Electronic Colloquium on Computational Complexity, Report No. 80 (2008)

Random low degree polynomials are hard to approximate

Ido Ben-Eliezer* Rani Hod† Shachar Lovett‡

July 3, 2008

Abstract

We study the problem of how well a typical multivariate polynomial can be approximated by lower degree polynomials over $\mathbb{F}_2$. We prove that, with very high probability, a random degree d polynomial has only an exponentially small correlation with all polynomials of degree $d-1$, for all degrees d up to $\Theta(n)$. That is, a random degree d polynomial does not admit good approximations of lesser degree.

In order to prove this, we prove far tail estimates on the distribution of the bias of a random low degree polynomial. As part of the proof, we also prove tight lower bounds on the dimension of truncated Reed–Muller codes.

1 Introduction

Two functions $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$ are said to be *ϵ-correlated* if

$$\Pr\left[f(x) = g(x)\right] \geq \frac{1+\epsilon}{2}.$$

The *correlation* of f and g is the maximal value of ϵ for which they are ϵ-correlated. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be *ϵ-correlated* with a set $F \subset \mathbb{F}_2^n \to \mathbb{F}_2$ of functions if it is ϵ-correlated with at least one function $g \in F$.

We are interested in functions that have a low correlation with the set of degree d polynomials; namely, functions that cannot be approximated by any polynomial of total degree at most d. How *complex* must such a function be? We use the most natural measure for complexity in these settings, which is the degree of the function when considered as a polynomial.

A simple probabilistic argument shows that for any constant $\delta < 1$ and for $d < \delta n$, a random function has an exponentially small correlation with degree d polynomials. However, a random function is complex as, with high probability, its degree is at least $n-2$. In this work, we study how well a random degree d polynomial can be approximated by any lower degree polynomial.

Contrarily, finding an explicit function f that has a correlation of ϵ with degree d polynomials, for degrees $d \geq \Omega(\log n)$, is an open problem even for constant $\epsilon < 1$. Moreover, finding such an explicit construction for degrees as small as $\operatorname{poly}\log n$ will have major consequences in circuit lower bounds. Such a function can give lower bounds and pseudorandom generators for the class ACC_0 of constant depth circuits with AND, OR and MOD gates. Currently no such non-trivial lower bounds are known.

*School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: idobene@tau.ac.il

†School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: ranihod@tau.ac.il

‡Weizmann Institute of Science, Rehovot, Israel. Email: shachar.lovett@weizmann.ac.il

ISSN 1433-8092

1.1 Our contribution

We show that, with very high probability, a random degree d polynomial has an exponentially small correlation with polynomials of lower degree. We prove this for degrees ranging from a constant up to $\delta_{\max} n$, where $0 < \delta_{\max} < 1$ is some constant. All results hold for large enough n.

We now state our main theorem.

Theorem 1. *There exist a constant $0 < \delta_{\max} < 1$ and constants $c, c' > 0$ such that the following holds. Let f be a random n-variate polynomial of degree $d \leq \delta_{\max} n$. The probability that f has a correlation of $2^{-cn/d}$ with polynomials of degree at most $d-1$ is at most $2^{-c'\binom{n}{\leq d}}$, where $\binom{n}{\leq d} = \sum_{i=0}^{d} \binom{n}{i}$.*

The main theorem is an easy corollary of the following lemma, which is the main technical contribution of the paper. We define the *bias* of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to be

$$bias(f) = \mathbb{E}_x \left[(-1)^{f(x)} \right] = \Pr\left[f(x) = 0\right] - \Pr\left[f(x) = 1\right] .$$

Lemma 2. *Let f be a random degree d polynomial for $d \leq (1-\epsilon)n$. Then,*

$$\Pr\left[|bias(f)| > 2^{-c_1 n/d} \right] \leq 2^{-c_2 \binom{n}{\leq d}} ,$$

where $0 < c_1, c_2 < 1$ depend only on ϵ.

Note that Lemma 2 holds for degrees up to $(1-\epsilon)n$, while we were only able to prove Theorem 1 for degrees up to $\delta_{\max} n$.

The Reed–Muller code $\mathcal{RM}(n,d)$ is the linear code of all polynomials (over $\mathbb{F}_2$) in n variables of total degree at most d. Interpreted in the language of Reed–Muller codes, Lemma 2 gives a tail estimate on the weight distribution of Reed–Muller codes.

The following proposition shows that the estimate in Lemma 2 is somewhat tight for degrees up to $n/2$.

