not\exists j}), \quad (7.15)$$

for every $j \in [s]$, where $A_{\exists j}$ and $A_{\not\exists j}$ are defined analogously to $A_{\exists s}$ and $A_{\not\exists s}$. This shows that $\Gamma$ is multilinear. Now we show that all monomials of $\Gamma$ are of the form $x_{\alpha_1} \cdots x_{\alpha_r}$ where $S_{\alpha_1}, \ldots, S_{\alpha_r}$ partition $\{1, \ldots, s\}$. For an arbitrary $\alpha \in A$, we will look at $D_{y_\alpha} \Gamma$ where $y_\alpha$ is only supported on $x_\alpha$. Let $\beta \in A$ be such that $S_{\beta}$ intersects $S_{\alpha}$, and let $\ell$ be an arbitrary element of $S_{\beta} \cap S_{\alpha}$. Using (7.15) for $j = \ell$ implies that $D_{y_\alpha} \Gamma$ does not depend on $x_\beta$. 143
This implies that all the monomials of \( \Gamma \) are of the form \( x_{\alpha_1} \cdots x_{\alpha_r} \), where \( S_{\alpha_1}, \ldots, S_{\alpha_r} \) are disjoint subsets of \([s]\). The fact that \( [s] = \bigcup_{i=1}^r S_{\alpha_i} \) follows from multilinearity of \( \Gamma \).  

By Claim 7.7.20, we have

\[
P(y_1, \ldots, y_s) = \sum_{\alpha_1, \ldots, \alpha_r} c_{\alpha_1, \ldots, \alpha_r} \cdot P_{i_{\alpha_1}, j_{\alpha_1}}((y_i)_{i \in S_{\alpha_1}}) \cdots P_{i_{\alpha_r}, j_{\alpha_r}}((y_i)_{i \in S_{\alpha_r}}),
\]

for field elements \( c_{\alpha_1, \ldots, \alpha_r} \), where \( y_i \)'s are vectors from \( \mathbb{F}^n \), \( \{\alpha_1, \ldots, \alpha_r\} \) are collections of elements of \( A \) for which \( S_{\alpha_1}, \ldots, S_{\alpha_r} \) partition \([s]\). Since \( P \) is invariant under permutations of \((y_1, \ldots, y_s)\), so should be the coefficients \( c_{\alpha_1, \ldots, \alpha_r} \) and therefore we may split this sum into orbits of the action of the permutation group, and conclude that \( P \) is a linear combination of basic symmetric monomials

\[
\text{Sym}^{m_1}(P_{i_{\beta_1}, j_{\beta_1}}) \ast \cdots \ast \text{Sym}^{m_l}(P_{i_{\beta_l}, j_{\beta_l}}),
\]

with \( m_1 i_{\beta_1} + \cdots + m_l i_{\beta_l} = s \). Note that \( i_{\alpha_1}, \ldots, i_{\alpha_l} \geq 2 \) because we had discarded all of the linear polynomials from our factor by restricting to the bounded index subspace \( V \). This finishes the proof of the existence of such a decomposition. It is left to show that we can find the decomposition algorithmically. This is simple using Lemma 7.2.5 for \( \Gamma \), since the factor is regular and provided that \( \gamma \) decreases suitably fast. Since \( \Gamma : \mathbb{F}^A \to \mathbb{F} \), we need only \( O(|A|(1)) \) queries to \( \Gamma \) to find explicit representation as a linear combination of monomials, and this can be guaranteed with high probability using Lemma 7.2.5.  

Theorem 7.7.2 now follows as a corollary.

**Proof of Theorem 7.7.2:** By Lemma 2.1.6 we know that \( DP \in \text{HCSM}_s(\mathbb{F}^n) \) and \( \text{bias}(DP) > \varepsilon \). By Theorem 7.7.13 we can find with probability \( 1 - \mu \) a bounded index \( V \) of \( \mathbb{F}^n \) and write \( DP \) as a linear combination (over \( \mathbb{F} \)) of a bounded number of expressions of
the form
\[ \text{Sym}^{m_1}(Q_1) \ast \cdots \ast \text{Sym}^{m_r}(Q_r), \tag{7.16} \]
for some \( m_1, \ldots, m_r \geq 1 \) and \( 2 \leq k_1, \ldots, k_r \leq s - 1 \) and \( Q_i \in \text{HCSM}_{k_i}(V) \) for \( i \in [r] \) with
\[ m_1 k_1 + \cdots + m_r k_r = s. \]

Now notice that by Lemma 7.7.9 for every \( Q_i \) we can find a degree-\( k_i \) polynomial \( \tilde{P}_i \) such that \( D \tilde{P}_i = Q_i \), and by Lemma 7.7.11 we can find a degree \( m_i k_i \) (non-classical) polynomial \( P_i \) for which \( \text{Sym}^{m_i}(D \tilde{P}_i) = D P_i \). Moreover if \( m_i \geq 2 \), \( P_i \) can be written as a function of a lower degree polynomial \( P^{(i)}_M \). Now notice that if \( r > 1 \) then by Lemma 7.7.7 for the polynomial \( \prod_{i=1}^{r} P_i \) we have
\[ D(\prod_{i=1}^{r} P_i) = \text{Sym}^{m_1}(Q_1) \ast \cdots \ast \text{Sym}^{m_r}(Q_r), \]
and \( \prod_{i=1}^{r} P_i \) is measurable in \( \{P_1, \ldots, P_r\} \). In the case when \( r = 1 \) then \( 2 \leq k_1 \leq s - 1 \) implies that \( m_1 \geq 2 \) which concludes the theorem over the subspace \( V \).

Notice that if \( V = \mathbb{F}^n \) we would be done. We have shown that we can find a degree-\( s \) polynomial \( Q \) and degree \( \leq s - 1 \) (non-classical) polynomials \( \tilde{Q}_1, \ldots, \tilde{Q}_C \) such that the statement of the theorem holds on a bounded index subspace \( V \). We can extend the statement to all of \( \mathbb{F}^n \) by a simple derivative trick. Let \( h_1, \ldots, h_K \in \mathbb{F}^n \) be the presentatives of \( \mathbb{F}^n/V \), where \( K = \frac{|\mathbb{F}|^n}{|V|} \). Suppose that \( x \) is a point in \( \mathbb{F}^n \), then there is \( i \in [K] \) such that \( x - h_i \in V \). This means that \( Q(x) = Q(x - h_i) - Q(x) - Q(x - h_i) = D_{-h_i} Q(x) - Q(x - h_i) \). Now notice that \( \deg(D_{-h_i} Q) \leq s - 1 \) and since \( x - h_i \) is in \( V \), \( Q(x - h_i) \) can be written as a function of \( \tilde{Q}_1, \ldots, \tilde{Q}_C \), and thus \( Q \) is measurable in \( \{D_{-h_1} Q, \ldots, D_{-h_K} Q\} \cup \{\tilde{Q}_1, \ldots, \tilde{Q}_C\} \). \( \square \)
7.8 Application: Polynomial Decompositions

Consider the following family of properties of functions over a finite field $\mathbb{F}$ of fixed prime order $p$.

**Definition 7.8.1.** Given a positive integer $k$, a vector of positive integers $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and a function $\Gamma : \mathbb{F}_p^k \to \mathbb{F}_p$, we say that a function $P : \mathbb{F}^n \to \mathbb{F}$ is $(k, \Delta, \Gamma)$-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 $(k, \Delta, \Gamma)$-decomposition.

Informally, a degree-structural property refers to a property from the family of $(k, \Delta, \Gamma)$-structured properties. Our main result in this section is that every degree-structural property can be decided in polynomial time:

**Theorem 7.8.2.** For every finite field $\mathbb{F}$ of prime order, positive integers $d$, $k$, every vector of positive integers $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and every function $\Gamma : \mathbb{F}^k \to \mathbb{F}$, there is a deterministic algorithm $A_{k, \Delta, \Gamma}$ 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 $(k, \Delta, \Gamma)$-decomposition of $P$ if one exists while otherwise returning NO.

