Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 87 (2014)

List decoding Reed-Muller codes over small fields

Abhishek Bhowmick*
Department of Computer Science
The University of Texas at Austin
bhowmick@cs.utexas.edu

Shachar Lovett †
Department of Computer Science and Engineering
University of California, San Diego
slovett@ucsd.edu

July 17, 2014

Abstract

The list decoding problem for a code asks for the maximal radius up to which any ball of that radius contains only a constant number of codewords. The list decoding radius is not well understood even for well studied codes, like Reed-Solomon or Reed-Muller codes.

Fix a finite field $\mathbb{F}$. The Reed-Muller code $\mathrm{RM}_{\mathbb{F}}(n, d)$ is defined by n-variate degree-d polynomials over $\mathbb{F}$. In this work, we study the list decoding radius of Reed-Muller codes over a constant prime field $\mathbb{F} = \mathbb{F}_p$, constant degree d and large n. We show that the list decoding radius is equal to the minimal distance of the code.

That is, if we denote by $\delta(d)$ the normalized minimal distance of $\mathrm{RM}_{\mathbb{F}}(n, d)$, then the number of codewords in any ball of radius $\delta(d) - \varepsilon$ is bounded by $c = c(p, d, \varepsilon)$ independent of n. This resolves a conjecture of Gopalan-Klivans-Zuckerman [STOC 2008], who among other results proved it in the special case of $\mathbb{F} = \mathbb{F}_2$; and extends the work of Gopalan [FOCS 2010] who proved the conjecture in the case of $d = 2$.

We also analyse the number of codewords in balls of radius exceeding the minimal distance of the code. For $e \leq d$, we show that the number of codewords of $\mathrm{RM}_{\mathbb{F}}(n, d)$ in a ball of radius $\delta(e) - \varepsilon$ is bounded by $\exp(c \cdot n^{d-e})$, where $c = c(p, d, \varepsilon)$ is independent of n. The dependence on n is tight. This extends the work of Kaufman-Lovett-Porat [IEEE Inf. Theory 2012] who proved similar bounds over $\mathbb{F}_2$.

The proof relies on several new ingredients: an extension of the Frieze-Kannan weak regularity to general function spaces, higher-order Fourier analysis, and an extension of the Schwartz-Zippel lemma to compositions of polynomials.

*Research supported in part by NSF Grant CCF-1218723.
†Supported by NSF CAREER award 1350481

ISSN 1433-8092

1 Introduction

The concept of *list decoding* was introduced by Elias [Eli57] and Wozencraft [Woz58] to decode *error correcting codes* beyond half the minimum distance. The objective of list decoding is to output all the codewords within a specified radius around the received word. After the seminal results of Goldreich and Levin [GL89] and Sudan [Sud97] which gave list decoding algorithms for the Hadamard code and the Reed-Solomon code respectively, there has been tremendous progress in designing list decodable codes. See the excellent surveys of Guruswami [Gur06, Gur04] and Sudan [Sud00].

List decoding has applications in many areas of computer science including hardness amplification in complexity theory [STV01, Tre03], derandomization [Vad12], construction of hard core predicates from one way functions [GL89, AGS03], construction of extractors and pseudorandom generators [TSZS01, SU05] and computational learning [KM93, Jac97]. Despite so much progress, the largest radius up to which list decoding is tractable is still a fundamental open problem even for well studied codes like Reed-Solomon (univariate polynomials) and Reed-Muller codes (multivariate polynomials). The goal of this work is to analyse Reed-Muller codes over small fields and small degree.

Reed-Muller codes (RM codes) were discovered by Muller in 1954. Fix a finite field $\mathbb{F} = \mathbb{F}_q$. Let $d \in \mathbb{N}$. The RM code $\mathrm{RM}_{\mathbb{F}}(n, d)$ is defined as follows. The message space consists of degree $\leq d$ polynomials in n variables over $\mathbb{F}$ and the codewords are evaluation of these polynomials on $\mathbb{F}^n$. Let $\delta_p(d)$ denote the normalized distance of $\mathrm{RM}_{\mathbb{F}}(n, d)$. Let $d = a(q-1) + b$ where $0 \leq b < q - 1$. We have

$$\delta_{\mathbb{F}}(d) = \frac{1}{q^a}\left(1 - \frac{b}{q}\right).$$

RM codes are one of the most well studied error correcting codes. Many of the applications in computer science involves low degree polynomials over small fields, namely RM codes. Given a received word $g : \mathbb{F}^n \to \mathbb{F}$ the objective is to output the list of codewords (e.g. low-degree polynomials) that lie within some distance of g. Typically we will be interested in regimes where list size is either independent of n or polynomial in the block length $\mathbb{F}^n$.

1.1 Previous Work

Let $\mathcal{P}_d(\mathbb{F}^n)$ denote the class of degree $\leq d$ polynomials $f : \mathbb{F}^n \to \mathbb{F}$. Let dist denote the normalized Hamming distance. For $\mathrm{RM}_{\mathbb{F}}(n, d)$, $\eta > 0$, let

$$\ell_{\mathbb{F}}(n, d, \eta) := \max_{g:\mathbb{F}^n \to \mathbb{F}} \left|\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leq \eta\}\right|.$$

Let $\mathrm{LDR}_{\mathbb{F}}(n, d)$ (short for *list decoding radius*) be the maximum η for which $\ell_{\mathbb{F}}(n, d, \eta - \varepsilon)$ is upper bounded by a constant depending only on $\varepsilon, |\mathbb{F}|, d$ for all $\varepsilon > 0$.

It is easy to see that $\mathrm{LDR}_{\mathbb{F}}(n, d) \leq \delta_{\mathbb{F}}(d)$. The difficulty lies in proving a matching lower bound. The first breakthrough result was in the setting of $d = 1$ over $\mathbb{F}_2$ (Hadamard Codes) where Goldreich and Levin showed that $\mathrm{LDR}_{\mathbb{F}_2}(n, 1) = \delta_{\mathbb{F}_2}(1) = 1/2$ [GL89]. Later, Goldreich, Rubinfield and Sudan [GRS00] generalized the field to obtain $\mathrm{LDR}_{\mathbb{F}}(n, 1) = \delta_{\mathbb{F}}(1) = 1 - 1/|\mathbb{F}|$. In the setting of $d < |\mathbb{F}|$, Sudan, Trevisan and Vadhan [STV01] showed that $\mathrm{LDR}_{\mathbb{F}}(n, d) \geq 1 - \sqrt{2d/|\mathbb{F}|}$ improving previous work by Arora and Sudan [AS03], Goldreich *et al* [GRS00] and Pellikaan and Wu [PW04]. A crucial result that was a bulding block in the multivariate setting was the problem of list decoding Reed-Solomon codes which was analysed by Sudan [Sud97] and Guruswami and Sudan [GS99]. The list decoding radius obtained above essentially attains the Johnson radius, which is a radius such that for any code over $\mathbb{F}$ with normalized minimum distance δ, the list decoding radius (LDR) is at least

$$\mathrm{J}_{\mathbb{F}}(\delta) := \left(1 - \frac{1}{|\mathbb{F}|}\right)\left(1 - \sqrt{1 - \frac{|\mathbb{F}|\delta}{|\mathbb{F}| - 1}}\right).$$

There have been few results that show list decodability beyond the Johnson radius [DGKS08, GKZ08].

In 2008, Gopalan, Klivans and Zuckerman [GKZ08] showed that $\mathrm{LDR}_{\mathbb{F}_2}(n,d) = \delta_{\mathbb{F}_2}(d)$. This beats the Johnson radius already for $d \geq 2$. The list decoding algorithm in [GKZ08] is a generalization of the Goldreich-Levin algorithm [GL89]. However their algorithm crucially depends on the fact that the ratio of minimum distance to unique decoding radius is equal to 2 which is the size of the field. Therefore, it does not generalize to higher fields (except for some special cases). They pose the following conjecture.

Conjecture 1 ([GKZ08])**.** *For all constants d and all fields $\mathbb{F}$, $\mathrm{LDR}_{\mathbb{F}}(n,d) = \delta_{\mathbb{F}}(d)$.*

An important contribution of [GKZ08] is an algorithm for list decoding that outputs the list of codewords up to radius η efficiently assuming $\ell_{\mathbb{F}}(n,d,\eta)$ is bounded.

It was also shown [GKZ08] that $\mathrm{LDR}_{\mathbb{F}}(n,d) \geq \frac{1}{2}\delta_{\mathbb{F}}(d-1)$ and this beats the Johnson radius already when d is large. It is believed [GKZ08, Gop10] that the hardest case is the setting of small d. An important step in this direction was taken in [Gop10] that considered quadratic polynomials and showed that $\mathrm{LDR}_{\mathbb{F}}(n,2) = \delta_{\mathbb{F}}(2)$ for all fields $\mathbb{F}$ and thus proved the conjecture for $d = 2$. In the setting of $\mathbb{F}_2$, Kaufman, Lovett and Porat [KLP10] showed tight list sizes for radii beyond the minimum distance.

1.2 Our Results

As mentioned before, the algorithmic problem of list decoding was reduced to the combinatorial problem in [GKZ08]. Our main theorem is a resolution of Conjecture 1 for prime fields. We note that prior to this, the conjecture was open even in the $d < |\mathbb{F}|$ case.