Proposition 3. *Let f be a random degree d polynomial for $d \leq (1/2 - \epsilon)n$. Then,*

$$\Pr\left[|bias(f)| > 2^{-c_1' n/d} \right] \geq 2^{-c_2' \binom{n}{\leq d}} ,$$

where $0 < c_1', c_2' < 1$ are constants depending only on ϵ.

As a part of the proof of Lemma 2, we give the following tight lower bound on the dimension of truncated Reed–Muller codes, which is of independent interest.

Lemma 4. *Let $x_1, \ldots, x_R$ be $R = 2^r$ distinct points in $\mathbb{F}_2^n$. Consider the linear space of degree d polynomials restricted to these points; that is, the space*

$$\{(p(x_1), \ldots, p(x_R)) : p \in \mathcal{RM}(n,d)\} .$$

The linear dimension of this space is at least $\binom{r}{\leq d}$.

We have recently learned that this lemma appeared earlier in [9, Theorem 1.5]. Our proof, on the other hand, is independent and has an algorithmic flavor.

1.2 Related Work

Reed–Muller codes are one of most studied objects in coding theory (for a background see, e.g., [10]). Determining the weight distribution is a long standing open problem for every $d \geq 3$. The weight distribution is completely known for $d = 2$ (see, for example, [2]) and some partial results are known also for $d = 3$. In the general case, there are estimates (see, e.g., [7, 8]) on the number of codewords with weight between w and $2.5w$, where $w = 2^{-d}$ is the minimal weight of the code.

The case of multilinear polynomials was considered by Alon et al. [1], who proved a tail estimate similar to Lemma 2 and used it to prove bounds on the size of distributions that fool low degree polynomials.

The Gowers Norm is a measure related to the approximability of functions by low degree polynomials. It was introduced by Gowers [3] in his seminal work on a new proof for Szemerédi's Theorem. Using the Gowers Norm machinery, it is easy to prove that a random polynomial of degree $d < \log n$ has a small correlation with lower degree polynomials. However, this approach fails for degrees exceeding $\log n$. Note that our result holds for degrees up to $\delta_{\max} n$.

Green and Tao [4] study the structure of biased multivariate polynomials. They prove that if their degree is at most the size of the field, then they must have structure — they can be expressed as a function of a constant number of lower degree polynomials. Kaufman and Lovett [6] strengthen this structure theorem for polynomials of every constant degree, removing the field size restriction.

The rest of the paper is organized as follows. Our main result, Theorem 1, is proved in Section 2. The proof of the lower bound on the bias (Proposition 3) is given in Section 3.

2 Proof of the Main Theorem

First we show that Theorem 1 follows directly from Lemma 2 by a simple counting argument.

Let f be a random degree d polynomial for $d \leq \delta_{\max} n$, where $\delta_{\max}$ will be determined later. For every polynomial g of degree at most $d-1$, $f - g$ is also a random degree d polynomial. By the union bound on all possible choices of g,

$$\Pr_{f \in \mathcal{RM}(n,d)} \left[\exists g \in \mathcal{RM}(n, d-1) : |bias(f-g)| \geq 2^{-c_1 n/d} \right] \leq 2^{\binom{n}{\leq d-1} - c_2 \binom{n}{\leq d}}$$

Choosing $\delta_{\max}$ to be a small enough constant, we get that there is a constant $c' > 0$ such that $c_2 \binom{n}{\leq d} - \binom{n}{\leq d-1} \geq c' \binom{n}{\leq d}$ for all $d \leq \delta_{\max} n$ (see, for example, [5, Exercise 1.14]).

We now move on to prove Lemma 2. The section is organized as follows. Lemma 2 is proved in Subsection 2.1, where the technical claims are postponed to Subsection 2.2. Lemma 4 is proved in Subsection 2.3.

2.1 Proof of Lemma 2

We need to prove that a random degree d polynomial has a very small bias with very high probability. Denote the dual code of $\mathcal{RM}(n,d)$ by $\mathcal{RM}(n,d)^{\perp}$. We start by correlating the moments of the bias of a random degree d polynomial to short words in $\mathcal{RM}(n,d)^{\perp}$.

Claim 2.1. *Fix $t \in \mathbb{N}$ and let $p \in \mathcal{RM}(n,d)$ and $x_1, \ldots, x_t \in \mathbb{F}_2^n$ be chosen independently and equiprobably. Then,*

$$\mathbb{E}\left[bias(p)^t\right] = \Pr\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^{\perp}\right],$$

where e_x for $x \in \mathbb{F}_2^n$ is the unit vector in $\mathbb{F}_2^{2^n}$, having 1 in position x and 0 elsewhere.

