Approximate Local Decoding of Cubic Reed-Muller Codes Beyond the List Decoding Radius

Pooya Hatami * Madhur Tulsiani †

October 9, 2017

Abstract

We consider the question of decoding Reed-Muller codes over $\mathbb{F}_2^n$ beyond their list-decoding radius. Since, by definition, in this regime one cannot demand an efficient exact list-decoder, we seek an approximate decoder: Given a word F and radii $r' > r > 0$, the goal is to output a codeword within radius r' of F, if there exists a codeword within distance r. As opposed to the list decoding problem, it suffices here to output any codeword with this property, since the list may be too large if r exceeds the list decoding radius.

Prior to our work, such decoders were known for Reed-Muller codes of degree 2, due to works of Wolf and the second author [FOCS 2011]. In this work we make the first progress on this problem for the degree 3 where the list decoding radius is $1/8$. We show that there is a constant $\delta = 1/2 - \sqrt{1/8} > 1/8$ and an efficient approximate decoder, that given query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$, such that F is within distance $r = \delta - \varepsilon$ from a cubic polynomial, runs in time polynomial in n and outputs with high probability a cubic polynomial which is at distance at most $r' = 1/2 - \varepsilon'$ from F, where ε' is a quasi polynomial function of ε.

*UT Austin pooyahat@gmail.com. Part of this work was conducted at DIMACS and partially enabled through support from the National Science Foundation under grant number CCF-1445755

†Toyota Technological Institute at Chicago. madhurt@ttic.edu. Work supported by NSF CCF-1254044

1 Introduction

Reed-Muller (RM) codes were discovered by Muller in 1954 and are one of the most well-studied families of codes in computer science. Let $\mathbb{F} = \mathbb{F}_q$ be a finite field. The codewords for the code $\mathsf{RM}_{\mathbb{F}}(n,d)$ are defined to be the evaluation tables of all polynomials in n variables with degree at most d, over the field $\mathbb{F}$. The length of each codeword is thus equal to q^n. If $d = a(q-1)+b$ for $0 \leqslant b < q-1$, then the *fractional Hamming distance* between any two codewords is at least

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

which is called the minimum distance of the code. In this paper, we will restrict our attention to RM codes over the field $\mathbb{F}_2$. We will refer to these codes as $\mathsf{RM}(n,d)$ and the distance as $\delta_d = 2^{-d}$.

Given the central importance of low-degree polynomials in theoretical computer science, Reed-Muller codes have found many applications in pseudorandomness [STV01, Tre03, TSZS06, SU05], cryptography [GL89, AGS03], learning theory [KM93, Jac97] and hardness of approximation [BGH+15]. We refer the reader to the excellent surveys [Tre04, V+12] for details and references to many other applications.

In several applications of low-degree RM codes, the parameter of interest is the number of variables n, and the block-length q^n is thought of as being of exponential size. A very useful property of RM codes here is that (for certain error-regimes), they are *locally decodable* i.e., it is possible to recover the polynomial from a possibly corrupted evaluation, in time $n^{O(d)}$, which is only poly-logarithmic in the block-length. The algorithms for local decoding are necessarily randomized (as they do not have time to read the entire codeword) and have query access to the codeword. This is the notion of decoding we will be concerned with in this work.

Unique and List Decoding

As with any error correcting code, Reed-Muller codes can be uniquely decoded within an error rate equal to half the minimum distance. An efficient (in the block-length) algorithm for this was given by Reed [MS77] and it was observed by [GKZ08] that this can be extended to the local decoding setting.

For error rates larger than half the minimum distance, unique decoding is no longer possible as the number of codewords within the required radius may be more than one. In such cases, one often considers the notion of *list decoding* defined by Elias [Eli57] and Wozencraft [Woz58], and studied by Sudan [Sud00] and Guruswami [Gur01, Gur06]. The goal here is to output a small list of codewords within the given error radius.

The celebrated Goldreich-Levin theorem [GL89] gave the first *local* list-decoding algorithm for $\mathsf{RM}(n,1)$, which is also known as the Hadamard code, showing that it can be list-decoded from error rates up to $1/2 - \varepsilon$ for any $\varepsilon > 0$. This was later generalized by Goldreich, Rubinfeld and Sudan [GRS00] to $\mathsf{RM}_{\mathbb{F}}(n,1)$ for other fields. For higher degree polynomials over large fields ($\mathsf{RM}_{\mathbb{F}}(n,d)$), it was shown by Sudan, Trevisan and Vadhan [STV01] (improving on [AS03]) that they can be locally list-decoded from error rate $1 - \sqrt{2d/|\mathbb{F}|}$.

A beautiful work of Gopalan, Klivans and Zuckermann [GKZ08] generalized the Goldreich-Levin algorithm to show that $\mathsf{RM}(n,d)$ can be list decoded from error rates $\delta_d - \varepsilon = 2^{-d} - \varepsilon$ for any $\varepsilon > 0$. This was later generalized by Bhowmick and Lovett

[BL15] to show local list decoding of $\mathsf{RM}_{\mathbb{F}}(n,d)$ within error radius $\delta_{d,\mathbb{F}} - \varepsilon$ for any $\varepsilon > 0$ and fixed $|\mathbb{F}|$. Since the number of codewords at distance $\delta_{d,\mathbb{F}}$ is known to be exponential in n [KLP12], this result is optimal for local list-decoding. Thus, the distance $\delta_{d,\mathbb{F}}$ is also said to be equal to the *list decoding radius* for $\mathsf{RM}_{\mathbb{F}}(n,d)$.

Decoding beyond the list-decoding radius

Gopalan, et. al. in [GKZ08] also gave a *global* list-decoder for $\mathsf{RM}(n,d)$, which works up to the Johnson radius at twice the minimum distance, defined as

$$J(2\delta_d) \;:=\; \frac{1}{2}\cdot\left(1-\sqrt{1-4\delta_d}\right) \;=\; \frac{1}{2}\cdot\left(1-\sqrt{1-2\delta_{d-1}}\right).$$

Their algorithm runs in time polynomial in the block-length $N = 2^n$, and outputs a list of size $\mathsf{poly}(N)$. Kaufman, Lovett and Porat [KLP12] later improved this global list-decoder up to distance $2 * \delta_d = \delta_{d-1}$, twice the list-decoding radius of $\mathsf{RM}(n,d)$.

For the case when the errors are assumed to be independently random, Dumer [Dum04] gave a local-decoding algorithm for $\mathsf{RM}(n,d)$ which works up to error rates $1/2 - \varepsilon$ for any constant $\varepsilon > 0$. A remarkable sequence of recent works [SSV17, ASW15, KKM+17] have also extended this result to global decoding algorithms for the case of large d (high-rate codes) and subconstant ε.

Note that since the list size is too large beyond the list-decoding radius, in these regimes the question of local list-decoding does not make sense. Motivated by decomposition theorems from higher-order Fourier analysis, [TW14] considered the following weaker problem for $\mathsf{RM}(n,2)$

Given query access to $F : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $\delta(F, \mathsf{RM}(n,2)) \leqslant 1/2 - \varepsilon$, find a codeword of $\mathsf{RM}(n,2)$ within distance $1/2 - \varepsilon'$ of F.

Note that since the list-decoding radius for degree 2 is $1/4$, the entire list of such codewords may be exponentially large as a function of n. However, the problem above only asks for *any one* such codeword. Moreover, the radius $1/2 - \varepsilon'$ for the output codeword is allowed to be different (larger) than the promised distance. Building on the work of Samorodnitsky [Sam07], who gave an algorithm for the testing version of the above problem, they gave a local algorithm for the *approximate* decoding question above, with $\varepsilon' = \exp(-1/\varepsilon^{O(1)})$. This was later improved to $\varepsilon' = \exp(-O((\log(1/\varepsilon))^4))$ by Ben-Sasson et al. [BSRZTW14].

Our Results

We extend the above result to the case of $d = 3$. In particular, we prove the following:

Theorem 1.1 (Main Theorem)**.** *There is an efficient algorithm, such that given query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ with the guarantee that*

$$\delta(F, \mathsf{RM}(n,3)) \;\leqslant\; \frac{1}{2} - \sqrt{\frac{1}{8}} - \varepsilon \;=\; J(1/4) - \varepsilon \;=\; J(2\cdot\delta_3) - \varepsilon\,,$$

runs in time $\mathsf{poly}(n, \exp((\log(1/\varepsilon))^{O(1)}), \log(1/\eta))$, and returns with probability at least $1-\eta$, a cubic polynomial $\tilde{P}$ such that $\delta(F,\tilde{P}) \leqslant 1/2 - \varepsilon'$ for $\varepsilon' = \exp\left(-(\log(1/\varepsilon))^{O(1)}\right)$.

Note that since the result of [TW14] can be thought of as starting with distance $J(2\delta_2) - \varepsilon$ ($J(2\delta_2) = 1/2$) the above is a direct generalization of the approximate local decoding result of [TW14]. The bound $J(2\delta_d)$ also arises naturally in our analysis (for all d) as described below, and in the results of [GKZ08].

The results on approximate local decoding (and testing) for the case of $d = 2$ rely heavily on the connection between quadratic polynomials and Gowers' U^3 uniformity norm (see Section 2 for the definition). In fact, the decoding algorithms are simply the algorithmic versions of the proofs of the inverse theorem for the U^3 norm by Green and Tao [GT08] and Samorodnitsky [Sam07]. Since it is known that,

$$\|(-1)^F\|_{U^3} \geqslant \sup_{\deg(P)\leqslant 2} \left\langle (-1)^F, (-1)^P \right\rangle = \sup_{\deg(P)\leqslant 2} \left| \mathop{\mathbf{E}}_{x\in\mathbb{F}_2^n} (-1)^{P-F} \right|,$$ [1]

the local decoding algorithm for $d = 2$ uses $\delta(F, \mathsf{RM}(n,2)) \leqslant 1/2 - \varepsilon$ to conclude $\|(-1)^F\|_{U^3} \geqslant \varepsilon$ and then passes entirely to reasoning in terms of the Gowers norm.

However, using this approach to extend the approximate local decoding results for $d = 3$ is problematic: the direct generalization of inverse conjecture to the higher-degree U^4 norms and cubic polynomials is known to be false [LMS08, GT09] and the proofs of the correct generalizations [GTZ11] are not combinatorial. In fact, it is a challenging open problem to give combinatorial proofs of the higher-degree inverse theorems.

However, we show that using the guarantee that there is indeed a cubic polynomial close to F (and not just the implication that the U^4 norm of $(-1)^F$ is large) one can indeed utilize techniques from the U^3 inverse theorem to obtain the above approximate decoding result. This indicates that the problem of approximate local decoding for RM codes may be easier than the problem of giving algorithmic or combinatorial versions of the inverse theorems for Gowers norms.

1.1 Overview of Proofs and Techniques

We first remark that instead of giving an algorithm that works with high-probability, it is sufficient to give an algorithm which outputs the desired polynomial Q with a small constant (depending on ε) probability. Since $\delta(F, Q)$ for the output Q can be estimated efficiently using sampling, we can simply repeat the algorithm using independent randomness to boost the success probability. In the rest of the overview, we only describe an algorithm that succeeds with some small probability.

Proofs for $d = 2$. All proofs of the inverse theorem for the U^3 norm (and the results on approximate local decoding) start with the following observation: If $P : \mathbb{F}_2^n \to \mathbb{F}_2$ is a degree-2 polynomial then the derivative

$$D_h P(x) := P(x+h) - P(x)$$

is a degree-1 polynomial in x of the form $\ell_h(x) + c_h$. Moreover, the map $\psi : h \mapsto \ell_h$ is a homomorphism (linear map) from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. The proofs show that this intuition can be used even if one is given access to $F : \mathbb{F}_2^n \to \mathbb{F}_2$ with $\delta(F, P) \leqslant 1/2 - \varepsilon$. In this case, one

[1] Also referred to as the direct theorem for U^3 norm.

uses the conclusion on the Gowers U^3 norm of $f = (-1)^F$ and the Fourier coefficients of f to design an approximate homomorphism φ satisfying

$$\mathop{\mathbf{Pr}}_{h_1,h_2\in\mathbb{F}_2^n}[\varphi(h_1)+\varphi(h_2)=\varphi(h_1+h_2)] \;\geqslant\; \varepsilon^{O(1)}\,.$$

Various tools from arithmetic combinatorics allow one to refine this weak linear structure of φ to obtain a *linear* map $\tau : \mathbb{F}_2^n \to \mathbb{F}_2^n$ which agrees with φ in ε' fraction of the inputs. This linear map τ can then be "integrated" to obtain a quadratic Q satisfying $\delta(Q,F) \leqslant 1/2 - \varepsilon'$.