Theorem 1. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta_{\mathbb{F}}(d) - \varepsilon) \leq c_{p,d,\varepsilon}.$$

Remark 1.1 (Algorithmic Implications)**.** *As mentioned above, using the reduction of algorithmic list decoding to combinatorial list decoding in [GKZ08] along with Theorem 1, for fixed prime fields, d and $\varepsilon > 0$, we now have list decoding algorithms in both the global setting (running time polynomial in $|\mathbb{F}|^n$) and the local setting (running time polynomial in n^d).*

Next, we study list sizes for radii which are larger than the minimal radius of the code. We give bounds which capture the correct exponent of n for all radii. This extends the results of Kaufman, Lovett and Porat [KLP10] who studied Reed-Muller codes over $\mathbb{F}_2$, to all prime fields.

Theorem 2. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $e \le d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta_{\mathbb{F}}(e) - \varepsilon) \le \exp\left(c_{p,d,\varepsilon} n^{d-e}\right)$$

Remark 1.2. *The exponent of n in Theorem 2 is tight, as the following example shows. Let $e = a(p-1) + b$ with $0 \le b < p - 1$. Consider polynomials of the form*

$$P(x) = \left(\prod_{i=i}^{a}(x_i^{p-1} - 1)\right)\left(\prod_{j=1}^{b}(x_{a+1} - j)\right)(x_{a+2} + Q(x_{a+3}, \dots, x_n))$$

for all polynomials Q of degree $d - e$. Observe that $\mathbf{Pr}[P(x) \ne 0] = \frac{1}{p^a}\left(1 - \frac{b}{p}\right)\left(1 - \frac{1}{p}\right) = \delta(e)(1 - 1/p)$. The number of such polynomials is $\exp(c' n^{d-e})$ for some $c' = c'_{p,d,e}$.

1.3 Proof overview

Previous results have mostly relied on the idea of local correction of the RM code. The work of [Gop10] uses (linear) Fourier analysis which does not seem to go beyond quadratic polynomials. We use tools from higher order Fourier analysis to resolve the conjecture. We think of $\mathbb{F} = \mathbb{F}_p, d, \varepsilon$ as constants. For a received word $g : \mathbb{F}^n \to \mathbb{F}$ our goal is to upper bound $|\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \le \eta\}|$. For simplicity of exposition, we assume in the proof overview that $d < |\mathbb{F}|$. The general case is somewhat more technical, as it requires the introduction of nonclassical polynomials.

A weak regularity (A low complexity proxy for the received word). The first step is an extension of the Frieze-Kannan weak regularity [FK99] which would allow us to move from an arbitrary received word g to a "low complexity" received word. We note that a somewhat similar idea appeared also in [TTV09].

Let X, Y be finite sets and let $P(Y) := \{f : Y \to \mathbb{R}_{\ge 0} : \sum_{y \in Y} f(y) = 1\}$ be the probability simplex over Y. We view functions $f : X \to P(Y)$ as randomized functions from X to Y. For $f, g : X \to P(Y)$ we define

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x\langle f(x), g(x)\rangle.$$

Given $\varepsilon > 0$, any function $g : X \to P(Y)$ and a collection F of functions $f : X \to P(Y)$, one can find a collection of $c := 1/\varepsilon^2$ functions $h_1, \dots, h_c \in F$ and a *proxy* $g_1 : X \to P(Y)$ for g, such that g_1 is determined by $h_1(x), \dots, h_c(x)$ and such that g_1 is indistinguishable from g with respect to F.

Lemma 3.1. *Let $g : X \to P(Y)$, $\varepsilon > 0$, and F be a collection of functions $f : X \to P(Y)$. Then there exist $c \le 1/\varepsilon^2$ functions $h_1, h_2, \dots, h_c \in F$ and a function $\Gamma : P(Y)^c \to P(Y)$ such that for all $f \in F$,*

$$|\mathbf{Pr}[g(x) = f(x)] - \mathbf{Pr}[\Gamma(h_1(x), h_2(x), \dots, h_c(x)) = f(x)]| \le \varepsilon.$$

In our case, $X = \mathbb{F}^n$, $Y = \mathbb{F}$ and $F = \mathcal{P}_d(\mathbb{F}^n)$. When F is a family of "deterministic" functions $f : X \to Y$, as it is in our case, we can obtain one-sided approximation using only deterministic functions $h_1, \dots, h_c$.

Corollary 3.3. *Let $g : X \to Y$, $\varepsilon > 0$, and F be a collection of functions $f : X \to Y$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in F$ such that for every $f \in F$, there is a function $\Gamma_f : Y^c \to Y$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

Strong regularity applied to $\mathcal{H}$. The collection of polynomials $\mathcal{H} = \{h_1, \ldots, h_c\} \subset \mathcal{P}_d(\mathbb{F}^n)$ defines a partition of the input space $\mathbb{F}^n$ into *atoms* $\{x \in \mathbb{F}^n : h_1(x) = a_1, \ldots, h_c(x) = a_c\}$. We next regularize $\mathcal{H}$. The objective of regularization is to further refine the partition into smaller atoms with the goal that the polynomials $h_1, \ldots, h_c$ are "pseudo-random". Formally, we require the polynomials to be inapproximable by lower degree polynomials, which is equivalent to having negligible Gowers uniformity norm. This ensures, for example, that for uniformly random X in $\mathbb{F}^n$, the distribution $(h_1(X), \ldots, h_c(X))$ is close to uniform over the atoms. This process of regularization was introduced by [GT09] and is now standard in higher-order Fourier analysis. Let $\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\} \subset \mathcal{P}_d(\mathbb{F}^n)$ be the regularized $\mathcal{H}$ that satisfies the above properties, where $c' = c'(p, d, c)$.

Structure of polynomials close to low complexity received words. Fix now an $f \in \mathcal{P}_d(\mathbb{F}^n)$ such that $\mathrm{dist}(f, g) \leq \delta_p(d) - \varepsilon$. We will show that f must be determined by $\mathcal{H}'$. That is,

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x))$$

for some $F : \mathbb{F}^{c'} \to \mathbb{F}$. This will bound the number of such functions by $p^{p^{c'}}$, which is independent of n.

In order to achieve that, we regularize the family of polynomials $\mathcal{H}' \cup \{f\}$. By choosing regularity parameters appropriately, we can assure that only f decomposes further,

$$f = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x))$$

where $\mathcal{H}'' = \{h_1, \ldots, h'_{c'}, h''_1, \ldots, h''_{c''}\}$ is regular. Moreover, for $G_f(h'_1(x), \ldots, h'_{c'}(x)) = \Gamma_f(h_1(x), \ldots, h_c(x))$, we know that

$$\mathbf{Pr}[f(x) = G_f(h'_1(x), \ldots, h'_{c'}(x))] \geq 1 - \delta_p(d) + \varepsilon/2.$$

The regularity of $\mathcal{H}''$ allows us to reduce the question to that of the structure of F vs G_f. We then show, by a variant of the Schwartz-Zippel lemma, that such an approximation can only exist when F does not depend on $h''_1, \ldots, h''_{c''}$. The bound for larger radii $\delta_{\mathbb{F}}(e) - \varepsilon$ with $e < d$ follows along similar lines. We show that in the decomposition above, since $\mathbf{Pr}[F = G_f] > 1 - \delta_{\mathbb{F}}(e) + \varepsilon/2$, this can only occur when $h''_1, \ldots, h''_{c''}$ have degree at most $d - e$. As the number of such polynomials is exponential in n^{d-e}, we derive similar bounds for the number of functions f.

2 Preliminaries

2.1 Notation

Let $\mathbb{N}$ denote the set of positive integers. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \ldots, n\}$. We use $y = x \pm \varepsilon$ to denote $y \in [x - \varepsilon, x + \varepsilon]$. Let $\mathbb{T}$ denote the torus $\mathbb{R}/\mathbb{Z}$. This is an abelian group under addition. For $n \in \mathbb{N}$, and $x, y \in \mathbb{C}^n$, let $\langle x, y \rangle := \sum_{i=1}^{n} x_i \overline{y_i}$ where $\overline{a}$ is the conjugate of a. Let $||x||_2 := \sqrt{\langle x, x \rangle}$.

Fix a prime field $\mathbb{F} = \mathbb{F}_p$. Let $|.|$ denote the natural map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\} \in \mathbb{Z}$. Let $e : \mathbb{T} \to \mathbb{C}$ be the map $e(x) := e^{2\pi i x}$. Let $e_p : \mathbb{F} \to \mathbb{C}$ be the map $e_p(x) = e(\frac{|x|}{p})$. For an integer $k \geq 0$, let $\mathbb{U}_k := \frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$. Note that $\mathbb{U}_k$ is a subgroup of $\mathbb{T}$. Let $\iota : \mathbb{F} \to \mathbb{U}_1$ be the bijection $\iota(a) = \frac{|a|}{p}$ (mod 1).

For a finite set X and $n \in \mathbb{N}$, with $f : X \to \mathbb{C}^n$, we write $\mathbb{E}_x f(x)$ to denote $\frac{1}{|X|} \sum_{x \in X} f(x)$. We define $||f||_2 := \sqrt{\mathbb{E}_x ||f(x)||_2^2}$. If $g : X \to \mathbb{C}^n$, we have $\langle f, g \rangle := \mathbb{E}_x \langle f(x), g(x) \rangle$. Let Y be a finite set. Let $P(Y) := \{f : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} f(y) = 1\}$ denote the probability simplex on Y. We shall write randomized functions by mapping them to the simplex. Thus, for $f, g : X \to P(Y)$ we define

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x \langle f(x), g(x) \rangle.$$

If $f : X \to Y$ is a deterministic function, then we embed Y into $P(Y)$ in the obvious way, and consider $f : X \to P(Y)$ with $f(x)_y = 1$ if $f(x) = y$ when viewed as a function to Y, and $f(x)_{y'} = 0$ for all $y' \in Y \setminus \{y\}$.