In favor of not interrupting the proof, we postpone the proof of Claim 2.1 and other technical claims to Subsection 2.2.

We proceed by introducing the following definitions. Fix d. For $x \in \mathbb{F}_2^n$ let $\mathrm{eval}_d(x)$ denote its *d-evaluation*; that is, a (row) vector in $\mathbb{F}_2^{\binom{n}{\leq d}}$ whose coordinates are the evaluation of all monomials of degree up to d at the point x. Formally,

$$\mathrm{eval}_d(x) = \left(\prod_{i\in I} x(i)\right)_{I\subset[n],|I|\leq d}.$$

For points $x_1, \ldots, x_t \in \mathbb{F}_2^n$ let $\mathcal{M}_d(x_1, \ldots, x_t)$ denote their *d-evaluation matrix*; this is a $t \times \binom{n}{\leq d}$ matrix whose ith row is the d-evaluation of x_i. We denote the rank of $\mathcal{M}_d(x_1, \ldots, x_t)$ by $\mathrm{rank}_d(x_1, \ldots, x_t)$. As this value is independent of the order of $x_1, \ldots, x_t$, we may refer without ambiguity to the *d-rank* of a set $S \subseteq \mathbb{F}_2^n$ by $\mathrm{rank}_d(S)$.

Claim 2.1 tells us that, in order to bound the moments of the bias of a random polynomial, we need to study the probability that a random word of length about[1] t is in $\mathcal{RM}(n,d)^\perp$.

Let $A = \mathcal{M}_d(x_1, \ldots, x_t)$. Note that $e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^\perp$ if and only if

$$p(x_1) + \cdots + p(x_t) = 0 \tag{1}$$

for any degree d polynomial p. Therefore, $e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^\perp$ if and only if the sum of the rows of A is zero. It is sufficient to satisfy (1) only on the monomial basis of the degree d polynomials; that is, verify that each column in A sums to zero.

We turn to bound the probability that the rows of A sum to the zero vector for random $x_1, \ldots, x_t \in \mathbb{F}_2^n$. For this we divide the n variables into two sets: V' of size $n' = \lceil n(1-1/d) \rceil$ and V'' of size $n'' = n - n'$. Let $\alpha = n''/n \approx 1/d$. Instead of requiring that *every* column of A sums to zero, we require this only for columns corresponding to monomials that contain exactly one variable from V'' (and thus up to $d-1$ variables from V').

For $i = 1, \ldots, t$ denote by x_i' $(\in \mathbb{F}_2^{n'})$ the restriction of $x_i \in \mathbb{F}_2^n$ to the variables in V'. The following claim bounds the probability that sum of A's rows is zero in terms of the $(d-1)$-rank of $x_1', \ldots, x_t'$.

Claim 2.2.

$$\Pr_{\{x_i\}}\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^\perp\right] \leq \mathbb{E}_{\{x_i'\}}\left[2^{-\mathrm{rank}_{d-1}(x_1',\ldots,x_t')\alpha n}\right].$$

To finish the proof, we provide a (general) lower bound on d-ranks of random vectors.

Claim 2.3. *For any constants $\beta < 1$ and $\delta < 1$, there exist constants $c > 0$ and $\eta > 1$ such that if $x_1, \ldots, x_t \in \mathbb{F}_2^n$ are chosen uniformly and independently, where $t \geq \eta\binom{n}{\leq d}$ and $d \leq \delta n$, then*

$$\Pr\left[\mathrm{rank}_d(x_1, \ldots, x_t) < \beta\binom{n}{\leq d}\right] \leq 2^{-c\binom{n}{\leq d+1}}.$$

We now put it all together, in order to complete the proof of Lemma 2. According to Claim 2.2, we have

$$\Pr_{\{x_i\}}\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^\perp\right] \leq \mathbb{E}_{\{x_i'\}}\left[2^{-\mathrm{rank}_{d-1}(x_1',\ldots,x_t')\alpha n}\right].$$

[1] We say "about t" as $x_1, \ldots, x_t$ might not be distinct.

Applying Claim 2.3 for $d-1$ and n' (instead of d and n in the claim statement), and assuming $t \geq \eta\binom{n'}{\leq d-1}$, we get that

$$\Pr\left[\mathrm{rank}_{d-1}(x'_1, \dots, x'_t) < \beta\binom{n'}{\leq d-1}\right] < 2^{-c\binom{n'}{\leq d}}.$$

Therefore,

$$\Pr_{\{x_i\}}\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^{\perp}\right] \leq 2^{-\beta\binom{n'}{\leq d-1}\alpha n} + 2^{-c\binom{n'}{\leq d}}.$$

Recalling that $n' = \lceil n(1-1/d) \rceil$ and $\alpha = 1 - n'/n = 1/d + O(1/n)$, we get that for any constant β (and $c = c(\beta)$) there is a constant c' such that

$$\Pr_{\{x_i\}}\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^{\perp}\right] \leq 2^{-c'\binom{n}{\leq d}}.$$

This is because $\binom{n'}{\leq d-1} = \Theta\left(\binom{n}{\leq d} d/n\right)$ and $\binom{n'}{\leq d} = \Theta\left(\binom{n}{\leq d}\right)$.

