Electronic Colloquium on Computational Complexity, Report No. 86 (2009)

Optimal testing of Reed-Muller codes

Arnab Bhattacharyya* Swastik Kopparty† Grant Schoenebeck‡ Madhu Sudan§
David Zuckerman¶

October 1, 2009

Abstract

We consider the problem of testing if a given function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is close to any degree d polynomial in n variables, also known as the Reed-Muller testing problem. Alon et al. [AKK+05] proposed and analyzed a natural 2^{d+1}-query test for this property and showed that it accepts every degree d polynomial with probability 1, while rejecting functions that are $\Omega(1)$-far with probability $\Omega(1/(d2^d))$. We give an asymptotically optimal analysis of their test showing that it rejects functions that are (even only) $\Omega(2^{-d})$-far with $\Omega(1)$-probability (so the rejection probability is a universal constant independent of d and n).

Our proof works by induction on n, and yields a new analysis of even the classical Blum-Luby-Rubinfeld [BLR93] linearity test, for the setting of functions mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2$. The optimality follows from a tighter analysis of counterexamples to the "inverse conjecture for the Gowers norm" constructed by [GT07, LMS08].

Our result gives a new relationship between the $(d+1)^{\text{st}}$-Gowers norm of a function and its maximal correlation with degree d polynomials. For functions highly correlated with degree d polynomials, this relationship is asymptotically optimal. Our improved analysis of the [AKK+05]-test also improves the parameters of an XOR lemma for polynomials given by Viola and Wigderson [VW07]. Finally, the optimality of our result also implies a "query-hierarchy" result for property testing of linear-invariant properties: For every function $q(n)$, it gives a linear-invariant property that is testable with $O(q(n))$-queries, but not with $o(q(n))$-queries, complementing an analogous result of [GKNR08] for graph properties.

*Computer Science and Artificial Intelligence Laboratory, MIT, abhatt@mit.edu. Work partially supported by a DOE Computational Science Graduate Fellowship and NSF Awards 0514771, 0728645, and 0732334.

†Computer Science and Artificial Intelligence Laboratory, MIT, swastik@mit.edu. Work was partially done while author was a summer intern at Microsoft Research New England and partially supported by NSF Grant CCF-0829672.

‡Department of Computer Science, University of California-Berkeley, grant@cs.berkeley.edu. Work was partially done while author was a summer intern at Microsoft Research New England and partially supported by a National Science Foundation Graduate Fellowship.

§Microsoft Research, One Memorial Drive, Cambridge, MA 02142, USA, madhu@microsoft.com.

¶Computer Science Department, University of Texas at Austin, diz@cs.utexas.edu. Work was partially done while the author consulted at Microsoft Research New England, and partially supported by NSF Grants CCF-0634811 and CCF-0916160.

ISSN 1433-8092

1 Introduction

We consider the task of testing if a Boolean function f on n bits, given by an oracle, is close to a degree d multivariate polynomial (over $\mathbb{F}_2$, the field of two elements). This specific problem, also known as the testing problem for the Reed-Muller code, was considered previously by Alon, Kaufman, Krivelevich, Litsyn, and Ron [AKK$^+$05] who proposed and analyzed a natural 2^{d+1}-query test for this task. In this work we give an improved, asymptotically optimal, analysis of their test. Below we describe the problem, its context, our results and some implications.

1.1 Reed-Muller Codes and Testing

The Reed-Muller codes are parameterized by two parameters: n the number of variables and d the degree parameter. The Reed-Muller codes consist of all functions from $\mathbb{F}_2^n \to \mathbb{F}_2$ that are evaluations of polynomials of degree at most d. We use $\mathrm{RM}(d, n)$ to denote this class, i.e., $\mathrm{RM}(d, n) = \{f : \mathbb{F}_2^n \to \mathbb{F}_2 | \deg(f) \leq d\}$.

The proximity of functions is measured by the (fractional Hamming) distance. Specifically, for functions $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$, we let the *distance* between them, denoted by $\delta(f, g)$, be the quantity $\Pr_{x \leftarrow_U \mathbb{F}_2^n}[f(x) \neq g(x)]$. For a family of functions $\mathcal{F} \subseteq \{g : \mathbb{F}_2^n \to \mathbb{F}_2\}$ let $\delta(f, \mathcal{F}) = \min\{\delta(f, g) | g \in \mathcal{F}\}$. We say f is δ-close to $\mathcal{F}$ if $\delta(f, \mathcal{F}) \leq \delta$ and δ-far otherwise.

Let $\delta_d(f) = \delta(f, \mathrm{RM}(d, n))$ denote the distance of f to the class of degree d polynomials. The goal of Reed-Muller testing is to "test", with "few queries" of f, whether $f \in \mathrm{RM}(d, n)$ or if f is far from $\mathrm{RM}(d, n)$. Specifically, for a function $q : \mathbb{Z}^+ \times \mathbb{Z}^+ \times (0, 1] \to \mathbb{Z}^+$, a *q-query tester* for the class $\mathrm{RM}(d, n)$ is a randomized oracle algorithm T that, given oracle access to some function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and a proximity parameter $\delta \in (0, 1]$, queries at most $q = q(d, n, \delta)$ values of f and accepts $f \in \mathrm{RM}(d, n)$ with probability 1, while if $\delta(f, \mathrm{RM}(d, n)) \geq \delta$ it rejects with probability at least, say, 1/2. The function q is the *query complexity* of the test and the main goal here is to make q as small as possible, as a function possibly of d, n and δ. We denote the test T run using oracle access to the function f by T^f

This task was already considered by Alon et al. [AKK$^+$05] who gave a tester with query complexity $O(\frac{d}{\delta} \cdot 4^d)$. This tester repeated a simple $O(2^d)$-query test, that we denote T_*, several times. Given oracle access to f, T_* selects a $(d+1)$-dimensional affine subspace A, and accepts if f restricted to A is a degree d polynomial. This requires 2^{d+1} queries of f (since that is the number of points contained in A). [AKK$^+$05] show that if $\delta(f) \geq \delta$ then T_* rejects f with probability $\Omega(\delta/(d \cdot 2^d))$. Their final tester then simply repeated T_* $O(\frac{d}{\delta} \cdot 2^d)$ times and accepted if all invocations of T_* accepted. The important feature of this result is that the number of queries is independent of n, the dimension of the ambient space. Alon et al. also show that any tester for $\mathrm{RM}(d, n)$ must make at least $\Omega(2^d + 1/\delta)$ queries. Thus their result was tight to within almost quadratic factors, but left a gap open. We close this gap in this work.

1.2 Main Result

Our main result is an improved analysis of the basic 2^{d+1}-query test T_*. We show that if $\delta_d(f) \geq 0.1$, in fact even if it's at least $0.1 \cdot 2^{-d}$, then in fact this basic test rejects with probability lower bounded by some *absolute constant*. We now give a formal statement of our main theorem.

Theorem 1 *There exists a constant $\epsilon_1 > 0$ such that for all d, n, and for all functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we have*

$$\Pr[T_*^f \text{ rejects}] \geq \min\{2^d \cdot \delta_d(f), \epsilon_1\}.$$

Therefore, to reject functions δ-far from $\mathrm{RM}(d, n)$ with constant probability, one can repeat the test T_* at most $O(1/\min\{2^d\delta_d(f), \epsilon_1\}) = O(1 + \frac{1}{2^d\delta})$ times, making the total query complexity $O(2^d + 1/\delta)$. This query complexity is asymptotically tight in view of the earlier mentioned lower bound in [AKK+05].

Our error-analysis is also asymptotically tight. Note that our theorem effectively states that functions that are accepted by T_* with constant probability (close to 1) are (very highly) correlated with degree d polynomials. To get a qualitative improvement one could hope that every function that is accepted by T_* with probability strictly greater than half is somewhat correlated with a degree d polynomial. Such stronger statements however are effectively ruled out by the counterexamples to the "inverse conjecture for the Gowers norm" given by [LMS08, GT07]. Since the analysis given in these works does not match our parameters asymptotically, we show (see Theorem 22 in Appendix A) how an early analysis due to the authors of [LMS08] can be used to show the asymptotic tightness of the parameters of Theorem 1.

Our main theorem (Theorem 1) is obtained by a novel proof that gives a (yet another!) new analysis even of the classical linearity test of Blum, Luby, Rubinfeld [BLR93]. We give more details on the proof in Section 1.6, but first we explain some of the context of our work and some implications.

1.3 Query-hierarchy for linear invariant properties

Our result falls naturally in the general framework of property testing [BLR93, RS96, GGR98]. Goldreich et al. [GKNR08] asked an interesting question in this broad framework: Given an ensemble of properties $\mathcal{F} = \{\mathcal{F}_N\}_N$ where $\mathcal{F}_N$ is a property of functions on domains of size N, which functions correspond to the query complexity of some property? That is, for a given complexity function $q(N)$, is there a corresponding property $\mathcal{F}$ such that $\Theta(q(N))$-queries are necessary and sufficient for testing membership in $\mathcal{F}_N$? This question is interesting even when we restrict the class of properties being considered.