2.2 Polynomials

Definition 2.1 (Derivative). *Given a function $f : \mathbb{F}^n \to \mathbb{T}$ and $a \in \mathbb{F}^n$, define the derivative of f in direction a as $D_a f : \mathbb{F}^n \to \mathbb{T}$ as $D_a f(x) = f(x + a) - f(x)$ for $x \in \mathbb{F}^n$.*

Definition 2.2 (Nonclassical Polynomial or Polynomial). *Let $d \in \mathbb{N}$. Then $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if for all $a_1, \ldots, a_{d+1}, x \in \mathbb{F}^n$,*

$$\left(D_{a_1} \ldots D_{a_{d+1}} f\right)(x) = 0. \tag{1}$$

The degree of f denoted by $\deg(f)$ is the smallest such $d \in \mathbb{N}$ for which the above holds. If the image of f lies in $\mathbb{U}_1$ then f is called a classical polynomial of degree d. When $d < |\mathbb{F}|$, it is known that all the polynomials of degree d satisfying (1) are classical polynomials. However, when $d \geq |\mathbb{F}|$, there exist nonclassical polynomials. We write $\mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$ to denote the class of degree $\leq d$ polynomials. Unless explicitly specified, a polynomial is a (potentially) nonclassical polynomial. The following lemma from [TZ11] characterizes polynomials.

Lemma 2.3 ([TZ11], Lemma 1.7). *Let $d \in \mathbb{N}$.*

- *A function $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if and only if $D_a f$ is a polynomial of degree $\leq d - 1$ for all $a \in \mathbb{F}^n$.*
- *A function $f : \mathbb{F}^n \to \mathbb{T}$ is a classical polynomial with $deg(f) \leq d$ if $f = \iota \circ P$ where $P : \mathbb{F}^n \to \mathbb{F}$ is of the form*

$$P(x_1, \ldots, x_n) = \sum_{0 \leq d_1, \ldots, d_n \leq p-1 : \sum_i d_i \leq d} c_{d_1, \ldots, d_n} \prod_{i=1}^{n} x_i^{d_i},$$

where $c_{d_1,\ldots,d_n} \in \mathbb{F}$ are unique.

- *A function $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial with deg$(f) \leq d$ if f is of the form*

$$f(x_1, \ldots, x_n) = \alpha + \sum_{0\leq d_1,\ldots,d_n \leq p-1, k\geq 0: \sum_i d_i \leq d-k(p-1)} \frac{c_{d_1,\ldots,d_n,k} \prod_{i=1}^n |x_i|^{d_i}}{p^{k+1}} \pmod 1,$$

  where $c_{d_1,\ldots,d_n,k} \in \{0, \ldots, p-1\}$ and $\alpha \in \mathbb{T}$ are unique. α is called the shift of f and the largest k such that some $c_{d_1,\ldots,d_n,k} \neq 0$ is the depth of f, denoted by $\mathrm{depth}(f)$. *Note that classical polynomials have 0 shift and 0 depth.*

- *If $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial with* $\mathrm{depth}(f) = k$, *then its image lies in a coset of $\mathbb{U}_{k+1}$.*

- *If $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial such that* $\deg(f) = d$ *and* $\mathrm{depth}(f) = k$, *then* $\deg(pf) = \max(d-p+1, 0)$ *and* $\mathrm{depth}(pf) = k-1$. *Also, if $c \in \{1, \ldots, p-1\}$ then the degree and depth of cf remain unchanged.*

Throughout the article, we assume without loss of generality that nonclassical polynomials have zero shift.

2.3 Rank and Polynomial Factors

Definition 2.4 (Rank)**.** *Let $d \in \mathbb{N}$ and $f : \mathbb{F}^n \to \mathbb{T}$. Then* $\mathrm{rank}_d(f)$ *is defined as the smallest integer r such that there exist polynomials $h_1, \ldots, h_r : \mathbb{F}^n \to \mathbb{T}$ of degree $\leq d-1$ and a function $\Gamma : \mathbb{T}^r \to \mathbb{T}$ such that $f(x) = \Gamma(h_1(x), \ldots, h_r(x))$. If $d = 1$, then the rank is 0 if f is a constant function and is ∞ otherwise. If f is a polynomial, then* $\mathrm{rank}(f) = \mathrm{rank}_d(f)$ *where* $d = \deg(f)$.

Definition 2.5 (Factor)**.** *Let X be a finite set. Then a factor $\mathcal{B}$ is a partition of the set X. The subsets in the partition are called atoms.*

For sets X and Y, and a factor $\mathcal{B}$ of X, a function $f : X \to P(Y)$ is said to be measurable with respect to $\mathcal{B}$ if it is constant on the atoms of $\mathcal{B}$. The average of f over $\mathcal{B}$ is $\mathbb{E}[f|\mathcal{B}] : X \to P(Y)$ defined as

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

where $\mathcal{B}(x)$ is the atom containing x. Clearly, $\mathbb{E}[f|\mathcal{B}]$ is measurable with respect to $\mathcal{B}$.

A collection of functions $h_1, \ldots, h_c : X \to Y$ defines a factor $\mathcal{B}$ whose atoms are $\{x \in X : h_1(x) = y_1, \ldots, h_c(x) = y_c\}$ for every $(y_1, \ldots, y_c) \in Y^c$. We use $\mathcal{B}$ to also denote the map $x \mapsto (h_1(x), \ldots, h_c(x))$. A function f is measurable with respect to a collection of functions if it is measurable with respect to the factor the collection defines.

Definition 2.6 (Polynomial Factor)**.** *A polynomial factor $\mathcal{B}$ is a factor defined by a collection of polynomials $\mathcal{H} = \{h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{T}\}$ and the factor is written as $\mathcal{B}_{\mathcal{H}}$. The degree of the factor is the maximum degree of $h \in \mathcal{H}$.*

Let $|\mathcal{B}|$ be the number of polynomials defining the factor. If $\mathrm{depth}(h_i) = k_i$ above, then we define $||\mathcal{B}|| := \prod_{i=1}^c p^{k_i+1}$ to be the number of (possibly empty) atoms.

Definition 2.7 (Rank and Regularity of Polynomial Factor)**.** *Let $\mathcal{B}$ be a polynomial factor defined by $h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{T}$ such that* $\mathrm{depth}(h_i) = k_i$ *for $i \in [c]$. Then, the rank of $\mathcal{B}$ is the least integer r such that there exists $(a_1, \ldots, a_c) \in \mathbb{Z}^c$, $\left(a_1 \mod p^{k_1+1}, \ldots, a_c \mod p^{k_c+1}\right) \neq (0, \ldots, 0)$ for which the linear combination $h(x) := \sum_{i=1}^{c} a_i h_i(x)$ has $\mathrm{rank}_d(h) \leq r$ where $d = \max_i \deg(a_i h_i)$. For a non decreasing function $r : \mathbb{N} \to \mathbb{N}$, a factor $\mathcal{B}$ is r-regular if its rank is at least $r(|\mathcal{B}|)$.*

Definition 2.8 (Semantic and Syntactic refinement)**.** *Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors on $\mathbb{F}^n$. A factor $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, denoted by $\mathcal{B}' \succeq_{syn} \mathcal{B}$ if the set of polynomials defining $\mathcal{B}$ is a subset of the set of polynomials defining $\mathcal{B}'$. It is a semantic refinement, denoted by $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

We will use the following regularity lemma proved in [BFH$^+$13].

Lemma 2.9 (Polynomial Regularity Lemma [BFH$^+$13])**.** *Let $r : \mathbb{N} \to \mathbb{N}$ be a non-decreasing function and $d \in \mathbb{N}$. Then there is a function $C_{r,d}^{(2.9)} : \mathbb{N} \to \mathbb{N}$ such that the following is true. Let $\mathcal{B}$ be 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 $\mathcal{B}'$ defined by polynomials $Q_1, \ldots, Q_{c'} : \mathbb{F}^n \to \mathbb{T}$ of degree at most d such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $c' \leq C_{r,d}^{(2.9)}(c)$.*

Moreover if $\mathcal{B} \succeq_{sem} \hat{\mathcal{B}}$ for some polynomial factor $\hat{\mathcal{B}}$ that has rank at least $r(c') + c' + 1$, then $\mathcal{B}' \succeq_{syn} \hat{B}$.

The next lemma shows that a regular factor has atoms of roughly equal size.

Lemma 2.10 (Size of atoms [BFH$^+$13])**.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of rank at least $r_d^{(2.10)}(\varepsilon)$ defined by polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ of degree at most d such that* $\mathrm{depth}(P_i) = k_i$ *for $i \in [c]$. For every $b \in \otimes_{i=1}^{c} \mathbb{U}_{k_i+1}$,*

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

Finally, we shall need the following lemma which shows that a function of high rank polynomials has the degree one expects.

Lemma 2.11 (Preserving degree [BFH$^+$13])**.** *Let $d > 0$ be an integer and let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ be polynomials of degree at most d that form a factor of* rank $\geq r_d^{(2.11)}(c)$. *Let $\Gamma : \mathbb{T}^c \to \mathbb{T}$ be an arbitrary function. Let $F : \mathbb{F}^n \to \mathbb{T}$ be defined by $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$, and assume that* $\deg(F) = d'$. *Then, for every collection of polynomials $Q_1, \ldots, Q_c : \mathbb{F}^n \to \mathbb{T}$ with* $\deg(Q_i) \leq \deg(P_i)$ *and* $\mathrm{depth}(Q_i) \leq \mathrm{depth}(P_i)$, *if $G : \mathbb{F}^n \to \mathbb{T}$ is defined by $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$, then* $\deg(G) \leq d'$.

