Electronic Colloquium on Computational Complexity, Report No. 182 (2010)

# An elementary proof of anti-concentration of polynomials in Gaussian variables

Shachar Lovett*
Institute for Advanced Study
Princeton, NJ
slovett@math.ias.edu

November 25, 2010

**Abstract**

Recently there has been much interest in polynomial threshold functions in the context of learning theory, structural results and pseudorandomness. A crucial ingredient in these works is the understanding of the distribution of low-degree multivariate polynomials evaluated over normally distributed inputs. In particular, the two important properties are exponential tail decay and anti-concentration.

In this work we study the latter property. The important work in this area is by Carbery and Wright, who gave a tight bound for anti-concentration of polynomials in normal variables. However, the proof of their result is quite complex. We give a weaker anti-concentration result which has an elementary proof, based on some convexity arguments, simple analysis and induction on the degree. Moreover, our proof technique is robust and extends to other distributions.

## 1 Introduction

There has been much interest recently in linear and polynomial threshold functions in the contexts of learning theory, structural results and pseudorandomness [BELY09, DHK$^+$10, DRST09, DSTW10, DGJ$^+$09, HKM09, Kan10, MZ10]. A crucial ingredient in the analysis of all these works is the understanding of the distribution of a low-degree multivariate polynomial evaluated over normally distributed inputs. The distribution of polynomials in normal variables has two important properties: on the one hand, their tails decay exponentially fast, while on the other hand these distributions are not too concentrated around any specific value.

*Supported by NSF grant DMS-0835373.



ISSN 1433-8092

This paper studies the latter property of anti-concentration of polynomials in normal variables. Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a polynomial of degree $d$, and assume it is normalized to have $\mathrm{Var}[f] = 1$ under the normal distribution. The main result in this area is a theorem of Carbery and Wright [CW01] that shows that for any $t \in \mathbb{R}$ and $\varepsilon > 0$,

$$\Pr_{\mathbf{x} \sim \mathcal{N}^n}[|f(\mathbf{x}) - t| \leq \varepsilon] \leq O(d) \cdot \varepsilon^{1/d}, \tag{1}$$

where $\mathcal{N} = \mathcal{N}(0,1)$ is a standard normal variable. This result is tight up to the hidden constant. The only major caveat with the result of Carbery and Wright is that its proof is quite complicated. The goal of this note is to demonstrate that a weaker version of an anti-concentration result has an elementary proof, based only on some convexity arguments, simple analysis and induction on the degree.

**Theorem 1.1.** *Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a degree $d$ polynomial, normalized to have $\mathrm{Var}[f] = 1$. Then for any $t \in \mathbb{R}$ and $\varepsilon > 0$,*

$$\Pr_{\mathbf{x} \sim \mathcal{N}^n}[|f(x) - t| \leq \varepsilon] \leq C_d \cdot \varepsilon^{1/c_d},$$

*where $C_d = O(d)^d$ and $c_d = O(d \cdot 4^d)$.*

Our proof technique is robust and extends to other distributions. Let $\mathcal{D}$ be a distribution over $\mathbb{R}$. We will require the distribution to have some anti-concentration property. Specifically, we require anti-concentration for quadratic polynomials which come from positive semi-definite matrices.

**Definition 1** (PSD anti-concentration property)**.** A distribution $\mathcal{D}$ has *PSD anti-concentration* if there exist $C, c > 0$ such that the following holds. Let $A$ be an $n \times n$ positive semi-definite matrix with $\mathrm{Tr}(A) = 1$. Then for any $\varepsilon > 0$,

$$\Pr_{\mathbf{x} \in \mathcal{D}^n}[\mathbf{x}^t A \mathbf{x} \leq \varepsilon] \leq C \cdot \varepsilon^c.$$

For example, the normal distribution has PSD anti-concentration with $c = 1/2$ (see Claim 4.2). Define $d\mathcal{D} := \mathcal{D} + \ldots + \mathcal{D}$ to be the distribution of the sum of $d$ independent elements sampled from $D$, and $\mathcal{D} - \mathcal{D}$ to be the distribution of the difference of two independent elements sampled from $D$.

**Theorem 1.2.** *Let $\mathcal{D}$ be a distribution over $\mathbb{R}$ such that $\mathcal{D} - \mathcal{D}$ has PSD anti-concentration. Then there exist $C_d, c_d > 0$ such that the following holds. Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a degree $d$ polynomial, normalized to have $\mathrm{Var}_{(d\mathcal{D})^n}[f] = 1$. Then for any $t \in \mathbb{R}$ and $\varepsilon > 0$,*

$$\Pr_{\mathbf{x} \sim (d\mathcal{D})^n}[|f(x) - t| \leq \varepsilon] \leq C_d \cdot \varepsilon^{1/c_d},$$

*where $c_d = O(d \cdot 2^{O(d)})$.*

In particular, Theorem 1.1 is an instance of Theorem 1.2 for $\mathcal{D} = \mathcal{N}(0, 1/d)$ such that $d\mathcal{D} = \mathcal{N}(0,1)$ and $\mathcal{D} - \mathcal{D} = \mathcal{N}(0, 2/d)$.