Approximate homomorphisms for $d = 3$. Note that for a polynomial P of any degree d, it is true that the map from h to the part of D_hP with degree *exactly* $d-1$ is still linear in the direction h. Thus, a possible approach to the approximate decoding question would be to construct an approximate φ as above, only given query access to F.

We start with the observation that when $\delta(F,P) \leqslant J(1/4) - \varepsilon$ for an unknown cubic polynomial P, it is indeed possible to construct such an approximate homomorphism φ, given query access to F. For the above distance, a simple application of Cauchy-Schwartz shows that

$$\mathop{\mathbf{E}}_{h\in\mathbb{F}_2^n}[\delta]\,(D_hF, D_hP) \;\leqslant\; \frac{1}{4} - \Omega(\varepsilon^2)\,.$$

In fact, for any d, when $\delta(F,P) \leqslant J(2\delta_d)$ for $\deg(P) \leqslant d$, the above gives $\mathbf{E}_{h\in\mathbb{F}_2^n}[\delta]\,(D_hF, D_hP) \leqslant \delta_{d-1} - \Omega(\varepsilon^2)$. This suggests an immediate way to construct φ which "guesses" the degree-$(d-1)$ part of D_hP: we simply apply the local list-decoding algorithm of [GKZ08] to D_hF and output the degree-$(d-1)$ part of a random polynomial from the list. The list size in [GKZ08] is $\mathrm{poly}(1/\varepsilon)$ for error-rate $\delta_d - \varepsilon^{O(1)}$. Thus, with probability at least $\varepsilon^{O(1)}$, we pick the degree-$(d-1)$ part of the "true" P from the list for points h_1, h_2, h_3, h_4 satisfying $h_1 + h_2 = h_3 + h_4$, which gives

$$\mathop{\mathbf{Pr}}_{\substack{\varphi\\ h_1+h_2=h_3+h_4}}[\varphi(h_1)+\varphi(h_2)=\varphi(h_3)+\varphi(h_4)] \;\geqslant\; \varepsilon^{O(1)}\,.$$

One can then refine this approximate homomorphism to obtain an affine map τ having significant agreement with φ. However, we need some extra properties to ensure τ can be integrated to a polynomial Q.

Symmetrization and integration For the case of $d = 2$, the map $h \mapsto \ell_h$ for a "true" quadratic polynomial P also has some symmetry properties in addition to being linear. If we think of $\ell(h)$ as given by Mh for a matrix M, then one can show that this matrix is symmetric and has zero-diagonal. The arithmetic combinatorics tools which refine φ to τ are not sensitive to these symmetry properties and one needs to modify the final τ to obtain these symmetries, before it can be integrated to a polynomial. For the case of $d = 2$ this argument is heavily Fourier analytic in nature and essentially uses the fact that row-rank of a matrix equals its column-rank.

For the case of $d = 3$, one can think of the map from h to a degree-2 polynomial as given by a tensor T, where the quadratic polynomial for D_hP is given by (say) $T(h,x,x)$. Then, to be able to integrate, the tensor needs to satisfy the symmetry property

$$T(h,x,x) \;=\; T(x,h,x) \;=\; T(x,x,h)\,.$$

However, as the arguments for the quadratic case do not carry over here, we need to use a different symmetry argument (inspired by the [GT08] proof of the U^3 inverse theorem) that we indeed have a "true" cubic polynomial P close to F. We denote the above map for P by ψ. Since the map ψ must satisfy all required symmetry properties, we can use it to show that if the map τ found by our algorithm has some agreement with ψ, then τ can also be (locally) symmetrised on the inputs where they agree. However, this poses a challenge since ψ is unknown and we need the guarantee that the homomorphism τ given by our algorithm to agree with this "hidden" homomorphism.

Decoding approximate and hidden homomorphisms We now return to the ideas from arithmetic combinatorics which refine the approximate homomorphism φ to obtain a linear map τ, which has significant agreement with τ. We additionally require that they agree on a significant fraction the "hidden" set

$$H \;=\; \{h \;\mid\; h \in B \text{ and } \varphi(h) = \psi(h)\}\;.$$

Here, B is also an unknown (large) set where D_hP is in the list of polynomials close to D_hF. We show that the algorithms from [TW14] and [BSRZTW14] can be modified to ensure agreement at least ε' on the set H, with probability at least ε', for $\varepsilon' = \exp\left(-(\log(1/\varepsilon))^{O(1)}\right)$. The argument in [TW14], was stated for a very special approximate homomorphism φ arising out of the Fourier coefficients of $f = (-1)^F$. Here, we give a more general version of the argument which is also sensitive to the hidden set H.

We remark that the parts about finding the approximate homomorphism φ and refining it to an affine τ can be carried out for all degrees d. However, the integration argument requires some steps which are currently only available for $d = 3$.

2 Preliminaries and Notation

Throughout the paper, we shall be using Latin letters such as x, y or z to denote elements of $\mathbb{F}_2^n$. Let $N := |\mathbb{F}_2^n| = 2^n$, and let $\delta_d = 1/2^d$ be the list-decoding radius of Reed-Muller codes of degree d over $\mathbb{F}_2^n$. Denote by e_k the vector with all entries equal to 0 except the kth entry equal to 1. For two functions $F, G : \mathbb{F}_2^n \to \{0,1\}$, define the distance between F and G to be $\delta(F,G) := \mathbf{Pr}_{x\in\mathbb{F}_2^n}(F(x) \neq G(x))$ and the correlation between F and G to be $\mathrm{Corr}(F,G) = |\mathbf{E}_{x\in\mathbb{F}_2^n}(-1)^{F(x)+G(x)}|$.

For any variable m, we denote by $\mathsf{poly}(m)$ a function that is of the form $O(m^{O(1)})$, and by $\mathsf{quasipoly}(m)$ a function of the form $O(2^{\log(m)^{O(1)}})$.

Given a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$, $\deg(F)$ denotes the degree of the unique multilinear polynomial that agrees with F. When a function F depends on disjoint sets of variables $x, y, \ldots$ we denote by $\deg_x(F)$ to be the maximum degree of F for any fixing of non-x variables.

For $F : \mathbb{F}_2^n \to \mathbb{F}_2$, define its additive derivative at direction $y \in \mathbb{F}_2^n$ at x to be $D_yF(x) := F(x+y) - F(x)$. For $f : \mathbb{F}_2^n \to \mathbb{R}$ define its multiplicative derivative at direction $y \in \mathbb{F}_2^n$ at x to be $\Delta_y f(x) := f(x)f(x+y)$. We correspond $\mathbb{F}_2$ with $\{0,1\}$ in the obvious manner. Thus, for example for $P : \mathbb{F}_2^n \to \mathbb{F}_2$ we have

$$\Delta_y(-1)^{P(x)} = (-1)^{D_yP(x)}.$$

A useful fact about the additive derivative is that if $\deg(P) = d$, then $\deg(D_y P) \leqslant d-1$ for any $y \in \mathbb{F}_2^n$.

2.1 Sampling

We shall be using the following standard probabilistic bounds without further mention.

Lemma 2.1 (Hoeffding bound for sampling)**.** *If $\mathbf{X}$ is a real valued random variable with $|\mathbf{X}| \leqslant 1$ and $\widehat{\mu}$ is the empirical average obtained from t samples, then*

$$\mathbf{Pr}\left[|\mathbf{E}\left[\mathbf{X}\right] - \widehat{\mu}| \ > \ \gamma\right] \ \leqslant \ \exp(-\Omega(\gamma^2 t)).$$

A Hoeffding-type bound can also be obtained for polynomial functions of ± 1-valued random variables.

Lemma 2.2 (Hoeffding bound for low-degree polynomials [O'D08])**.** *Suppose that $\mathbf{F} = \mathbf{F}(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ is a polynomial of degree d in random variables $\mathbf{X}_1, \ldots, \mathbf{X}_N$ taking value ± 1, then*

$$\mathbf{Pr}\left[|\mathbf{F} - \mathbf{E}\left[\mathbf{F}\right]| \ > \ \gamma\right] \ \leqslant \ \exp\left(-\Omega\left(d \cdot (\gamma/\sigma)^{2/d}\right)\right),$$

where $\sigma = \sqrt{\mathbf{E}\left[\mathbf{F}^2\right] - \mathbf{E}\left[\mathbf{F}\right]^2}$ is the standard deviation of $\mathbf{F}$.

2.2 Results from Additive Combinatorics

For a set $A \subseteq \mathbb{F}_2^n$, we write $A + A$ for the set of elements $a + a'$ such that $a, a' \in A$. More generally, the k-fold *sumset*, denoted by kA, consists of all k-fold sums of elements of A.

The Balog-Szemerédi-Gowers theorem, a fundamental result in additive combinatorics, states that if a set contains many additive quadruples, that is, elements a_1, a_2, a_3, a_4 such that $a_1 + a_2 = a_3 + a_4$, then a large subset of it must have small sumset.

Theorem 2.3 (Balog-Szemerédi-Gowers [Gow98])**.** *Suppose that $A \subseteq \mathbb{F}_2^n$ contains at least $|A|^3/K$ additive quadruples. Then there exists a subset $A' \subseteq A$ of size $|A'| \geqslant K^{-C}|A|$ with the property that $|A' + A'| \leqslant K^C |A'|$.*

We will also need the following version of Plünnecke's inequality.

Lemma 2.4 (Plünnecke's Inequality)**.** *Let $B \subseteq \mathbb{F}_2^n$ be such that $|B + B| \leqslant K|B|$ for some $K > 1$. Then for any positive integer k, we have $|kB| \leqslant K^k |B|$.*

We shall also require the notion of a *Freiman homomorphism*. The map l is said to be a Freiman 2-homomorphism if $x + y = z + w$ implies $l(x) + l(y) = l(z) + l(w)$. More generally, a Freiman homomorphism of order k is a map l such that $x_1 + x_2 + \cdots + x_k = x_1' + x_2' + \cdots + x_k'$ implies that $l(x_1) + \cdots + l(x_k) = l(x_1') + \cdots + l(x_k')$. The order of the Freiman homomorphism measures the degree of linearity of l; in particular, a truly linear map is a Freiman homomorphism of all orders.

2.3 Gowers Uniformity Norms and Inverse Theorems.

Here we recall the definition of Gowers uniformity norms introduced by Gowers in [Gow98].

Definition 2.5. *Let G be any finite abelian group. For any positive integer $k \geqslant 2$ and any function $f : G \to \mathbb{C}$, define the U^k-*norm *by the formula*

$$\|f\|_{U^k}^{2^k} = \mathop{\mathbf{E}}_{x,h_1,\ldots,h_k \in G} \prod_{\omega \in \{0,1\}^k} C^{|\omega|} f(x + \omega \cdot h),$$

where $\omega \cdot h$ is shorthand for $\sum_i \omega_i h_i$, and $C^{|\omega|} f = f$ if $\sum_i \omega_i$ is even and $\overline{f}$ otherwise.

In the special case $k = 2$, a computation shows that

$$\|f\|_{U^2} = \|\widehat{f}\|_{l^4}, \tag{1}$$

and hence any argument using the U^2 norm is essentially equivalent to using ordinary Fourier analysis. In the case $k = 3$, the U^3 norm counts the number of additive octuples "contained in" f, that is, we average over the product of f at all eight vertices of a 3-dimensional parallelepiped in G.

These uniformity norms satisfy a number of important properties: they are clearly nested

$$\|f\|_{U^2} \leqslant \|f\|_{U^3} \leqslant \|f\|_{U^4} \leqslant \ldots$$

and can be defined inductively

$$\|f\|_{U^{k+1}} = \left(\mathop{\mathbf{E}}_x \|f_x\|_{U^k}^{2^k}\right)^{1/2^{k+1}}, \tag{2}$$

where $k \geqslant 2$ and the function f_x stands for the assignment $f_x(y) = f(y)\overline{f(x+y)}$. Thinking of the function f as a complex exponential (a phase function), we can interpret the function f_x as a kind of *discrete derivative* of f.

It follows straight from a simple but admittedly ingenious sequence of applications of the Cauchy-Schwarz inequality that if the balanced function $1_A - \alpha$ of a set $A \subseteq G$ of density α has small U^k norm, then A contains the expected number of arithmetic progressions of length $k+1$, namely $\alpha^{k+1}|G|^2$. This fact makes the uniformity norms interesting for number-theoretic applications.

In computer science they have been used in the context of probabilistically checkable proofs (PCP) [ST06], communication complexity [VW07], as well as in the analysis of pseudo-random generators that fool low-degree polynomials [BV10].

In many applications, being small in the U^k norm is a desirable property for a function to have. What can we say if this is not the case? It is not too difficult to verify that $\|f\|_{U^k} = 1$ if and only if f is a polynomial phase function of degree $k-1$, i.e. a function of the form $\omega^{p(x)}$ where p is a polynomial of degree $k-1$ and ω is an appropriate root of unity. But does every function with large U^k norm look like a polynomial phase function of degree $k-1$?