For completely general properties this question is easy to solve. For graph properties [GKNR08] et al. show that for every efficiently computable function $q(N) = O(N)$ there is a graph property for which $\Theta(q(N))$ queries are necessary and sufficient (on graphs on $\Omega(\sqrt{N})$ vertices). Thus this gives a "hierarchy theorem" for query complexity.

Our main theorem settles the analogous question in the setting of "affine-invariant" properties. Given a field $\mathbb{F}$, a property $\mathcal{F} \subseteq \{\mathbb{F}^n \to \mathbb{F}\}$ is said to be affine-invariant if for every $f \in \mathcal{F}$ and affine map $A : \mathbb{F}^n \to \mathbb{F}^n$, the composition of f with A, i.e, the function $f \circ A(x) = f(A(x))$, is also in $\mathcal{F}$. Affine-invariant properties seem to be the algebraic analog of graph-theoretic properties and generalize most natural algebraic properties (see Kaufman and Sudan [KS08]).

Since the Reed-Muller codes form an affine-invariant family, and since we have a tight analysis for their query complexity, we can get the affine-invariant version of the result of [GKNR08]. Specifically, given any (reasonable) query complexity function $q(N)$ consider N that is a power of two and consider the class of functions on $n = \log_2 N$ variables of degree at most $d = \lceil \log_2 q(N) \rceil$.

We have that membership in this family requires $\Omega(2^d) = \Omega(q(N))$-queries, and on the other hand $O(2^d) = O(q(N))$-queries also suffice, giving an ensemble of properties $\mathcal{P}_N$ (one for every $N = 2^n$) that is testable with $\Theta(q(N))$-queries.

Theorem 2 *For every $q : \mathbb{N} \to \mathbb{N}$ that is at most linear, there is a linear invariant property that is testable with $O(q(n))$ queries (with one-sided error) but is not testable in $o(q(n))$ queries (even with two-sided error). Namely, this property is membership in* $\mathrm{RM}(\lceil \log_2 q(n) \rceil, n)$.

1.4 Gowers norm

A quantity closely related to the rejection probability for T_* also arises in some of the recent results in additive number theory, under the label of the *Gowers norm*, introduced by Gowers [Gow98, Gow01].

To define this norm, we first consider a related test $T_0^f(k)$ which, given parameter k and oracle access to a function f, picks $x_0, a_1, \ldots, a_k \in \mathbb{F}_2^n$ uniformly and independently and accepts if f restricted to the affine subspace $x_0 + \mathrm{span}(a_1, \ldots, a_k)$ is a degree $k-1$ polynomial. Note that since we don't require $a_1, \ldots, a_k$ to be linearly independent, T_0 sometimes (though rarely) picks a subspace of dimension $k-1$ or less. When $k = d+1$, if we condition on the event that $a_1, \ldots, a_k$ are linearly independent, $T_0(d+1)$ behaves exactly as T_*. On the other hand when $a_1, \ldots, a_k$ do have a linear dependency, $T_0(k)$ accepts with probability one. In Proposition 19 we show that when $n \geq d+1$, the probability that $a_1, \ldots, a_{d+1}$ are linearly independent is lower bounded by a constant and so the rejection probability of $T_0(d+1)$ is lower bounded by a constant multiple of the rejection probability of T_* (for every function f). The test T_0 has a direct relationship with the Gowers norm.

In our notation, the Gowers norm can be defined as follows. For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, the k^{th}-Gowers norm of f, denoted $\|f\|_{U^k}$, is given by the expression

$$\|f\|_{U^k} \stackrel{\text{def}}{=} \left(\Pr[T_0^f(k) \text{ accepts}] - \Pr[T_0^f(k) \text{ rejects}]\right)^{\frac{1}{2^k}},$$

Gowers [Gow01] (see also [GT05]) showed that the "correlation" of f to the closest degree d polynomial, i.e., the quantity $1 - 2\delta_d(f)$, is at most $\|f\|_{U^{d+1}}$. The well-known Inverse Conjecture for the Gowers Norm states that some sort of converse holds: if $\|f\|_{U^{d+1}} = \Omega(1)$, then the correlation of f to some degree d polynomial is $\Omega(1)$, or equivalently $\delta_d(f) = 1/2 - \Omega(1)$. (That is, if the acceptance probability of T_0 is slightly larger than $1/2$, then f is at distance slightly smaller than $1/2$ from some degree d polynomial.) Lovett et al. [LMS08] and Green and Tao [GT07] disproved this conjecture, showing that the symmetric polynomial S_4 has $\|S_4\|_{U^4} = \Omega(1)$ but the correlation of S_4 to any degree 3 polynomial is exponentially small. This still leaves open the question of establishing tighter relationships between the Gowers norm $\|f\|_{U^{d+1}}$ and the maximal correlation of f to some degree d polynomial. The best analysis known seems to be in the work of [AKK$^+$05] whose result can be interpreted as showing that there exists $\epsilon > 0$ such that if $\|f\|_{U^{d+1}} \geq 1 - \epsilon/4^d$, then $\delta_d(f) = O(4^d(1 - \|f\|_{U^{d+1}}))$.

Our results show that when the Gowers norm is close to 1, there is actually a tight relationship between the Gowers norm and distance to degree d. More precisely (Theorem 18), there exists $\epsilon > 0$ such that if $\|f\|_{U^{d+1}} \geq 1 - \epsilon/2^d$, then $\delta_d(f) = \Theta(1 - \|f\|_{U^{d+1}})$.

1.5 XOR lemma for low-degree polynomials

One application of the Gowers norm and the Alon et al. analysis to complexity theory is an elegant "hardness amplification" result for low-degree polynomials, due to Viola and Wigderson [VW07]. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $\delta_d(f)$ is noticeably large, say ≥ 0.1. Viola and Wigderson showed how to use this f to construct a $g : \mathbb{F}_2^m \to \mathbb{F}_2$ such that $\delta_d(g)$ is significantly larger, around $\frac{1}{2} - 2^{-\Omega(m)}$. In their construction, $g = f^{\oplus t}$, the t-wise XOR of f, where $f^{\oplus t} : (\mathbb{F}_2^n)^t \to \mathbb{F}_2$ is given by:

$$f^{\oplus t}(x_1, \ldots, x_t) = \sum_{i=1}^{t} f(x_i).$$

In particular, they showed that if $\delta_d(f) \geq 0.1$, then $\delta_d(f^{\oplus t}) \geq 1/2 - 2^{-\Omega(t/4^d)}$. Their proof proceeded by studying the rejection probabilities of T_* on the functions f and $f^{\oplus t}$. The analysis of the rejection probability of T_* given by [AKK$^+$05] was a central ingredient in their proof. By using our improved analysis of the rejection probability of T_* from Theorem 1 instead, we get the following improvement.

Theorem 3 *Let ϵ_1 be as in Theorem 1. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Then*

$$\delta_d(f^{\oplus t}) \geq \frac{1 - (1 - 2\min\{\epsilon_1/4, 2^{d-2} \cdot \delta_d(f)\})^{t/2^d}}{2}.$$

In particular, if $\delta_d(f) \geq 0.1$, then $\delta_d(f^{\oplus t}) \geq 1/2 - 2^{-\Omega(t/2^d)}$.

1.6 Technique

The heart of our proof of the main theorem (Theorem 1) is an inductive argument on n, the dimension of the ambient space. While proofs that use induction on n have been used before in the literature on low-degree testing (see, for instance, [BFL91, BFLS91, FGs$^+$96]), they tend to have a performance guarantee that degrades significantly with n. Indeed no inductive proof was known even for the case of testing linearity of functions from $\mathbb{F}_2^n \to \mathbb{F}_2$ that showed that functions at $\Omega(1)$ distance from linear functions are rejected with $\Omega(1)$ probability. (We note that the original analysis of [BLR93] as well as the later analysis of [BCH$^+$96] do give such bounds - but they do not use induction on n.) In the process of giving a tight analysis of the [AKK$^+$05] test for Reed-Muller codes, we thus end up giving a new (even if weaker) analysis of the linearity test over $\mathbb{F}_2^n$. Below we give the main idea behind our proof.

Consider a function f that is δ-far from every degree d polynomial. For a "hyperplane", i.e., an $(n-1)$-dimensional affine subspace A of $\mathbb{F}_2^n$, let $f|_A$ denote the restriction of f to A. We first note that the test can be interpreted as first picking a random hyperplane A in $\mathbb{F}_2^n$ and then picking a random $(d+1)$-dimensional affine subspace A' within A and testing if $f|_{A'}$ is a degree d polynomial. Now, if on every hyperplane A, $f|_A$ is still δ-far from degree d polynomials then we would be done by the inductive hypothesis. In fact our hypothesis gets weaker as $n \to \infty$, so that we can even afford a few hyperplanes where $f|_A$ is not δ-far. The crux of our analysis is when $f|_A$ is close to some degree d polynomial P_A for several (but just $O(2^d)$) hyperplanes. In this case we manage to "sew" the different polynomials P_A (each defined on some $(n-1)$-dimensional subspace within $\mathbb{F}_2^n$)

into a degree d polynomial P that agrees with *all* the P_A's. We then show that this polynomial is close to f, completing our argument.