In [13], a similar result was obtained for the special case of $d < |\mathbb{F}|$, using our algorithmic Green-Tao regularity lemma that applies in the high characteristic case.

The algorithm itself is quite simple. Given a polynomial $P : \mathbb{F}^n \to \mathbb{F}$, we first use Corollary 7.7.3 to write it as a function of a uniform factor $\{Q_1, \ldots, Q_m\}$ of non-classical polynomials, i.e. $P(x) = G(Q_1(x), \ldots, Q_m(x))$ where $m = O(1)$ and $G : \mathbb{T}^m \to \mathbb{F}$. Now, the proof technique of [13] (similar reasoning also in proof of Theorem 1.7 of [15]) shows that the
only way $P$ can have a $(k, \Delta, \Gamma)$-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 classical polynomial of degree at most $\Delta_i$. 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.

Another natural way to view the algorithm is as an induction on the dimension. The proof shows that there exists $i \in [n]$ such that $P$ is $(k, \Delta, \Gamma)$-structured if and only if $P|_{x_i=0}$ is $(k, \Delta, \Gamma)$-structured. We can keep doing this until the number of variables reaches a constant (that is dependent only on $\mathbb{F}$, $k$, $\Delta$ and $\Gamma$), and moreover, at every step, in polynomial time, we can check whether this base case has been reached. When the base case is reached, in constant time, we can compute a $(k, \Delta, \Gamma)$-structure if one exists. Finally, the proof also shows a way to lift up the recovered decomposition back to the original input on $n$ variables. Thus, the algorithm is another instance of the “restrict-solve-lift” paradigm that is used in many algebraic algorithms, such as those for factoring.
CHAPTER 8
NORMS DEFINED BY LINEAR FORMS

In this chapter we study norms that are defined by averaging over a set of linear forms. The results in this chapter are based on the joint work with Hamed Hatami and Shachar Lovett. An example of such norms is the Gowers norms. Such norms can be defined over any abelian group, but here similar to the rest of this thesis, we will only be interested in the case where the group is $\mathbb{F}^n$, where $\mathbb{F} = \mathbb{F}_p$ for a prime $p$. For the most general averages that we could define, we also allow the terms $f(L_i(X))$ to have non-negative real valued powers.

Definition 8.0.3. Let $\mathcal{L} = \{L_1, ..., L_l\}$ be a family of linear forms in $\mathbb{F}^k$, and $\alpha, \beta \in \mathbb{R}_{\geq 0}$. For a function $f : \mathbb{F}^n \to \mathbb{C}$, define

$$t_{\mathcal{L}, \alpha, \beta}(f) \overset{\text{def}}{=} \mathbb{E} \left[ \prod_{i=1}^{l} f(L_i(X))^{\alpha_i} \bar{f}(L_i(X))^{\beta_i} \right],$$

where $X \in (\mathbb{F}^n)^k$ is chosen uniformly at random. Letting $T := (\mathcal{L}, \alpha, \beta)$, for a function $f : \mathbb{F}^n \to \mathbb{C}$ we write

$$\|f\|_T \overset{\text{def}}{=} t_{\mathcal{L}, \alpha, \beta}(f)^{1/|\alpha, \beta|},$$

where $|\alpha, \beta| \overset{\text{def}}{=} \sum_{i=1}^{l} (\alpha_i + \beta_i)$.

Remark 8.0.4. For $\|\cdot\|_T$ to be well defined over complex valued functions, we will require $\alpha_i - \beta_i$ to be an integer for every $i$. Notice that in the study of real functions we will not have conjugations, namely we can assume without loss of generality that $\beta = 0$ and we will not make any further assumptions about $\alpha$.

The next theorem to be proved in Section 8.4 states that in order for $t_{\mathcal{L}, \alpha, \beta}$ to be a norm, $\alpha$ and $\beta$ must be one of two specific forms.
Informal Statement (Theorem 8.2.2). Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm. Then either there is an $s \geq 1$ such that

$$t_{\mathcal{L}, \alpha, \beta}(f) = \mathbb{E} \left[ \prod_{i=1}^{l} |f(L_i(X))|^{s/2} \right],$$

or for every $i$, $\{\alpha_i, \beta_i\} = \{0, 1\}$. We call norms of the former case type (I) norms and norms of the latter case type (II) norms.

In order to prove Theorem 8.2.2, in Section 8.3 we develop tools such as Gowers type Cauchy-Schwarz inequality and a Hölder type inequality for $\|\cdot\|_T$ norms. This approach follows the same lines as [48].

Example. Gowers $U^2$ norm over $\{G \to \mathbb{R}\}$ can be defined by $(\mathcal{L}, \alpha, \beta)$, where $\mathcal{L} = \{(1,0,0),(1,1,0),(1,0,1),(1,1,1)\}$, $\alpha = (1,1,1,1)$, and $\beta = (0,0,0,0)$, and therefore it is a type (I) norm. Namely

$$\|f\|_{U^2} = \|f\|_{\mathcal{L}, \mathcal{M}}.$$

Later we investigate the properties of type (I) and type (II) norms over the space of bounded valued complex functions. Lemma 8.5.2 is a simple observation that type (II) norms are equivalent to $L_1$ norm. Let $\mathbb{D}$ denote the complex unit disc $\mathbb{D} \overset{\text{def}}{=} \{c \in \mathbb{C} : |c| \leq 1\}$.

Informal Statement (Lemma 8.5.2). Let $T = (\mathcal{L}, \alpha, \beta)$, be such that $\|\cdot\|_T$ is a type (II) norm. Then for every $g : \mathbb{F}^n \to \mathbb{D}$ we have

- $\|g\|_1 \leq \|g\|_T$, and

- $\|g\|_T \leq \|g\|_{1/|\alpha, \beta|}^{(1/|\alpha, \beta|)}$.

In Theorem 8.4.3 we prove that every type (I) norm must be invariant under affine transformations, which means we can assume that the first coordinate of the linear forms defining the norm is 1. This allows us to find a positive integer $z = z(\mathcal{L}, \mathcal{M})$ for which we prove that every type (I) norm is essentially equivalent to the Gowers norm $\|\cdot\|_{U^{z+1}}$.  

149
Informal Statement (Theorem 8.5.4). Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a type (I) norm and $\mathcal{L} \cup \mathcal{M}$ is of complexity $\leq p$. Then there exist an integer $z$ and functions $\delta_1, \delta_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for every $n$ and $f : \mathbb{F}^n \to \mathbb{D}$,

i. If $\|f\|_T \leq \varepsilon$ then $\|f\|_{U^{z+1}} \leq \delta_1(\varepsilon)$, and

ii. If $\|f\|_{U^{z+1}} \leq \varepsilon$ then $\|f\|_T \leq \delta_2(\varepsilon)$.

8.1 Notation

Recall the definition of type (I) and (II) norms from the statement of Theorem 8.2.2 as stated in the introduction. Since a type (II) norm is defined only by $\mathcal{L}$ and $s$, from now on we denote a type (II) norm by $\|\cdot\|_{T_{s}}$, where $T = (\mathcal{L}, s)$. In the case of type (I) norms, namely when for all $i$, $\{\alpha_i, \beta_i\} = \{0, 1\}$, we drop the $\alpha$ and $\beta$ and work with two collections of linear forms $\mathcal{L} = \{L_1, ..., L_l\}$ and $\mathcal{M} = \{M_1, ..., M_m\}$ and write

$$t_{\mathcal{L}, \mathcal{M}}(f) = \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{l} f(L_i(X)) \prod_{i=1}^{m} \bar{f}(M_i(X)) \right],$$

moreover for $T = (\mathcal{L}, \mathcal{M})$ define

$$\|f\|_T = t_{\mathcal{L}, \mathcal{M}}(f)^{1/(l+m)}.$$  

From now on by $T = (\mathcal{L}, \alpha, \beta)$ we mean that $\mathcal{L} = \{L_1, ..., L_l\}$ is a family of linear forms and $\alpha, \beta \in \mathbb{R}_{\geq 0}^l$. Also by $T = (\mathcal{L}, \mathcal{M})$, we mean that $\mathcal{L} = \{L_1, ..., L_l\}$ and $\mathcal{M} = \{M_1, ..., M_m\}$ are families of linear forms over $k$ variables where $k$ is a fixed positive integer.