We thus proved that there is a constant c' such that

$$\mathbb{E}_{f \in \mathcal{RM}(n,d)}\left[bias(f)^t\right] \leq 2^{-c'\binom{n}{\leq d}},$$

for $t = \eta\binom{n'}{\leq d-1} = \Theta\left(\binom{n}{\leq d-1}\right)$. Hence, $tn/d \leq c''\binom{n}{\leq d}$ for some constant c''.

By Markov inequality, for small enough $c_1 > 0$ such that $c_2 = c' - c''c_1 > 0$,

$$\Pr\left[|bias(f)| \geq 2^{-c_1 n/d}\right] \leq 2^{tc_1 n/d - c'\binom{n}{\leq d}} \leq 2^{(c''c_1 - c')\binom{n}{\leq d}} \leq 2^{-c_2\binom{n}{\leq d}}.$$

2.2 Proofs of technical claims

Proof of Claim 2.1. Write p as

$$p(x) = \sum_{I \subset [n], |I| \leq d} \alpha_I \prod_{i \in I} x(i),$$

where $x(i)$ denotes the ith coordinate of $x \in \mathbb{F}_2^n$. As p was chosen uniformly, all α_I are uniform and independent over $\mathbb{F}_2$. Therefore,

$$\begin{aligned}
\mathbb{E}_p\left[(bias(p))^t\right] &= \mathbb{E}_p\left[\prod_{j=1}^{t} \mathrm{bias}(p)\right] \\
&= \mathbb{E}_{\{\alpha_I\}}\left[\prod_{j=1}^{t} \mathbb{E}_{x_j}\left[(-1)^{\sum_I \alpha_I \prod_{i \in I} x_j(i)}\right]\right] \\
&= \mathbb{E}_{\{x_j\}}\left[\prod_{I} \mathbb{E}_{\alpha_I}\left[(-1)^{\alpha_I(\sum_{j=1}^{t} \prod_{i \in I} x_j(i))}\right]\right] \\
&= \mathbb{E}_{\{x_j\}}\left[\prod_{I} \mathbf{1}_{\{\sum_{j=1}^{t} \prod_{i \in I} x_j(i) = 0\}}\right] \\
&= \Pr_{\{x_j\}}\left[\forall I \sum_{j=1}^{t} \prod_{i \in I} x_j(i) = 0\right] \\
&= \Pr_{\{x_j\}}\left[e_{x_1} + \cdots + e_{x_t} \in \mathcal{RM}(n,d)^{\perp}\right].
\end{aligned}$$

$\square$

Proof of Claim 2.2. Let $A' = \mathcal{M}_{d-1}(x'_1, \ldots, x'_t)$ be the $t \times \binom{n'}{\le d-1}$ sub-matrix of A corresponding to monomials of degree at most $d-1$ in variables from V'. Let $\mathcal{E}$ be the event in which every column of A corresponding to a monomial that contains exactly one variable from V'' sums to zero. It is easy to see that this event is equivalent to the event that every column of A' is orthogonal to the set of vectors $\{(x_1(i), \ldots, x_t(i)) : i \in V''\}$.

Fix the variables in V'; this determines A'. As the variables in V'' are independent of those in V', the probability of $\mathcal{E}$ (given A') is

$$\left(2^{-\mathrm{rank}(A')}\right)^{|V''|} = 2^{-\mathrm{rank}(A')\alpha n} = 2^{-\mathrm{rank}_{d-1}(x'_1,\ldots,x'_t)\alpha n}.$$

This holds for every assignment for variables of V', hence the result follows. $\square$

Proof of Claim 2.3. Let $B = \mathcal{M}_d(x_1, \ldots, x_t)$ be the $t \times \binom{n}{\le d}$ d-evaluation matrix of the random $x_1, \ldots, x_t \in \mathbb{F}_2^n$. We need to bound the probability that $\mathrm{rank}(B) < \beta\binom{n}{\le d}$.