To stress the novelty of our proof, note that this is not a "self-correction" argument as in [AKK$^+$05], where one defines a natural function that is close to P, and then works hard to prove it is a polynomial of appropriate degree. In contrast, our function is a polynomial by construction and the harder part (if any) is to show that the polynomial is close to f. Moreover, unlike other inductive proofs, our main gain is in the fact that the new polynomial P has degree no greater than that of the polynomials given by the induction.

Organization of this paper: We prove our main theorem, Theorem 1, in Section 2 assuming three lemmas, two of which study the rejection probability of the k-dimensional affine subspace test, and another that relates the rejection probability of the basic $(d+1)$-dimensional affine subspace test to that of the k-dimensional affine subspace test. These three lemmas are proved in the following section, Section 3. We give the relationship to the Gowers norm in Section 4, and we prove our improved hardness amplification theorem, Theorem 3, in Section 5. Finally, we show the tightness of our main theorem in the appendix.

2 Proof of Main Theorem

In this section we prove Theorem 1. We start with an overview of our proof. Recall that a k-flat is an affine subspace of dimension k, and a hyperplane is an $(n-1)$-flat.

The proof of the main theorem proceeds as follows. We begin by studying a variant of the basic tester T_*, which we call $T_{d,k}$ or the *k-flat test*. For an integer $k \geq d+1$, $T^f_{d,k}$ picks a uniformly random k-flat in $\mathbb{F}_2^n$, and accepts if and only if the restriction of f to that flat has degree at most d. In this language, the tester T_* of interest to us is $T_{d,d+1}$. To prove Theorem 1, we first show that for $k \approx d+10$, the tester $T^f_{d,k}$ rejects with constant probability if $\delta_d(f)$ is $\Omega(2^{-d})$ (see Lemma 7). We then relate the rejection probabilities of $T^f_{d,k}$ and T^f_* (see Lemma 8).

The central ingredient in our analysis is thus Lemma 7 which is proved by induction on n, the dimension of the ambient space. Recall that we want to show that the two quantities (1) $\delta_d(f)$ and (2) $\Pr[T^f_{d,k}$ rejects], are closely related. We consider what happens to f when restricted to some hyperplane A. Denote such a restriction by $f|_A$. For a hyperplane A we consider the corresponding two quantities (1) $\delta_d(f|_A)$ and (2) $\Pr[T^{f|_A}_{d,k}$ rejects]. The inductive hypothesis tells us that these two quantities are closely related for each A. Because of the local nature of tester $T_{d,k}$, it follows easily that $\Pr[T^f_{d,k}$ rejects] is the average of $\Pr[T^{f|_A}_{d,k}$ rejects] over all hyperplanes A. The main technical content of Lemma 7 is that there is a similar tight relationship between $\delta_d(f)$ and the numbers $\delta_d(f|_A)$ as A varies over all hyperplanes A. This relationship suffices to complete the proof. The heart of our analysis focuses on the case where for many hyperplanes (about 2^k of them, independent of n), the quantity $\delta_d(f|_A)$ is very small (namely, for many A, there is a polynomial P_A of degree d that is very close to $f|_A$). In this case, we show how to "sew" together the polynomials P_A to get a polynomial P on $\mathbb{F}_2^n$ that is also very close to f. In contrast to prior approaches which yield a polynomial P with larger degree than that of the P_A's, our analysis crucially preserves this degree, leading to the eventual tightness of our analysis.

We now turn to the formal proof.

2.1 Preliminaries

We begin by formally introducing the k-flat test and some related notation.

Definition 4 (k-flat test $T_{d,k}$) *The test $T^f_{d,k}$ picks a random k-flat $A \subseteq \mathbb{F}_2^n$ and accepts if and only if $f|_A$ (f restricted to A) is a polynomial of degree at most d.*

The rejection probability of $T^f_{d,k}$ is denoted $\mathsf{Rej}_{d,k}(f)$. In words, this is the probability that $f|_A$ is not a degree d polynomial when A is chosen uniformly at random among all k-flats of F_2^n.

Although we don't need it for our argument, we note that $T_* = T_{d,d+1}$ accepts if and only if the 2^{d+1} evaluations $f|_A$ sum to 0.

The following folklore proposition shows that for $k \geq d+1$, $T_{d,k}$ has perfect completeness.

Proposition 5 *For every $k \geq d+1$, $\delta_d(f) = 0$ if and only if $\mathsf{Rej}_{d,k}(f) = 0$.*

2.2 Key Lemmas

We now state our three key lemmas, and then use them to finish the proof of Theorem 1. The first is a simple lemma that says if the function is sufficiently close to a degree d polynomial, then the rejection probability is linear in its distance from degree d polynomials.

Lemma 6 *For every k, ℓ, d such that $k \geq \ell \geq d+1$, if $\delta(f) = \delta$ then $\mathsf{Rej}_{d,k}(f) \geq 2^\ell \cdot \delta \cdot (1-(2^\ell-1)\delta)$. In particular, if $\delta \leq 2^{-(d+2)}$ then $\mathsf{Rej}_{d,k}(f) \geq \min\{\frac{1}{8}, 2^{k-1} \cdot \delta\}$.*

The next lemma is the heart of our analysis and allows us to lower bound the rejection probability when the function is bounded away from degree d polynomials.

Lemma 7 *There exist positive constants $\beta < 1/4, \epsilon_0, \gamma$ and c such that the following holds for every d, k, n, such that $n \geq k \geq d+c$. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $\delta(f) \geq \beta \cdot 2^{-d}$. Then $\mathsf{Rej}_{d,k}(f) \geq \epsilon_0 + \gamma \cdot 2^d/2^n$.*

The final lemma relates the rejection probabilities of different dimensional tests.

Lemma 8 *For every n, d and $k \geq k' \geq d+1$, and every $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we have*

$$\mathsf{Rej}_{d,k'}(f) \geq \mathsf{Rej}_{d,k}(f) \cdot 2^{-(k-k')}.$$

Given the three lemmas above, Theorem 1 follows easily as shown below.

Proof of Theorem 1: Let ϵ_0 and c be as in Lemma 7. We prove the theorem for $\epsilon_1 = \epsilon_0 \cdot 2^{-(c-1)}$. First note that if $\delta(f) \leq 2^{-(d+2)}$, then we are done by Lemma 6. So assume $\delta(f) \geq 2^{-(d+2)} \geq \beta \cdot 2^{-d}$, where β is the constant from Lemma 7. By Lemma 7, we know that $\mathsf{Rej}_{d,d+c}(f) \geq \epsilon_0$. Lemma 8 now implies that $\mathsf{Rej}_{d,d+1}(f) \geq \epsilon_0 \cdot 2^{-(c-1)}$, as desired. ■

3 Analysis of the k-flat test

Throughout this section we fix d, so we suppress it in the subscripts and simply use $\delta(f) = \delta_d(f)$ and $\mathsf{Rej}_k(f) = \mathsf{Rej}_{d,k}(f)$.

3.1 Lemma 6: When f is close to $\mathrm{RM}(d, n)$

Recall that we wish to prove

Lemma 6 (recalled): *For every k, ℓ, d such that $k \geq \ell \geq d + 1$, if $\delta(f) = \delta$ then $\mathsf{Rej}_k(f) \geq 2^\ell \cdot \delta \cdot (1 - (2^\ell - 1)\delta)$. In particular, if $\delta \leq 2^{-(d+2)}$ then $\mathsf{Rej}_k(f) \geq \min\{\frac{1}{8}, 2^{k-1} \cdot \delta\}$.*

Proof of Lemma 6: The main idea is to show that with good probability, the flat will contain exactly one point where f and the closest degree d polynomial differ, in which case the test will reject. The main claim we prove is that $\mathsf{Rej}_\ell(f) \geq 2^\ell \cdot \delta \cdot (1 - (2^\ell - 1)\delta)$. The first part then follows from the monotonicity of the rejection probability, i.e., $\mathsf{Rej}_k(f) \geq \mathsf{Rej}_\ell(f)$ if $k \geq \ell$. The second part follows by setting $\ell = k$ if $\delta \leq 2^{-(k+1)}$ and ℓ such that $2^{-(\ell+2)} < \delta \leq 2^{-(\ell+1)}$ otherwise. In the former case, we get $\mathsf{Rej}_k(f) \geq 2^{-(k-1)} \cdot \delta$ while in the latter case we get $\mathsf{Rej}_k(f) \geq \mathsf{Rej}_\ell(f) \geq \frac{1}{8}$. We thus turn to proving $\mathsf{Rej}_\ell(f) \geq 2^\ell \cdot \delta \cdot (1 - (2^\ell - 1)\delta)$.