3 Weak Regularity

Let X and Y be finite sets. Recall that $P(Y) := \{f : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} f(y) = 1\}$ is the probability simplex on Y. As mentioned before, we shall write randomized functions by mapping them to the simplex. Thus for $f, g : X \to P(Y)$ we have

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x \langle f(x), g(x) \rangle.$$

Lemma 3.1. *Let $g : X \to P(Y)$, $\varepsilon > 0$, and F be a collection of functions $f : X \to P(Y)$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \dots, h_c \in F$ and a function $\Gamma : P(Y)^c \to P(Y)$ such that for all $f \in F$,*

$$|\mathbf{Pr}[g(x) = f(x)] - \mathbf{Pr}[\Gamma(h_1(x), h_2(x), \dots, h_c(x)) = f(x)]| \leq \varepsilon.$$

Proof. We construct $\mathcal{H} = \{h_1, \dots, h_c\} \subseteq F$ such that, if $\mathcal{B}_{\mathcal{H}}$ is the factor of X induced by $\mathcal{H}$, then for all $f \in F$

$$|\mathbf{Pr}[\mathbb{E}[g|\mathcal{B}_{\mathcal{H}}] = f(x)] - \mathbf{Pr}[g(x) = f(x)]| \leq \varepsilon.$$

We then set $\Gamma : P(Y)^c \to P(Y)$ so that $\Gamma(h_1(x), \dots, h_c(x)) = \mathbb{E}[g|\mathcal{B}_{\mathcal{H}}]$. In the following we shorthand $g_{\mathcal{H}} = \mathbb{E}[g|\mathcal{B}_{\mathcal{H}}]$. We consider the following variant of the Frieze-Kannan weak regularity algorithm [FK99].

- Initialize $\mathcal{H} = \emptyset$
- While there exists $f \in F$ such that $|\mathbf{Pr}[g_{\mathcal{H}}(x) = f(x)] - \mathbf{Pr}[g(x) = f(x)]| > \varepsilon$
  - Update $\mathcal{H} = \mathcal{H} \cup \{f\}$

The lemma follows from the following claim, which shows that we update $\mathcal{H}$ at most $1/\varepsilon^2$ times. Let $\|g_{\mathcal{H}}\|_2^2 := \mathbb{E}_x \|g_{\mathcal{H}}(x)\|_2^2$.

Claim 3.2. *Consider any stage in the algorithm, with $\mathcal{H}$ being the set of functions at that stage, and $f \in F$ being the new function added to $\mathcal{H}$. Then*

- $0 \leq \|g_{\mathcal{H}}\|^2 \leq 1$*;*
- $\|g_{\mathcal{H} \cup \{f\}}\|^2 \geq \|g_{\mathcal{H}}\|^2 + \varepsilon^2$*.*

Proof. The first part of the claim is trivial as $g_{\mathcal{H}}$ maps to $P(Y)$. For the second part, observe that $\langle g_{\mathcal{H} \cup \{f\}} - g_{\mathcal{H}}, g_{\mathcal{H}} \rangle = 0$ and thus

$$\|g_{\mathcal{H} \cup \{f\}}\|_2^2 = \|g_{\mathcal{H}}\|_2^2 + \|g_{\mathcal{H} \cup \{f\}} - g_{\mathcal{H}}\|_2^2$$

We will show that $\|g_{\mathcal{H} \cup \{f\}} - g_{\mathcal{H}}\|_2^2 \geq \varepsilon^2$. We have

$$\begin{aligned}
\varepsilon &< |\mathbf{Pr}[g_{\mathcal{H}}(x) = f(x)] - \mathbf{Pr}[g(x) = f(x)]| \\
&= |\mathbb{E}_x \langle f(x), g_{\mathcal{H}}(x) \rangle - \mathbb{E}_x \langle f(x), g(x) \rangle| \\
&= \left|\mathbb{E}_x \langle f(x), g_{\mathcal{H}}(x) \rangle - \mathbb{E}_x \langle f(x), g_{\mathcal{H} \cup \{f\}}(x) \rangle\right| \qquad \text{(as } f \text{ is measurable with respect to } \mathcal{B}_{\mathcal{H} \cup \{f\}}) \\
&= \left|\mathbb{E}_x \langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H} \cup \{f\}}(x) \rangle\right| \\
&\leq \mathbb{E}_x \left|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H} \cup \{f\}}(x) \rangle\right|.
\end{aligned}$$

Now, as $f : X \to P(Y)$, for every $x \in X$, $\|f(x)\|_2 \leq 1$. Thus, by the Cauchy-Schwartz inequality, for every $x \in X$, we have

$$|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H} \cup \{f\}}(x) \rangle| \leq \|f(x)\|_2 \|g_{\mathcal{H} \cup \{f\}}(x) - g_{\mathcal{H}}(x)\|_2 \leq \|g_{\mathcal{H} \cup \{f\}}(x) - g_{\mathcal{H}}(x)\|_2$$

Thus, by another application of the Cauchy-Schwartz inequality, we have

$$\varepsilon^2 \leq \mathbb{E}_x \left|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H} \cup \{f\}}(x) \rangle\right|^2 \leq \|g_{\mathcal{H} \cup \{f\}} - g_{\mathcal{H}}\|_2^2.$$

□

This finishes the proof of the lemma. □

The following corollary for deterministic functions $f : X \to Y$ allows to obtain one-sided deterministic estimates. This simplifies some of the arguments later on.

Corollary 3.3. *Let $g : X \to Y$, $\varepsilon > 0$, and F be a collection of functions $f : X \to Y$. Then there exist $c \le 1/\varepsilon^2$ functions $h_1, h_2, \dots, h_c \in F$ such that for every $f \in F$, there is a function $\Gamma_f : Y^c \to Y$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \dots, h_c(x)) = f(x)] \ge \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

Proof. Applying Lemma 3.1 to F we may assume the existence of $h_1, \dots, h_c : X \to Y$ and $\Gamma : Y^C \to P(Y)$ such that for any $f \in F$,

$$|\mathbf{Pr}[f(x) = \Gamma(h_1(x), \dots, h_c(x))] - \mathbf{Pr}[f(x) = g(x)]| \le \varepsilon.$$

Let $A_{y_1,\dots,y_c} = \{x \in X : h_1(x) = y_1, \dots, h_c(x) = y_c\}$ be an atom defined by $h_1, \dots, h_c$. Given $f \in F$, define $\Gamma_f : Y^c \to Y$ by letting $\Gamma_f(y_1, \dots, y_c)$ to be the most common value that f attains on $A_{y_1,\dots,y_c}$. Then

$$\begin{aligned}
&\mathbf{Pr}[f(x) = \Gamma_f(h_1(x), \dots, h_c(x))] \\
&= \sum_{y_1,\dots,y_c \in Y} \mathbf{Pr}[x \in A_{y_1,\dots,y_c}] \cdot \max_{y^* \in Y} \mathbf{Pr}[f(x) = y^* | x \in A_{y_1,\dots,y_c}] \\
&\ge \sum_{y_1,\dots,y_c \in Y} \mathbf{Pr}[x \in A_{y_1,\dots,y_c}] \cdot \mathbf{Pr}[f(x) = \Gamma(y_1, \dots, y_c) | x \in A_{y_1,\dots,y_c}] \\
&= \mathbf{Pr}[f(x) = \Gamma(h_1(x), \dots, h_c(x))] \ge \mathbf{Pr}[f(x) = g(x)] - \varepsilon.
\end{aligned}$$

□

4 Proof of Theorem 1

Fix a prime field $\mathbb{F} = \mathbb{F}_p$. For $d \in \mathbb{N}$, we shorthand $\delta(d) = \delta_{\mathbb{F}}(d)$. We restate Theorem 1.

Theorem 1. *Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta(d) - \varepsilon) \le c_{p,d,\varepsilon}.$$

We prove Theorem 1 in the remainder of this section. Let $g : \mathbb{F}^n \to \mathbb{U}_1$ be a received word where we identify $\mathbb{F}$ with $\mathbb{U}_1$. Apply Corollary 3.3 with $X = \mathbb{F}^n$, $Y = \mathbb{U}_1$, $F = \mathrm{Poly}_{\le d}(\mathbb{F}^n \to \mathbb{U}_1)$ and approximation parameter $\varepsilon/2$ to obtain $\mathcal{H} = \{h_1, \dots, h_c\} \subseteq F$, $c \le 4/\varepsilon^2$ such that, for every $f \in F$, there is a function $\Gamma_f : \mathbb{U}_1^c \to \mathbb{U}_1$ satisfying

$$\mathbf{Pr}[\Gamma_f(h_1(x), h_2(x), \dots, h_c(x)) = f(x)] \ge \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2.$$

Let $r_1, r_2 : \mathbb{N} \to \mathbb{N}$ be two non decreasing functions to be specified later, and let $C^{(2.9)}_{r,d}$ be as given in Lemma 2.9. We will require that for all $m \geq 1$,

$$r_1(m) \geq r_2(C^{(2.9)}_{r_2,d}(m+1)) + C^{(2.9)}_{r_2,d}(m+1) + 1. \tag{2}$$

As a first step, we r_1-regularize $\mathcal{H}$ by Lemma 2.9. This gives an r_1-regular factor $\mathcal{B}'$ of degree at most d, defined by polynomials $h'_1, \ldots, h'_{c'} : \mathbb{F}^n \to \mathbb{T}$, such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$, $c' \leq C^{(2.9)}_{r_1,d}(c)$ and $\mathrm{rank}(\mathcal{B}') \geq r_1(c')$. We denote $\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\}$. Note that $\mathcal{H}'$ can have nonclassical polynomials as a result of the regularization. Let $\mathrm{depth}(h'_i) = k_i$ for $i \in [c']$. Let $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{U}_1$ be defined such that

$$\Gamma_f(h_1(x), \ldots, h_c(x)) = G_f(h'_1(x), \ldots, h'_{c'}(x)).$$

Then

$$\mathbf{Pr}[G_f(h'_1(x), h'_2(x), \ldots, h'_{c'}(x)) = f(x)] \geq \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2. \tag{3}$$

Next, given any classical polynomial $f : \mathbb{F}^n \to \mathbb{T}$ of degree at most d, we will show that if $\mathbf{Pr}[f(x) \neq g(x)] \leq \delta(d) - \varepsilon$, then f is measurable with respect to $\mathcal{H}'$ and this would upper bound the number of such polynomials by $p^{||\mathcal{B}'||} = p^{\prod_{i\in[c']} p^{k_i+1}}$ and as $c' = c'(p, d, \varepsilon)$ and $k_i \leq \left\lfloor \frac{d-1}{p-1} \right\rfloor$ this is independent on n.

