Optimal Testing of Reed-Muller Codes*

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

January 30, 2010

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 problem of testing Reed-Muller codes. We are interested in determining the query-complexity of distinguishing with constant probablity between the case where f is a degree d polynomial and the case where f is $\Omega(1)$-far from all degree d polynomials. Alon et al. [AKK+05] proposed and analyzed a natural 2^{d+1}-query test T_0, and showed that it accepts every degree d polynomial with probability 1, while rejecting functions that are $\Omega(1)$-far with probability $\Omega(1/(d2^d))$. This leads to a $O(d4^d)$-query test for degree d Reed-Muller codes.

We give an asymptotically optimal analysis of T_0, showing that it rejects functions that are $\Omega(1)$-far with $\Omega(1)$-probability (so the rejection probability is a universal constant independent of d and n). In particular, this implies that the query complexity of testing degree d Reed-Muller codes is $O(2^d)$.

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 $\mathbb{F}_2$. Our results also imply a "query hierarchy" result for property testing of affine-invariant properties: For every function $q(n)$, it gives an affine-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.

---

*This is a brief overview of the results in the paper [BKS+09].

†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.

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.

2 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.

3 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 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]. Below we explain some of the context of our work and some implications.

4 Query hierarchy for affine-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 an affine-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* $\mathrm{RM}(\lceil \log_2 q(n) \rceil, n)$.

5 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. It turns out 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

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

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

6 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 $\delta_d(f^{\oplus t}) \geq 1/2 - 2^{-\Omega(t/2^d)}$.

7 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, FGL+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.

The proofs of the theorems mentioned above may be found in our paper [BKS+09].

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.

[BKS+09] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. *ECCC Technical Report*, TR09-086, October 2009.

[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.

[FGL+96] Uriel Feige, Shafi Goldwasser, László Lovász, 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.