Let $g \in \mathrm{RM}(d, n)$ be a polynomial achieving $\delta(f) = \delta(f, g)$. Consider a random ℓ-flat A of $\mathbb{F}_2^n$. We think of the points of A as generated by picking a random full-rank matrix $M \in \mathbb{F}_2^{n \times \ell}$ and a random vector $b \in \mathbb{F}_2^n$, and then letting $A = \{a_x \stackrel{\mathrm{def}}{=} Mx + b \mid x \in \mathbb{F}_2^\ell\}$. Thus the points of A are indexed by elements of $\mathbb{F}_2^\ell$.

For $x \in \mathbb{F}_2^\ell$, let E_x be the event that "$f(a_x) \neq g(a_x)$". Further let F_x be the event that "$f(a_x) \neq g(a_x)$ and $f(a_y) = g(a_y)$ for every $y \neq x$". We note that if any of the events F_x occurs (for $x \in \mathbb{F}_2^\ell$), then the ℓ-flat test rejects f. This is because distinct degree d polynomials differ in at least 2^{-d} fraction of points, so they cannot differ in exactly one point if $\ell > d$.

We now lower bound the probability of $\cup_x F_x$. Using the fact that a_x is distributed uniformly over $\mathbb{F}_2^n$ and a_y is distributed uniformly over $\mathbb{F}_2^n - \{a_x\}$, we note that $\Pr[E_x] = \delta$ and $\Pr[E_x \text{ and } E_y] \leq \delta^2$. We also have $\Pr[F_x] \geq \Pr[E_x] - \sum_{y \neq x} \Pr[E_x \text{ and } E_y] \geq \delta - (2^\ell - 1) \cdot \delta^2$. Finally, noticing that the events F_x are mutually exclusive we have that $\Pr[\cup_i F_i] = \sum_i \Pr[F_i] \geq 2^\ell \cdot \delta \cdot (1 - (2^\ell - 1) \cdot \delta)$, as claimed.

■

3.2 Lemma 7: When f is bounded away from $\mathrm{RM}(d, n)$

The main idea of the proof of Lemma 7 is to consider the restrictions of f on randomly chosen "hyperplanes", i.e., $(n-1)$-flats. If on an overwhelmingly large (to be quantified in the proof) fraction of hyperplanes, our function is far from degree d polynomials, then the inductive hypothesis suffices to show that f will be rejected with high probability (by the k-flat test). The interesting case is when the restrictions of f to several hyperplanes are close to degree d polynomials. In Lemma 10 we use the close polynomials on such hyperplanes to construct a polynomial that has significant agreement with f on the union of the hyperplanes.

We start by first fixing some terminology. We say A and B are *complementary* hyperplanes if $A \cup B = \mathbb{F}_2^n$. Recalling that a hyperplane is the set of points $\{x \in \mathbb{F}_2^n | L(x) = b\}$ where $L : \mathbb{F}_2^n \to \mathbb{F}_2$ is a nonzero linear function and $b \in \mathbb{F}_2$, we refer to L as the linear part of the hyperplane. We say that hyperplanes $A_1, \ldots, A_\ell$ are linearly independent if the corresponding linear parts are independent. The following proposition lists some basic facts about hyperplanes that we use. The proof is omitted.

Proposition 9 (Properties of hyperplanes)

1. *There are exactly $2^{n+1} - 2$ distinct hyperplanes in $\mathbb{F}_2^n$.*
2. *Among any $2^\ell - 1$ distinct hyperplanes, there are at least ℓ independent hyperplanes.*
3. *There is an affine invertible transform that maps independent hyperplanes $A_1, \ldots, A_\ell$ to the hyperplanes $x_1 = 0, x_2 = 0, \ldots, x_\ell = 0$.*

We are now ready to prove Lemma 7. We first recall the statement.

Lemma 7 (recalled): *There exist positive constants $\beta < 1/4, \epsilon_0, \gamma$ and c such that the following holds for every d, k, n, such that $n \geq k \geq d + c$. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $\delta(f) \geq \beta \cdot 2^{-d}$. Then $\mathsf{Rej}_{d,k}(f) \geq \epsilon_0 + \gamma \cdot 2^d/2^n$.*

Proof of Lemma 7: We prove the lemma for every $\beta < 1/24$, $\epsilon_0 < 1/8$, $\gamma \geq 72$, and c such that $2^c \geq \max\{4\gamma/(1 - 8\epsilon_0), \gamma/(1 - \epsilon_0), 2/\beta\}$. (In particular, the choices $\beta = 1/25$, $\epsilon_0 = 1/16$, $\gamma = 72$ and $c = 10$ work.)

The proof uses induction on $n - k$. When $n = k$ we have $\mathsf{Rej}_k(f) = 1 \geq \epsilon_0 + \gamma \cdot 2^{d-k}$ as required, because $2^c \geq \frac{\gamma}{1-\epsilon_0}$. So we move to the inductive step.

Let $\mathcal{H}$ denote the set of hyperplanes in $\mathbb{F}_2^n$. Let $N = 2(2^n - 1)$ be the cardinality of $\mathcal{H}$. Let $\mathcal{H}^*$ be the set of all the hyperplanes $A \in \mathcal{H}$ such that $\delta(f|_A, \mathrm{RM}(d, n-1)) < \beta \cdot 2^{-d}$. Let $K = |\mathcal{H}^*|$.

Now because a random k-flat of a random hyperplane is a random k- flat:

$$\mathsf{Rej}_k(f) = \mathbb{E}_{A \in \mathcal{H}}[\mathsf{Rej}_k(f|_A)].$$

By the induction hypothesis, for any $A \in \mathcal{H} \setminus \mathcal{H}^*$, we have

$$\mathsf{Rej}_k(f|_A) \geq \epsilon_0 + \gamma \cdot \frac{2^d}{2^{n-1}}.$$

Thus,

$$\mathsf{Rej}_k(f) \geq \epsilon_0 + \gamma \cdot \frac{2^d}{2^{n-1}} - K/N.$$

We now take cases on whether K is large or small:

1. **Case 1:** $K \leq \gamma \cdot 2^d$.
   In this case, $\mathsf{Rej}_k(f) \geq \epsilon_0 + \gamma \cdot 2^d/2^{n-1} - K/N \geq \epsilon_0 + \gamma \cdot 2^d/2^n$ as desired.

2. **Case 2:** $K > \gamma \cdot 2^d$.

   Lemma 10 (below) shows that in this case, $\delta(f) \leq \frac{3}{2}\beta \cdot 2^{-d} + 9/(\gamma 2^d) \stackrel{\text{def}}{=} \delta_0$, provided $\beta \cdot 2^{-d} < 2^{-(d+2)}$ (which holds since $\beta < 1/24 < 1/4$).

   Note that since $\beta < 1/24$ and $9/\gamma < 1/8$, we get $\delta_0 < 2^{-(d+2)}$ and so Lemma 6 implies that $\mathsf{Rej}_k(f) \geq \min\{2^{k-1} \cdot \delta(f), \frac{1}{8}\} \geq \min\{2^{k-1} \cdot \beta \cdot 2^{-d}, \frac{1}{8}\}$. We verify both quantities above are at least $\epsilon_0 + \gamma/2^{(c+1)} \geq \epsilon_0 + \gamma 2^d/2^n$. The condition $1/8 > \epsilon_0 + \gamma/2^{c+1}$ follows from the fact that $2^c \geq 4\gamma/(1 - 8\epsilon_0)$. To verify the second condition, note that $2^{k-1} \cdot \beta \cdot 2^{-d} \geq 2^{c-1}\beta \geq 1$ since $2^c \geq 2/\beta$.

We thus conclude that the rejection probability of f is at least $\epsilon_0 + \gamma \cdot 2^d/2^n$ as claimed. ■

Lemma 10 *For $f : \mathbb{F}_2^n \to \mathbb{F}_2$, let $A_1, \ldots, A_K$ be hyperplanes such that $f|_{A_i}$ is α-close to some degree d polynomial on A_i. If $K > 2^{d+1}$ and $\alpha < 2^{-(d+2)}$, then $\delta(f) \leq \frac{3}{2}\alpha + 9/K$.*

Proof Let P_i be the degree d polynomial such that $f|_{A_i}$ is α-close to P_i.