Fix some $b \le \beta\binom{n}{\le d}$, and let us consider the event that the first b rows of B span the entire row span of B. Denote by V the linear space spanned by the first b rows of B. Since all rows of B are d-evaluations of some points in $\mathbb{F}_2^n$, we need to study the maximum number of d-evaluations contained in a linear subspace of dimension b.

Assume there are at least 2^r distinct d-evaluations in V. By Lemma 4, $\dim(V) \ge \binom{r}{\le d}$. Assume further that $\mathrm{rank}(B) < \beta\binom{n}{\le d}$; we get that

$$\beta\binom{n}{\le d} > \mathrm{rank}(B) \ge \dim(V) \ge \binom{r}{\le d}.$$

By Claim 2.4, $r \le n(1-\gamma/d)$, where γ is a constant depending only on β. In other words, out of the 2^n d-evaluations of all points in $\mathbb{F}_2^n$, at most $2^{n(1-\gamma/d)}$ fall in V and hence the probability that a random d-evaluation is in V is at most $2^{-\gamma n/d}$.

Assume the number of rows t is at least $\eta\binom{n}{\le d}$ for some $\eta > 1$. The probability that all the remaining rows of B are in V is at most

$$\left(2^{-\gamma n/d}\right)^{t-b} \le 2^{-(\eta-\beta)\binom{n}{\le d}\gamma n/d} \le 2^{-\gamma\rho(\eta-\beta)\binom{n}{\le d+1}},$$

where the last inequality follows from the fact that there exists a constant $\rho > 0$ such that $(n/d)\binom{n}{\le d} \ge \rho\binom{n}{\le d+1}$ for all n, d.

Choosing η large enough (as a function of β), we get that when we union bound over all possible ways to choose at most $\beta\binom{n}{\le d}$ rows out of $t \ge \eta\binom{n}{\le d}$, the probability that any of them spans the rows of B is at most $2^{-c\binom{n}{\le d+1}}$, where c depends only on β. $\square$

Claim 2.4. *For any $\beta, \delta < 1$, there is a constant $\gamma = \gamma(\beta, \delta)$ such that if $1 \le d \le \delta n$ and $r \ge d$ satisfy $\beta\binom{n}{\le d} \ge \binom{r}{\le d}$ then $r \le n(1-\gamma/d)$.*

Proof. We bound

$$\frac{1}{\beta} \le \binom{n}{\le d}\Big/\binom{r}{\le d} \le \max_{0\le i\le d}\binom{n}{i}\Big/\binom{r}{i} = \binom{n}{d}\Big/\binom{r}{d} \le \left(\frac{n-d}{r-d}\right)^d = \left(1 + \frac{n-r}{r-d}\right)^d.$$

Taking logarithms and assuming for the sake of contradiction that $r > n(1-\gamma/d)$, we get

$$\ln(1/\beta) \le d\ln\left(1 + \frac{n-r}{r-d}\right) \le \frac{d(n-r)}{r-d} < \frac{\gamma n}{r-d} < \frac{\gamma}{r/n-\delta} < \frac{\gamma}{1-\delta+\gamma/d}.$$

This can be made false by picking, e.g., $\gamma = (1-\delta)\ln(1/\beta)$. $\square$

2.3 Proof of Lemma 4

Restating the lemma in terms of d-evaluations, we need to show that for every subset $S \subseteq \mathbb{F}_2^n$ of size $R = 2^r$, $\mathrm{rank}_d(S) \geq \binom{r}{\leq d}$. Let $S = \{x_1, \ldots, x_{2^r}\}$ be the set of points. We simplify S by applying a sequence of transformations that do not increase its d-rank until we arrive to the linear space $\mathbb{F}_2^r \times \{0\}^{n-r}$.

We now define our basic non-linear transformation Π, mapping the set S to a set $\Pi(S)$ of equal size and not greater d-rank. Informally, Π tries to set the first bit of each element in S to zero, unless this results in an element already in S (and in this case Π keeps the element unchanged).

For $y = (y_1, \ldots, y_{n-1}) \in \mathbb{F}_2^{n-1}$, denote by $0y$ and $1y$ the elements $(0, y_1, \ldots, y_{n-1})$ and $(1, y_1, \ldots, y_{n-1})$ in $\mathbb{F}_2^n$, respectively. Extend this notation to sets by writing $0T = \{0y : y \in T\}$, $1T = \{1y : y \in T\}$ for a set $T \subseteq \mathbb{F}_2^{n-1}$.