Fix such a classical polynomial f. Appealing again to Lemma 2.9, we r_2-regularize $\mathcal{B}_f := \mathcal{B}' \cup \{f\}$. We get an r_2-regular factor $\mathcal{B}'' \succeq_{syn} \mathcal{B}'$ defined by the collection $\mathcal{H}'' = \{h'_1, \ldots, h'_{c'}, h''_1, \ldots, h''_{c''}\} \subseteq \mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$. Note that it is a syntactic refinement of $\mathcal{B}'$ as by our choice of r_1,

$$\mathrm{rank}(\mathcal{B}') \geq r_1(c') \geq r_2(C^{(2.9)}_{r_2,d}(c'+1)) + C^{(2.9)}_{r_2,d}(c'+1) + 1 \geq r_2(|\mathcal{B}''|) + |\mathcal{B}''| + 1.$$

We will choose r_2 such that for all $m \geq 1$,

$$r_2(m) = \max\left(r_d^{(2.10)}\left(\frac{\varepsilon/4}{\left(p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}\right)^m} \right), r_d^{(2.11)}(m) \right). \tag{4}$$

Let $\mathrm{depth}(h''_j) = l_j$ for $j \in [c'']$ and denote $S := \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \times \otimes_{j=1}^{c''} \mathbb{U}_{l_j+1}$. Since f is measurable with respect to $\mathcal{B}''$, there exists $F : S \to \mathbb{U}_1$ such that

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)).$$

We next show that we can have each polynomial in the factor have a disjoint set of inputs, and still obtain more or less the same approximation factor.

Claim 4.1. *Let x^i, y^j, $i \in [c'], j \in [c'']$ be pairwise disjoint sets of $n \in \mathbb{N}$ variables each. Let $n' = n(c' + c'')$. Let $\tilde{f} : \mathbb{F}^{n'} \to \mathbb{U}_1$ and $\tilde{g} : \mathbb{F}^{n'} \to \mathbb{U}_1$ be defined as*

$$\tilde{f}(x) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''}))$$

and

$$\tilde{g}(x) = G_f(h'_1(x^1), \ldots, h_{c'}(x^{c'})).$$

Then $\deg(\tilde{f}) \leq d$ and

$$\left| \mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] - \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \right| \leq \varepsilon/4.$$

Proof. The bound $\deg(\tilde{f}) \le \deg(f) \le d$ follows from Lemma 2.11 since $r_2(|\mathcal{H}''|) \ge r_d^{(2.11)}(|\mathcal{H}''|)$. To establish the bound on $\mathbf{Pr}[\tilde{f} = \tilde{g}]$, for each $s \in S$ let

$$p_1(s) = \mathbf{Pr}_{x \in \mathbb{F}^n}[(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = s].$$

Applying Lemma 2.10 and since our choice of r_2 satisfies $\mathrm{rank}(\mathcal{H}'') \ge r_d^{(2.10)}(\varepsilon/4|S|)$, we have that p_1 is nearly uniform over S,

$$p_1(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

Similarly, let

$$p_2(s) = \mathbf{Pr}_{x^1,\ldots,x^{c'},y^1,\ldots,y^{c''} \in \mathbb{F}^n}[(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})) = s].$$

Note that the rank of the collection of polynomials $\{h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})\}$ defined over $\mathbb{F}^{n'}$ cannot be lower than that of $\mathcal{H}''$. Applying Lemma 2.10 again gives

$$p_2(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

For $s \in S$, let $s' \in \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1}$ be the restriction of s to first c' coordinates, that is, $s' = (s_1, \ldots, s_{c'})$. Thus

$$\begin{aligned}
\mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] &= \sum_{s \in S} p_2(s) 1_{F(s) = G_f(s')} \\
&= \sum_{s \in S} p_1(s) 1_{F(s) = G_f(s')} \pm \varepsilon/4 \\
&= \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \pm \varepsilon/4.
\end{aligned}$$

□

So, we obtain that

$$\mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] \ge \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), \ldots, h'_{c'}(x))] - \varepsilon/4 \ge 1 - \delta(d) + \varepsilon/4.$$

Next, we need the following variant of the Schwartz-Zippel lemma [Sch80, Zip79].

Claim 4.2. *Let $d, n_1, n_2 \in \mathbb{N}$. Let $f_1 : \mathbb{F}^{n_1+n_2} \to \mathbb{F}$ and $f_2 : \mathbb{F}^{n_1} \to \mathbb{F}$ be such that* $\deg(f_1) \le d$ *and*

$$\mathbf{Pr}[f_1(x_1, \ldots, x_{n_1+n_2}) = f_2(x_1, \ldots, x_{n_1})] > 1 - \delta(d)$$

Then, f_1 does not depend on $x_{n_1+1}, \ldots, x_{n_1+n_2}$.

Proof. We will show that f_1 does not depend on $z = x_{n_1+n_2}$ say. The proof for any other variable is similar. Recall that $\delta(d) := \frac{1}{p^a}\left(1 - \frac{b}{p}\right)$ where $d = a \cdot (p-1) + b$. Let $f_1(x) = \sum_{k=0}^{d'} c_k z^k$ where $c_k \in \mathbb{F}[x_1, \ldots, x_{n_1+n_2-1}]$ and $d' \le \min\{d, p-1\}$. Then $(f_1 - f_2)(x) = c_0 - f_2(x) + \sum_{k=1}^{d'} c_k z^k$. We will show that $d' \ge 1$ will lead to a contradiction. Let $\deg(c_{d'}) = d''$. Note that $d'' + d' \le d$. Then,

$$\mathbf{Pr}[(f_1 - f_2)(x) = 0] \le \mathbf{Pr}[c_{d'} = 0] + (1 - \mathbf{Pr}[c_{d'} = 0])(1 - \delta(d')) \le 1 - \delta(d'')\delta(d').$$

We will show that for any $d \geq 1$ and any $1 \leq c \leq p-1$, we have $\delta(c)\delta(d-c) \geq \delta(d)$ and this will show that $\mathbf{Pr}[(f_1 - f_2)(x) = 0] \leq 1 - \delta(d' + d'') \leq 1 - \delta(d)$ which leads to a contradiction. Thus, f_1 will not depend on z. We will now show that

$$\delta(c)\delta(d-c) \geq \delta(d) \tag{5}$$

Let $d = a \cdot (p-1) + b$.

Case 1: $0 \leq c \leq b$

$$\begin{aligned}(5) \quad &\Leftrightarrow \quad \left(1 - \frac{c}{p}\right) \frac{1}{p^a} \left(1 - \frac{b-c}{p}\right) \geq \frac{1}{p^a} \left(1 - \frac{b}{p}\right) \\ &\Leftrightarrow \quad b \geq c\end{aligned}$$

Case 2: $b < c \leq p-1$

$$\begin{aligned}(5) \quad &\Leftrightarrow \quad \left(1 - \frac{c}{p}\right) \frac{1}{p^{a-1}} \left(\frac{1+c-b}{p}\right) \geq \frac{1}{p^a} \left(1 - \frac{b}{p}\right) \\ &\Leftrightarrow \quad (c-b)\left(1 - \frac{c+1}{p}\right) \geq 0\end{aligned}$$

which is true by hypothesis. □

Now apply Claim 4.2 to $f_1 = \tilde{f}, f_2 = \tilde{g}, n_1 = nc', n_2 = nc''$. We obtain that $\tilde{f}$ does not depend on $y^1, \ldots, y^{c''}$. Hence,

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(h_1'(x^1), \ldots, h_{c'}'(x^{c'}), C_1, \ldots, C_{c''})$$

where $C_j = h_j''(0) \in \mathbb{U}_{l_j+1}$ for $j \in [c'']$. If we substitute $x^1 = \ldots = x^{c'} = x$ we get that

$$f(x) = F(h_1'(x), \ldots, h_{c'}'(x), h_1''(x), \ldots, h_{c''}''(x)) = F(h_1'(x), \ldots, h_{c'}'(x), C_1, \ldots, C_{c''}),$$

which shows that f is measurable with respect to $\mathcal{H}'$, as claimed.

5 Proof of Theorem 2

Theorem 2. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $e \leq d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta(e) - \varepsilon) \leq \exp\left(c_{p,d,\varepsilon} n^{d-e}\right).$$

The proof follows along the same lines as that of Theorem 1. It will rely on the following lemma which generalizes Claim 4.2.