Claim 11 *If $4\alpha < 2^{-d}$ then for every pair of hyperplanes A_i and A_j, we have $P_i|_{A_i \cap A_j} = P_j|_{A_j \cap A_i}$.*

Proof If A_i and A_j are complementary then this is vacuously true. Otherwise, $|A_i \cap A_j| = |A_i|/2 = |A_j|/2$. So $\delta(f|_{A_i \cap A_j}, P_i|_{A_i \cap A_j}) \leq 2\delta(f|_{A_i}, P_i) \leq 2\alpha$ and similarly $\delta(f|_{A_i \cap A_j}, P_j|_{A_i \cap A_j}) \leq 2\alpha$. So $\delta(P_i|_{A_i \cap A_j}, P_j|_{A_i \cap A_j}) \leq 4\alpha < 2^{-d}$. But these are both degree d polynomials and so if their proximity is less than 2^{-d} then they must be identical. ■

Let $\ell = \lfloor \log_2(K + 1) \rfloor$. Thus $\ell > d$. By Proposition 9 there are at least ℓ linearly independent hyperplanes among $A_1, \ldots, A_K$. Without loss of generality let these be $A_1, \ldots, A_\ell$. Furthermore, by an affine transformation of coordinates, for $i \in [\ell]$ let A_i be the hyperplane $\{x \in \mathbb{F}_2^n \mid x_i = 0\}$. For $i \in [\ell]$ extend P_i to a function on all of $\mathbb{F}_2^n$ by making P_i independent of x_i. We will sew together $P_1, \ldots, P_\ell$ to get a polynomial close to f.

Let us write all functions from $\mathbb{F}_2^n \to \mathbb{F}_2$ as polynomials in n variables $x_1, \ldots, x_\ell$ and $\mathbf{y}$ where $\mathbf{y}$ denotes the last $n - \ell$ variables. For $i \in [\ell]$ and $S \subseteq [\ell]$, let $P_{i,S}(\mathbf{y})$ be the monomials of P_i which contain x_i for $i \in S$, and no x_j for $j \in [\ell]$§. That is, $P_{i,S}(\mathbf{y})$ are polynomials such that $P_i(x_1, \ldots, x_\ell, \mathbf{y}) = \sum_{S \subseteq [\ell]} P_{i,S}(\mathbf{y}) \prod_{j \in S} x_j$. Note that the degree of $P_{i,S}$ is at most $d - |S|$. (In particular, if $|S| > d$, then $P_{i,S} = 0$.) Note further that since P_i is independent of x_i, we have that $P_{i,S} = 0$ if $i \in S$.

Claim 12 *For every $S \subseteq [\ell]$ and every $i, j \in [\ell] - S$, $P_{i,S}(\mathbf{y}) = P_{j,S}(\mathbf{y})$.*

Proof Note that $P_i|_{A_i \cap A_j}(\mathbf{x}, \mathbf{y}) = \sum_{S \subseteq [\ell] - \{i,j\}} P_{i,S}(\mathbf{y}) \prod_{m \in S} x_m$. Similarly $P_j|_{A_i \cap A_j}(\mathbf{x}, \mathbf{y}) = \sum_{S \subseteq [\ell] - \{i,j\}} P_{j,S}(\mathbf{y}) \prod_{m \in S} x_m$. Since the two functions are equal (by Claim 11), we have that every pair of coefficients of $\prod_{m \in S} x_m$ must be the same. We conclude that $P_{i,S} = P_{j,S}$. ■

Claim 12 above now allows us to define, for every $S \subsetneq [\ell]$, the polynomial $P_S(\mathbf{y})$ as the unique polynomial $P_{i,S}$ where $i \notin S$. We define

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

By construction, the degree of P is at most d. This is the polynomial that we will eventually show is close to f.

Claim 13 *For every $i \in [K]$, $P|_{A_i} = P_i|_{A_i}$.*

Proof First note that for each $i \in [\ell]$, $P|_{A_i} = P_i|_{A_i}$. This is because the coefficients of the two polynomials become identical after substituting $x_i = 0$ (recall that A_i is the hyperplane $\{x \in \mathbb{F}_2^n \mid x_i = 0\}$).

Now consider general $i \in [K]$. For any point $x \in A_i \cap (\bigcup_{j=1}^{\ell} A_j)$, letting $j^* \in [\ell]$ be such that $x \in A_{j^*}$, we have $P_i(x) = P_{j*}(x)$ (by Claim 11) and $P_{j^*}(x) = P(x)$ (by what we just showed, since $j^* \in [\ell]$). Thus P and P_i agree on all points in $A_i \cap (\bigcup_{j=1}^{\ell} A_j)$. Now since $\ell > d$, we have that $|A_i \cap (\bigcup_{j=1}^{\ell} A_j)|/|A_i| \geq 1 - 2^{-\ell} > 1 - 2^{-d}$, and since $P|_{A_i}$ and $P_i|_{A_i}$ are both degree d polynomials, we conclude that $P|_{A_i}$ and $P_i|_{A_i}$ are identical. Thus for all $i \in [K]$, $P|_{A_i} = P_i|_{A_i}$. ■

We will show below that P is close to f, by considering all the hyperplanes $A_1, \ldots, A_K$. If these hyperplanes uniformly covered F_2^n, then we could conclude $\delta(f, P) \leq \alpha$, as f is α-close to P on each hyperplane. Since the A_i don't uniformly cover $\mathbb{F}_2^n$, we'll argue that almost all points are covered approximately the right number of times, which will be good enough. To this end, let

$$\textsc{Bad} = \{z \in \mathbb{F}_2^n | z \text{ is contained in less than } K/3 \text{ of the hyperplanes } A_1, \ldots, A_K\}.$$

Let $\tau = |\textsc{Bad}|/2^n$.

Claim 14 *$\delta(f, P) \leq 3/2 \cdot \alpha + \tau$.*

Proof Consider the following experiment: Pick $z \in \mathbb{F}_2^n$ and $i \in [K]$ uniformly and independently at random and consider the probability that "$z \in A_i$ and $f(z) \neq P_i(z)$".

On the one hand, we have

$$\begin{aligned}
&\Pr_{z,i}[z \in A_i \ \&\ f(z) \neq P_i(z)] \\
&\leq \max_i \Pr_z[z \in A_i] \cdot \Pr_z[f(z) \neq P_i(z) | z \in A_i] \\
&\leq \frac{1}{2} \cdot \alpha
\end{aligned}$$

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

$$
\begin{aligned}
&\Pr_{z,i}[z \in A_i \ \&\ f(z) \neq P_i(z)] \\
&= \Pr_{z,i}[z \in A_i \ \&\ f(z) \neq P(z)] \\
&\geq \Pr_{z,i}[z \in A_i \ \&\ f(z) \neq P(z) \ \&\ z \notin \textsc{Bad}] \\
&= \Pr_{z}[f(z) \neq P(z) \ \&\ z \notin \textsc{Bad}] \cdot \Pr_{z,i}[z \in A_i | f(z) \neq P(z) \ \&\ z \notin \textsc{Bad}] \\
&\geq \Pr_{z}[f(z) \neq P(z) \ \&\ z \notin \textsc{Bad}] \cdot \min_{z: z \notin \textsc{Bad} \ \&\ f(z) \neq P(z)} \Pr_{i}[z \in A_i] \\
&\geq (\delta(f,P) - \tau) \cdot \min_{z: z \notin \textsc{Bad}} \Pr_{i}[z \in A_i] \\
&\geq (\delta(f,P) - \tau) \cdot \frac{1}{3}
\end{aligned}
$$

We thus conclude that $(\delta(f,P) - \tau)/3 \leq \alpha/2$ yielding the claim. ■

Claim 15 $\tau \leq 9/K$.

Proof The proof is a straightforward "pairwise independence" argument, with a slight technicality to handle complementary hyperplanes.

Consider a random variable z distributed uniformly over $\mathbb{F}_2^n$. For $i \in [K]$, let Y_i denote the random variable that is $+1$ if $z \in A_i$ and -1 otherwise. Note that $z \in \textsc{Bad}$ if and only if $\sum_i Y_i \leq -K/3$ and so $\tau = \Pr[\sum_i Y_i \leq -K/3]$. We now bound this probability.

For every i, note that $E[Y_i] = 0$ and $\mathrm{Var}[Y_i] = 1$. Notice further that if A_i and A_j are not complementary hyperplanes, then Y_i and Y_j are independent and so $E[Y_i Y_j] = 0$, while if they are complementary, then $E[Y_i Y_j] = -1 \leq 0$. We conclude that $E[\sum_i Y_i] = 0$ and $\mathrm{Var}[\sum_i Y_i] \leq K$. Using Chebychev's bound, we conclude that $\tau = \Pr[\sum_i Y_i \leq -K/3] \leq \mathrm{Var}(\sum_i Z_i)/(K^2/9) \leq 9/K$. ■

The lemma follows from the last two claims above. ■

3.3 Lemma 8: Relating different dimensional tests

Lemma 16 *Let $k \geq d+1$ and let $f : \mathbb{F}_2^{k+1} \to \mathbb{F}_2$ have degree greater than d. Then* $\mathsf{Rej}_{d,k}(f) \geq 1/2$.