## 1.1 Proof overview

We sketch the proof for normal variables. Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a degree $d$ polynomial with $\mathrm{Var}[f] = 1$. Assume for now for the simplicity of the exposition that $f$ is multilinear and homogeneous of degree $d$. That is,

$$f(x) = \sum_{I \subset [n]: |I|=d} f_I \prod_{i \in I} x_i,$$

where $\sum |f_I|^2 = \mathrm{Var}[f] = 1$.

The proof is established by first reducing to the special family of set-multilinear polynomials. Let $\mathbf{y}^1, \ldots, \mathbf{y}^d \in \mathbb{R}^n$ be $d$ sets of variables, where $\mathbf{y}^j = (y_1^j, \ldots, y_n^j)$. A polynomial $g(\mathbf{y}^1, \ldots, \mathbf{y}^d)$ is *set-multilinear of degree* $d$ if

$$g(\mathbf{y}^1, \ldots, \mathbf{y}^d) = \sum_{I=(i_1,\ldots,i_d) \in [n]^d} g_I \prod_{j=1}^{d} y_{i_j}^j,$$

that is, any monomial of $g$ contains exactly one variable from each one of $\mathbf{y}^1, \ldots, \mathbf{y}^d$. The advantage of reducing to set-multilinear polynomials is that bounds for such polynomials are amenable to induction on the degree.

### 1.1.1 Reduction to set-multilinear polynomials

The reduction uses directional derivatives. For $\mathbf{y} \in \mathbb{R}^n$ define the derivative of $f$ in direction $\mathbf{y}$ to be $(\Delta_{\mathbf{y}} f)(\mathbf{x}) := f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})$. We define iterated derivatives in directions $\mathbf{y}^1, \ldots, \mathbf{y}^k \in \mathbb{R}^n$ by $\Delta_{\mathbf{y}^1,\ldots,\mathbf{y}^k} f = \Delta_{\mathbf{y}^1} \ldots \Delta_{\mathbf{y}^k} f$. It is not hard to verify (Claim 3.2) that as $f$ is a degree $d$ polynomial, if we derive it in directions $\mathbf{y}^1, \ldots, \mathbf{y}^d$ we get

$$\Delta_{\mathbf{y}^1,\ldots,\mathbf{y}^d} f(\mathbf{x}) = \sum_{I=(i_1,\ldots,i_d) \in [n]^d} f_I \prod_{j=1}^{d} y_{i_j}^j, \tag{2}$$

where $f_I$ denotes the corresponding coefficient of $f$ for the (unordered) set $I$. In particular, $\Delta_{\mathbf{y}^1,\ldots,\mathbf{y}^d} f$ is a constant function (i.e. it does not depend on $\mathbf{x}$), and is a set-multilinear polynomial of degree $d$.

The next ingredient is a convexity argument. Fix a distribution $\mathcal{D}$ over $\mathbb{R}^n$. Let $\{X^{i,j} \in \mathcal{D}\}_{i \in [d], j \in \{0,1\}}$ be independently chosen, and for each $I \in \{0,1\}^d$ define a random variable

$$\overline{X}_I = \sum_{i \in [d]} X^{i,I_i}.$$

Let also $W^1, \ldots, W^d \sim \mathcal{D}$ be independently chosen. An iterated application of the Cauchy-Schwarz inequality (Claim 3.3) shows that for any subset $S \subset \mathbb{R}^n$ we have

$$\Pr[\forall I \in \{0,1\}^d, \overline{X}_I \in S] \geq \Pr[W^1 + \ldots + W^d \in S]^{2^d}. \tag{3}$$

We now apply these as follows. Let $S = \{\mathbf{x} \in \mathbb{R}^n : |f(\mathbf{x}) - t| \le \varepsilon\}$. Let $\mathcal{D} = \mathcal{N}(0, 1/d)$ so that $\overline{X}_I \sim \mathcal{N}(0, 1)$ for all $I \in \{0,1\}^d$, and also $W^1 + \ldots + W^d \sim \mathcal{N}(0, 1)$. We thus have

$$\Pr_{X \in \mathcal{N}^n}[|f(X) - t| \le \varepsilon] \le \Pr[\forall I \in \{0,1\}^d, |f(\overline{X}_I) - t| \le \varepsilon]^{1/2^d}.$$