It turns out that any function with large U^k norm correlates with a polynomial phase function of degree $k-1$. This is known as the inverse theorem for the U^k norm, proved by Green and Tao [GT08] for $k = 3$ and $p > 2$, Samorodnitsky [Sam07] for $k = 3$ and $p = 2$, and Bergelson, Tao and Ziegler [BTZ10, TZ10] for $k > 3$ (where in the latter case the above definition of a "polynomial phase function" needs to be slightly modified). We shall restrict our attention to the case $k = 3$ in this paper, which we can state as follows.

Theorem 2.6 (Global Inverse Theorem for U^3 [GT08], [Sam07]). *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^3} \geqslant \varepsilon$. Then there exists a quadratic form q and a vector b such that*

$$|\mathop{\mathbf{E}}_x f(x)\omega^{q(x)+b\cdot x}| \geqslant \exp(-O(\varepsilon^{-C})).$$

3 Proof of The Main Theorem

Theorem 1.1 (restated). *There is an efficient algorithm, such that given query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ with the guarantee that*

$$\delta(F, \mathsf{RM}(n,3)) \;\leqslant\; \frac{1}{2} - \sqrt{\frac{1}{8}} - \varepsilon \;=\; J(1/4) - \varepsilon \;=\; J(2\cdot\delta_3) - \varepsilon\,,$$

runs in time $\mathsf{poly}(n, \exp((\log(1/\varepsilon))^{O(1)}), \log(1/\eta))$, and returns with probability at least $1-\eta$, a cubic polynomial $\tilde{P}$ such that $\delta(F,\tilde{P}) \leqslant 1/2 - \varepsilon'$ for $\varepsilon' = \exp\left(-(\log(1/\varepsilon))^{O(1)}\right)$.

3.1 Random Derivatives

The first step of the proof relies on the observation that if F is sufficiently close to a cubic polynomial P, then the distance of D_yF from D_yP averaged over y is small. We prove this more generally for any degree d in the next lemma.

Lemma 3.1. *For $\varepsilon > 0$, there is $\varepsilon'(\varepsilon, d) \geqslant \varepsilon^2$ such that the following holds. Let $F, P : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $\delta(F,P) \leqslant \frac{1}{2}(1 - \sqrt{1-2\delta_{d-1}}) - \varepsilon$, then*

$$\mathop{\mathbf{E}}_{y\in\mathbb{F}_2^n}\left[\delta(D_yF, D_yP)\right] \;\leqslant\; \delta_{d-1} - \varepsilon'\,.$$

Note that the bound on $\delta(F,P)$ is the well-known Johnson bound applied to $2\delta_d$ which is strictly larger than δ_d, and equals $\frac{1}{2} - \sqrt{\frac{1}{8}} - \varepsilon$ when $d = 3$ as in the statement of Theorem 1.1. The above lemma is a special case of the following claim by choosing $1-\alpha = \frac{1}{2}\cdot\left(1-\sqrt{1-2\delta_{d-1}}\right) - \varepsilon$.

Claim 3.2. *Let $F, P : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $\delta(F,P) \leqslant 1-\alpha$, where we assume $\alpha > 1/2$. Then we have*

$$\mathop{\mathbf{E}}_{y\in\mathbb{F}_2^n}\left[\delta(D_yf, D_yP)\right] \;\leqslant\; 2\alpha(1-\alpha).$$

Proof. Let $A := \{x : f(x) = P(x)\}$, hence $|A| \geqslant \alpha 2^n$. We are interested in

$$\begin{aligned}
&1 - \mathop{\mathbf{E}}_{y\in\mathbb{F}_2^n}\left[\mathop{\mathbf{Pr}}_x\left(F(x+y)+F(x) = P(x+y)+P(x)\right)\right]\\
=\;& 1 - \mathop{\mathbf{E}}_y \frac{1}{2^n}\left(|A\cap(A+y)| + |\overline{A}\cap(\overline{A}+y)|\right)\\
\leqslant\;& 1 - \alpha^2 - (1-\alpha)^2.
\end{aligned}$$

□

Suppose that $\delta(F,P) < \frac{1}{2} - \sqrt{\frac{1}{8}} - \varepsilon$ for some $\varepsilon > 0$. Pick $\varepsilon'(\varepsilon,3)$ such that by Claim 3.2 $\mathbf{E}_{y\in\mathbb{F}_2^n}\,\delta(D_yF,D_yP) \leqslant \frac{1}{4} - \varepsilon'$. Now define

$$B := \left\{ y \in \mathbb{F}_2^n \;\mid\; \delta(D_yf, D_yP) \leqslant \frac{1}{4} - \frac{\varepsilon'}{2} \right\}.$$

We call directions $y \in B$ as good directions for F. We will need the fact that the set B is large and contains many 4-tuples y_1, y_2, y_3, y_4 such that $y_1 + y_2 = y_3 + y_4$ (has high additive energy).

Claim 3.3. *Let B be as defined above. Then, $|B| \geqslant 3\varepsilon' \cdot 2^n$, and*

$$\mathop{\mathbf{Pr}}_{y_1+y_2=y_3+y_4}\left[\bigwedge_{i=1}^{4}(y_i \in B)\right] \;\geqslant\; \left(3\varepsilon'\right)^4.$$

Proof. By Markov's inequality we have that

$$1 - \frac{|B|}{2^n} \;=\; \mathop{\mathbf{Pr}}_{y\in\mathbb{F}_2^n}\left[\delta(D_yf,D_yP) > \frac{1}{4} - \frac{\varepsilon'}{2}\right] \;\leqslant\; \frac{\frac{1}{4}-\varepsilon'}{\frac{1}{4}-\frac{\varepsilon'}{2}} \;=\; 1 - \frac{\varepsilon'/2}{\frac{1}{4}-\frac{\varepsilon'}{2}} \;\leqslant\; 1 - 3\varepsilon'.$$

Thus, we have that $\mathbf{E}_{y\in\mathbb{F}_2^n}\left[\mathbb{1}_B(y)\right] \geqslant 3\varepsilon'$. To bound the additive energy, note that

$$\mathop{\mathbf{E}}_{y_1+y_2=y_3+y_4}\left[\prod_{i=1}^{4}\mathbb{1}_B(y_i)\right] \;=\; \sum_{\beta\in\mathbb{F}_2^n}\left(\widehat{\mathbb{1}_B}(\beta)\right)^4 \;\geqslant\; \left(\widehat{\mathbb{1}_B}(0)\right)^4 \;\geqslant\; \left(3\varepsilon'\right)^4.$$

□

3.2 The Quadratic Part of the Derivative

The set B as defined above has the nice property that for $y \in B$ we have $\delta(D_yF, D_yP) < \delta_{d-1} - \frac{\varepsilon'}{2}$; this means that D_yP will be in the list of D_yF at radius $\delta_{d-1} - \frac{\varepsilon'}{2}$, and the list-size will be a constant determined by ε. We will use the local list-decoding algorithm of Gopalan, et. al. [GKZ08] in order to guess D_yP.

Theorem 3.4 ([GKZ08])**.** *There is an algorithm that given $d > 0$, and $\varepsilon, \eta > 0$ and query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$, runs in time* $\mathsf{poly}(1/\varepsilon, \log(1/\eta), n)$ *and with probability $1 - \eta$, returns a list of size* $\mathsf{poly}(1/\varepsilon)$ *containing all degree d polynomials P such that $\delta(F,P) \leqslant \delta_d - \varepsilon$.*

Using the above algorithm, we can provide query access to a random map $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$, defined as follows for a given vector $h \in \mathbb{F}_2^n$ as input:

1. Efficiently list-decode D_hF up to distance $\frac{1}{4} - \frac{\varepsilon'}{2}$ and let $\mathcal{L}_h$ denote the set of all quadratic polynomials on the list.
2. If $\mathcal{L}_h = \emptyset$ then let $\varphi(h) := \bar{0}$. Note that all such directions must belong to $\mathbb{F}_2^n \backslash B$, assuming the $\mathcal{L}_h$ is the correct list.
3. If $\mathcal{L}_h \neq \emptyset$, we pick a quadratic $Q_y \in_R \mathcal{L}_h$, uniformly at random and let $\varphi(y)$ be the upper-triangular matrix defined by $\varphi(y)_{ij}$ being the coefficient of x_ix_j in $Q_y(x)$, when $i < j$. In other words, $\varphi(y)$ represents the quadratic part of Q_y.

From here on, we will condition on the list-decoder succeeding for every single time we run the decoder. We are able to do so via a union bound, as the running time of the algorithm in Theorem 3.4 has polylogarithmic dependence on the error probability. Denote by $\psi(h)$ the upper-triangular matrix representing the quadratic part of $D_h P(x)$. We say $h \in \mathbb{F}_2^n$ is *lucky* if $h \in B$ and $\varphi(h) = \psi(h)$. Note that for every $h \in B$, $D_h P \in \mathcal{L}_h$ and since $|\mathcal{L}_h| \leqslant \mathsf{poly}(1/\varepsilon)$,

$$\forall h \in B \qquad \mathop{\mathbf{Pr}}_{\varphi}[h \text{ is lucky}] \;=\; \mathop{\mathbf{Pr}}_{\varphi}[\varphi(h) = \psi(h)] \;\geqslant\; \mathsf{poly}(\varepsilon)\,.$$

Given a map φ, define $G_\varphi := \{(x, \varphi(x)) : x \in \mathbb{F}_2^n\}$, thus $|G_\varphi| \leqslant 2^n$. Note that the map $\psi : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ is a homomorphism since the quadratic part of $D_h P$ is a linear function of h. Thus, φ is an approximate homomorphism on h that are lucky. This is captured by the following claim.

Claim 3.5. *Let φ be the random map as defined above. Then*

$$\mathop{\mathbf{E}}_{\varphi, h_1+h_2=h_3+h_4}\left[\mathbb{1}_{(h_1,\varphi(h_1))+(h_2,\varphi(h_2))=(h_3,\varphi(h_3))+(h_4,\varphi(h_4))} \cdot \mathbb{1}_{h_1,\ldots,h_4 \text{ are lucky}}\right] \;\geqslant\; \mathsf{poly}(\varepsilon)\,.$$

Proof. By Claim 3.3 and Lemma 3.1, we have that $h_1, \ldots, h_4 \in B$ with probability at least $\varepsilon^{O(1)}$. Also, each h_i is lucky (independently) with probability $\varepsilon^{O(1)}$ over the randomness in the choice of φ. Since the map ψ is a homomorphism, in this case, we have that

$$\varphi(h_1) + \varphi(h_2) \;=\; \psi(h_1) + \psi(h_2) \;=\; \psi(h_3) + \psi(h_4) \;=\; \varphi(h_3) + \varphi(h_4)\,,$$

which proves the claim. □

Applying Markov inequality to the conclusion of Claim 3.5, we obtain the following.

Theorem 3.6. *Given query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ that is $J(2\delta_3) - \varepsilon$ close to a cubic polynomial P. There is a distribution $\mathcal{D}$ on functions $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ which satisfies the following.*

- *For every h, $\varphi(h)$ is selected independently at random, and this query can be computed in $\mathsf{poly}(n, \varepsilon)$ time.*
- *Defining B as above, for every $h \in B$, $\mathbf{Pr}_{\varphi\sim\mathcal{D}}(\varphi(h) = \psi(h)) \geqslant \mathsf{poly}(\varepsilon)$.*
- *There is a constant $c = \varepsilon^{O(1)}$ such that*

$$\mathop{\mathbf{Pr}}_{\varphi}\left[\mathop{\mathbf{Pr}}_{h_1+h_2=h_3+h_4}\left[\begin{array}{c} h_1, \ldots, h_4 \text{ are lucky, and} \\ (h_1, \varphi(h_1)) + (h_2, \varphi(h_2)) = (h_3, \varphi(h_3)) + (h_4, \varphi(h_4)) \end{array}\right] \geqslant c\right] \geqslant c.$$

3.3 Finding a Homorphism

Theorem 3.7 (Hidden Homomorphism Decoding)**.** *There is an algorithm that given query access to a random map $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ given by Theorem 3.6, runs in time $O_\varepsilon(n^{O(1)})$ and with probability $\mathsf{quasipoly}(\varepsilon)$ (over the randomness of the algorithm and φ) outputs an affine map $\tau : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ such that*

$$\mathop{\mathbf{Pr}}_{h}[\tau(h) = \varphi(h) = \psi(h) \text{ and } h \in B] \;\geqslant\; \mathsf{quasipoly}(\varepsilon)\,.$$

3.4 Finding a Correlating Cubic

By Theorem 3.7 we have found an affine map $\tau : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ such that

$$\mathop{\mathbf{Pr}}_{h}\left[\tau(h) = \varphi(h) = \psi(h) \text{ and } h \in B\right] \geqslant \mathsf{quasipoly}(\varepsilon). \tag{3}$$

Let $Q(x,x) := x^t\tau(0)x$, and denote by $\tau' := \tau - \tau(0)$. Since, τ' is a linear map, it naturally corresponds to an $n \times n \times n$ tensor which we henceforth denote by T. Define

$$S := \{h \in \mathbb{F}_2^n \mid \tau(h) = \psi(h)\} = \left\{h \mid \tau'(h) + \psi(h) = \tau(0)\right\}.$$

Noting that $S \supseteq \{h \mid \tau(h) = \varphi(h) = \psi(h) \text{ and } h \in B\}$, Eq. (3) implies that $|S| \geqslant \mathsf{quasipoly}(\varepsilon) \cdot 2^n$. Moreover, since both τ' and ψ are linear maps, S is an affine subspace of $\mathbb{F}_2^n$.