Definition 8.1.1. We will define $r_{\mathcal{L}, \mathcal{M}}(f)$ and $r_{\mathcal{L}, \alpha, \beta}(f)$ similar to $t_{\mathcal{L}, \mathcal{M}}$ and $t_{\mathcal{L}, \alpha, \beta}$,

$$r_{\mathcal{L}, \mathcal{M}}(f)(X) \overset{\text{def}}{=} \left[ \prod_{i=1}^{l} f(L_i(X)) \prod_{i=1}^{m} \bar{f}(M_i(X)) \right],$$

150
and
\[ r_{\mathcal{L}, \alpha, \beta}(f)(X) \overset{\text{def}}{=} \prod_{i=1}^{l} f(L_i(X))^{\alpha_i} \bar{f}(L_i(X))^{\beta_i} . \]

This allows one to simply write
\[ t_{\mathcal{L}, \alpha, \beta}(f) = \mathbb{E}_{X \in (\mathbb{F}^n)^k} [ r_{\mathcal{L}, \alpha, \beta}(f)(X) ] , \]

and
\[ t_{\mathcal{L}, \mathcal{M}}(f) = \mathbb{E}_{X \in (\mathbb{F}^n)^k} [ r_{\mathcal{L}, \mathcal{M}}(f)(X) ] . \]

### 8.2 Norms defined by linear forms

Let \( \mathcal{L} = \{ L_1, ..., L_l \} \) be a system of linear forms on \( k \) variables. Analytic averages such as \( \mathbb{E} \left[ \prod_i f(L_i(X)) \right] \) have been subject of interest in several works, see for example [51, 37]. In this section we will consider a generalized version of such averages and study these averages in the cases when they obey norm properties. Namely letting \( T = (\mathcal{L}, \alpha, \beta) \), where \( \alpha, \beta \in \mathbb{R}^l_\geq 0 \), we study norms of the following form:

\[ \| f \|_T = \mathbb{E}_{x \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{l} f(L_i(X))^{\alpha_i} \prod_{i=1}^{l} \bar{f}(L_i(x))^{\beta_i} \right] . \]  

(8.1)

We will assume that for every \( i \), \( \alpha_i + \beta_i \neq 0 \), since otherwise \( L_i \) is redundant.

**Remark 8.2.1.** For \( \| \cdot \|_T \) in Eq. (8.1) to be well defined over complex functions, we need to assume that for all \( i \in [l] \), \( \alpha_i - \beta_i \in \mathbb{Z} \). Notice that in the case of real functions there is no need for \( \beta \) and \( T = (\mathcal{L}, \alpha) \) would be sufficient to define such norms, where \( \alpha \in \mathbb{R}_{\geq 0} \).

In the following theorem we show that for \( \| \cdot \|_T \) to be a norm, \( T \) has to be one of only two types.

151
Theorem 8.2.2. Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm. Then one of the following holds

(i) For every $i \in [l]$, $\{\alpha_i, \beta_i\} = \{0, 1\}$, or

(ii) There exists $s \geq 1$ such that for every $i \in [l]$, $\alpha_i = \beta_i = s/2$.

We call norms of the former case type (I) norms and norms of the latter case, type (II) norms.

We prove the above theorem by extending ideas from the study of hypergraph normings by the first author [48]. One of the main tools that is used in [48] to prove a characterization theorem for hypergraph normings is certain Hölder type inequalities for such norms. We prove Theorem 8.2.2 by showing that similar Hölder type inequalities also hold for norms defined by systems of linear forms. We state and prove these inequalities in Section 8.3. Finally we present the proof of Theorem 8.2.2 in Section 8.4.

8.3 Hölder type inequalities

Definition 8.3.1. Let $f_1, f_2 : \mathbb{F}^n \to \mathbb{C}$ be two functions. The tensor of $f_1$ and $f_2$, denoted by $f_1 \otimes f_2 : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{C}$ is defined as follows

$$f_1 \otimes f_2(x_1, x_2) := f_1(x_1) \cdot f_2(x_2).$$

Moreover by $f^\otimes l$ we mean the $l$-th tensor power of $f$, $f^\otimes l \overset{\text{def}}{=} f \otimes f \otimes \ldots \otimes f : \mathbb{F}^{nl} \to \mathbb{C}$.

It is a simple observation that $\|\cdot\|_T$ norms are multiplicative with respect the above tensor products.

Claim 8.3.2. Let $T = (\mathcal{L}, \alpha, \beta)$. For functions $f_1, f_2 : \mathbb{F}^n \to \mathbb{C}$ we have

$$t_{\mathcal{L}, \alpha, \beta}(f_1 \otimes f_2) = t_{\mathcal{L}, \alpha, \beta}(f_1) \cdot t_{\mathcal{L}, \alpha, \beta}(f_2).$$
Proof. This follows from the independence of entries of \( x \), and the fact that conjugation is distributive over both addition and multiplication. \( \square \)

A very similar argument shows that

\[
E \left[ \prod_{i=1}^{l} f_i \otimes \overline{f_i}(L_i(X))^{\alpha_i} \otimes g_i(L_i(X))^{\beta_i} \right] = E \left[ \prod_{i=1}^{l} f_i(L_i(X))^{\alpha_i} \overline{f_i}(L_i(X))^{\beta_i} \right]^2 \tag{8.2}
\]

**Theorem 8.3.3** (Hölder Type Inequality). Let \( T = (\mathcal{L}, \alpha, \beta) \). If \( \| \cdot \|_T \) is a semi-norm, then for every \( a \in [l] \), if \( \alpha_a \neq 0 \) then

\[
E \left[ g(L_a(X)) \cdot \prod_{i=1}^{l} f_i(L_i(X))^{\alpha_i-e_a(i)} \overline{f_i}(L_i(X))^{\beta_i} \right] \leq \| f \|_{T}^{\| \alpha, \beta \|-1} \| g \|_{T}, \tag{8.3}
\]

and if \( \beta_a \neq 0 \) then

\[
E \left[ \overline{g}(L_a(X)) \cdot \prod_{i=1}^{l} f_i(L_i(X))^{\alpha_i} \overline{f_i}(L_i(X))^{\beta_i-e_a(i)} \right] \leq \| f \|_{T}^{\| \alpha, \beta \|-1} \| g \|_{T}, \tag{8.4}
\]

where \( e_a \) is the vector which is equal to 1 on the \( a \)-th entry and 0 everywhere else. Moreover if there exists \( a \in [l] \) such that at least one of Eq. (8.3) or Eq. (8.4) holds for every pair of functions \( f \) and \( g \), then \( \| \cdot \|_T \) satisfies the triangle inequality.