Lemma 5.1. *Fix $d \geq e \geq 1, \varepsilon > 0$. There exists $r_{d,\varepsilon}^{(5.1)} \in \mathbb{N}$ such that the following holds. Let $f_1 : \mathbb{F}^{n_1+n_2} \to \mathbb{U}_1$ be a classical polynomial of degree at most d. Assume that*

- *There exist $f_2 : \mathbb{F}^{n_1} \to \mathbb{U}_1$ be such that* $\mathbf{Pr}[f_1(x,y) = f_2(x)] \geq 1 - \delta(e) + \varepsilon$.
- *There exists a polynomial $h : \mathbb{F}^{n_2} \to \mathbb{U}_{k+1}$ of degree at most d such that the factor it defines has rank at least $r^{(5.1)}_{d,\varepsilon}$, and a function $\Gamma : \mathbb{F}^{n_1} \times \mathbb{U}_{k+1} \to \mathbb{U}_1$, such that*
$$f_1(x,y) = \Gamma(x, h(y)).$$
- *The dependence on the depth of h is nontrivial: $f_1(x,y)$ cannot be written as $\Gamma'(x, p \cdot h(y))$ for any $\Gamma' : \mathbb{F}^{n_1} \times \mathbb{U}_k \to \mathbb{U}_1$.*

Then $\deg(h) \leq d - e$.

We first prove Theorem 2 assuming Lemma 5.1.

Proof of Theorem 2 assuming Lemma 5.1. The initial part of the proof is as in Theorem 1. Assume that $n > r^{(5.1)}_{d,\varepsilon/4}$ otherwise the theorem is trivially true. Let $f, g : \mathbb{F}^n \to \mathbb{U}_1$ with $\deg(f) \leq d$ and $\mathrm{dist}(f,g) \leq \delta(e) - \varepsilon$. For non decreasing functions $r_1, r_2 : \mathbb{N} \to \mathbb{N}$, chosen as in the proof of Theorem 1, we have an r_1-regular $\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\}$ and an r_2-regular $\mathcal{H}'' = \mathcal{H}' \cup \{h''_1, \ldots, h''_{c''}\}$ where each h'_i, h''_i is a nonclassical polynomial of degree $\leq d$, such that the following holds.

Let $\mathrm{depth}(h'_i) = k_i$ for $i \in [c']$ and $\mathrm{depth}(h''_j) = l_j$ for $j \in [c'']$. Since f is measurable with respect to $\mathcal{H}''$, there exists $F : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \times \otimes_{j=1}^{c''} \mathbb{U}_{l_j+1} \to \mathbb{U}_1$ such that
$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)).$$
We may assume that for all $i \in [c'']$, the depth of h''_i is minimal, in the sense that we cannot replace h''_i with $p \cdot h''_i$ and change F accordingly to still compute f (if this is not the case, then replace h''_i with $p \cdot h''_i$ whenever possible; this only reduces the degree of h''_i and the new factor has rank at least that of the original factor). Also, there exists a function $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{U}_1$ such that
$$\mathbf{Pr}[G_f(h'_1(x), \ldots, h'_{c'}(x)) = f(x)] \geq 1 - \delta(e) + \varepsilon/2.$$
We will show that this implies that $\deg(h''_i) \leq d - e$ for all $i \in [c'']$. Let $\mathcal{B}''$ be the factor defined by $\mathcal{H}''$. As the number of polynomials of degree $d - e$ is exponential in n^{d-e}, the number of functions f is controlled by the product of the number of composing functions F, which is $p^{||\mathcal{B}''||} = p^{\left(\prod_{i \in [c']} p^{k_i+1}\right)\left(\prod_{j \in [c'']} p^{l_j+1}\right)} = c_1(p, d, \varepsilon)$, and the number of choices for $h''_1, \ldots, h''_{c''}$, which is $\exp(c_2 c'' n^{d-e})$. This amounts to at most $\exp(cn^{d-e})$ for some $c = c(p, d, \varepsilon)$, as claimed.

To prove the bound on the degrees of $h''_1, \ldots, h''_{c''}$, define, as in the proof of Theorem 1, x^i, y^j for $i \in [c'], j \in [c'']$ to be pairwise disjoint sets of $n \in \mathbb{N}$ variables. Let $n' = n(c' + c'')$. Define $\tilde{f} : \mathbb{F}^{n'} \to \mathbb{U}_1$ and $\tilde{g} : \mathbb{F}^{n'} \to \mathbb{U}_1$ as
$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''}))$$
and
$$\tilde{g}(x^1, \ldots, x^{c'}) = G_f(h'_1(x^1), \ldots, h_{c'}(x^{c'})).$$
Then, by Claim 4.1, $\deg(\tilde{f}) \leq d$ and $\mathbf{Pr}[\tilde{f} = \tilde{g}] \geq 1 - \delta(e) + \varepsilon/4$.

We next apply Lemma 5.1 to show that $\deg(h''_j) \le d - e$ for all $j \in [c'']$. To see that for say, $j = c''$, let $k = \mathrm{depth}(h''_{c''})$, $n_1 = n(c' + c'' - 1), n_2 = n, h(y) = h''_{c''}(y)$ and $\Gamma : \mathbb{F}^{n_1} \times \mathbb{U}_{k+1} \to \mathbb{U}_1$ given by

$$\Gamma((x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''-1}), \alpha) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''-1}(y^{c''-1}), \alpha).$$

so that

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = \Gamma((x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''-1}), h''_{c''}(y^{c''})).$$

If we make sure that $r_2(m) \ge r^{(5.1)}_{d,\varepsilon/4}$ for all $m \ge 1$, then we establish all the requirements for Lemma 5.1. Hence we deduce that $\deg(h''_{c''}) \le d - e$ as claimed. $\square$

5.1 Proof of Lemma 5.1

We prove Lemma 5.1 in this section. Fix $d \ge e \ge 1$ and $\varepsilon > 0$. Let $r = r^{(5.1)}_{d,\varepsilon}$ be large enough to be chosen later. We first show that we can replace h with a simple polynomial of the same degree and depth, which would allow us to simplify the analysis.

Let $\mathrm{depth}(h) = k$ and let $A = \deg(h) - (p-1)k$. Define $\tilde{h} : \mathbb{F}^{rA} \to \mathbb{U}_{k+1}$ as follows. Let $z = (z_{1,1}, \ldots, z_{r,A}) \in \mathbb{F}^{rA}$ and define

$$\tilde{h}(z) := \frac{\sum_{i=1}^{r} \prod_{j=1}^{A} z_{i,j}}{p^{k+1}}. \tag{6}$$

Note that $\tilde{h}$ and h are both polynomials of the same degree and depth. Define $\tilde{f}_1 : \mathbb{F}^{n_1+rA} \to \mathbb{U}_1$ as

$$\tilde{f}_1(x, z) = \Gamma(x, \tilde{h}(z)).$$

We will show that we may analyze $\tilde{f}_1$ instead of f_1 to obtain the upper bound on $\deg(h)$. To simplify the presentation, denote $Z_i := \prod_{j=1}^{A} z_{i,j}$ for $i \in [r]$. First, we argue that if r is chosen large enough then both $h, \tilde{h}$ are nearly uniform over $\mathbb{U}_{k+1}$.

Claim 5.2. *If r is chosen large enough then for all $\alpha \in \mathbb{U}_{k+1}$,*

$$\mathbf{Pr}_{y \in \mathbb{F}^{n_2}}[h(y) = \alpha] = p^{-(k+1)}(1 \pm \varepsilon/2)$$

and

$$\mathbf{Pr}_{z \in \mathbb{F}^{rA}}[\tilde{h}(z) = \alpha] = p^{-(k+1)}(1 \pm \varepsilon/2).$$

Proof. The proof for h follows from Lemma 2.10 by choosing $r \ge r^{2.10}_{d}\left(\frac{\varepsilon}{2p^{k+1}}\right)$. The proof for $\tilde{h}$ follows by a simple Fourier calculation. Let $\omega = \exp(2\pi i/p^{k+1})$. We have $\mathbf{Pr}[Z_i = 0], \mathbf{Pr}[Z_i = 1] \ge p^{-A} \ge p^{-d}$. One can verify that this implies that for any nonzero $c \in \mathbb{Z}_{p^{k+1}}$, $\mathbb{E}[\omega^{cZ_i}] \le 1 - \eta$ for $\eta = p^{-O(d)}$. As $Z_1, \ldots, Z_r$ are independent we have $\mathbb{E}\left[\omega^{c(Z_1 + \ldots + Z_r)}\right] \le (1-\eta)^r$. Hence if we choose r large enough so that $(1-\eta)^r < (\varepsilon/2)p^{-(k+1)}$ then, for any $a \in \mathbb{Z}_{p^{k+1}}$,

$$\begin{aligned} \mathbf{Pr}[Z_1 + \ldots + Z_r = a \pmod{p^{k+1}}] &= p^{-(k+1)}\left(1 + \sum_{c \in \mathbb{Z}_{p^{k+1}} \setminus \{0\}} \omega^{-ac} \cdot \mathbb{E}\left[\omega^{c(Z_1 + \ldots + Z_r)}\right]\right) \\ &= p^{-(k+1)}(1 \pm \varepsilon/2). \end{aligned}$$

$\square$

This implies that $f_2(x)$ is also well approximates $\tilde{f}_1(x,z)$.