We take the input to τ' to be the first mode of the tensor i.e., τ' is the quadratic polynomial given by $T(h,x,x)$. We observe that T must obey the following "symmetries".

Claim 3.8. *For every $a \in \mathbb{F}_2^n$ and $y \in S$, there exists an affine form $\ell_{y,a}$ such that for every $x \in S + a$,*

$$T(x,y,x) + T(x,x,y) + T(y,x,x) = T(y,x,x) + \ell_{y,a}(x).$$

Proof. By the definition of S, for every $h \in S$ there exists an affine map $\ell'_h(x)$ such that

$$x^t\psi(h)x = D_hP(x) = T(h,x,x) + Q(x,x) + \ell'_h(x).$$

Thus deriving along x at directions y_1 and y_2 we get that for every $h \in S$ there exists an affine $\ell'_h(x)$ such that

$$(\forall y_1, y_2 \in \mathbb{F}_2^n)\;\; D_{y_1}D_{y_2}\ell'_h(x) + D_{y_1}D_{y_2}T(h,x,x) + D_{y_1}D_{y_2}Q(x,x) = D_{y_1}D_{y_2}D_hP(x).$$

which since $D_{y_1}D_{y_2}\ell'_h(x) \equiv 0$ simplifies to

$$T(h,y_1,y_2) + T(h,y_2,y_1) + Q(y_1,y_2) + Q(y_2,y_1) = D_{y_1}D_{y_2}D_hP(x) \tag{4}$$

By commutativity of the derivative operator, we get that, for every $h, y_1, y_2 \in S$

$$T(h,y_1,y_2) + T(h,y_2,y_1) + Q(y_1,y_2) + q(y_2,y_1) = T(y_2,h,y_1) + T(y_2,y_1,h) + Q(h,y_1) + Q(y_1,h).$$

Thus letting $h = y$ and $y_1 = y_2 = x - a \in S$, we have

$$T(x-a,y,x-a) + T(x-a,x-a,y) + Q(x-a,y) + Q(y,x-a) = 0.$$

Observing that $\deg_x(Q(x-a,y) + Q(y,x-a)) \leqslant 1$, $deg_x(T(x,y,x) - T(x-a,y,x-a)) \leqslant 1$ and $deg_x(T(x,x,y) - T(y,x-a,x-a)) \leqslant 1$ concludes the claim. □

Note that if T' denotes the tensor corresponding to the map ψ (where $\psi(y)$ corresponds to the quadratic part of D_yP for the *true* polynomial P) then T' satisfies the above symmetry for all $x, y \in \mathbb{F}_2^n$.

We decode a cubic polynomial P' from the "symmetrised" version of T. Define a polynomial $M(h,x) := T(h,x,x) + T(h,x,h) + T(x,x,h)$, where M is linear in h and quadratic in x.

Claim 3.9. *There exists a cubic polynomial $P' : \mathbb{F}_2^n \to \mathbb{F}_2$ efficiently computable from M, such that*

$$\deg_x(D_h P'(x) - M(h,x))) \leqslant 1.$$

Namely, there exists an affine $\ell''_h(x)$ such that $D_h P'(x) = M(h,x) + \ell''_h(x)$.

Proof. Suppose that

$$M(h,x) = \sum_{i,j,k \in [n]} M_{i,j,k} h_i x_j x_k.$$

It follows from the definition of M that for every $i,j,k \in [n]$, the coefficients of $h_i x_j x_k$, $x_i h_j x_k$ and $x_k h_i x_j$ are the same; we denote this value by $A_{i,j,k}$. It is easy to check that $P'(x) := \sum_{i<j<k} A_{i,j,k} x_i x_j x_k$ satisfies the claim. □

Suppose that $h \in S$ is such that $\delta(D_h f, D_h P) \leqslant \frac{1}{4} - \varepsilon'$. By a simple averaging argument, there exists $a \in \mathbb{F}_2^n$ such that

$$\delta(D_h f|_{S+a}, D_h P|_{S+a}) \leqslant \frac{1}{4} - \varepsilon'. \tag{5}$$

Note that by Claim 3.8, $\deg_x(M(h,x) - T(h,x,x)) \leqslant 1$ for all $h \in S$ and $x \in S + a$. Thus by Claim 3.9 and Eq. (3) we can find P' such that for all $h \in S$, there exists an affine $\ell_{h,a}$ such that

$$(\forall x \in S + a) \;\; D_h P(x) - D_h P'(x) = \ell''_{h,a}(x). \tag{6}$$

Thus, over the coset $S + a$, $D_h f - D_h P'$ has distance at most $\frac{1}{4} - \varepsilon'$ from $\ell''_{h,a}$. By a standard averaging argument, this implies that there exists an affine function ℓ'_h such that

$$\mathrm{Corr}(D_h f - D_h P', \ell_h) \geqslant \frac{|S|}{2^n} \cdot \left(\frac{1}{4} + \varepsilon'\right) = \Omega(\mathsf{quasipoly}(\varepsilon)).$$

It follows that

$$\|(-1)^{D_h f - D_h P'}\|_{U^2} = \Omega(\mathsf{quasipoly}(\varepsilon)).$$

Finally, observing from Eq. (3) that the above inequality holds with probability $\mathsf{quasipoly}(\varepsilon)$ over $h \in \mathbb{F}_2^n$, it follows from Eq. (2) that

$$\|(-1)^{f-P'}\|_{U^3} = \left(\mathop{\mathbf{E}}_h \|(-1)^{D_h f - D_h P'}\|_{U^2}^4\right)^{1/8} \geqslant \mathsf{quasipoly}(\varepsilon).$$

Finally, we can use the quasipolynomial quadratic Goldreich-Levin theorem of [BSRZTW12] below, with $g := (-1)^{f-P'}$, to find a quadratic polynomial Q' such that f has correlation at least $\mathsf{quasipoly}(\mathsf{quasipoly}(\varepsilon)) = \mathsf{quasipoly}(\varepsilon)$ with $P' + Q'$.

Theorem 3.10. *There exists a randomized algorithm that given $\varepsilon, \delta > 0$ and oracle access to a function $f : \mathbb{F}_2^n \to \{-1,1\}$, runs in time $O_{\delta,\varepsilon}(n^{O(1)})$, and either outputs a quadratic $Q(x)$ or $\perp$. The algorithm satisfies the following guarantee*

- *If $\|(-1)^F\|_{U^3} \geqslant \varepsilon$, then with probability at least $1 - \delta$ it finds a quadratic form q such that $\langle f, (-1)^Q \rangle \geqslant \mathsf{quasipoly}(\varepsilon)$.*
- *The probability that the algorithm outputs a quadratic form Q with $\langle f, (-1)^Q \rangle \leqslant \frac{\eta}{2}$ is at most δ.*

3.5 Putting Things Together

Within the previous sections we have developed an algorithm that given query access to a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ with the guarantee that $\delta(F, RM(\mathbb{F}_2^n, 3)) \leqslant \frac{1}{2} - \sqrt{\frac{1}{8}} - \varepsilon$, runs in time $O_{\delta,\varepsilon}(n^{O(1)})$, and returns with probability $\mathsf{quasipoly}(\varepsilon)$ a cubic polynomial $\tilde{P} := P' + Q'$ such that $Corr(f, \tilde{P}) \geqslant \mathsf{quasipoly}(\varepsilon)$.

Applying the above algorithm $O(\frac{\log(1/\eta)}{\mathsf{quasipoly}(\varepsilon)})$, and each time using Lemma 2.1 to test whether $Corr(f, \tilde{P}) \geqslant \mathsf{quasipoly}(\varepsilon)$ given the desired algorithm in Theorem 1.1.

4 Decoding an Approximate Homomorphism

Theorem 3.7 (restated). *There is an algorithm that given query access to a random map $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ given by Theorem 3.6, runs in time $O_\varepsilon(n^{O(1)})$ and with probability $\mathsf{quasipoly}(\varepsilon)$ (over the randomness of the algorithm and φ) outputs an affine map $\tau : \mathbb{F}_2^n \to \mathbb{F}_2^{n^2}$ such that*

$$\mathop{\mathbf{Pr}}_{h}[\tau(h) = \varphi(h) = \psi(h) \text{ and } h \in B] \;\geqslant\; \mathsf{poly}(\varepsilon)\,.$$

In this section, we start with a (randomized) function $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^m$, which has significant agreement with an *unknown* homomorphism $\psi : \mathbb{F}_2^n \to \mathbb{F}_2^m$. In particular, we have that for a set B of size at least (say) $\varepsilon \cdot 2^n$, we have that

$$\forall x \in B \qquad \mathop{\mathbf{Pr}}_{\varphi}[\varphi(x) = \psi(x)] \;\geqslant\; \varepsilon\,. \tag{7}$$

This also implies that φ is an approximate homomorphism, satisfying

$$\mathop{\mathbf{Pr}}_{\varphi}\;\mathop{\mathbf{Pr}}_{x_1+x_2=x_3+x_4}[\varphi(x_1) + \varphi(x_2) = \varphi(x_3) + \varphi(x_4)] \;\geqslant\; \varepsilon^8\,. \tag{8}$$

We φ refine it to an *affine* map τ_φ, which has a non-trivial agreement with φ, and also with the "hidden" homomorphism ψ. Note that we do not have any access to ψ, except through the guarantee in Eq. (7). However, the output τ is still required to have a significant agreement with ψ, at least with a small constant probability (depending on ε).

Also note that since φ is randomized, τ_φ is actually a random variable which depends on the choice of φ. We will give an algorithmic procedure for computing τ_φ for a given φ. We assume that $\varphi(x)$ is chosen independently at random for each $x \in \mathbb{F}_2^n$. This allows us to fix a φ and give a procedure for computing τ_φ by querying $\varphi(x)$ at a small number of inputs. This can be easily combined with a random choice of φ since the value of $\varphi(x)$ at each of the query points can be generated independently. For the rest of the section, we fix the randomness in the choice of φ and take it to be a fixed deterministic function satisfying

$$\mathop{\mathbf{Pr}}_{x_1+x_2=x_3+x_4}[\varphi(x_1) + \varphi(x_2) = \varphi(x_3) + \varphi(x_4)] \;\geqslant\; \varepsilon^8 \quad \& \quad \mathop{\mathbf{Pr}}_{x\in B}[\varphi(x) = \psi(x)] \;\geqslant\; \varepsilon\,.$$

We will also consider the (unknown) set H defined as

$$H \;:=\; \{x \in B \;\mid\; \varphi(x) = \psi(x)\}\,.$$

By the previous assumptions, we have $|H| \geqslant \varepsilon^2 \cdot 2^n$. The map τ (for which we will suppress the dependence on φ) will still be random as the points at which φ is queried will be chosen randomly.

The procedure described in this section (for finding an affine map τ that has significant agreement with φ) is implicit in the works on quadratic Goldreich-Levin theorems [TW14, BSRZTW14], where it was described for a special φ based on the Fourier coefficients of a given function. Here, we describe the procedure for any approximate homomorphism φ. We also show that (a slight modification of) this procedure ensures a significant agreement with the hidden homomorphism ψ, with at least a small constant probability.

4.1 Reduction to the Case of Approximate Linearity

Given a map $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^m$ as above, we modify it to obtain a map φ' which satisfies $\mathbf{Pr}_{x,y}\left[\varphi'(x) + \varphi'(y) = \varphi'(x+y)\right] \geqslant \varepsilon$. For a given $z \in \mathbb{F}_2^n$, define

$$\varphi_z(x) \;:=\; \varphi(x+z) + \varphi(z)\,.$$

Note that if $z \in H$, then for all $x \in H + z$, we have

$$\varphi_z(x) \;=\; \varphi(x+z) + \varphi(z) \;=\; \psi(x+z) + \psi(z) \;=\; \psi(x)\,.$$

The following claims show that one can take φ' to be φ_z for a random z. Also, it is sufficient to find an affine map which has significant agreement with φ'.