Proof. We first prove the fact that each of Eq. (8.3) or Eq. (8.4) implies the triangle inequality. Assume that \( f, g : \mathbb{F}^n \to \mathbb{C} \) and that Eq. (8.3) holds for \( a \). Then

\[
t_{\mathcal{L},\alpha,\beta}(f + g) = E \left[ (g(L_a(X)) + f(L_a(X))) \cdot \prod_{i=1}^{l} (f + g)(L_i(X))^{\alpha_i-e_a(i)}(f + g)(L_i(X))^{\beta_i} \right] \\
\leq t_{\mathcal{L},\alpha,\beta}(f + g)^{1-1/\| \alpha, \beta \|} (\| f \|_T + \| g \|_T)
\]

which simplifies to the triangle inequality for \( \| \cdot \|_T \). The proof for Eq. (8.4) is identical.
For the other direction, assume $f, g : \mathbb{F}^n \to \mathbb{C}$ are functions with $\|f\|_T, \|g\|_T \leq \infty$. Since $\| \cdot \|_T$ is a semi-norm, for every $\delta \geq 0$

$$\|f + \delta g\|_T \leq \|f\|_T + \delta \|g\|_T,$$

thus

$$\frac{d\|f + \delta g\|_T}{d\delta}\bigg|_{0} \leq \|g\|_T. \quad (8.5)$$

Now expanding the derivative

$$\frac{d\|f + \delta g\|_T}{d\delta} = \frac{\|f + \delta g\|^{1-|\alpha, \beta|}}{|\alpha, \beta|} \cdot \text{E} \left[ \sum_{i=1}^{l} \left( \alpha_i g(L_i(x)) r_{L, \alpha-e_i, \beta}(f + \delta g)(x) + \beta_i \bar{g}(L_i(x)) r_{L, \alpha, \beta-e_i}(f + \delta g)(x) \right) \right].$$

Thus by Eq. (8.5)

$$\frac{\|f\|^{1-|\alpha, \beta|}}{|\alpha, \beta|} \cdot \text{E} \left[ \sum_{i=1}^{l} \left( \alpha_i g(L_i(x)) r_{L, \alpha-e_i, \beta}(f)(x) + \beta_i \bar{g}(L_i(x)) r_{L, \alpha, \beta-e_i}(f)(x) \right) \right] \leq \|g\|_T,$$

which can be rearranged to

$$\frac{1}{|\alpha, \beta|} \cdot \text{E} \left[ \sum_{i=1}^{l} \left( \alpha_i g(L_i(x)) r_{L, \alpha-e_i, \beta}(f)(x) + \beta_i \bar{g}(L_i(x)) r_{L, \alpha, \beta-e_i}(f)(x) \right) \right] \leq \|f\|_{T}^{1-|\alpha, \beta|} \|g\|_T. \quad (8.6)$$

Replacing $f$ and $g$ by $(f \otimes \bar{f})^\otimes m$ and $(g \otimes \bar{g})^\otimes m$ for an integer $m > 0$, by Claim 8.3.2 and
Eq. (8.2) we have

\[
\frac{1}{|\alpha, \beta|} \left[ \sum_{i=1}^{l} \left( \alpha_i | \mathbb{E}[g(L_i(x))r_{L, \alpha-e_i, \beta}(f)(x)] \right)^{2m} + \beta_i \left| \mathbb{E}[\tilde{g}(L_i(x))r_{L, \alpha, \beta-e_i}(f)(x)] \right|^{2m} \right] \leq \left( \|f\|_{T} \|\alpha, \beta\| - 1 \|g\|_{T} \right)^{2m}.
\]

The proof follows from the fact that this inequality holds for all \( m > 0 \) and that

\[
|\alpha, \beta| = \sum_{i=1}^{l} \alpha_i + \sum_{i=1}^{m} \beta_i.
\]

One simple corollary to the above theorem is that the absolute value of the average of a function \( f \) is bounded by its \( \|\cdot\|_T \) norm.

**Corollary 8.3.4.** Let \( T = (\mathcal{L}, \alpha, \beta) \) be such that \( \|\cdot\|_T \) is a norm. For every \( g : \mathbb{F}^n \rightarrow \mathbb{C} \) we have

\[
|\mathbb{E}[g]| \leq \|g\|_{T}.
\]

**Proof.** Define \( f(x) = 1 \) for all \( x \in \mathbb{F}^n \). Assume there exists \( i \in [l] \) such that \( \alpha_i \neq 0 \), then we have

\[
|\mathbb{E}[g(x)]| = \left| \mathbb{E} \left[ g(L_i(x))r_{\mathcal{L}, \alpha-e_i, \beta}(f)(x) \right] \right| \leq \|g\|_{T},
\]

where the inequality follows from Theorem 8.3.3. The case when for all \( i \in [l] \), \( \alpha_i = 0 \) is handled similarly using the second inequality from Theorem 8.3.3. \( \square \)

Using the above claim we have the following corollary to Theorem 8.3.3.
Corollary 8.3.5. Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm. Then for every $i \in [l]$,

$$\alpha_i + \beta_i \geq 1.$$ 

Proof. Assume for contradiction that $\alpha_i + \beta_i < 1$. By Eq. (8.3) we have

$$\left| \mathbb{E} \left[ g(L_i(X)) \cdot r_{\mathcal{L}, \alpha - e_i, \beta}(f)(X) \right] \right| \leq \|f\|_T^{[\alpha,\beta]-1} \|g\|_T. \quad (8.7)$$

The next claim shows that there is $n$ and $Y \in (\mathbb{F}^n)^k$ such that $L_i(Y) \neq L_j(Y)$ for all $j \neq i$.

Claim 8.3.6. Let $\mathcal{L} = \{L_1, \ldots, L_l\}$ be a family of linear forms over $k$ variables. For every $i \in [l]$, there exists an integer $n$ and $X \in (\mathbb{F}^n)^k$ such that

$$L_i(X) \notin \{L_j(X) : j \neq i\}.$$ 

Proof. For every $j \neq i$, let $A_j$ denote the set of solutions to $L_i(X) = L_j(X)$. Since $L_i - L_j \neq 0$ we have that $|A_j| \leq p^{n(k-1)}$. Thus

$$\left| \bigcup_{j \neq i} A_j \right| \leq l \cdot p^{n(k-1)},$$

which for large enough $n$ is strictly less than $\left| (\mathbb{F}^n)^k \right| = p^{nk}$. \qed

Let $\omega = L_i(Y)$, and for $\varepsilon > 0$ define $f : \mathbb{F}^n \to \mathbb{C}$ as follows

$$f(x) = \begin{cases} 
\varepsilon & x = \omega \\
1 & \text{Otherwise}
\end{cases}$$

Defining $g(x) := 1$ we have $\|f\|_T \leq 1$ and $\|g\|_T = 1$. One the other hand
Choosing $\varepsilon$ small enough this contradicts Eq. (8.7). \hfill $\square$

The next lemma is an analogue of Gowers Cauchy-Schwarz, for averages over arbitrary linear forms that satisfy the norm properties.

**Lemma 8.3.7 (Cauchy-Schwarz Type Inequality).** Let $T = (\mathcal{L}, \mathcal{M})$, such that $\|\cdot\|_T$ is a semi-norm. Then for every set of $m + l$ functions $f_1, \ldots, f_l, g_1, \ldots, g_m : \mathbb{F}^n \to \mathbb{C}$, we have

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{l} f_i(L_i(X)) \prod_{i=1}^{m} g_i(M_i(X)) \right] \right| \leq \prod_{i=1}^{l} \|f_i\|_T \prod_{j=1}^{m} \|g_j\|_T.$$ 

**Proof.** Assume for contradiction that there exist functions $f_1, \ldots, f_l, g_1, \ldots, g_m$ such that the inequality does not hold. One can normalize the functions such that

$$\left| \mathbb{E}_{X \in (\mathbb{F}^n)^k} \left[ \prod_{i=1}^{l} f_i(L_i(X)) \prod_{i=1}^{m} g_i(M_i(X)) \right] \right| = c > 1,$$

and for every $i \in [l]$, $t_{\mathcal{L}, \mathcal{M}}(f_i) \leq 1$, and for every $j \in [m]$, $t_{\mathcal{L}, \mathcal{M}}(g_j) \leq 1$. Let $h$ be such that $c^{2h/(m+l)} \geq m + l$. Taking $h$-th tensor power of $f_i \otimes \tilde{f}_i$ and $g_i \otimes g_j$,

$$t_{\mathcal{L}+\mathcal{M}}(\sum_{i=1}^{l} (f_i \otimes \tilde{f}_i)^{\otimes h} + \sum_{i=1}^{m} (g_i \otimes \tilde{g}_i)^{\otimes h}) =$$

$$\mathbb{E} \left[ \prod_{i=1}^{l} \left( \sum_{j=1}^{l} (f_i \otimes \tilde{f}_i)^{\otimes h}(L_i(X)) + \sum_{j=1}^{m} (g_i \otimes \tilde{g}_i)^{\otimes h}(L_i(X)) \right) \right] \cdot$$