Proof Assume for contradiction that there is a strict majority of hyperplanes A on which $f|_A$ has degree d. Then there exists two complementary hyperplanes A and $\bar{A}$ such that $f|_A$ and $f|_{\bar{A}}$ both have degree d. We can interpolate a polynomial P of degree at most $d+1$ that now equals f everywhere. If P is of degree d, we are done, so assume P has degree exactly $d+1$ and let P_h be the homogenous degree $d+1$ part of P (i.e,, $P = P_h + Q$ where $\deg(Q) \leq d$ and P_h is homogenous). Now consider all hyperplanes A such that $f|_A = P|_A$ has degree at most d. Since these form a strict majority, there are at least $\frac{1}{2}(2^{k+2}-2)+1 > 2^{k+1}-1$ such hyperplanes. It follows that there are at least $k+1 \geq d+2$ linearly independent hyperplanes such that this condition holds. By an affine transformation we can assume these hyperplanes are of the form $x_1 = 0, \ldots, x_{d+2} = 0$. But then $\prod_{i=1}^{d+2} x_i$ divides P_h which contradicts the fact that the degree of P_h is at most $d+1$. ■

Lemma 17 *Let $n \geq k \geq d+1$ and let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ have degree greater than d. Then $\mathsf{Rej}_{d,k}(f) \geq 2^{k-n}$.*

Proof The proof is a simple induction on n. The base case of $n = k$ is trivial. Now assume for $n-1$. Pick a random hyperplane A. With probability at least $1/2$ (by the previous lemma), $f|_A$ is not a degree d polynomial. By the inductive hypothesis, a random k-flat of A will now detect that $f|_A$ is not of degree d with probability 2^{k-n+1}. We conclude that a random k-flat of $\mathbb{F}_2^n$ yields a function of degree greater than d with probability at least 2^{k-n}. ■

We now have all the pieces needed to prove Lemma 8.

Lemma 8 (recalled): *For every n, d and $k \geq k' \geq d+1$, and every $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we have*

$$\mathsf{Rej}_{d,k'}(f) \geq \mathsf{Rej}_{d,k}(f) \cdot 2^{-(k-k')}.$$

Proof of Lemma 8: We view the k'-flat test as the following process: first pick a random k-flat A_1 of $\mathbb{F}_2^n$, then pick a random k'-flat A of A_1, and accept iff $f|_A$ is a degree d polynomial. Note that this is completely equivalent to the k'-flat test.

To analyze our test, we first consider the event that $f|_{A_1}$ is not a degree d polynomial. The probability that this happens is $\mathsf{Rej}_{d,k}(f)$. Now conditioned on the event that $f|_{A_1}$ is not a degree d polynomial, we can now use Lemma 17 to conclude that the probability that $(f|_{A_1})|_A$ is not a degree d polynomial is at least $2^{-(k-k')}$. We conclude that $\mathsf{Rej}_{d,k'}(f) \geq \mathsf{Rej}_{d,k}(f) \cdot 2^{-(k-k')}$. The lemma follows. ■

4 Gowers norms

Our main theorem can be interpreted as giving a tight relationship between the Gowers norm of a function f and its proximity to some low degree polynomial. In this section we describe this relationship.

We start by recalling the definition of the test $T_0^f(k)$ and the Gowers norm $\|f\|_{U^k}$. On oracle access to function f, the test $T_0(k)$ picks x_0 and directions $a_1, \ldots, a_k$ uniformly and independently in $\mathbb{F}_2^n$ and accepts if and only if $f|_A$ is a degree $k-1$ polynomial, where $A = \{x_0 + \mathrm{span}(a_1, \ldots, a_k)\}$. The Gowers norm is given by the expression

$$\|f\|_{U^k} \stackrel{\mathrm{def}}{=} (\Pr[T_0^f(k) \text{ accepts}] - \Pr[T_0^f(k) \text{ rejects}])^{\frac{1}{2^k}}.$$

Our main quantity of interest is the correlation of f with degree d polynomials, i.e., the quantity $1 - 2\delta_d(f)$.

Our theorem relating the Gowers norm to the correlation is given below.

Theorem 18 *There exists $\epsilon > 0$ such that if $\|f\|_{U^{d+1}} \geq 1 - \epsilon/2^d$, then $\delta_d(f) = \Theta(1 - \|f\|_{U^{d+1}})$.*

To prove the theorem we first relate the rejection probability of the test T_0 with that of the test T_*.

Proposition 19 *For every $n \geq d+1$ and for every f, $\Pr[T_0^f(d+1)\text{rejects}] \geq \frac{1}{4} \cdot \Pr[T_*^f\text{rejects}]$.*

Proof We show that with probability at least 1/4, the a_i are linearly independent. Consider picking d independent vectors $a_1, \ldots, a_d$ in $\mathbb{F}_2^n$. For fixed $\beta_1, \ldots, \beta_d \in \mathbb{F}_2$ (not all zero), the probability that $\sum_i \beta_i a_i = 0$ is at most 2^{-n}. Taking the union bound over all sequences $\beta_1, \ldots, \beta_d$ we find that the probability that $a_1, \ldots, a_d$ have a linear dependency is at most $2^{d-n} \geq \frac{1}{2}$ if $n \geq d+1$. For any fixed $a_1, \ldots, a_d$, the probability that $a_{d+1} \in \text{span}(a_1, \ldots, a_d)$ is also at most $\frac{1}{2}$. Thus we find with probability at least 1/4, the vectors $a_1, \ldots, a_{d+1}$ are linearly independent provided $n \geq d+1$. The proposition follows since the rejection probability of $T_0^f(d+1)$ equals the rejection probability of T_*^f times the probability that $a_1, \ldots, a_{d+1}$ are linearly independent. ■

We are now ready to prove Theorem 18.

Proof of Theorem 18: The proof is straightforward given our main theorem and the work of Gowers et al. [Gow01, GT05]. As mentioned earlier, Gowers already showed that $1 - 2\delta_d(f) \leq \|f\|_{U^{d+1}}$ [Gow01, GT05], i.e., $\delta_d(f) \geq (1 - \|f\|_{U^{d+1}})/2$.

For the other direction, suppose $\|f\|_{U^{d+1}} = 1 - \gamma$, where $\gamma \leq \epsilon/2^d$ for small enough ϵ. Let ρ denote the rejection probability of $T_0^f(d+1)$. By Proposition 19 we have $\rho \geq \frac{1}{4} \cdot \mathsf{Rej}_{d,d+1}(f)$. By choosing ϵ small enough, we also have $1 - 2\rho = \|f\|_{U^{d+1}}^{2^{d+1}} > 1 - \epsilon_1/2$, i.e., $\rho < \epsilon_1/4$, so $\mathsf{Rej}_{d,d+1}(f) < \epsilon_1$. Thus, by Theorem 1,

$$\begin{aligned}
\delta_d(f) &\leq \frac{1}{2^d}\mathsf{Rej}_{d,d+1}(f)(f) \\
&\leq \frac{1}{2^{d-2}}\rho \\
&= \frac{1}{2^{d-1}}(1 - \|f\|_{U^{d+1}}^{2^{d+1}}) \\
&= \frac{1}{2^{d-1}}(1 - (1-\gamma)^{2^{d+1}}) \\
&\leq \frac{1}{2^{d-1}}(1 - (1 - O(2^{d+1}\gamma))) \\
&= O(\gamma),
\end{aligned}$$

as required. ■

5 XOR lemma for low-degree polynomials

A crucial feature of the test T_0 (that is not a feature of the k-flat test for $k > d+1$) is that the rejection probability of $f^{\oplus t}$ can be exactly expressed as a rapidly growing (in t) function of the rejection probability of f. Let $\mathsf{Rej}_d^0(f)$ denote the rejection probability of $T_0^f(d+1)$. Then we have:

Proposition 20

$$(1 - 2\mathsf{Rej}_d^0(f^{\oplus t})) = (1 - 2\mathsf{Rej}_d^0(f))^t.$$

Proof We first note that the proposition is equivalent to showing that $\|f^{\oplus t}\|_{U^{d+1}} = (\|f\|_{U^{d+1}})^t$. It is a standard fact (e.g., Fact 2.6 in [VW07]) that for functions f, g on disjoint sets of inputs, $\|f(x) + g(y)\|_{U^{d+1}} = \|f(x)\|_{U^{d+1}} \cdot \|g(y)\|_{U^{d+1}}$. This immediately yields the proposition. ■

We also use the following well-known relationship between the Gowers norm and the correlation of a function to the class of degree d polynomials. (We state it in terms of the rejection probability of the test T_0.)

Lemma 21 ([Gow01, GT05])

$$1 - 2\delta_d(g) \leq (1 - 2\mathsf{Rej}_d^0(g))^{\frac{1}{2^d}}.$$

We are now ready to prove Theorem 3 which we recall below.

Theorem 3 (recalled): *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Then*

$$\delta_d(f^{\oplus t}) \geq \frac{1 - (1 - 2\min\{\epsilon_1/4, 2^{d-2} \cdot \delta_d(f)\})^{t/2^d}}{2}.$$

In particular, if $\delta_d(f) \geq 0.1$, then $\delta_d(f^{\oplus t}) \geq \frac{1 - 2^{-\Omega(t/2^d)}}{2}$.