We define the following three sets in $\mathbb{F}_2^{n-1}$.

$$\begin{aligned}
T_* &= \{y \in \mathbb{F}_2^{n-1} : \ 0y \in S \text{ and } 1y \in S\}\,, \\
T_0 &= \{y \in \mathbb{F}_2^{n-1} : \ 0y \in S \text{ and } 1y \notin S\}\,, \\
T_1 &= \{y \in \mathbb{F}_2^{n-1} : \ 0y \notin S \text{ and } 1y \in S\}\,.
\end{aligned}$$

Writing S as

$$S = 0T_* \cup 1T_* \cup 0T_0 \cup 1T_1\,,$$

we define $\Pi(S)$ to be

$$\Pi(S) = 0T_* \cup 1T_* \cup 0T_0 \cup 0T_1\,;$$

namely, we set to zero the first bit of all the elements in $1T_1$. It is easy to see that $|\Pi(S)| = |S|$ as $\Pi(S)$ introduces no collisions.

Proposition 5. $\mathrm{rank}_d(\Pi(S)) \leq \mathrm{rank}_d(S)$.

Proof. It will be easier to prove this using an alternative definition for $\mathrm{rank}_d(S)$.

Let $(x_1, \ldots, x_{2^r})$ be some ordering of S. For a degree d polynomial $p \in \mathcal{RM}(n,d)$, let $v_p \in \mathbb{F}_2^{2^r}$ be the evaluation of p on the points of S

$$v_p = (p(x_1), p(x_2), \ldots, p(x_{2^r}))\,.$$

Consider the linear space of vectors v_p for all $p \in \mathcal{RM}(n,d)$. The dimension of this space is exactly $\mathrm{rank}_d(S)$, as the monomials used in the definition of d-rank form a basis for the space of polynomials.

But now, instead of the dimension, consider the co-dimension. We call a point x_i, $1 \leq i \leq 2^r$, *dependent* if there are coefficients $\alpha_1, \ldots, \alpha_{i-1} \in \mathbb{F}_2$ such that for all degree d polynomials

$$p(x_i) = \sum_{j=1}^{i-1} \alpha_j p(x_j)\,.$$

We thus expressed $\mathrm{rank}_d(S)$ as the number of independent points in S, which is the same as the difference between $|S| = 2^r$ and the number of dependent points in S. To prove that $\mathrm{rank}_d(\Pi(S)) \leq \mathrm{rank}_d(S)$, it suffices to show that Π maps dependent points in S to dependent images in $\Pi(S)$. Let us consider an ordering of S in which the elements of $1T_1$ come last. Since all other points in S are mapped to themselves by Π, it is clear that dependent points in S appearing before $1T_1$ are also dependent in $\Pi(S)$. It remains to prove the claim for points in $1T_1$.

Let $t_1 = |T_1|$ and let $y_1, \ldots, y_{t_1}$ be some ordering of T_1. Assume $1y_i \in S$ is dependent and we will show that $0y_i \in \Pi(S)$ is also dependent. By definition, there exist coefficients $\alpha_y, \beta_y, \gamma_y, \delta_y$ such that, for any degree d polynomial,

$$p(1y_i) = \sum_{y \in T_*} \alpha_y p(0y) + \sum_{y \in T_*} \beta_y p(1y) + \sum_{y \in T_0} \gamma_y p(0y) + \sum_{y_j \in T_1 : j < i} \delta_{y_j} p(1y_j)\,.$$

Each polynomial $p \in \mathcal{RM}(n, d)$ can be uniquely decomposed as

$$p(x_1, \ldots, x_n) = x_1 p'(x_2, \ldots, x_n) + p''(x_2, \ldots, x_n)\,,$$

where $p' \in \mathcal{RM}(n-1, d-1)$ and $p'' \in \mathcal{RM}(n-1, d)$. Moreover, for every $y \in \mathbb{F}_2^{n-1}$, we have that $p(0y) = p''(y)$ and $p(1y) = p'(y) + p''(y)$. Since p' and p'' are independent, we can decompose the dependency of $p(1y_i)$ into its p' and p'' components as follows.

$$p'(y_i) = \sum_{y \in T_*} \beta_y p''(y) + \sum_{y_j \in T_1 : j < i} \delta_{y_j} p'(y_j)\,, \tag{2}$$

$$p''(y_i) = \sum_{y \in T_*} (\alpha_y + \beta_y) p''(y) + \sum_{y \in T_0} \gamma_y p''(y) + \sum_{y_j \in T_1 : j < i} \delta_{y_j} p''(y_j)\,. \tag{3}$$

We now move to consider $\Pi(S)$. Every $1y_i$ for $y_i \in T_1$ is mapped to $0y_i$, so we should only consider the p'' component for T_1's elements. Also, by the definition of T_* and T_0, for each $y \in T_* \cup T_0$, $0y \in S \cap \Pi(S)$. By (3), for any $p \in \mathcal{RM}(n, d)$,

$$p(0y_i) = \sum_{y \in T_*} (\alpha_y + \beta_y) p(0y) + \sum_{y \in T_0} \gamma_y p(0y) + \sum_{y_j \in T_1 : j < i} \delta_{y_j} p(0y_j)\,,$$

that is, $0y_i$ is also dependent in $\Pi(S)$.