$$\prod_{i=1}^{m} \left( \sum_{j=1}^{l} (f_i \otimes \tilde{f}_i)^{\otimes h}(M_i(X)) + \sum_{j=1}^{m} (g_i \otimes \tilde{g}_i)^{\otimes h}(M_i(X)) \right).$$
which by expanding is equal to

\[
\sum_{\sigma, \pi} \left| \mathbb{E} \left[ \prod_{i: \sigma(i) \leq m} (f_{\sigma(i)})^{\otimes h}(L_i(X)) \prod_{i: \sigma(i) > m} (g_{\sigma(i)})^{\otimes h}(L_i(X)) \right] \right|^2,
\]

where the sum is over \( \sigma \in [m+l]^l \) and \( \pi \in [m+l]^m \). Using Claim 8.3.2, and Eq. (8.2) it is easy to see that this value is greater than or equal to \( c^2 h \). On the other hand \( \left| (f_i \otimes \tilde{f}_i)^{\otimes h} \right|_T \leq 1 \), and \( \left| (g_i \otimes \tilde{g}_i)^{\otimes h} \right|_T \leq 1 \). This contradicts the triangle inequality as

\[
\left| \sum_{i=1}^l (f_i \otimes \tilde{f}_i)^{\otimes h} + \sum_{i=1}^m (g_i \otimes \tilde{g}_i)^{\otimes h} \right|_T \geq c^{2h/(m+l)} > m + l
\]

\[
\geq \left\| (f_i \otimes \tilde{f}_i)^{\otimes h} \right\|_T + \left\| (g_i \otimes \tilde{g}_i)^{\otimes h} \right\|_T.
\]

\[\square\]

**Lemma 8.3.8 (Cauchy-Schwarz type inequality for nonnegative functions).** Let

\( T = (\mathcal{L}, \alpha, \beta) \) be such that \( \|\cdot\|_T \) is a norm, and \( \{T_i\}_{i \in [m]} \) with \( T_i = (\mathcal{L}, \alpha^i, \beta^i) \), be such that for each \( i, j \), \( \alpha^i_j + \beta^i_j > 0 \). Moreover let \( f_1, \ldots, f_m : \mathbb{F}^n \to \mathbb{R}_{\geq 0} \), then we have

\[
\mathbb{E}_X \left[ \prod_{i=1}^m r_{\mathcal{L}, \alpha^i, \beta^i}(f_i(X)) \right] \leq \prod_{i=1}^m \|f_i\|_{T, [\alpha^i, \beta^i]} \]

**Proof.** The proof is very similar to the proof of Lemma 8.3.7, with the difference that we use the boost up argument using \( f_i^{\otimes h} \) for large enough \( h \). \( \square \)

### 8.4 A characterization theorem

Now we have the necessary tools to prove the following characterization theorem.
Theorem 8.2.2 (restated). Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm. Then one of the following holds

(i) For every $i \in [l]$, $\{\alpha_i, \beta_i\} = \{0, 1\}$, or

(ii) There exists $s \geq 1$ such that for every $i \in [l]$, $\alpha_i = \beta_i = s/2$.

Proof. Assume $a \in [l]$ and $b \in [l]$ contradict the statement of the Theorem, namely $\{\alpha_a, \beta_a\} \neq \{0, 1\}$ and $\alpha_b \neq \beta_b$. Moreover assume that $\alpha_b > \beta_b$. The case when $\beta_b > \alpha_b$ can be handled in a similar fashion. Let $\varepsilon$ be a sufficiently small constant. The following lemma states that for large enough $n$ there exists $h : \mathbb{F}^n \to \{-1, 1\}$ with $\|h\|_T \leq \varepsilon$ and $\|E h^c\| \leq \varepsilon$, where $c = \alpha_b - \beta_b$.

Lemma 8.4.1. Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm, and $\alpha_i \neq \beta_i$ for some $i$. For every $\varepsilon \geq 0$, there exists $n \geq 0$ and $h : \mathbb{F}^n \to \{(-1)^{1/|\alpha_i - \beta_i|}, 1\}$ such that

- $\|h\|_T \leq \varepsilon$, and

- Letting $g(x) = h(x)^{|\alpha_i - \beta_i|}$, we have $\|E g\| \leq \varepsilon$.

We defer the proof of this Lemma to later in the section. Define $f$ as

$$f(x) := \begin{cases} 1 & h(x) = 1 \\ 0 & \text{Otherwise} \end{cases}$$

Since $\|E h^c\| \leq \varepsilon$ we have that $\|f\|_1 \geq 1/3$. By definition of $f$ and the fact that either $\alpha_a - 1 \neq 0$ or $\beta_a \neq 0$ we have

$$\left| \mathbb{E} h(L_i(X)) \cdot r_{\mathcal{L}, \alpha - e_a, \beta}(f(X)) \right| = t_{\mathcal{L}, \alpha, \beta}(f) \geq 3^{-|\alpha, \beta|}, \quad (8.8)$$

where the inequality follows from the following simple observation.
Claim 8.4.2. Let $T = (\mathcal{L}, \alpha, \beta)$ be such that $\|\cdot\|_T$ is a norm. Then for every $f : \mathbb{F}^n \to \mathbb{R}_{\geq 0}$ we have

$$\|f\|_T \geq \|f\|_1$$

Proof. This is an immediate corollary to Theorem 8.3.3. Letting $g := 1$, we have $\|g\|_T^{-1} = 1$, and

$$\|f\|_1 = \left| \mathbf{E} [f(L_1(X)) \cdot r_{\mathcal{L}_1, \alpha - e_1, \beta}(g)(X)] \right| \leq \|f\|_T.$$

On the other hand since $\alpha_a \neq 0$, by Theorem 8.3.3 we have

$$\left| \mathbf{E}_X h(L_i(X)) \cdot r_{\mathcal{L}_a, \alpha - e_a, \beta}(f)(X) \right| \leq \|h\|_T \leq \varepsilon,$$

which for small enough $\varepsilon$ contradicts Eq. (8.8).

Now we prove that if $\alpha = \beta$ then there is a universal $s$ such that for every $i \in [l]$, $\alpha_i = \beta_i = s/2$. Letting

$$s := \max_i \{\alpha_i + \beta_i\},$$

we will prove that $\|\cdot\|_{T'}$, for $T' = (\mathcal{L}, \frac{\alpha}{s}, \frac{\beta}{s})$, is a norm. This combined with Corollary 8.3.5 proves our claim. To prove that $\|\cdot\|_{T'}$ is a norm we will use the reverse direction of Theorem 8.3.3 for $a$ such that $\alpha_a + \beta_a = s$. Notice that

$$\left| \mathbf{E} \left[ g(L_a(X)) r_{\mathcal{L}_a, \alpha - e_a, \beta}(f)(X) \right] \right| \leq \mathbf{E} \left[ r_{\mathcal{L}_a, s \cdot e_a, \frac{\alpha}{s}}(\|g\|^{1/s}(X)) \cdot r_{\mathcal{L}_a, s \cdot e_a, \frac{\beta}{s}}(\|f\|^{1/s}(X)) \right]$$

$$\leq \|f\|^{1/s} \|\alpha, \beta\|_T^{-s} \|g\|^{1/s} \|\alpha, \beta\|_T \|\alpha, \beta\|_T^{-1} \|g\|_{T'},$$

where the last inequality follows from Lemma 8.3.8. The reason we could exchange from $f$
to $|f|$ and back is that for every $i$, $\alpha_i = \beta_i$.  