Corollary 5.3. *$\mathbf{Pr}[\tilde{f}_1(x,z) = f_2(x)] \geq \mathbf{Pr}[f_1(x,y) = f_2(x)] - \varepsilon/2 \geq 1 - \delta(e) + \varepsilon/2$ where $x \in \mathbb{F}^{n_1}, y \in \mathbb{F}^{n_2}, z \in \mathbb{F}^{rA}$ are chosen uniformly and independently.*

Proof. Claim 5.2 implies that the statistical distance between $h(y)$ and $\tilde{h}(z)$ is at most $\varepsilon/2$. Hence for every fixed x, $|\mathbf{Pr}[\Gamma(x, h(y)) = f_2(x)] - \mathbf{Pr}[\Gamma(x, \tilde{h}(z)) = f_2(x)]| \leq \varepsilon/2$. $\square$

We next argue that by choosing r large enough, we can guarantee that $\tilde{f}_1$ has degree at most d.

Claim 5.4. *If r is chosen large enough then $\deg(\tilde{f}_1) \leq \deg(f_1) \leq d$.*

Proof. By Claim 5.2, if r is chosen large enough then $h(y), \tilde{h}(z)$ attain all possible values in $\mathbb{U}_{k+1}$. For every $\alpha \in \mathbb{U}_{k+1}$, let $f_\alpha(x) := \Gamma(x, \alpha)$. Note that as there exists some $y_\alpha \in h^{-1}(\alpha)$ then $f_\alpha(x) = f_1(x, y_\alpha)$ is a (classical) polynomial in x of degree at most d.

We have $f_1(x,y) = \Gamma(x, h(y)) = \Gamma'((f_\alpha(x) : \alpha \in \mathbb{U}_{k+1}), h(y))$ for some $\Gamma' : \mathbb{F}^{p^{k+1}} \times \mathbb{U}_{k+1} \to \mathbb{F}$. Let $\mathcal{H} = \{f_\alpha(x) : \alpha \in \mathbb{U}_{k+1}\}$ and for $r_1 : \mathbb{N} \to \mathbb{N}$ a growth function to be specified later, let $\mathcal{H}' = \{g_1(x), \ldots, g_c(x)\}$ be the result of r_1-regularizing $\mathcal{H}$ by Lemma 2.9. Then

$$f_1(x,y) = \Gamma''(g_1(x), \ldots, g_c(x), h(y))$$

for some $\Gamma'' : \mathbb{F}^c \times \mathbb{U}_{k+1} \to \mathbb{F}$. Hence also

$$\tilde{f}_1(x,z) = \Gamma(x, \tilde{h}(z)) = \Gamma''(g_1(x), \ldots, g_c(x), \tilde{h}(z)).$$

We next apply Lemma 2.11 to bound the degree of $\tilde{f}_1$. This requires to assume that $r_1(c) \geq r_d^{(2.11)}(c+1)$ and $r \geq r_d^{(2.11)}(C_{r_1,d}^{(2.9)}(p^{k+1}) + 1)$. We obtain that

$$\deg(\tilde{f}_1) = \deg(\Gamma''(g_1(x), \ldots, g_c(x), \tilde{h}(z))) \leq \deg(\Gamma''(g_1(x), \ldots, g_c(x), h(y))) = \deg(f_1) = d.$$

$\square$

We next analyze the specific properties of $\tilde{h}$. Recall that we set $Z_i := \prod_{j=1}^{A} z_{i,j}$ so that $\tilde{h}(z) = \frac{\sum Z_i}{p^{k+1}}$. Since $\tilde{h}$ depends only on $W = \sum Z_i \mod p^{k+1}$, let the digits of $W \mod p^{k+1}$ in base p, be represented by classical polynomials $W_0(z), \ldots, W_k(z) : \mathbb{F}^{rA} \to \mathbb{F}$. Then, we can express $\tilde{f}_1(x,z)$ as

$$\tilde{f}_1(x,z) = \Gamma(x, \tilde{h}(z)) = \Gamma'(x, W_0(z), W_1(z), \ldots, W_k(z)) \tag{7}$$

for some $\Gamma' : \mathbb{F}^{n_1} \times \mathbb{F}^{k+1} \to \mathbb{U}_1$. Recall that we assumed that Γ depends nontrivially on the depth of its second argument. This implies that Γ' depends nontrivially on its last input (i.e. $W_k(z)$). As $\tilde{f}_1$ is a classical polynomial, and each W_i take values in $\mathbb{F}$, identifying $\mathbb{U}_1$ with $\mathbb{F}$, we can decompose

$$\tilde{f}_1(x,z) = \sum_{0 \leq d_0, \ldots, d_k \leq p-1} f_{d_0,\ldots,d_k}(x) \prod_{i=0}^{k} W_i(z)^{d_i}, \tag{8}$$

where $f_{d_0,\ldots,d_k} \in \mathbb{F}[x]$ is a classical polynomial. We next argue that $\deg(f_{d_0,\ldots,d_k})$ cannot be too large.

Lemma 5.5. $\deg(f_{d_0,\ldots,d_k}) \leq d - A\sum_{i=0}^{k} p^i d_i$ *for all* $0 \leq d_0, \ldots, d_k \leq p-1$.

We will require a few simple claims first. The ℓ-th symmetric polynomial in $Z = (Z_1, \ldots, Z_r)$, for $1 \leq \ell \leq r$, is a classical polynomial of degree ℓ defined as

$$S_\ell(Z) = \sum_{1 \leq i_1 < \ldots < i_\ell \leq r} \prod_{j=1}^{\ell} Z_{i_j}.$$

For $0 \leq i \leq k$, define $W'_i : \mathbb{F}^{rA} \to \mathbb{F}$ by $W'_i(z) := S_{p^i}(Z)$. The following claim follows immediately from Lucas theorem [Luc78].

Claim 5.6. *Let* $z \in \{0,1\}^{rA}$*. Then,* $W_i(z) = W'_i(z)$ *for* $i = 0, \ldots, k$*.*

Proof. If $z \in \{0,1\}^{rA}$ then $Z \in \{0,1\}^r$. Lucas theorem implies that the i-th least significant digit (starting at 0) of $W = Z_1 + \ldots + Z_r$ in base p is given by $\binom{Z_1+\ldots+Z_r}{p^i} \mod p = S_{p^i}(Z)$. □

For every polynomial $P \in \mathbb{F}[z]$, define $\mathrm{ML}(P)$ to be the multilinearization of P. That is, it is obtained by replacing each $z_{i,j}^a$ by $z_{i,j}$ for all $a \geq 1$ and all $i \in [r], j \in [A]$. Note that $\mathrm{ML}(P)(z) = P(z)$ for all $z \in \{0,1\}^{rA}$.

Claim 5.7. *Let* $P, Q : \mathbb{F}^{rA} \to \mathbb{F}$ *be two polynomials such that* $P(z) = Q(z)$ *for all* $z \in \{0,1\}^{rA}$*. Then* $\mathrm{ML}(P) \equiv \mathrm{ML}(Q)$*.*

Proof. Let $n = rA$. It is easy to see that a multilinear polynomial $f : \mathbb{F}^n \to \mathbb{F}$ satisfies $f(z) = 0$ for all $z \in \{0,1\}^n$ if and only if $f \equiv 0$. Therefore, for every polynomial $P : \mathbb{F}^n \to \mathbb{F}$, $\mathrm{ML}(P)$ is the unique multilinear polynomial that agrees with P on $\{0,1\}^n$. Let $R : \mathbb{F}^n \to \mathbb{F}$ be defined as $R := P - Q$. Then by linearity, $\mathrm{ML}(R) :\equiv \mathrm{ML}(P) - \mathrm{ML}(Q)$. As $\mathrm{ML}(R) = 0$ for all $z \in \{0,1\}^n$, $\mathrm{ML}(R) \equiv 0$ which implies $\mathrm{ML}(P) \equiv \mathrm{ML}(Q)$. □

Proof of Lemma 5.5. For $D = \sum_{i=0}^{k} p^i d_i$, define

$$W^{(D)}(z) := \prod_{i=0}^{k} W_i(z)^{d_i}, \qquad W'^{(D)}(z) := \prod_{i=0}^{k} W'_i(z)^{d_i}.$$

By Claim 5.6 and Claim 5.7, we can define a common multilinearization of $W^{(D)}$ and $W'^{(D)}$ by

$$M^{(D)} := \mathrm{ML}\left(W^{(D)}\right) = \mathrm{ML}\left(W'^{(D)}\right).$$

Let $m'(z) = \prod_{i=1}^{D} Z_i = \prod_{i=1}^{D}\prod_{j=1}^{A} z_{i,j}$ be a monomial. The coefficient of m' in $W'^{(D)}$ is equal to the coefficient of $\prod_{i=1}^{D} Z_i$ in $\prod_{i=0}^{k} S_{p^i}(Z)^{d_i}$, which is equal to the number of partitions of a set of size D to d_0 sets of size 1, d_1 sets of size p, d_2 sets of size p^2, up to d_k sets of size p^k. This is given by

$$\prod_{i=0}^{k}\prod_{j=1}^{d_i} \binom{jp^i + d_{i+1}p^{i+1} + \ldots + d_k p^k}{p^i},$$

which by Lucas theorem is equal modulo p to $\prod_{i=0}^{k}(d_i!) \neq 0 \mod p$.

Owing to the above, we have $\deg(M^{(D)}) \leq \deg(W'^{(D)}) = AD$. Also, since $m'(z)$ is of maximal degree, it also remains in $M^{(D)}$ after multilinearization. Define

$$\bar{f}_1(x,z) := \sum_{0 \leq d_0,\ldots,d_k \leq p-1} f_{d_0,\ldots,d_k}(x) M^{(D)}(z).$$

Then, we have $\deg(\bar{f}_1) \leq \deg(\tilde{f}_1) \leq d$.