Claim 4.1. *Let $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be as above. Then, we have*

$$\mathop{\mathbf{Pr}}_{z\in H}\left[\mathop{\mathbf{Pr}}_{x,y\in H+z}\left[\varphi_z(x) + \varphi_z(y) = \varphi_z(x+y)\right] \geqslant \varepsilon^2/2\right] \;\geqslant\; \varepsilon^2/2\,.$$

Proof. We note that

$$\begin{aligned}\varphi_z(x) + \varphi_z(y) = \varphi_z(x+y) \;&\Leftrightarrow\; \varphi(x+z) + \varphi(y+z) = \varphi(x+y+z) + \varphi(z)\\ &\Leftrightarrow\; \psi(x+y+z) = \varphi(x+y+z)\end{aligned}$$

using the facts that $z, x+z, y+z \in H$ and ψ is linear. Moreover, we have

$$\begin{aligned}\mathop{\mathbf{Pr}}_{z,x+z,y+z\in H}\left[x+y+z \in H\right] \;&=\; \mathop{\mathbf{Pr}}_{x,y,z\in H}\left[x+y+z \in H\right]\\ &=\; \frac{\mathbf{E}_{x,y,z\in\mathbb{F}_2^n}\left[\mathbb{1}_H(x)\mathbb{1}_H(y)\mathbb{1}_H(z)\mathbb{1}_H(x+y+z)\right]}{\mathbf{E}_{x,y,z\in\mathbb{F}_2^n}\left[\mathbb{1}_H(x)\mathbb{1}_H(y)\mathbb{1}_H(z)\right]}\\ &\geqslant\; \frac{1}{\varepsilon^6}\cdot\sum_{\beta\in\mathbb{F}_2^n}\left(\widehat{\mathbb{1}_H}(\beta)\right)^4 \;\geqslant\; \varepsilon^8/\varepsilon^6 \;=\; \varepsilon^2\,.\end{aligned}$$

since $x+y+z \in H$ implies $\varphi(x+y+z) = \psi(x+y+z)$, this gives

$$\mathop{\mathbf{E}}_{z\in H}\mathop{\mathbf{E}}_{x,y\in H+z}\left[\mathbb{1}_{\varphi_z(x)+\varphi_z(y)=\varphi_z(x+y)}\right] \;\geqslant\; \varepsilon^2\,.$$

An averaging argument then proves the claim. □

Since a random $z \in \mathbb{F}_2^n$ lies in H with probability ε^2, with probability $\varepsilon^4/2$ a random $z \in \mathbb{F}_2^n$ satisfies the conclusions of Claim 4.1. We asume in the rest of the argument that we have selected such a z. For $\varphi' := \varphi_z$, $H_0 := H + z$, we have $\varphi'(x) = \psi(x)$ for all $x \in H_0$. Also, since $|H_0| \geqslant \varepsilon^2 \cdot 2^n$, by Claim 4.1, we have

$$\mathop{\mathbf{Pr}}_{x,y\in\mathbb{F}_2^n}\left[\varphi'(x) + \varphi'(y) = \varphi'(x+y)\right] \;\geqslant\; \varepsilon^6\,.$$

The following claim shows that it suffices to find an affine map having significant agreement with φ_z for such a z.

Claim 4.2. *If $\tau' : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is an affine map satisfying $\mathbf{Pr}_{x\in\mathbb{F}_2^n}[\tau'(x) = \varphi_z(x) = \psi(x)] \geqslant \gamma$ for some $z \in H$ and $\gamma > 0$, then the map $\tau(x) = \tau'(x) + \tau'(z) + \tau'(0) + \varphi(z)$ satisfies $\mathbf{Pr}_{x\in\mathbb{F}_2^n}[\tau(x) = \varphi(x) = \psi(x)] \geqslant \gamma$.*

Proof. The claim follows by observing that

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{x\in\mathbb{F}_2^n}\left[\varphi_z(x) = \tau'(x) = \psi(x)\right] &= \mathop{\mathbf{Pr}}_{x\in\mathbb{F}_2^n}\left[\varphi(x+z) + \varphi(z) = \tau'(x) = \psi(x)\right] \\
&= \mathop{\mathbf{Pr}}_{x\in\mathbb{F}_2^n}\left[\varphi(x) + \varphi(z) = \tau'(x+z) = \psi(x+z)\right] \\
&= \mathop{\mathbf{Pr}}_{x\in\mathbb{F}_2^n}\left[\varphi(x) + \varphi(z) = \tau'(x) + \tau'(z) + \tau'(0) = \psi(x) + \varphi(z)\right],
\end{aligned}$$

where the last equality uses the fact that τ' is an affine map, ψ is linear, and $\varphi(z) = \psi(z)$ since $z \in H$. □

For the rest of this section, we will refer to φ' as φ (to avoid notational clutter) and assume that it satisfies the above properties. Note that the above reduction works with probability at least $\varepsilon^4/2$.

4.2 Applying the Balog-Szemerédi-Gowers Theorem

As in [Sam07, TW14] we first apply the Balog-Szemerédi-Gowers (BSG) theorem to the set

$$A_\varphi := \{(x, \varphi(x)) \;:\; x \in \mathbb{F}_2^n\}.$$

For any set A that has *some* additive structure, the Balog-Szemerédi-Gowers theorem allows us to find a large subset $A' \subseteq A$ with small doubling. We state the following version from [BS94].

Theorem 4.3 (Balog-Szemerédi-Gowers Theorem [BS94])**.** *Let $A \subseteq \mathbb{F}_2^n$ be such that $\mathbf{Pr}_{a_1,a_2\in A}[a_1 + a_2 \in A] \geqslant \rho$. Then there exists $A' \subseteq A$, $|A'| \geqslant \rho \cdot |A|$ such that $|A' + A'| \leqslant (2/\rho)^8 \cdot |A|$.*

We are interested in finding the set A'_φ which results from applying Theorem 4.3 to the set A_φ. However, since the set A'_φ is of exponential size, we do not have time to write down the entire set (even if we can find it). Instead, we will need an efficient algorithm for testing membership in the set. We follow the algorithmic version in [TW14], based on the proofs in [SSV05, Vio11]. We will modify the algorithm to ensure that (with significant probability) the set A' has a significant intersection with the set

$$H_\varphi := \{(x, \varphi(x)) \;:\; x \in H_0\} = \{(x, \psi(x)) \;:\; x \in H_0\}.$$

In this proof one constructs a graph on the set A_φ and then selects a subset of the neighborhood of a random vertex as A'_φ, after removing certain problematic vertices. It can be deduced that the set A'_φ can be found in time polynomial in the size of the graph. However, as discussed above, this time is still *exponential* in n and hence we instead develop a (randomized) membership oracle for A'_φ. This will be an approximate oracle and we will only be able to say that the output of the oracle is sandwiched between two sets $A^{(1)}_\varphi \subseteq A^{(2)}_\varphi$.

We first define a graph $G_\varphi = (A_\varphi, E_\varphi)$ where the edge set E_φ is defined as

$$E_\varphi := \{(x, \varphi(x)), (y, \varphi(y)) \;\mid\; \varphi(x) + \varphi(y) = \varphi(x+y)\}.$$

Let the size of the vertex set be $N = 2^n$. Claim 4.1 implies that G_φ has density at least $\rho = \varepsilon^6$. Moreover, there exists a set H_φ of size at least $\varepsilon^2 \cdot N$, which has density at least $\rho^{1/3} = \varepsilon^2$.

The proof by in [SSV05] considers the neighborhood of a random vertex u and removes vertices that have too few neighbors in common with other vertices in the graph. For a vertex $u = (x, \varphi(x)) \in A_\varphi$, they define the following sets.

$$\begin{aligned}
N(u) &:= \{v \,:\, (u,v) \in E_\gamma\} \\
S(u) &:= \left\{ v \in N(u) \,:\, \mathop{\mathbf{Pr}}_{v_1}\left[v_1 \in N(u) \text{ and } |N(v) \cap N(v_1)| \leqslant \rho^3 N \right] \geqslant \rho^2 \right\} \\
&= \left\{ v \in N(u) \,:\, \mathop{\mathbf{Pr}}_{v_1}\left[v_1 \in N(u) \text{ and } \mathop{\mathbf{Pr}}_{v_2}\left[v_2 \in N(v) \cap N(v_1) \right] \leqslant \rho^3 \right] > \rho^2 \right\} \\
T(u) &:= N(u) \setminus S(u) \\
&= \left\{ v \in N(u) \,:\, \mathop{\mathbf{Pr}}_{v_1}\left[v_1 \in N(u) \text{ and } \mathop{\mathbf{Pr}}_{v_2}\left[v_2 \in N(v) \cap N(v_1) \right] \leqslant \rho^3 \right] \leqslant \rho^2 \right\}
\end{aligned}$$

It is shown in [SSV05] (see also [Vio11]) that if the graph has density ρ, then picking $A'_\varphi = T(u)$ for a random vertex u is a good choice.

Lemma 4.4. *Let G_φ have density at least ρ, and set $A'_\varphi = T(u)$ for a random vertex u. Then, with probability at least $\rho/2$ over the choice of u, the set A'_φ satisfies*

$$|A'_\varphi| \geqslant \rho \cdot N \quad \textit{and} \quad |A'_\varphi + A'_\varphi| \leqslant (2/\rho)^8 \cdot N.$$

We now translate the condition for membership in the set $T(u)$ into an algorithm.

`BSG-Test` (u, v, ρ_1, ρ_2) (Approximate test to check if $v \in T(u)$)

- Let $u = (x, \varphi(x))$ and $v = (y, \varphi(y))$.
- Sample $(z_1, \varphi(z_1)), \ldots, (z_r, \varphi(z_r))$.
- For each $i \in [r]$, sample $(w_1^{(i)}, \varphi(w_1^{(i)})), \ldots, (w_s^{(i)}, \varphi(w_s^{(i)}))$.
- If $(u, v) \notin E_\varphi$, then output 0.
- For $i \in [r], j \in [s]$, let

$$\begin{aligned}
X_i &= \mathbb{1}_{\{\varphi(x)+\varphi(z_i)=\varphi(x+z_i)\}} \\
Y_{ij} &= \mathbb{1}_{\left\{\varphi(y)+\varphi\left(w_j^{(i)}\right)=\varphi\left(x+w_j^{(i)}\right)\right\}} \\
Z_{ij} &= \mathbb{1}_{\left\{\varphi(z_i)+\varphi\left(w_j^{(i)}\right)=\varphi\left(z_i+w_j^{(i)}\right)\right\}}
\end{aligned}$$

- For each i, take $B_i = 1$ if $\frac{1}{s}\sum_j Y_{ij} \cdot Z_{ij} \leqslant \rho_1$ and 0 otherwise.
- Answer 1 if $\frac{1}{r}\sum_i X_i \cdot B_i \leqslant \rho_2$ and 0 otherwise.

Choice of parameters for `BSG-Test`**:** Recall that $\rho = \varepsilon^6$, and let $\rho_1 = 21\rho^3/80$ and $\rho_2 = 19\rho^2/20$. Given an error parameter η, we take r and s to be $\mathrm{poly}(1/\rho, \log(1/\eta))$, so that with probability at least $1-\eta$, the error in the last two estimates is at most $\rho^3/100$.

As in [TW14], we "sandwich" the elements on which `BSG-Test` answers 1 between a large set and a set with small doubling.

Lemma 4.5. *Let $\eta > 0$ and let the parameters ρ_1, ρ_2, r, s be chosen as above. Then for every $u = (x, \varphi(x))$, there exist two sets $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$ such that the output of* `BSG-Test` *satisfies the following with probability at least $1-\eta$.*

- $\texttt{BSG-Test}(u, v, \gamma_1, \rho_1, \rho_2) = 1 \implies v \in A_\varphi^{(2)}(u)$.
- $\texttt{BSG-Test}(u, v, \gamma_1, \rho_1, \rho_2) = 0 \implies v \notin A_\varphi^{(1)}(u)$.