Proof of Lemma 8.4.1. Set $a := |\alpha_i - \beta_i|$. Let $A_i$ be the set of vectors $X$ such that $L_i(X)$ is equal to some $L_j(X)$, namely

$$A := \left\{ X \in (F^n)^k : (\exists j \neq i) \ (L_i(X) = L_j(X)) \right\}.$$

As we saw in the proof of Corollary 8.3.5,

$$|A| \leq l \cdot p^n(k-1). \quad (8.9)$$

Let $h : F^n \to \{-1, 1\}$ be a random function which is a result of an independent Bernoulli process. Since $\| \cdot \|_T$ is a norm, then for every $h$ we know $t_{\mathcal{L},\alpha,\beta}(h) \in \mathbb{R}_{\geq 0}$. We will look at the following expected value

$$\mathbb{E}_h t_{\mathcal{L},\alpha,\beta}(h) = \mathbb{E}_h \mathbb{E}_X r_{\mathcal{L},\alpha,\beta}(h)(X)$$

$$= \mathbb{E}_h \mathbb{E}_X r_{\mathcal{L},\alpha,\beta}(h)(X)1_A(X) + \mathbb{E}_h \mathbb{E}_X r_{\mathcal{L},\alpha,\beta}(h)(X)1_{\overline{A}}(X).$$

By Eq. (8.9) we have

$$\mathbb{E}_h \mathbb{E}_X r_{\mathcal{L},\alpha,\beta}(h)(X)1_A(X) \leq \frac{l}{p^n}.$$

Moreover

$$\mathbb{E}_h \mathbb{E}_X r_{\mathcal{L},\alpha,\beta}(h)(X)1_{\overline{A}}(X) = 0,$$

which follows from the independence of $h(L_i(X))$ from the other terms. Thus

$$\mathbb{E}_h t_{\mathcal{L},\alpha,\beta}(h) \leq l \cdot p^{-n}.$$
Moreover by Chernoff bound we have that
\[
\Pr_h \left( \left| \mathbb{E}_x h(x)^n \right| > p^{-\frac{1}{n^{1/3}}} \right) \leq 2e^{-\frac{2}{2n^{1/3}}}.
\]

Therefore for large enough \( n \) there exists \( h : \mathbb{F}^n \to \{-1, 1\} \) with both \( \|h\|_T \leq \varepsilon \) and \( |\mathbb{E}_x h(x)| \leq \varepsilon \).

Here we prove further that all type (I) norms must be invariant under affine transformations.

**Theorem 8.4.3 (Affine Transformations).** Every type (I) norm \( \|\cdot\|_T \) defined by a system of linear forms is invariant under affine transformations.

**Proof.** For \( b \in \mathbb{F}^n \) define
\[
f^{+b}(x) \overset{\text{def}}{=} f(x + b).
\]
We want to show that \( \|f\|_T = \|f^{+b}\|_T \). For every \( t \in \mathbb{F} \) and \( j \in [n] \) define
\[
f^j_t(X) \overset{\text{def}}{=} f(x + te_j),
\]
then for every \( \alpha \in \mathbb{F} \) by Cauchy-Schwarz type inequality we have
\[
\|f^j_t\|_T^n \leq \prod_{i=0}^{p-1} \|f^j_{t+\alpha_i}\|_T^{c_i}, \tag{8.10}
\]
where \( c_i \) is the number of linear forms \( L_a \) such that \( L_{a,1} = i \) plus the number of linear forms \( M_b \) such that \( M_{b,1} = i \). Notice that this inequality holds for every \( t \) and \( \alpha \), thus for every \( f \) one can move the \( f^j_t \) with maximum \( \|f^j_t\|_T \) to the left hand side of Eq. (8.10), which implies invariance of norm under transformations of type \( te_j \). Since this holds for every \( j \), \( \|\cdot\|_T \) is invariant under any affine transformation. \( \square \)
8.5 Topology

Theorem 8.2.2 states that norms defined by linear forms can only be of two types, type (I) and type (II). In this subsection we study the topology that such norms induce on bounded functions on finite dimensional vector spaces. We prove every type (I) norm \( \| \cdot \|_T \) is essentially equivalent to Gowers norm of a proper order, where the order depends only on the system of linear forms defining \( \| \cdot \|_T \). We also show that type (II) norms are equivalent to \( L_1 \) norm and thus \( L_p \) norms. For a formal statement we have to first define what we mean by equivalence of norms over bounded functions.

Definition 8.5.1 (Equivalence of Norms on Bounded Functions). We say that two norms \( \| \cdot \|_A \) and \( \| \cdot \|_B \) are equivalent on the space of bounded functions \( \{ \mathbb{F}^n \to \mathbb{D}, n < \infty \} \), if for every sequence of functions \( \{ f_i : \mathbb{F}^{n_i} \to \mathbb{D} \}_{i,n_i \in \mathbb{N}} \) we have

\[
\lim_{i \to \infty} \| f_i \|_A = 0 \quad \text{if and only if} \quad \lim_{i \to \infty} \| f_i \|_B = 0.
\]

Type (II) norms turn out to be easy to handle. The following lemma relates type (II) norms to the \( L_1 \) norm.

Lemma 8.5.2. Let \( T = (\mathcal{L}, s) \), be such that \( \| \cdot \|_T \) is a type (II) norm. Then for every \( g : \mathbb{F}^n \to \mathbb{D} \) we have

- \( \| g \|_1 \leq \| g \|_T \), and
- \( \| g \|_T \leq \| g \|_1^{(1/ls)} \).

Proof. The first part follows from Theorem 8.3.3 letting \( f(x) := 1 \) for all \( x \in \mathbb{F}^n \), and the second part is a simple consequence of the fact that \( \| g \|_\infty \leq 1 \).

This immediately implies the following corollary.
Corollary 8.5.3. Let $T = (\mathcal{L}, s)$, be such that $\|\cdot\|_T$ is a type (II) norm. Then $\|\cdot\|_T$ is equivalent to $L_1$ norm on bounded complex functions.

Let $\|\cdot\|_T$ be a type (I) norm. We show that there exists some integer $z > 0$ such that for every bounded function $f$, $\|f\|_T$ is small if and only if $\|f\|_{U^{z+1}}$ is small. This is the statement of the next theorem which we will prove in Section 8.6. Unfortunately for technical reasons we have to make an assumption that $p$, the characteristic of the underlying field, is larger than the complexity of the system of linear forms defining $T$.

Theorem 8.5.4. Let $T = (\mathcal{L}, \mathcal{M})$ be such that $\|\cdot\|_T$ is a type (I) norm and $\mathcal{L} \cup \mathcal{M}$ is of Cauchy-Schwarz complexity $s \leq p$. Then there exist an integer $z \leq s$ and functions $\delta_1, \delta_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for every $n$ and $f : \mathbb{F}^n \to \mathbb{D},$

i. If $\|f\|_T \leq \varepsilon$ then $\|f\|_{U^{z+1}} \leq \delta_1(\varepsilon)$, and

ii. If $\|f\|_{U^{z+1}} \leq \varepsilon$ then $\|f\|_T \leq \delta_2(\varepsilon)$.

This in particular implies that $\|\cdot\|_T$ is equivalent to Gowers’ $U^{z+1}$ norm.

Corollary 8.5.5. Let $T = (\mathcal{L}, \mathcal{M})$, be such that $\|\cdot\|_T$ is a type (I) norm. Then there exists $z > 0$ such that $\|\cdot\|_T$ is equivalent to Gowers’ $U^{z+1}$ norm on the space of bounded complex functions.

8.6 Type (I) norms

In the case of type (I) norms, we may write $T = (\mathcal{L}, \mathcal{M})$ where $|L| = l$, and $|M| = m$, defining the norm by

$$t_{\mathcal{L},\mathcal{M}}(f) = \mathbb{E} \left[ \prod_{i=1}^{l} f(L_i(X)) \prod_{i=1}^{m} \bar{f}(M_i(X)) \right],$$

and

164
\[ \| f \|_T^{m+l} = t_{\mathcal{LM}}(f). \]

By Theorem 8.4.3 we know that that \( \mathcal{L} \) and \( \mathcal{M} \) are systems of affine linear forms. Theorem 8.5.4 states a relation between type (I) norms and Gowers’ \( U^{z+1} \) norm of a function for some integer \( z > 0 \). Here we will show how \( z \) is derived for two given systems of linear forms \( \mathcal{L}, \mathcal{M} \) and in Lemma 8.6.2 and Corollary 8.6.3 will show that in the case when the linear forms are affine such \( z(\mathcal{L}, \mathcal{M}) \) exists.