Now, suppose that the lemma is false. Let $D = \sum p^i d_i$ be maximal such that $\deg(f_{d_0,\ldots,d_k}) > d - AD$. Note that D corresponds to a unique tuple $(d_0, \ldots, d_k)$. Let $m(x)$ be any monomial in $f_{d_0,\ldots,d_k}(x)$ with maximal degree, and recall that $m'(z) = \prod_{i=1}^{D} Z_i = \prod_{i=1}^{D} \prod_{j=1}^{A} z_{i,j}$. Hence, the monomial $m(x)m'(z)$, whose degree is larger than d, has a nonzero coefficient in $f_{d_0,\ldots,d_k}(x)M^{(D)}(z)$ as noted above. We will show it has a zero coefficient in any other $f_{d'_0,\ldots,d'_k}(x)M^{(D')}(z)$ with $(d'_0,\ldots,d'_k) \neq (d_0,\ldots,d_k)$, $D' = \sum_i p^i d'_i$ which will contradict the fact that $\deg(\bar{f}_1) \leq d$.

So, let $(d'_0,\ldots,d'_k) \neq (d_0,\ldots,d_k)$ and let $D' = \sum p^i d'_i$. Note that necessarily $D' \neq D$. If $D' > D$ then by maximality of D, $\deg\left(f_{d'_0,\ldots,d'_k}\right) \leq d - AD' < d - AD$ and hence $m(x)$ cannot appear in $f_{d'_0,\ldots,d'_k}(x)$. If $D' < D$ then $\deg\left(M^{(D')}\right) = AD' < AD$ and hence $m'(z)$ cannot appear in $M^{(D')}(z)$. □

Let $w = (w_0, \ldots, w_k) \in \mathbb{F}^{k+1}$ be new variables, and define $f'_1 : \mathbb{F}^{n_1+k+1} \to \mathbb{F}$ by

$$f'_1(x,w) = \Gamma'(x, w_0, \ldots, w_k) = \sum_{0 \leq d_0,\ldots,d_k \leq p-1} f_{d_0,\ldots,d_k}(x) \prod_{i=0}^{k} w_i^{d_i}. \tag{9}$$

We next argue that f'_1 is also well approximated by f_2.

Claim 5.8. $\mathbf{Pr}[f'_1(x,w) = f_2(x)] \geq \mathbf{Pr}[\tilde{f}_1(x,z) = f_2(x)] - \varepsilon/4 \geq 1 - \delta(e) + \varepsilon/4$, *where* $x \in \mathbb{F}^{n_1}, z \in \mathbb{F}^{rA}, w \in \mathbb{F}^{k+1}$ *are uniformly and independently distributed.*

Proof. By Claim 5.2, the distribution of $\tilde{h}$ is $\varepsilon/4$-close close in statistical distance to the uniform distribution over $\mathbb{U}_{k+1}$, hence the distribution of $(W_0(z), \ldots, W_k(z))$ is $\varepsilon/4$-close in statistical distance to the uniform distribution over $\mathbb{F}^{k+1}$. □

To conclude the proof of Lemma 5.1, expand $f'_1 - f_2$ as

$$f'_1(x,w) - f_2(x) = \sum_{i=0}^{d'} c_i(x, w_0, \ldots, w_{k-1}) w_k^i$$

where $c_i \in \mathbb{F}[x, w_0, \ldots, w_{k-1}]$, $d' \leq \min(d, p-1)$ and $c_{d'} \neq 0$. We have that $d' \geq 1$ since Γ' depends on $W_k(z)$. Also, by Lemma 5.5, for $i \geq 1$ we have $\deg(c_i) \leq d - Ap^k i$. To see this, suppose not. Consider the expansion in (9). Then, for some $d_0, \ldots d_{k-1}$, $\deg(f_{d_0,\ldots d_{k-1},i}) + \sum_{j=0}^{k-1} d_j > d - Ap^k i$, which implies that

$$\deg(f_{d_0,\ldots d_{k-1},i}) > d - \sum_{j=0}^{k-1} d_j - Ap^k i \geq d - A \sum_{j=0}^{k-1} d_j p^j - Ap^k i,$$

which is a contradiction to Lemma 5.5. Hence

$$\begin{aligned}\mathbf{Pr}[f_1'(x,w) = f_2(x)] &\leq \mathbf{Pr}[c_{d'} = 0] + (1 - \mathbf{Pr}[c_{d'} = 0])(1 - \delta(d')) \\ &\leq 1 - \delta(d - Ap^k d')\delta(d') \leq 1 - \delta(d - d'(Ap^k - 1)),\end{aligned}$$

where the last inequality was established in Claim 4.2. So, as we established that $\delta(d-d'(Ap^k-1)) < \delta(e)$ and $d' \geq 1$ we must have $Ap^k - 1 < d - e$, and hence $Ap^k \leq d - e$. Now, recall that $\deg(h) = \deg(\tilde{h}) = A + (p-1)k$ and it is a simple exercise to verify that $A + (p-1)k \leq Ap^k$ for all $A \geq 1, k \geq 0$. We thus showed that $\deg(h) \leq d - e$, as claimed.

6 Open Problems

Theorem 1 and Theorem 2 establish that over any fixed prime field $\mathbb{F}_p$ and any fixed $e \leq d$ and $\varepsilon > 0$, the number of degree d polynomials in a any ball of radius $\delta(e) - \varepsilon$ is at most $\exp(cn^{d-e})$ for some $c = c(p, d, \varepsilon)$, which in particular resolves the conjecture raised in [GKZ08] when $e = d$.

However, the bounds on c which we obtain are of Ackermann-type, which seem far from optimal. This leaves open the question of obtaining better bounds. This may require a different approach, as currently higher-order Fourier analysis does not seem to provide better bounds. We also leave as an open problem the question of extending our work to non-prime fields, and note that the missing ingredient is an extension of the higher-order Fourier analytic techniques to non prime fields.

References

[AGS03] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. In *Proc. 44^{th} IEEE Symposium on Foundations of Computer Science (FOCS'03)*, 2003.

[AS03] S. Arora and M. Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003.

[BFH+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *STOC*, pages 429–436, 2013.

[DGKS08] Irit Dinur, Elena Grigorescu, Swastik Kopparty, and Madhu Sudan. Decodability of group homomorphisms beyond the johnson bound. In *STOC*, pages 275–284, 2008.

[Eli57] P. Elias. List decoding for noisy channels. Technical Report 335, Research Laboratory of Electronics, MIT, 1957.

[FK99] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[GKZ08] P. Gopalan, A. Klivans, and D. Zuckerman. List decoding Reed-Muller codes over small fields. In *Proc. 40^{th} ACM Symposium on the Theory of Computing (STOC'08)*, pages 265–274, 2008.

[GL89] O. Goldreich and L. Levin. A hard-core predicate for all one-way functions. In *Proc. 21^{st} ACM Symposium on the Theory of Computing*, pages 25–32, 1989.

[Gop10] P. Gopalan. A Fourier-analytic approach to Reed-Muller decoding. In *Proc. 51^{st} IEEE Symp. on Foundations of Computer Science (FOCS'10)*, pages 685–694, 2010.

[GRS00] O. Goldreich, R. Rubinfeld, and M. Sudan. Learning polynomials with queries: The highly noisy case. *SIAM J. Discrete Math.*, 13(4):535–570, 2000.

[GS99] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and Algebraic-Geometric codes. *IEEE Transactions on Information Theory*, 45(6):1757–1767, 1999.

[GT09] B. Green and T. Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contrib. Discrete Math*, 4(2):1–36, 2009.

[Gur04] V. Guruswami. *List Decoding of Error-Correcting Codes*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004.

[Gur06] V. Guruswami. *Algorithmic Results in List Decoding*, volume 2 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers, 2006.

[Jac97] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55:414–440, 1997.

[KLP10] T. Kaufman, S. Lovett, and E. Porat. Weight distribution and list-decoding size of Reed-Muller codes. In *Innovations in Computer Science (ICS'10)*, pages 422–433, 2010.

[KM93] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM Journal of Computing*, 22(6):1331–1348, 1993.

[Luc78] Edouard Lucas. Thorie des fonctions numriques simplement priodiques. *American Journal of Mathematics*, 1(2):pp. 184–196, 1878.

[PW04] R. Pellikaan and X. Wu. List decoding of q-ary Reed-Muller codes. *IEEE Transactions on Information Theory*, 50(4):679–682, 2004.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verication of polynomial identities. *Journal of the ACM*, 27:701–717, 1980.

[STV01] M. Sudan, L. Trevisan, and S. P. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. Syst. Sci.*, 62(2):236–266, 2001.

[SU05] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *J. ACM*, 52(2):172–216, 2005.

[Sud97] M. Sudan. Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180–193, 1997.

[Sud00] M. Sudan. List decoding: Algorithms and applications. *SIGACT News*, 31(1):16–27, 2000.

[Tre03] L. Trevisan. List-decoding using the XOR lemma. In *Proc. 44^{th} IEEE Symposium on Foundations of Computer Science (FOCS'03)*, page 126, 2003.

[TSZS01] A. Ta-Shma, D. Zuckerman, and S. Safra. Extractors from Reed-Muller codes. In *Proc. 42^{nd} IEEE Symp. on Foundations of Computer Science (FOCS'01)*, pages 638–647, 2001.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 126–136. IEEE, 2009.

[TZ11] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *ArXiv e-prints*, January 2011.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Woz58] J. Wozencraft. List decoding. Technical Report 48:90-95, Quarterly Progress Report, Research Laboratory of Electronics, MIT, 1958.

[Zip79] R E. Zippel. Probabilistic algorithms for sparse polynomials. *Proceedings of EUROSAM*, pages 216–226, 1979.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de