Therefore, we have established that $\mathrm{rank}_d(\Pi(S)) \leq \mathrm{rank}_d(S)$. □

We now combine our basic Π with invertible linear transformations to define a wider class of simplifying transformations. For any $u, v \in \mathbb{F}_2^n$ such that their inner product $\langle u, v \rangle = 1$, we define the mapping $\Pi_{u,v}$ as follows. Informally, $\Pi_{u,v}$ tries to add v to elements x of S for which $\langle u, x \rangle = 1$, unless this results in an element already in S. In other words, if both x and $x + v$ are in S, then $\Pi_{u,v}(S)$ maps them both to themselves. Otherwise, if just one of them is in S, it maps it to x if $\langle u, x \rangle = 0$, and to $x + v$ if $\langle u, x + v \rangle = 0$. This is well defined as $\langle u, v \rangle = 1$. Note that $\Pi_{e_1, e_1} \equiv \Pi$.

Formally, let A be an $n \times n$ invertible matrix such that $e_1^T A = u$ and $A^{-1} e_1 = v$. We can construct such invertible A since $\langle u, v \rangle = 1$ by setting the first row of A to be u and the remaining rows of A to be a basis for the $(n-1)$-dimensional space normal to v. Define $\Pi_{u,v} = A^{-1} \Pi A$.

Observe that invertible affine transformations do not change the d-rank of a set, as they act as permutations on the set of degree d polynomials. Combining this with Proposition 5, we get that $\Pi_{u,v}$ maintains the size of S without increasing the d-rank.

We now use a sequence of $\Pi_{u,v}$ applications to transform the set S into the linear space $V = \mathbb{F}_2^r \times \{0\}^{n-r}$ spanned by the first r unit vectors $e_1, \ldots, e_r$. We say that $x \in S$ is *good* if $x \in V$, and is *bad* otherwise. If all the elements of S are good then $S = V$ since all the elements of S are distinct. Otherwise, let $x \in S$ be some bad element and let $x' \in V \setminus S$. Since $x \notin V$, there must be some index $r < i \leq n$ such that $x_i = 1$; set $u = e_i$ and $v = x + x'$.

We show that applying $\Pi_{u,v}$ maps x to x' and does not affect any good elements, thus increasing the number of good elements. First see that $\langle u, v \rangle = v_i = x_i + x'_i = 1 + 0 = 1$

since $x' \in V$ so $\Pi_{u,v}$ is well defined. See also that as $\langle u, x\rangle = x_i = 1$ and $x + v \notin S$, $\Pi_{u,v}$ will add v to x, transforming it to $x' \in V$. Also, any good element y is unchanged by $\Pi_{u,v}$ since $\langle u, y\rangle = y_i = 0$. In total, the number of good elements increased by at least one.

We repeat this until all elements are good, that is, until S is transformed to V, establishing that $\mathrm{rank}_d(S) \geq \mathrm{rank}_d(V)$. To finish the proof, observe that the restriction of polynomials in $\mathcal{RM}(n, d)$ to points in a linear space of dimension r is exactly $\mathcal{RM}(r, d)$. Since $|\mathcal{RM}(r, d)| = \binom{r}{\leq d}$ (see [10]), we get that for any set S of size 2^r,

$$\mathrm{rank}_d(S) \geq \binom{r}{\leq d},$$

as required.

3 Proof of Proposition 3

Let $d < \gamma n$ for constant $\gamma < 1/2$. We define a set of polynomials with measure of at least $2^{-c'_2\binom{n}{\leq d}}$ such that all polynomials in this set have a bias of at least $2^{-c'_1 n/d}$ (for constants c'_1, c'_2). That is, we will prove

$$\Pr_{f\in\mathcal{RM}(n,d)}\left[bias(f) \geq 2^{-c'_1 n/d}\right] \geq 2^{-c'_2\binom{n}{\leq d}}.$$

Similar to the proof of Theorem 1, we divide the n variables into two sets: V' of size $n' = \lceil n/d\rceil$ and V'' of size $n'' = n - n'$. Consider the set of monomials of degree at most d that are multilinear in V' (and thus have degree at most $d - 1$ in V'').