**Definition 8.6.1.** Let \( T = (\mathcal{L}, \mathcal{M}) \). Define \( z(\mathcal{L}, \mathcal{M}) \) as maximum \( d \geq 1 \) for which

\[
L_1 \otimes^d + L_2 \otimes^d + \cdots + L_l \otimes^d - M_1 \otimes^d - M_2 \otimes^d - \cdots - M_m \otimes^d = 0 \mod p,
\]

where \( L_i \otimes^d \) is the \( d \)-th tensor power of \( L_i \). Existence of such \( z \) is the subject of Corollary 8.6.3.

**Lemma 8.6.2.** Let \( L_1, \ldots, L_m \in \mathbb{F}^k \) be a set of distinct linear forms. Assume that \( \alpha \in \mathbb{F}^m \setminus \{0\} \).

There exists \( z \) such that

\[
\sum_i \alpha_i L_i ^\otimes^z \neq 0 \mod p.
\]

**Proof.** Let \( \alpha_L = \alpha_i \) if \( L = L_i \) and \( \alpha_L = 0 \) otherwise, where \( L \) ranges over all linear forms over \( k \) variables, namely \( \mathbb{F}^k \). So \( \alpha \in \mathbb{F}^p^k \) and if no such \( z \) exists then \( \sum_{L \in \mathbb{F}^k} \alpha_L \prod_{j=1}^k L_j^{z_j} = 0 \) for all \( w \in \mathbb{F}^k \).

Let \( M \) be the \( p^k \times p^k \) matrix over \( F \) defined as

\[
M_{L,w} = \prod_{j=1}^k L_j^{z_j}.
\]

(note that we also allow \( w = (0, 0, \ldots, 0) \); and \( 0^0 = 1, 0^a = 0 \) for \( 1 \leq a \leq p - 1 \). Then if no such \( z \) exists then \( M \alpha = 0 \). However, we can factor \( M \) as the \( k \)-th tensor power of the \( p \times p \) matrix \( M' \) given by \( (M')_{a,b} = a^b \mod p \). Then \( M' \) is a Vandermonde matrix and hence \( det(M') \neq 0 \). So also \( det(M) = det(M')^k \neq 0 \), hence we must have \( \alpha = 0 \). \( \square \)
Corollary 8.6.3. Let $L_1, ..., L_l, M_1, ..., M_m \in \mathbb{F}^k$ be a set of distinct affine linear forms. There exists a maximal $z > 0$ such that for every $d \leq z$,

$$\sum_i L_i^\otimes d - \sum_j M_j^\otimes d \equiv 0 \mod p.$$ 

Proof. Since

$$\sum_i L_i^\otimes z - \sum_j M_j^\otimes z \equiv \sum_i L_i^\otimes z + \sum_j (p - 1) \cdot M_j^\otimes z \mod p,$$

by Lemma 8.6.2 there exists a finite maximal $z$ such that

$$\sum_i L_i^\otimes z - \sum_j M_j^\otimes z \equiv 0 \mod p.$$ 

This completes the proof because for an affine linear form $L$, $L^\otimes d$ is a sub-vector of $L^\otimes z$ for $d \leq z$.

Claim 8.6.4. If $\|\cdot\|_T$ is a norm then $z(\mathcal{L}, \mathcal{M}) > 0$, namely

$$\sum_{i=1}^l L_i - \sum_{i=1}^m M_i \equiv 0 \mod p.$$ 

Proof. Assume for contradiction that $z(\mathcal{L}, \mathcal{M}) = 0$. We look at the linear Fourier characters $\chi_q$, for $q \in \mathbb{F}^n$, notice that since $\sum_{i=1}^l L_i - \sum_{i=1}^m M_i \not\equiv 0 \mod p$, thus $\|\chi_q\|_T = 0$, unless $q = 0$. This is in contradiction with $\|\cdot\|_T$ being a norm.

Theorem 8.5.4 (restated). Let $T = (\mathcal{L}, \mathcal{M})$ be such that $\|\cdot\|_T$ is a type (I) norm and $\mathcal{L} \cup \mathcal{M}$ is of Cauchy-Schwarz complexity $s \leq p$. Then there exist an integer $z = z(\mathcal{L}, \mathcal{M}) \leq s$ and functions $\delta_1, \delta_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for every $n$ and $f : \mathbb{F}^n \to \mathbb{D}$,

1. If $\|f\|_T \leq \varepsilon$ then $\|f\|_{U^z+1} \leq \delta_1(\varepsilon)$, and
ii. If $\|f\|_{U^{z+1}} \leq \varepsilon$ then $\|f\|_T \leq \delta_2(\varepsilon)$.

**Remark 8.6.5.** Notice that $z(L, M) \leq s \leq p$, where $s$ is the Cauchy-Schwarz complexity of the system of linear forms. This is because for any $s < d$ and a high enough rank polynomial $P$ of degree $\leq d+1$, we would have $\|e_{\mathbb{F}}(P)\|_{U^{s+1}} \leq \varepsilon$. But at the same time $\|e_{\mathbb{F}}(P)\|_{L,M} = 1$, which would contradict Lemma 2.4.3.

We will prove Theorem 8.5.4 in two parts through the next two subsections, namely Lemma 8.6.6 and Theorem 8.6.8.

### 8.6.1 Theorem 8.5.4, the Direct Part

**Lemma 8.6.6** (Direct Theorem). Let $T = (L, M)$ be such that $\|\cdot\|_T$ is a norm. For every $\varepsilon$ there is a $\delta_1(\varepsilon)$ such that for every $n$, for every function $f : \mathbb{F}^n \to \mathbb{D}$, with $\|f\|_T \leq \delta_1(\varepsilon)$ we have $\|f\|_{U^{z(L,M)+1}} \leq \varepsilon$.

**Proof.** Assume that $\|f\|_T \leq \varepsilon$, then for every polynomial $q$ of degree $d \leq z$

$$|\langle f, q \rangle| \leq \|f \cdot q\|_T = \|f\|_T \leq \varepsilon,$$

where the first inequality follows by Corollary 8.3.4 and the equality is the subject of the following claim.

**Claim 8.6.7.** Let $T = (L, M)$. Let $f : \mathbb{F}^n \to \mathbb{D}$, and $q$ be a non-classical polynomial of degree $d$ with $d \leq z(L, M)$. We have

$$t_{L,M}(f \cdot e_{\mathbb{F}}(q)) = t_{L,M}(f).$$

**Proof.** Since $d \leq z(L, M)$
\[ r_{\mathcal{L},\mathcal{M}}(f \cdot e_F(q)) = r_{\mathcal{L},\mathcal{M}}(f) \prod_{i=1}^{l} e_F(q(L_i(X))) \prod_{i=1}^{m} e_F(-q(M_i(X))) = r_{\mathcal{L},\mathcal{M}}(f), \]

where the last equality is due to the fact that all the coefficients of monomials of \( \sum_{i=1}^{l} q(L_i(X)) - \sum_{i=1}^{m} q(M_i(X)) \) are zero since \( d \leq z(\mathcal{L},\mathcal{M}) \).

We conclude by the inverse theorem for Gowers Norms (Theorem 2.1.10).

### 8.6.2 Theorem 8.5.4, the Inverse Part

Here we will prove the following inverse theorem.

**Theorem 8.6.8 (Inverse Theorem).** Let \( T = (\mathcal{L},\mathcal{M}) \) be such that \( \| \cdot \|_T \) is a norm. Assume that \( \mathcal{L} \cup \mathcal{M} \) has Cauchy-Schwarz complexity \( d \leq p \). For every \( \varepsilon \) there exists \( \delta_2(\varepsilon) \) such that for every function \( f : \mathbb{F}^n \to \mathbb{D} \), with \( \| f \|_{U^{z(\mathcal{L},\mathcal{M})} + 1} \leq \delta_2(\varepsilon) \) we have \( \| f \|_T \leq \varepsilon \).