Moreover, with probability at least $\rho/3$ over the choice of u, we have

$$|A_\varphi^{(1)}(u) \cap H_\varphi| \geqslant (\rho/3) \cdot N \quad \text{and} \quad |A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)| \leqslant (2/\rho)^8 \cdot N.$$

Proof. To deal with the approximate nature the test, we define the following family of sets

$$T(u, \rho_1, \rho_2) := \left\{ v \in N(u) \ : \ \mathop{\mathbf{Pr}}_{v_1}\left[v_1 \in N(u) \ \&\ \mathop{\mathbf{Pr}}_{v_2}[v_2 \in N(v) \cap N(v_1)] \leqslant \rho_1 \right] \leqslant \rho_2 \right\}$$

It is easy to check that for $\rho_1', \rho_2' > 0$, one has the inclusion

$$T(u, \rho_1, \rho_2) \ \subseteq \ T(u, \rho_1 - \rho_1', \rho_2 + \rho_2')\,.$$

We define the sets $A_\varphi^{(1)}(u)$ and $A_\varphi^{(2)}(u)$ as follows.

$$\begin{aligned} A_\varphi^{(1)}(u) \ &:= \ T(u, 11\rho^3/40, 9\rho^2/10) \\ A_\varphi^{(2)}(u) \ &:= \ T(u, \rho^3/4, \rho^2) \end{aligned}$$

By the monotonicity property noted above, we have that $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$. Also, by choice of the parameters r, s, with probability at least $1-\eta$, the error in all estimates used in `BSG-Test` is at most $\rho^3/100$. Hence, we find that with probability at least $1-\eta$, if `BSG-Test` answers 1, then the input is in $A_\varphi^{(2)}$ and if `BSG-Test` answers 0, then it is not in $A_\varphi^{(1)}$. It remains to prove the bounds on the size and doubling of these sets.

The proof that $A_\varphi^{(2)}(u)$ has small growth is identical to the one in [SSV05, Vio11] (with slightly different parameters). We include it below for completeness.

Claim 4.6. *Let $|A_\varphi^{(2)}(u)| > (\rho/6) \cdot N$. Then, we have $|A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)| \leqslant (2/\rho)^8 \cdot N$*

Proof. Consider any $v_1, v_2 \in A_\varphi^{(2)}(u) = T(u, \rho^3/4, \rho^2)$. By definition, we have

$$\left| \left\{ w \in A_\varphi^{(2)}(u) \ \mid \ |N(v_i) \cap N(w)| \leqslant (\rho^3/4)N \right\} \right| \ \leqslant \ \rho^2 N \qquad \text{for } i = 1, 2\,.$$

Thus, there exist at least $(\rho/6 - 2\rho^2) \cdot N \geqslant (\rho/12) \cdot N$ vertices $w \in A_\varphi^{(2)}(u)$ such that $\left| \left\{ w \in A_\varphi^{(2)}(u) \ \mid \ |N(v_i) \cap N(w)| \right\} \right| \geqslant \rho^3/4$ for both $i = 1, 2$. For any such w, and $w_i \in$

$N(v_i) \cap N(w)$, (v_1, w_1, w, w_2, v_2) is a length-4 path in G_φ. Thus, we have at least $(\rho^7/192) \cdot N^3$ paths of length 4 from v_1 to v_2, any two of which differ in at least one vertex.

For any such path (v_1, w_1, w, w_2, v_2), we have by definition of G_φ,

$$v_1 + v_2 \;=\; (v_1 + w_1) + (w_1 + w) + (w + w_2) + (w_2 + v_2)\,,$$

where each of the terms on the right are in $A_\varphi = \{(x, \varphi(x)) \;\mid\; x \in \mathbb{F}_2^n\}$. Since, no two paths agree on all vertices, we have

$$\left|A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)\right| \cdot (\rho^7/192) \cdot N^3 \;\leqslant\; \left|A_\varphi\right|^4 \quad\Longrightarrow\quad \left|A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)\right| \;\leqslant\; (2/\rho)^8 \cdot N\,,$$

which proves the claim. □

To show the lower bound on the size of $A_\varphi^{(2)}(u)$, we will show that $|A_\varphi^{(1)}(u) \cap H_\varphi| \geqslant (\rho/6) \cdot N$. Since $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$, this suffices for the proof. This will require a slight modification of the argument from [SSV05], to take care of the (hidden) set H_φ.

Claim 4.7. *With probability at least $\rho^{2/3}/3$ over the choice of a random $u \in H_\varphi$, we have*

$$\left|A_\varphi^{(1)}(u) \cap H_\varphi\right| \;\geqslant\; (\rho^{2/3}/3) \cdot N$$

Proof. We first show an upper bound on the expected size of the set $S'(u)$ defined as

$$\begin{aligned} S'(u) &:= N(u) \setminus T(u, 11\rho^3/40, 9\rho^2/10) \\ &= \left\{ v \in N(u) \;:\; \mathop{\mathbf{Pr}}_{v_1}\left[v_1 \in N(u) \;\&\; \mathop{\mathbf{Pr}}_{v_2}\left[v_2 \in N(v) \cap N(v_1)\right] \leqslant 11\rho^3/40 \right] \geqslant 9\rho^2/10 \right\}. \end{aligned}$$

We call a pair (v, v_1) *bad* if $|N(v) \cap N(v_1)| \leqslant 11\rho^3 N/40$. Note that for each of the $N(N-1)$ choices for (v, v_1), if they form a bad pair, then a random $u \in A_\varphi$ is in $N(v) \cap N(v_1)$ with probability at most $11\rho^3/40$. Hence,

$$\mathop{\mathbf{E}}_{u \in H_\varphi}\left[\#\left\{\text{bad pairs } (v, v_1) \textit{ mid } v \in N(u) \;\&\; v_1 \in N(u)\right\}\right] \;\leqslant\; \frac{N}{|H_\varphi|} \cdot N^2 \cdot \frac{11\rho^3}{40} \;\leqslant\; \frac{11\rho^{8/3}N^2}{40}\,,$$

using $|H_\varphi| \geqslant \rho^{1/3} \cdot N$. Using Markov's inequality, this gives that

$$\mathop{\mathbf{E}}_{u \in H_\varphi}\left[|S'(u)|\right] \;\leqslant\; \frac{11\rho^{8/3}N^2/40}{9\rho^2 N/10} \;=\; (11\rho^{2/3}/36) \cdot N$$

Finally, using the fact that H_φ has density at least $\rho^{1/3}$, we get

$$\begin{aligned} \mathop{\mathbf{E}}_{u \in H_\varphi}\left[\left|A_\varphi^{(1)}(u) \cap H_\varphi\right|\right] \;&=\; \mathop{\mathbf{E}}_{u \in H_\varphi}\left[\left|(N(u) \setminus S'(u)) \cap H_\varphi\right|\right] \\ &\geqslant\; \mathop{\mathbf{E}}_{u \in H_\varphi}\left[\left|N(u) \cap H_\varphi\right|\right] - \mathop{\mathbf{E}}_{u \in H_\varphi}\left[\left|S'(u)\right|\right] \\ &\geqslant\; \rho^{1/3} \cdot (\rho^{1/3} \cdot N) - (11\rho^{2/3}/36) \cdot N \;\geqslant\; (2\rho^{2/3}/3) \cdot N\,. \end{aligned}$$

An averaging argument then proves the claim. □

To prove the lemma, we observe that since $|H_\varphi| \geqslant \rho^{1/3} \cdot N$, a random $u \in A_\varphi$ satisfies the conclusion of Claim 4.7 with probability at least $\rho^{1/3} \cdot (\rho^{2/3}/3) = \rho/3$. □

4.3 Finding a Good Model Set

We will use the routine BSG-Test described in Section 4.2, and assume that we have chosen a good vertex u guarantee in Lemma 4.5 for BSG-Test. Since the output of BSG-Test is contained in a set $A_\varphi^{(2)}$ with doubling $\rho^{-O(1)}$, we will apply Sanders' quasi-polynomial Bogolyubov lemma, which lets us conclude that $4A_\varphi^{(2)}$ contains a large subspace V. If the subspace V was indeed inside $A_\varphi^{(2)}$, we have a subspace with points of the form $(x, \varphi(x))$. Thus, φ must be a linear map when restricted to points in V.

However, since the Bogolyubov lemma only finds a subspace inside the set $4A_\varphi^{(2)}$, an element of the subspace is of the form $(x_1 + x_2 + x_3 + x_4, \varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4))$. Unlike tuples of the form $(x, \varphi(x))$, the second half of the tuple $(\varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4))$ may not uniquely depend on the first $(x_1 + x_2 + x_3 + x_4)$.

Since it will be important to have this uniqueness property from our subspace, we restrict our sets to get new sets $A_\varphi'^{(1)} \subseteq A_\varphi'^{(2)}$. These restrictions will satisfy the following property: for all tuples x_1, x_2, x_3, x_4 and x_1', x_2', x_3', x_4' satisfying $x_1 + x_2 + x_3 + x_4 = x_1' + x_2' + x_3' + x_4'$, we also have $\varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4) = \varphi(x_1') + \varphi(x_2') + \varphi(x_3') + \varphi(x_4)'$. In other words, φ is a Freiman 4-homomorphism on the first n coordinates of $A_\varphi'^{(2)}$. We will, in fact, need to ensure that it is a Freiman 8-homomorphism in order to obtain a truly linear map.

This step is often called *finding a good model*, and appears (in non-algorithmic form) as Lemma 6.2 in [GT08]. The version described here is the algorithmic version developed in [TW14] (Lemma 37).

We will intersect the original sets with a subspace defined using a random linear map $\Gamma : \mathbb{F}_2^n \to \mathbb{F}_2^r$ and a random element $c \in \mathbb{F}_2^r$ (for $r = O(\log(1/\varepsilon))$). We shall apply the restriction $\Gamma(\varphi(x)) = c$ to the elements $v = (x, \varphi(x))$ on which BSG-Test outputs 1.

> Model-Test (v, Γ, c)
>
> - Let $v = (y, \varphi(y))$.
> - Answer 1 if BSG-Test returns 1 on v and $\Gamma(\varphi(y)) = c$, and 0 otherwise.

We also define

$$A_\varphi'^{(j)} := \left\{(x, \varphi(x)) \in A_\varphi^{(j)} \mid \Gamma(\varphi(x)) = c\right\} \quad \text{for } j \in \{1, 2\}$$

Recall that a map φ is called a *Freiman 8-homomorphism* on its domain if for any $x_1, \ldots, x_8$ with the property that $x_1 + x_2 + x_3 + x_4 = x_5 + x_6 + x_7 + x_8$, we also have $\varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4) = \varphi(x_5) + \varphi(x_6) + \varphi(x_7) + \varphi(x_8)$. We re-state the lemma from [TW14] which proves the required properties of the above procedure.

Lemma 4.8 (Lemma 37 from [TW14])**.** *Let the calls to* BSG-Test *in* Model-Test *be with a good choice of parameters u, ρ_1, ρ_2 and with error parameter $\eta > 0$. Then, there exist two sets $A_\varphi'^{(1)} \subseteq A_\varphi'^{(2)}$ for which the output of* Model-Test *on input $v = (y, \varphi(y))$ satisfies the following with probability $1 - \eta$.*

- Model-Test$(v, \Gamma, c) = 1 \implies v \in A_\varphi'^{(2)}$.
- Model-Test$(v, \Gamma, c) = 0 \implies v \notin A_\varphi'^{(1)}$.

Moreover, with probability at least $\theta = \varepsilon^{O(1)}$ over the choice of Γ and c, we have

$$\left|A_\varphi'^{(1)} \cap H_\varphi\right| \geqslant \theta \cdot N \quad \text{and} \quad \varphi \text{ is a Freiman 8-homomorphism on } A^{(2)},$$

where we denote the projection of $A_\varphi'^{(2)}$ onto the first n coordinates by $A^{(2)}$.

We remark that [TW14] only proved a lower bound on the size of $A_\varphi'^{(1)}$. However, this is simply proves using linearity of expectation and since here we know that $\left|A_\varphi^{(1)} \cap H_\varphi\right| \geqslant (\rho/3) \cdot N$ from Lemma 4.5, the same argument also implies a lower bound on $|A_\varphi'^{(1)} \cap H_\varphi|$.

4.4 Obtaining an Affine Function on a Subspace

We now identify an affine function that has significant agreement with φ. However, we only be able to identify it on a *coset of a subspace* V such that $\mathsf{cod}(V) = \mathsf{polylog}(1/\varepsilon)$.

Lemma 4.9. *Let φ be as above and let the parameters for* `BSG-Test` *and* `Model-Test` *be such that they satisfy the guarantees of Lemma 4.5 and Lemma 4.8. Let ε be as above. Then there exists an algorithm running in time $O_\varepsilon(n^{O(1)})$ which outputs with probability at least $\mathsf{quasipoly}(\varepsilon)$ a subspace V of codimension at most $(\log(1/\varepsilon))^{O(1)}$, $c_1 \in \mathbb{F}^m$, and an affine map $x \mapsto \tau(x)$ satisfying $\mathbf{Pr}_{x \in V + c_1}[\tau(x) = \varphi(x)] \geqslant \varepsilon^{O(1)}$.*

Throughout the argument that follows, we shall assume that we have already chosen good parameters for `BSG-Test` and `Model-Test` so that the conclusions of Lemma 4.5 and Lemma 4.8 hold. We also assume that we have access to a good function φ.