Proof of Theorem 3: By Theorem 1 and Proposition 19,

$$\mathsf{Rej}_d^0(f) \geq \min\{\epsilon_1/4, 2^{d-2} \cdot \delta_d(f)\}.$$

Thus by Proposition 20,

$$(1 - 2\mathsf{Rej}_d^0(f^{\oplus t}))^{\frac{1}{2^d}} = (1 - 2\mathsf{Rej}_d^0(f))^{\frac{t}{2^d}} \leq (1 - 2\min\{\epsilon_1/4, 2^{d-2} \cdot \delta_d(f)\})^{\frac{t}{2^d}}.$$

Finally, Lemma 21 shows that

$$\delta_d(f^{\oplus t}) \geq \frac{1 - \left(1 - 2\min\{\epsilon_1/4, 2^{d-2} \cdot \delta_d(f)\}\right)^{\frac{t}{2^d}}}{2}.$$

■

Acknowledgments

Thanks to Alex Samorodnitsky and Shachar Lovett for sharing some of the unpublished parts of their work [LMS08] and allowing us to present parts of their proof in Appendix A. Thanks to Alex also for numerous stimulating discussions from the early stages of this work, and to Jakob Nordström for bringing some of the authors together on this work.

References

[AB01] Noga Alon and Richard Beigel. Lower bounds for approximations by low degree polynomials over Z_m. In *IEEE Conference on Computational Complexity*, pages 184–187, 2001.

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

[BCH+96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos Kiwi, and Madhu Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, November 1996.

[BCJ+06] Morgan V. Brown, Neil J. Calkin, Kevin James, Adam J. King, Shannon Lockard, and Robert C. Rhoades. Trivial Selmer groups and even partitions of a graph. *INTEGERS*, 6, December 2006.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the 23rd ACM Symposium on the Theory of Computing*, pages 21–32, New York, 1991. ACM Press.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993. Earlier version in STOC'90.

[BM88] Richard P. Brent and Brendan D. McKay. On determinants of random symmetric matrices over $\mathbb{Z}_m$. *ARS Combinatoria*, 26A:57 – 64, 1988.

[FGs+96] Uriel Feige, Shafi Goldwasser, László Lová sz, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[GKNR08] Oded Goldreich, Michael Krivelevich, Ilan Newman, and Eyal Rozenberg. Hierarchy theorems for property testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(097), 2008.

[Gow98] William T. Gowers. A new proof of Szeméredi's theorem for arithmetic progressions of length four. *Geometric Functional Analysis*, 8(3):529–551, 1998.

[Gow01] William T. Gowers. A new proof of Szeméredi's theorem. *Geometric Functional Analysis*, 11(3):465–588, 2001.

[GT05] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3 norm. *arXiv.org:math/0503014*, 2005.

[GT07] Ben Green and Terence Tao. The distribution of polynomials over finite fields, i with applications to the Gowers norms. Technical report, http://arxiv.org/abs/0711.3191v1, November 2007.

[KS08] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 403–412, New York, NY, USA, 2008. ACM.

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In Richard E. Ladner and Cynthia Dwork, editors, *STOC*, pages 547–556. ACM, 2008.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996.

[VW07] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Computational Complexity, 2007. CCC '07. Twenty-Second Annual IEEE Conference on*, pages 141–154, June 2007.

A Tightness of main theorem

In this section we show that our main theorem cannot be improved asymptotically. Specifically, we show that there is a constant $\alpha > 1/2$ such that for infinitely many d, for sufficiently large n, there exists a function $f = f_{d,n} : \mathbb{F}_2^n \to \mathbb{F}_2$ that passes the degree d AKKLR test (i.e., the $(d+1)$-flat test) with probability α (i.e., strictly greater than half) while being almost uncorrelated with degree d polynomials.

Our example comes directly from the works of [LMS08, GT07]. In particular, the function $f_{d,n}$ is simply the degree $d+1$ symmetric polynomial over n variables (defined formally below). When $d+1$ is a power of two, then [LMS08, GT07] (who in turn attribute the ideas to [AB01]) already show that this function is far from every degree d polynomial. To complete our theorem we only need to show that this function passes the $(d+1)$-flat test with probability noticeably greater than 1/2. [LMS08, GT07] also analyzed this quantity, but the published versions only show that this function passes the $(d+1)$-flat test with probability $1/2+\epsilon(d)$ where $\epsilon(d) \to 0$ as $d \to \infty$. However, it turns out that an early (unpublished) proof by the authors of [LMS08] can be used to show that the acceptance probability is $1/2+\epsilon$ where ϵ is an absolute constant. For completeness we include a complete proof here.

We start with the definition of the counterexample functions. For positive integers d, n with $d \leq n$, let $S_{d,n} : \mathbb{F}_2^n \to \mathbb{F}_2$ be given by

$$S_{d,n}(x_1, \ldots, x_n) = \sum_{I \subseteq [n], |I|=d} \prod_{i \in I} x_i.$$

Theorem 22 *Let $d + 1 = 2^t$ for some integer $t \geq 2$. Then, for every $\epsilon > 0$, there exists n_0 such that for every $n \geq n_0$, the following hold:*

1. $\delta_d(S_{d+1,n}) \geq 1/2 - \epsilon$.
2. $\mathsf{Rej}_{d,d+1}(S_{d+1,n}) \leq 1/2 - 2^{-7} + \epsilon$.

Theorem 22 follows immediately from the following two lemmas.

Lemma 23 ([GT07, Theorem 11.3]) *Let $d + 1 = 2^t$ for some integer $t \geq 0$. Then, for every $\epsilon > 0$, for sufficiently large n, we have $\delta_d(S_{d+1,n}) \geq 1/2 - \epsilon$.*

Lemma 24 *For every $d \geq 3$ and $\epsilon > 0$, for sufficiently large n, we have $\mathsf{Rej}_{d,d+1}(S_{d+1,n}) \leq 1/2 - 2^{-7} + \epsilon$.*

We prove Lemma 24 in the rest of this section. We stress again that this approach is from an unpublished version of [LMS08], and we include it for completeness.

We start with some notation. For $x, a_1, \ldots, a_{d+1} \in \mathbb{F}_2^n$, let $I(x, a_1, \ldots, a_{d+1}) = 1$ if the $(d+1)$-flat test rejects $S_{d+1,n}$ when picking the affine subspace $x + \text{span}(a_1, \ldots, a_{d+1})$. Note that $\mathsf{Rej}_{d,d+1}(S_{d+1,n}) = \mathbb{E}_{x,a_1,\ldots,a_{d+1}}[I(x, a_1, \ldots, a_{d+1})]$, where the expectation is taken over $x, a_1, \ldots, a_{d+1}$ picked uniformly and independently from $\mathbb{F}_2^n$, conditioned on $a_1, \ldots, a_{d+1}$ being linearly independent.

Lemma 25 *For $x, a_1, \ldots, a_{d+1} \in \mathbb{F}_2^n$, let $M \in \mathbb{F}_2^{(d+1)\times n}$ be the matrix whose i^{th} row is a_i. Then $I(x, a_1, \ldots, a_{d+1}) = 1$ if and only if $M \cdot M^T$ is of full rank.*

Note that the acceptance of the $(d+1)$-flat is independent of x; this is explained in the proof.

Proof For each $I \subseteq [n]$, $|I| = d + 1$, we define the polynomial $f_I(x) = \prod_{i \in I} x_i$. Note that

$$S_{d+1,n} = \sum_{I \subseteq [n], |I|=d+1} f_I.$$

Now, the $(d+1)$-flat test accepts $S_{d+1,n}$ when picking the affine subspace $x + \text{span}(a_1, \ldots, a_{d+1})$ if and only if

$$\sum_{J \subseteq [d+1]} S_{d+1,n}(x + \sum_{j \in J} a_j) = 0. \tag{1}$$

Note that the acceptance of the $(d+1)$-flat test is independent of x. This is because $S_{d+1,n}$ is a degree $d+1$ polynomial, so it's $(d+1)$st derivative (the output of the $(d+1)$-flat test) is constant. Hence, the test accepts if and only if

$$\sum_{J\subseteq[d+1]} S_{d+1,n}(\sum_{j\in J} a_j) = 0, \tag{2}$$

which can be rewritten as

$$\sum_{I\subseteq[n],|I|=d+1}\ \sum_{J\subseteq[d+1]} f_I(\sum_{j\in J} a_j) = 0.$$

For a fixed I, we focus our attention on the expression $\sum_{J\subseteq[d+1]} f_I(\sum_{j\in J} a_j)$. By definition, this equals $\sum_{J\subseteq[d+1]}\prod_{i\in I}(\sum_{j\in J} a_{j,i})$ which in turn equals $\sum_{J\subseteq[d+1]}(-1)^{d+1-|J|}\prod_{i\in I}(\sum_{j\in J} a_{j,i})$ (since $1=-1$ in $\mathbb{F}_2$). By Ryser's formula, this equals $\mathrm{perm}(M_I)$, where perm is the permanent function, and M_I is the $(d+1)\times(d+1)$ submatrix of M formed by the columns of I.