**Proof.** First of all by Theorem 8.4.3 we may assume that \( \mathcal{L} \) and \( \mathcal{M} \) are systems of affine linear forms, namely we can assume that for every \( i \in [l] \) and \( j \in [m] \), \( L_i,1 = 1 \) and \( M_j,1 = 1 \). Following is a list of the constants we will be working with.

1. Let \( z := z(\mathcal{L},\mathcal{M}) \).
2. Let \( \delta' = (\varepsilon/3)^{m+l} \).
3. Let \( r(C) := \tau_{p,d} \left( (\varepsilon/3)^{m+l} p^{-C} \right) \), where \( \tau_{p,d}(\cdot) \) is the function in Lemma 2.2.20.

Let \( f : \mathbb{F}^n \to \mathbb{D} \) be a function with \( \| f \|_{U^{z+1}} \leq \delta_2 = (\varepsilon/3)^{m+l} p^{-C^*} \), where \( C^* = C_{\max}(p,d,\delta',r(\cdot)) \) is the constant from Theorem 2.3.1. By Theorem 2.3.1 and Remark 2.3.2, we can write

\[ f = f_1 + f_2, \]
such that

\[ f_1 = \mathbf{E}(f|\mathcal{B}), \quad \text{and} \quad \|f_2\|_{U^{d+1}} \leq \delta', \]

where $\mathcal{B}$ is a polynomial factor of degree at most $d$, complexity $C \leq C_{\max}(p, d, \delta', r(\cdot))$, and rank at least $r(C)$. Moreover $\mathcal{B}$ is defined by homogeneous polynomials.

By Lemma 2.4.3,

\[ \|f_2\|_T \leq \|f_2\|_{U^{z+1}}^{1/m+1} \leq \varepsilon/3. \] (8.11)

Now we will bound $\|f_1\|_T$. Since $f_1 = \Gamma(P_1, \ldots, P_C)$, for a set of homogeneous polynomials $P_i \in \text{Poly}_{\leq s}$, we can write the degree $s$ Fourier decomposition of $f$,

\[ f_1(x) = \sum_{\gamma \in \mathbb{F}^C} \hat{\Gamma}(\gamma) e_{\mathbb{F}} \left( \sum_{i=1}^C \gamma(i) P_i(x) \right). \]

Therefore

\[ \|f_1\|_T = \left\| \sum_{\gamma \in \mathbb{F}^C} \hat{\Gamma}(\gamma) e_{\mathbb{F}} \left( \sum_{i=1}^C \gamma(i) P_i(x) \right) \right\|_T \leq \sum_{\gamma \in \mathbb{F}^C} |\hat{\Gamma}(\gamma)| \cdot \left\| e_{\mathbb{F}} \left( \sum_{i=1}^C \gamma(i) P_i(X) \right) \right\|_T, \] (8.12)

where the last inequality is the triangle inequality and positive homogeneity for the $\|\cdot\|_T$ norm. Looking at the terms in the RHS of Eq. (8.12), for each $\gamma \in \mathbb{F}^C$ one of the following holds:

(i) $(\sum_{i=1}^C \gamma(i) P_i(X))$ is of degree $\leq z$.

(ii) $(\sum_{i=1}^C \gamma(i) P_i(X))$ has multinomial of degree higher than $z$.

**Case (i):** By Remark 2.2.6 and the fact that $\|f\|_{U^{z+1}} \leq (\varepsilon/3)^{m+l} p^{-C} \leq (\varepsilon/3)^{m+l} p^{-C}$, using the direct theorem for Gowers norms we have that the terms satisfying (i) contribute at most $(\varepsilon/3)^{m+l}$ to the RHS of Eq. (8.12).
Case (ii): Since the degree of such Polynomials is higher than $z$, by Lemma 2.2.20 we have

$$
\sum_{j=1}^{l} \sum_{i=1}^{C} \gamma(i)P_i(L_j(X)) - \sum_{j=1}^{m} \sum_{i=1}^{C} \gamma(i)P_i(M_j(X)) \not\equiv 0,
$$

and that the terms satisfying this condition contribute at most $(\varepsilon/3)^{m+l}$ to the RHS of Eq. (8.12).

Finally adding everything up and using the triangle inequality we conclude

$$
\|f\|_T \leq \|f_1\|_T + \|f_2\|_T \leq 2(\varepsilon/3)^{m+l} + \varepsilon/3 \leq \varepsilon,
$$

since $m + l \geq 1$. \qed
CHAPTER 9

CONCLUSION

**Algebraic property testing.** In Chapter 6 we proved the following characterization theorem for affine invariant properties defined by induced affine constraints.

**Theorem 6.4.5 (restated).** For any integer \( d > 0 \) and (possibly infinite) fixed collection \( \mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\} \) of induced affine constraints, each of complexity \( \leq d \), there are functions \( q_\mathcal{A} : (0, 1) \rightarrow \mathbb{Z}^+ \), \( \delta_\mathcal{A} : (0, 1) \rightarrow (0, 1) \) and a tester \( T \) which, for every \( \varepsilon > 0 \), makes \( q_\mathcal{A}(\varepsilon) \) queries, accepts \( \mathcal{A} \)-free functions and rejects functions \( \varepsilon \)-far from \( \mathcal{A} \)-free with probability at least \( \delta_\mathcal{A}(\varepsilon) \). Moreover, \( q_\mathcal{A} \) is a constant if \( \mathcal{A} \) is of finite size.

This implies that every locally characterized affine-invariant property is proximity-obliviously testable (Theorem 6.1.1), however in the case of testability, our main theorem implies that every subspace hereditary property of bounded complexity is testable (Theorem 6.3.3). This leaves the following conjecture of [17] open for the case of unbounded complexity.

**Conjecture 6.3.2 [17] (restated).** Every subspace hereditary property is testable.

Affine invariant properties of unbounded complexity are those which can only be defined by an infinite collection of induced affine constraints with unbounded Cauchy-Schwarz complexity. We are not aware of any natural property of unbounded complexity, and moreover suspect that the answer to the above conjecture should be in the negative.

Another interesting question is whether Theorem 6.4.5 can be extended to linear-invariant properties. The proof of Theorem 6.4.5 makes use of near-equidistribution of regular polynomial factors over affine collections of linear forms (Theorem 2.2.21). Although this near-equidistribution property was not known for general collections of linear forms at the time in [15], we now have such a general equidistribution (Theorem 4.1.1, [50]). However, it is still unclear how to adapt the rest of our proof to the linear-invariant case.
Reed-Muller codes beyond list decoding radius. Our decoder in Section 7.5.3 only works for a $P$ which is a polynomial of degree $d < p$. It is a fascinating problem to find out if this can be extended to general functions. A first step to this program would be to get a deeper understanding of cubic polynomials. The known testers and decoders in this case follow the same steps as the proofs of the inverse theorems for the Gowers norms [68, 79, 18]. However, in some cases, even a combinatorial proof of such inverse theorems is not yet available. It has been seen, in the context of property testing, that one can bypass such obstacles by considering algorithms that require the inverse theorems and decomposition theorems only in their analysis [52, 15]. It is therefore plausible to hope for such algorithms for the decoding question as well.

Norms defined by linear forms. In Chapter 8, we gave necessary conditions for a collection of linear forms defining a norm (Theorem 8.2.2). A great and possibly hard open problem is to give a full characterization of collections of linear forms that define a norm, and we think of our result as a very small step towards this. Theorem 8.2.2 shows that a system of linear forms that defines a norm must be one of only two types, one of which (type (II)) is easily seen to be essentially equivalent to $L_p$ norms. Next, we prove under the assumption of $|F|$ being sufficiently large, that type (I) norms are essentially the same as some Gowers norm (Theorem 8.5.4). We believe that the assumption on the size of the field is unnecessary, meaning that every type (I) norm is equivalent to a Gowers norm, and we leave this as an open problem.
REFERENCES