We first show that the number of such monomials is only a constant factor smaller than the number of all monomials of degree at most d. The number of monomials we consider is

$$\binom{n'}{1}\binom{n''}{\leq d-1} \geq \frac{n}{d}\binom{n(1-1/d)}{d-1}.$$

There exists a constant $c_\gamma > 0$ such that if $d < \gamma n$ then

$$\binom{n(1-1/d)}{d-1} \geq c_\gamma\binom{n}{d-1} \qquad \text{and also} \qquad \binom{n}{d} \geq c_\gamma\binom{n}{\leq d}.$$

Hence the number of monomials multilinear in V' is at least $c_\gamma^2\binom{n}{\leq d}$.

Let $\mathcal{L}$ be the linear space of polynomials on these monomials, $|\mathcal{L}| \geq 2^{c_\gamma^2\binom{n}{\leq d}}$. Consider a random polynomial $f \in \mathcal{L}$. Since each monomial of f has exactly one variable in V', we can decompose f as the sum of products of a variable from V' and a random degree $d-1$ polynomial from V''. That is, if $V' = \{x_1, \ldots, x_{n'}\}$ and $V'' = \{x_{n'+1}, \ldots, x_n\}$, we can write

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n'} x_i g_i(x_{n'+1}, \ldots, x_n).$$

We now show f has an expected bias of $2^{-n'} \geq 2^{-n/d}$. Consider a partial assignment to the variables $x_1, \ldots, x_{n'}$ of V'. If all of them are zero, then $f(0, \ldots, 0, x_{n'+1}, \ldots, x_n) \equiv 0$, and hence has bias 1. In all other cases, we are left with a random degree $d - 1$ polynomial in the variables from V'' and as such it has bias 0 (e.g., since the constant term is random). Thus,

$$\mathbb{E}_{f\in\mathcal{L}}[bias(f)] = 1 \cdot \Pr\left[\forall 1 \leq i \leq n' : x_i = 0\right] + 0 \cdot \Pr\left[\exists 1 \leq i \leq n' : x_i \neq 0\right] = 2^{-n'},$$

and we get that

$$\Pr\left[bias(f) > 2^{-(n'+1)} \;\middle|\; f \in \mathcal{L}\right] > 2^{-(n'+1)}.$$

We conclude that there is a constant c_2' such that

$$\Pr\left[bias(f) > 2^{-(n/d+1)}\right] \geq \Pr\left[f \in \mathcal{L}\right] \cdot \Pr\left[bias(f) > 2^{-(n/d+1)} \;\middle|\; f \in \mathcal{L}\right] \geq 2^{-c_2'\binom{n}{\leq d}}.$$

Acknowledgements. We would like to thank Simon Litsyn and Michael Krivelevich for helpful discussions. We thank Noga Alon for pointing out an alternative proof of Lemma 4. The third author would like to thank his advisor, Omer Reingold, for helpful discussions.

References

[1] N. Alon, I. Ben-Eliezer and M. Krivelevich, *Small sample spaces cannot fool low degree polynomials*, proceedings of the 12th International Workshop on Randomization and Computation (RANDOM 2008), to appear.

[2] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, **Covering Codes**, North–Holland, 1997.

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

[4] B. Green and T. Tao, *The distribution of polynomials over finite fields, with applications to the Gowers Norm*, submitted, 2007.

[5] S. Jukna, **Extremal Combinatorics with Applications in Computer Science**, Springer–Verlag, 2001.

[6] T. Kaufman and S. Lovett, *Average case to worst case reduction for polynomials*, proceedings of the 49th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2008), to appear.

[7] T. Kasami and N. Tokura, *On the weight structure of Reed–Muller codes*, IEEE Transactions on Information Theory 16(6):752–759, 1970.

[8] T. Kasami, N. Tokura and S. Azumi, *On the weight enumeration of weights less than* $2.5d$ *of Reed–Muller codes*, Information and Control 30(4):380–395, 1976.

[9] P. Keevash and B. Sudakov, *Set systems with restricted cross-intersections and the minimum rank of inclusion matrices*, SIAM J. of Discrete Mathematics 18(4):713–727, 2005.

[10] J. MacWilliams and N. Sloane, **The Theory of Error Correcting Codes**, North–Holland Publishing Co., Amsterdam, 1977.

[11] E. Viola and A. Wigderson, *Norms, XOR lemmas, and lower bounds for* $GF(2)$ *polynomials and multiparty protocols*, proceedings of the 22nd IEEE Conference on Computational Complexity, pp. 141–154, 2007.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/