Thus, the left hand side of Equation (1) equals

$$\sum_{I\subseteq[n],|I|=d+1} \mathrm{perm}(M_I).$$

Since we are working over $\mathbb{F}_2$, we have that $\mathrm{perm}(M_I) = \det(M_I) = \det(M_I)^2$. Thus,

$$\begin{aligned}\sum_{I\subseteq[n],|I|=d+1} \mathrm{perm}(M_I) &= \sum_{I\subseteq[n],|I|=d+1} \det(M_I)^2\\ &= \det(MM^T). \qquad \text{by the Cauchy-Binet formula}\end{aligned}$$

We thus conclude that $I(x, a_1, \ldots, a_{d+1}) = 1$ if and only if MM^T is nonsingular.

■

We thus turn our attention to the probability that for a randomly chosen matrix M, the matrix $M\cdot M^T$ is of full rank. We first note the following fact on the distribution of $M\cdot M^T$ when M is chosen uniformly from the space of full rank matrices.

Lemma 26 *Let $A, B \in \mathbb{F}_2^{(d+1)\times(d+1)}$ be random variables generated as follows: A is a symmetric matrix chosen uniformly at random, and $B = M\cdot M^T$ where M is a random $(d+1)\times n$ matrix chosen uniformly from matrices of rank $d+1$. Then the total variation distance between A and B is $O(2^{d-n})$.*

Proof Let E be the event that the rows of M, along with the vector $\mathbf{1}$ (the vector which is 1 in each coordinate) are all linearly independent. Note that the probability of E is at least $1-2^{d+1-n}$. We will now show that the distribution of $B|E$ is $\exp(-n)$-close to the distribution of A. This will complete the proof.

Let the rows of M be $a_1, \ldots, a_{d+1}$. We pick them one at at time. Having picked $a_1, \ldots, a_i$, the new entries of B that get determined by a_{i+1} are the entries $B_{i+1,j}$ for all $j \le i+1$ (these determine the entries $B_{j,i+1}$). If a_{i+1} is picked uniformly from $\mathbb{F}_2^n$, then by the linear independence of $a_1, \ldots, a_i, \mathbf{1}$, we see that the bits

- $B_{i+1,1} = \langle a_{i+1}, a_1 \rangle$,
- $B_{i+1,2} = \langle a_{i+1}, a_2 \rangle$,
- $\ldots$,
- $B_{i+1,i} = \langle a_{i+1}, a_i \rangle$,
- $B_{i+1,i+1} = \langle a_{i+1}, a_{i+1} \rangle = \langle a_{i+1}, \mathbf{1} \rangle$,

are all uniformly random and independent, as required. However, since we have conditioned on E, a_{i+1} is not picked uniformly from $\mathbb{F}_2^n$, but picked uniformly from $\mathbb{F}_2^n \setminus \mathrm{span}\{a_1, \ldots, a_i, \mathbf{1}\}$. Still, this distribution of a_{i+1} is 2^{i+1-n}-close to the uniform distribution over $\mathbb{F}_2^n$, and as a consequence, the distribution of the bits $B_{i+1,1}, \ldots, B_{i+1,i+1}$ is $O(2^{d-n})$-close to the distribution of uniform and independent random bits.

To summarize, the entries of the matrix B are exposed in $d+1$ rounds. The bits $B_{i,j}$ for $j < i$ are exposed in round i, and their distribution, conditioned on the bits exposed in all the previous rounds, is $O(2^{d-n})$-close to that of uniform and independent random bits. This implies the desired claim on the distribution of B. ■

The final lemma shows that the random symmetric matrix $A \in \mathbb{F}_2^{(d+1)\times(d+1)}$ is full rank with probability bounded away from $1/2$ by some constant independent of d. This seems to be a well-analyzed problem and [BCJ+06, Theorem 4.14] (see also [BM88] for related work) already proves this fact; in particular, they show that if $k \geq 3$, a random symmetric k-by-k matrix over $\mathbb{F}_2$ is full rank with probability at most $7/16$. For completeness, we include a simple proof that establishes a weaker bound on the probability of non-singularity.

Lemma 27 *For $k \geq 4$, the probability that a random symmetric matrix $A \in \mathbb{F}_2^{k\times k}$ has full rank is at most $1/2 - 2^{-7}$.*

Proof Let A_i denote the $i \times i$ submatrix containing the first i rows and columns of A. We consider the probability that A has full rank, conditioned upon various choices of A_{k-1}. The first claim below shows that the probability of this event is at most half if the rank of A_{k-1} is either $k-1$ or $k-2$; and zero if the rank of A_{k-1} is at most $k-3$. We then argue in the next claim that the probability that A_{k-1} has rank at most $k-3$ is bounded below by a positive constant independent of k. The lemma follows immediately.

Claim 28 *Fix $B \in \mathbb{F}_2^{(k-1)\times(k-1)}$. The following hold:*

1. *If $A_{k-1} = B$ and $\mathrm{rank}(B) \leq k-3$, then $\mathrm{rank}(A) < k$.*
2. *If $\mathrm{rank}(B) = k-1$ then $\Pr_A[\mathrm{rank}(A) = k | A_{k-1} = B] \leq 1/2$.*
3. *If $\mathrm{rank}(B) = k-2$ then $\Pr_A[\mathrm{rank}(A) = k | A_{k-2} = B] \leq 1/2$.*

Proof Note that for every i, we have $\mathrm{rank}(A_i) \le \mathrm{rank}(A_{i-1}) + 2$ since A_i may be obtained from A_{i-1} by first adding a column and then a row, and each step may increase the rank by at most 1. Part (1) follows immediately.

For part (2), fix $a_{k,1}, \ldots, a_{k,k-1}$ and consider a random choice of $a_{k,k}$. Since A_{k-1} has full rank, there is a unique linear combination of the $k-1$ rows of A_{k-1} that generates the row $\langle a_{k,1}, \ldots, a_{k,k-1} \rangle$. A has full rank only if $a_{k,k}$ does not equal the same linear combination of $a_{1,k}, \ldots, a_{k-1,k}$, and the probability of this event is at most $1/2$.

Finally for part (3), assume for notational simplicity that A_{k-2} has full rank and the $(k-1)$th row of A is linearly dependent on the first $k-2$ rows. Now consider the addition of a kth row to A_{k-1} consisting of $a_{k,1}, \ldots, a_{k,k-1}$. Note that a necessary condition for A to have full rank is that the newly added row is linearly independent of the first $k-2$ rows of A_{k-1} (otherwise, the rank of the first $k-1$ columns of A is only $k-2$). But again (as in Part (2)), there is a unique linear combination of the rows of A_{k-2} that generates the row $\langle a_{k,1}, \ldots, a_{k,k-2} \rangle$. The probability that $a_{k,k-1}$ equals this linear combination applied to the $(k-1)$-th column of A_{k-1} is at least $1/2$. ■

Claim 29 $\Pr_A[\mathrm{rank}(A_{k-1}) \le k-3] \ge 2^{-6}$.

Proof We start with the subclaim that for every ℓ, we have $\Pr[\mathrm{rank}(A_{\ell+1}) = \mathrm{rank}(A_\ell) | A_\ell] = 2^{(\mathrm{rank}(A_\ell)-\ell-1)}$. To see this, let $I \subseteq [\ell]$ be such that A_ℓ restricted to rows in I has full rank (and so $|I| = \mathrm{rank}(A_\ell)$). Then A_ℓ restricted to rows and columns of I also has full rank. (All the rows not in I are in the span of the rows that are in I, and thus, by symmetry, all columns not in I are in the span of the columns in I.) Fix $a_{\ell+1,j}$ for $j \in I$ and note that there is unique linear combination of the rows of I in A_ℓ such that they yield $a_{\ell+1,j}$ for $j \in I$. This linear combination determines a unique setting for the remaining $\ell + 1 - |I|$ entries of the $(\ell+1)^{\mathrm{th}}$ row of $A_{\ell+1}$, if the the rank of $A_{\ell+1}$ is to equal the rank of A_ℓ. The probability of this unique setting occurring equals $2^{|I|-\ell-1}$. The subclaim follows.

Now the claim follows easily. Let m be the smallest integer $\le k-2$ such that $\mathrm{rank}(A_m) \ge k-4$. If such an m does not exist, then $\mathrm{rank}(A_{k-1}) \le k-3$. Otherwise, m exists, $m \ge k-4$ and $\mathrm{rank}(A_m) \le k-3$. Using the subclaim above, we have for every $\ell \in \{m, \ldots, k-2\}$, it is the case that $\Pr[\mathrm{rank}(A_{\ell+1}) = \mathrm{rank}(A_\ell) | A_\ell] \ge 2^{(k-4)-\ell-1}$. Combining the claims for $\ell \in [m, k-2]$ (and recalling that $m \ge k-4$), we get $\Pr[\mathrm{rank}(A_{k-1}) \le k-3] \ge 2^{-6}$. ■

Given the claims above, the lemma follows immediately.

■

ECCC ISSN 1433-8092

http://eccc.hpi-web.de