To find the subspace V we apply Bogolyubov's lemma to the set identified by the procedure `Model-Test`. We shall look at the second half of the tuples in this subspace (coordinates $n+1$ to $m+n$) to find a linear choice function. Let $A^{(1)}$ and $A^{(2)}$ denote the projection onto the first n coordinates of the sets $A_\varphi'^{(1)}$ and $A_\varphi'^{(2)}$ given by Lemma 4.8. Also, recall that the set H_0 is the projection of H_φ onto the first n coordinates. Since the last m coordinates are a function (namely φ) of the first n coordinates, we also have $|A^{(1)} \cap H_0| \geqslant \theta \cdot N$, for $\theta = \varepsilon^{O(1)}$ as in Lemma 4.8. The Bogolyubov-Ruzsa lemma by Sanders [San10] shows that there exists a subspace V of co-dimension $(\log(1/\theta))^4$, such that $V \subseteq 4A^{(1)} \subseteq \mathbb{F}_2^n$.

We will need an algorithmic version of the lemma, with query access to the (random) function $h : \mathbb{F}_2^n \to \{0,1\}$ defined by $h(y) = 1$ if $\texttt{Model-Test}((y, \varphi(y)), \Gamma, c) = 1$ and 0 otherwise. The error parameter η' for `Model-Test` is taken to be η/n^4. We shall apply the following algorithmic version from [BSRZTW14] with queries to h and with error parameter $\eta_1 = \eta/20$.

Lemma 4.10 (Algorithmic Quasipolynomial Bogolyubov-Ruzsa Lemma [BSRZTW14])**.** *There exists a randomized algorithm* `Bogolyubov` *with input parameters $\eta_1 \geqslant \alpha > 0$ which, given oracle access to a function $h : \mathbb{F}_2^n \to \{0,1\}$ with $\mathbf{E}\, h \geqslant \alpha$, outputs a subspace $V_0 \subseteq \mathbb{F}_2^n$ of codimension at most $O(\log^4(1/\alpha))$ (by giving a basis for $V_0^\perp$) such that with probability at least $1 - \eta_1$, we have $h * h * h * h(v) > \alpha^4/2$ for each $v \in V_0$. The algorithm runs in time $2^{O(\log^4(1/\alpha))} \cdot \mathsf{polylog}(1/\eta_1) \cdot n^3 \log^2 n$.*

If h was indeed the indicator function of a set A, we would have that the support of $h * h * h * h$ is contained in $4A$, implying $V_0 \subseteq 4A$. However, we only have that with

probability $1-\eta'$ for each input x, the inequality $1_{A^{(1)}}(x) \leqslant h(x) \leqslant 1_{A^{(2)}}(x)$ holds. We can then define a function

$$h' \;:=\; \min\{\mathbb{1}_{A^{(2)}}, \max\{h, \mathbb{1}_{A^{(1)}}\}\}$$

If t denotes the running time of the algorithm `Bogolyubov`, then with probability $1-\eta'\cdot t$, it only reads outputs from the function h' and hence with probability $1-\eta_1-\eta'\cdot t$, the output subspace satisfies

$$\forall v \in V \qquad \mathbb{1}_{A^{(2)}} * \mathbb{1}_{A^{(2)}} * \mathbb{1}_{A^{(2)}} * \mathbb{1}_{A^{(2)}}(v) \;\geqslant\; h' * h' * h' * h'(v) \;>\; \theta^4/2\,,$$

which implies $V_0 \in 4A^{(2)}$. Thus, each element $x \in V_0$ can we written as $x_1+x_2+x_3+x_4$ for $x_1,x_2,x_3,x_4 \in A^{(2)}$. The following claim (well-known) shows that the set

$$Z_0 := \left\{ (x_1+x_2+x_3+x_4, \varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)) \;\middle|\; \begin{array}{c} x_1+x_2+x_3+x_4 \in V_0, \\ x_1,x_2,x_3,x_4 \in A^{(2)} \end{array} \right\}$$

is also a subspace of $\mathbb{F}_2^{n+m}$. Observe that the value of $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)$ is uniquely determined by $x_1+x_2+x_3+x_4$.

Claim 4.11. *There exists a linear map $\zeta : V_0 \to \mathbb{F}_2^n$ satisfying the following property. For any $x_1,x_2,x_3,x_4 \in A^{(2)}$ such that $x_1+x_2+x_3+x_4 \in V_0$, we have $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(x_1+x_2+x_3+x_4)$. Thus, the set Z_0 can be written as $Z_0 = \{(x,\zeta(x)) : x \in V_0\}$ and is a subspace of $\mathbb{F}_2^{n+m}$.*

Proof. We first show that the value of $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)$ is uniquely determined by $x_1+x_2+x_3+x_4$. By Lemma 4.8, we know that φ is a Freiman 8-homomorphism on $A^{(2)}$ and hence it is also a Freiman 4-homomorphism. In particular, if for $x_1,x_2,x_3,x_4 \in A^{(2)}$ and $x_1',x_2',x_3',x_4' \in A^{(2)}$, we have that $x_1+x_2+x_3+x_4 = x_1'+x_2'+x_3'+x_4'$, then it also holds that $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \varphi(x_1')+\varphi(x_2')+\varphi(x_3')+\varphi(x_4')$. Thus, we can write the set Z_0 as $\{(x,\zeta(x)) : x \in V_0\}$, where ζ is some function on V. We next show that ζ must be a linear function.

We begin by showing that $\zeta(0)=0$. Since $0 \in V_0$, we must have elements $x_1,x_2,x_3,x_4 \in A^{(2)}$ with the property that $x_1+x_2+x_3+x_4=0$, in other words, $x_1+x_2 = x_3+x_4$. But since φ is also a Freiman 2-homomorphism, we get that $\varphi(x_1)+\varphi(x_2) = \varphi(x_3)+\varphi(x_4)$, which implies that $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(0) = 0$.

Since φ is a Freiman 8-homomorphism on $A^{(2)}$ and $V_0 \subseteq 4A^{(2)}$, it follows that ζ is a Freiman 2-homomorphism on V_0. Since V_0 is closed under addition, for $x,y \in V_0$ we can write $x+y = 0+(x+y)$ with all four summands in V_0. Since ζ is a 2-homomorphism, we get that $\zeta(x)+\zeta(y) = \zeta(0)+\zeta(x+y) = \zeta(x+y)$. □

We would like to use the linear map ζ to obtain the choice function on a coset of the space V_0. However, we do not *know* the function ζ. We get around by generating random tuples $(x_1+x_2+x_3+x_4, \varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4))$ such that $x_1+x_2+x_3+x_4$ as well as each x_i are in $A^{(2)}$. We show that for sufficiently many samples, the sampled points span a large subspace V of V_0. Since $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(x_1+x_2+x_3+x_4)$ on V_0, we will be able to obtain the desired linear map on the subspace V.

We sample a point as follows. For the j^{th} sample, we generate four pairs $(x_1^j,\varphi(x_1^j)),\ldots,(x_4^j,\varphi(x_4^j))$. We accept the sample if all four pairs are accepted by `Model-Test` and if $x_1^j+x_2^j+x_3^j+x_4^j \in V_0$. If a sample is accepted, we store the point $y^j = x_1^j+x_2^j+x_3^j+x_4^j$ and $\zeta(y^j) = \varphi(x_1^j)+\varphi(x_2^j)+\varphi(x_3^j)+\varphi(x_4^j)$.

Note that membership in V_0 can be tested efficiently since we know the basis for $V_0^\perp$. We first estimate the probability that a point $(y, \zeta(y))$ for $y \in V_0$ is accepted by the above test. This also gives a bound on the number of samples to be tried so that at least $t = O(n^2)$ samples are accepted.

Claim 4.12. *For $y \in V_0$, the probability that a sample is accepted by the above procedure and the stored pair is equal to $(y, \zeta(y))$ is at least $\theta^4/4N$. Moreover, for some sufficiently large constant C, the probability that out of $\exp\left(C \cdot \log(1/\theta)^4\right) \cdot t \cdot \log(10/\eta)$ samples fewer than t are accepted is at most $\eta/10$.*

Proof. Since the function $h(x) = 1$ exactly when `Model-Test` accepts $(x, \varphi(x))$, the probability that a sample $(x_1, \varphi(x_1)), \ldots, (x_4, \varphi(x_4))$ is accepted and that $x_1 + x_2 + x_3 + x_4 = y$, is equal to

$$\mathbf{Pr}\left[\bigwedge_{i=1}^{4}(h(x_i) = 1) \wedge (x_1 + x_2 + x_3 + x_4 = y)\right] \;=\; \frac{1}{N} \cdot \mathop{\mathbf{E}}_{h, x_1+x_2+x_3+x_4=y}\left[\prod_{i=1}^{4} h(x_i)\right]$$

As before, we consider the function $h' = \max\{1_{A^{(1)}}, \min\{h, 1_{A^{(2)}}\}\}$ satisfying that for each x, $\mathbf{Pr}\left[h(x) \neq h'(x)\right] \leqslant \eta'$ (where η' is the error parameter in Lemma 4.10), and that $h' * h' * h' * h'(x) > \theta^4/2$ for each $x \in V_0$. We can now estimate the above expectation as

$$\begin{aligned}
&\mathop{\mathbf{E}}_{h, x_1+x_2+x_3+x_{=}y}\left[h(x_1)h(x_2)h(x_3)h(x_4)\right] \\
&\geqslant \mathop{\mathbf{Pr}}_{h, x_1+x_2+x_3+x_4=y}\left[\wedge_{i=1}^{4}(h(x_i) = h'(x_i))\right] \cdot \mathop{\mathbf{E}}_{h, x_1, x_2, x_3}\left[h'(x_1)h'(x_2)h'(x_3)h'(y + x_1 + x_2 + x_3)\right] \\
&\geqslant (1 - 4\eta') \cdot h' * h' * h' * h'(y) \\
&\geqslant (1 - 4\eta') \cdot (\theta^4/2) \;\geqslant\; \theta^4/4.
\end{aligned}$$

The last inequality exploited the fact that $h' * h' * h' * h'(y) \geqslant \theta^4/2$ for $y \in V_0$.

The probability that a sample is accepted is equal to the probability that one selects a pair $(y, \zeta(y))$ for *some* $y \in V_0$. This is least $(|V_0|/N) \cdot (\theta^4/2) = \exp\left(-O((\log(1/\theta))^4)\right) \cdot (\theta^4/2)$. The bound on the probability of accepting fewer than t samples is then given by a Hoeffding bound. □

Let $(y^1, \zeta(y^1)), \ldots, (y^t, \zeta(y^t))$ be t stored points corresponding to t samples accepted by the above procedure. It is easy to check that the projection onto the first n coordinates of these points must span a large subspace of V_0.

Claim 4.13. *Let $(y^1, \zeta(y^1)), \ldots, (y^t, \zeta(y^t))$ be t points stored according to the above procedure. For $t = n^2 + \log(10/\eta)$, the probability that $\mathsf{cod}(\langle y^1, \ldots, y^t\rangle) \geqslant \mathsf{cod}(V_0) + \log(4/\theta^4)$ is at most $\eta/10$.*

Proof. Let $k = \mathsf{cod}(V_0) + 4\log(4/\theta)$ and let S be any subspace of codimension k. The probability that a sample $(x_1, \varphi(x_1)), \ldots, (x_4, \varphi(x_4))$ is accepted and has $x_1 + x_2 + x_3 + x_4 = y$ for a specific $y \in S$ is at most $1/N$. Thus, the probability that an accepted sample $(y^j, \zeta(y^j))$ has $y^j \in S$, conditional on being accepted, is at most $(|S|/N)/((|V_0|/N) \cdot (\theta^4/2))$. Thus, the probability that all t stored points lie in *any* subspace of co-dimension k is at most

$$\left(\frac{|S|/N}{(|V_0|/N) \cdot (\theta^4/2)}\right)^t \cdot \#\{\text{subspaces of co-dimension } k\} \;=\; \left(\frac{\theta^4/4}{\theta^4/2}\right)^t \cdot 2^{n(n-k)} \;\leqslant\; 2^{-t} \cdot 2^{n^2},$$

which is at most $\eta/10$ for $t = n^2 + \log(10/\eta)$. □

Let $V = \mathsf{Span}\left(y^1, \ldots, y^t\right)$. The above claim shows that with high probability, the codimension of V satisfies $\mathsf{cod}(V) = \exp(O(\log(1/\theta))^4)$. From the way the samples were generated, we also know $\zeta(y^1), \ldots, \zeta(y^t)$. Since ζ is a linear function by Claim 4.11, we can extend it to a linear transform $x \mapsto \tau(x)$ such that $\forall x \in V, \tau(x) = \zeta(x)$.

Let $Z \subseteq 4A_\varphi'^{(2)}$ be the subspace $\{(x, \tau(x)) \mid x \in V\}$. We now show that there is a coset of Z in a significant fraction of points are of the form $(x, \varphi(x)) \in A_\varphi'^{(2)}$.

Claim 4.14. *The sets $Z + A_\varphi'^{(1)}$ and $Z + A_\varphi'^{(2)}$ both consist of at most $(1/\theta) \cdot (N/|Z|)$ cosets of Z. Hence, for some $c \in A_\varphi'^{(1)}$ we have*

$$\left|(Z + c) \cap A_\varphi'^{(2)}\right| \geqslant \left|(Z + c) \cap \left(A_\varphi'^{(1)} \cap H_\varphi\right)\right| \geqslant \theta^2 \cdot |Z|\,.$$

Proof. Since $Z \subseteq 4A_\varphi'^{(2)}$ and $A_\varphi'^{(1)} \subseteq A_\varphi'^{(2)}$, we have that

$$Z + A_\varphi'^{(1)} \subseteq Z + A_\varphi'^{(2)} \subseteq 5A_\varphi'^{(2)} \subseteq 5A_\varphi^{(2)}.$$

The last inclusion follows from the fact that $A_\varphi'^{(2)}$ was obtained by intersecting $A_\varphi^{(2)}$ (given by Lemma 4.5) with a subspace.

We know from Lemma 4.5 that $|A_\varphi^{(2)} + A_\varphi^{(2)}| \leqslant (2/\rho)^8 \cdot N \leqslant (2/\rho)^8 \cdot (6/\rho) \cdot |A_\varphi^{(2)}|$. Lemma 2.4 (Plünnecke's inequality) then gives that $|5A_\varphi^{(2)}| \leqslant (6/\rho)^{45} \cdot |A_\varphi^{(2)}| \leqslant (1/\theta) \cdot |A_\varphi^{(2)}| \leqslant (1/\theta) \cdot N$. Thus, $|Z + A_\varphi'^{(2)}| \leqslant (1/\theta) \cdot N$ and it is the union of at most $(1/\theta) \cdot (N/|Z|)$ cosets.

Since $A_\varphi'^{(1)} \cap H_\varphi \subseteq Z + A_\varphi'^{(1)}$, there must exist at least one coset $Z + c$ for $c \in A_\varphi'^{(1)}$, such that

$$\left|(Z + c) \cap \left(A_\varphi'^{(1)} \cap H_\varphi\right)\right| \geqslant \frac{|A_\varphi'^{(1)} \cap H_\varphi|}{(1/\theta) \cdot (N/|Z|)} \geqslant \theta^2 \cdot |Z|\,,$$

where the last inequality used the fact that $|A_\varphi'^{(1)} \cap H_\varphi| \geqslant \theta \cdot N$, as guaranteed by Lemma 4.8. □

Since $Z + A_\varphi'^{(1)}$ is contained in at most $(1/\theta) \cdot (N/|Z|) = \exp O((\log(1/\theta))^4)$ cosets of Z, we can simply pick a coset at random and it must satisfy the conclusion of Claim 4.14 with probability at least $\exp\left(-O((\log(1/\theta))^4)\right)$. To find all cosets, we first sample random points from $A_\varphi'^{(1)}$ as above using `Model-Test`. Since the desired coset has large size, $\exp O((\log(1/\theta))^4)$ samples suffice to identify a collection of cosets, containing the desired one (with high probability). We omit the details. We can now combine the previous argument to prove Lemma 4.9.

Proof of Lemma 4.9: We follow the steps described above to find the subspace V_0, and subsequently the subspace V together with the transformation τ. This immediately yields the subspace $Z = \{(x, \tau(x)) : x \in V\}$. With probability $\exp -O((\log(1/\theta))^4)$, we can find a $c = (c_1, c_2) \in \mathbb{F}^{n+m}$ such that a fraction of at least $\theta^2/2$ of points $(y + c_1, \tau(y) + c_2)$ in the coset $Z + (c_1, c_2)$ are of the form $(x, \varphi(x))$ for $(x, \varphi(x)) \in A_\varphi'^{(1)} \cap H_\varphi$, which gives $\psi(x) = \varphi(x) = \tau(x) + \tau(c_1) + c_2$ for these points. Thus, we have

$$\mathop{\mathbf{E}}_{x \in c_1 + V}\left[\tau(x) + \tau(c_1) + c_2 = \varphi(x) = \psi(x)\right] \geqslant (\theta^2/2) \geqslant \varepsilon^{O(1)}.$$

The errors in the application of Bogolyubov's lemma and in Claim 4.12, and Claim 4.13 add up to $\eta/2 < \eta$. The running time is dominated by the $C\exp(O((\log(1/\theta))^4)) \cdot t \cdot \log(10/\eta)$ calls to `Model-Test` in Claim 4.12 for $t = O(n^2)$. Since each call to `Model-Test` takes $n^{O(1)} \cdot \mathrm{poly}(1/\varepsilon) \cdot \log(\eta/n^4)$ time, the total running time is $n^{O(1)} \cdot \exp(O((\log(1/\theta))^4)) \cdot \log^2(1/\eta)$. □

Acknowledgements

We thank Arnab Bhattacharyya, Julia Wolf and Abhishek Bhowmick for many illuminating discussions. The second author would also like to thank the Simons Institute at UC Berkeley for their generous hospitality and stimulating atmosphere, which led to some of the ideas in this work.

References

[AGS03] Adi Akavia, Shafi Goldwasser, and Shmuel Safra, *Proving hard-core predicates using list decoding*, FOCS, 2003, pp. 146–157. 1

[AS03] Sanjeev Arora and Madhu Sudan, *Improved low degree testing and its applications*, Combinatorica **23** (2003), no. 3, 365–426. 1

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson, *Reed–muller codes for random erasures and errors*, IEEE Transactions on Information Theory **61** (2015), no. 10, 5229–5252. 2

[BGH+15] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer, *Making the long code shorter*, SIAM Journal on Computing **44** (2015), no. 5, 1287–1324. 1

[BL15] Abhishek Bhowmick and Shachar Lovett, *The list decoding radius of Reed-Muller codes over small fields*, Proceedings of the 47th ACM Symposium on Theory of Computing, ACM, 2015, pp. 277–285. 2

[BS94] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica **14** (1994), 263–268, 10.1007/BF01212974. 15

[BSRZTW12] Eli Ben-Sasson, Noga Ron-Zewi, Madhur Tulsiani, and Julia Wolf, *Algorithmic proofs of almost periodicity and applications*, Manuscript, 2012. 12

[BSRZTW14] _____, *Algorithmic proofs of almost periodicity and applications*, Proceedings of the 41st International Colloquium on Automata, Languages and Programming, 2014. 2, 5, 14, 20

[BTZ10] V. Bergelson, T. Tao, and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of* $\mathbb{F}^\omega$, Geom. Funct. Anal. **16** (2010), no. 6, 1539–1596. 7

[BV10] Andrej Bogdanov and Emanuele Viola, *Pseudorandom bits for polynomials*, SIAM J. Comput. **39** (2010), no. 6, 2464–2486. 7

[Dum04] Ilya Dumer, *Recursive decoding and its performance for low-rate Reed-Muller codes*, IEEE Transactions on Information Theory **50** (2004), no. 5, 811–823. 2

[Eli57] Peter Elias, *List decoding for noisy channels*, Tech. Report 335, Research Laboratory of Electronics, MIT, 1957. 1

[GKZ08] P. Gopalan, A.R. Klivans, and D. Zuckerman, *List-decoding Reed-Muller codes over small fields*, Proceedings of the 40th ACM Symposium on Theory of Computing, 2008, pp. 265–274. 1, 2, 3, 4, 9

[GL89] O. Goldreich and L. Levin, *A hard-core predicate for all one-way functions*, Proceedings of the 21st ACM Symposium on Theory of Computing, 1989, pp. 25–32. 1

[Gow98] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), no. 3, 529–551. 6, 7

[GRS00] Oded Goldreich, Ronitt Rubinfeld, and Madhu Sudan, *Learning polynomials with queries: The highly noisy case*, SIAM Journal on Discrete Mathematics **13** (2000), no. 4, 535–570. 1

[GT08] B.J. Green and T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153. MR 2391635 (2009g:11012) 3, 5, 7, 8, 19

[GT09] Ben Green and Terence Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, Contributions to Discrete Mathematics **4** (2009), no. 2. 3

[GTZ11] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers u^4-norm*, Glasgow Mathematical Journal **53** (2011), no. 1, 1–50. 3

[Gur01] Venkatesan Guruswami, *List decoding of error-correcting codes*, Ph.D. thesis, MIT, 2001. 1

[Gur06] _____, *Algorithmic results in list decoding*, Foundations and Trends in Theoretical Computer Science **2** (2006), no. 2. 1

[Jac97] Jeffrey C Jackson, *An efficient membership-query algorithm for learning DNF with respect to the uniform distribution*, Journal of Computer and System Sciences **3** (1997), no. 55, 414–440. 1

[KKM+17] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D Pfister, Eren Sasoglu, and Rudiger Urbanke, *Reed-Muller codes achieve capacity on erasure channels*, IEEE Transactions on Information Theory (2017). 2

[KLP12] Tali Kaufman, Shachar Lovett, and Ely Porat, *Weight distribution and list-decoding size of reed–muller codes*, IEEE Transactions on Information Theory **58** (2012), no. 5, 2689–2696. 2

[KM93] Eyal Kushilevitz and Yishay Mansour, *Learning decision trees using the fourier spectrum*, SIAM Journal on Computing **22** (1993), no. 6, 1331–1348. 1

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky, *Inverse conjecture for the Gowers norm is false*, Proceedings of the 40th ACM Symposium on Theory of Computing, 2008, pp. 547–556. 3

[MS77] F.J. MacWilliams and N.J.A. Sloane, *The theory of error-correcting codes*, North-Holland, 1977. 1

[O'D08] R. O'Donnell, *Some topics in analysis of Boolean functions*, STOC, 2008, pp. 569–578. 6

[Sam07] A. Samorodnitsky, *Low-degree tests at large distances*, Proceedings of the 39th ACM Symposium on Theory of Computing, 2007, pp. 506–515. 2, 3, 7, 8, 15

[San10] Tom Sanders, *On the Bogolyubov-Ruzsa lemma*, Anal. PDE (2010). 20

[SSV05] B. Sudakov, E. Szemerédi, and V.H. Vu, *On a question of Erdös and Moser*, Duke Mathematical Jounal **129** (2005), no. 1, 129–155. 15, 16, 17, 18

[SSV17] Ramprasad Saptharishi, Amir Shpilka, and Ben Lee Volk, *Efficiently decoding Reed–Muller codes from random errors*, IEEE Transactions on Information Theory **63** (2017), no. 4, 1954–1960. 2

[ST06] Alex Samorodnitsky and Luca Trevisan, *Gowers uniformity, influence of variables, and PCPs*, STOC, 2006, pp. 11–20. 7

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan, *Pseudorandom generators without the XOR lemma*, Journal of Computer and System Sciences **62** (2001), no. 2, 236–266. 1

[SU05] Ronen Shaltiel and Christopher Umans, *Simple extractors for all min-entropies and a new pseudorandom generator*, Journal of the ACM **52** (2005), no. 2, 172–216. 1

[Sud00] Madhu Sudan, *List decoding: algorithms and applications*, SIGACT News **31** (2000), no. 1, 16–27. 1

[Tre03] Luca Trevisan, *List-decoding using the XOR Lemma*, Proceedings of the 44th IEEE Symposium on Foundations of Computer Science, 2003, pp. 126–135. 1

[Tre04] ———, *Some applications of coding theory in computational complexity*, Quaderni di Matematica **13** (2004), 347–424, arXiv:cs.CC/0409044. 1

[TSZS06] Amnon Ta-Shma, David Zuckerman, and Shmuel Safra, *Extractors from reed–muller codes*, Journal of Computer and System Sciences **72** (2006), no. 5, 786–812. 1

[TW14] Madhur Tulsiani and Julia Wolf, *Quadratic Goldreich–Levin theorems*, SIAM Journal on Computing **43** (2014), no. 2, 730–766. 2, 3, 5, 14, 15, 17, 19, 20

[TZ10] T. Tao and T. Ziegler, *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, Analysis and PDE **3** (2010), 1–20. 7

[V+12] Salil P Vadhan et al., *Pseudorandomness*, Foundations and Trends® in Theoretical Computer Science **7** (2012), no. 1–3, 1–336. 1

[Vio11] Emanuele Viola, *Selected results in additive combinatorics: An exposition*, Graduate Surveys, no. 3, Theory of Computing Library, 2011. 15, 16, 17

[VW07] Emanuele Viola and Avi Wigderson, *Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols*, IEEE Conference on Computational Complexity, 2007. 7

[Woz58] John M Wozencraft, *List decoding*, Quarterly Progress Report **48** (1958), 90–95. 1