Define a polynomial $h(\{X^{i,j}\}) := \sum_{I \in \{0,1\}^d} (-1)^{|I|} f(\overline{X}_I)$. Note that if $|f(\overline{X}_I) - t| \le \varepsilon$ for all $I \in \{0,1\}^d$, then $|h(\{X^{i,j}\})| \le 2^d \varepsilon$. We thus get

$$\Pr_{X \in \mathcal{N}^n}[|f(X) - t| \le \varepsilon] \le \Pr[|h(\{X^{i,j}\})| \le 2^d \varepsilon]^{1/2^d}. \tag{4}$$

We now study the structure of $h$. Define $X' := \sum_{i=1}^d X^{i,0}$ and $Y^i := X^{i,1} - X^{i,0}$. It is not hard to verify that

$$h(\{X^{i,j}\}) = (\Delta_{Y^1,\ldots,Y^d} f)(X') = g(Y^1, \ldots, Y^d).$$

Moreover, note that $Y^1, \ldots, Y^d \in \mathcal{N}(0, 2/d)$ and are independent. Thus, we obtained the bound

$$\begin{aligned} \Pr_{X \in N^n}[|f(X) - t| \le \varepsilon] &\le \Pr_{Y^1,\ldots,Y^d \in \mathcal{N}(0,2/d)^n}[|g(Y^1, \ldots, Y^d) - t| \le 2^d \varepsilon]^{1/2^d} \\ &= \Pr_{Z^1,\ldots,Z^d \in \mathcal{N}^n}[|g(Z^1, \ldots, Z^d) - t(d/2)^{d/2}| \le (2d)^{d/2} \varepsilon]^{1/2^d}, \end{aligned} \tag{5}$$

where the last equality follows from the multilinearity of $g$. We thus reduced an anti-concentration bound for $f$ to that of a set-multilinear polynomial $g$ (with the same degree and somewhat worse parameters). We note that the analysis presented in this overview is for multilinear $f$; for general $f$ the analysis is somewhat more complicated. One needs to study $f$ in the basis of the Hermite polynomials, which are the orthogonal polynomials under the normal distribution. Also, one needs to handle the scenario where most of the mass of the coefficients of $f$ belongs to monomials of degree less than $d$, which causes some further complications.

### 1.1.2 A bound for set-multilinear polynomials

Let $B_d^{ML}(\varepsilon)$ denote the maximal probability that a set-multilinear polynomial of degree $d$ and variance 1 lies in an interval $(t - \varepsilon, t + \varepsilon)$. We prove a bound for set-multilinear polynomials by induction on the degree.

Let $g(\mathbf{x}^1, \ldots, \mathbf{x}^d) = \sum_{I \in [n]^d} g_I \prod_{j=1}^d x_{i_j}^j$ be a set-multilinear polynomial of degree $d$. We consider fixings of the last variable $\mathbf{x}^d = \mathbf{z}$. Define

$$g_{\mathbf{z}}(\mathbf{x}^1, \ldots, \mathbf{x}^{d-1}) = g(\mathbf{x}^1, \ldots, \mathbf{x}^{d-1}, \mathbf{z}).$$

For every $\mathbf{z} \in \mathbb{R}^n$ the function $g_{\mathbf{z}}$ is a set-multilinear polynomial of degree $d-1$, and of variance $\mathrm{Var}[g_{\mathbf{z}}] = \sum_{i_1,\ldots,i_{d-1} \in [n]} |\sum_{i_d \in [n]} g_{i_1,\ldots,i_d} z_{i_d}|^2$. Denote $\|g_{\mathbf{z}}\|_2 = \sqrt{\mathrm{Var}[g_{\mathbf{z}}]}$ and note that we have by the induction hypothesis that

$$\Pr_{\mathbf{x}^1,\ldots,\mathbf{x}^{d-1} \in \mathcal{N}^n}[|g_{\mathbf{z}}(\mathbf{x}^1, \ldots, \mathbf{x}^{d-1}) - t| \le \varepsilon] \le B_{d-1}^{ML}(\varepsilon / \|g_{\mathbf{z}}\|_2).$$

We now average over $\mathbf{z} \in \mathcal{N}^n$. If $\|g_{\mathbf{z}}\|_2 \leq \sqrt{\varepsilon}$ we use the bound guaranteed by $B^{ML}_{d-1}(\cdot)$; otherwise we use the trivial bound 1. We thus get that

$$\begin{aligned} Pr[|g(\mathbf{x}^1, \ldots, \mathbf{x}^d) - t| \leq \varepsilon] &= \mathbb{E}_{\mathbf{z} \in \mathcal{N}^n}[\Pr[|g_{\mathbf{z}}(\mathbf{x}^1, \ldots, \mathbf{x}^{d-1}) - t| \leq \varepsilon]] \\ &\leq B^{ML}_{d-1}(\sqrt{\varepsilon}) + \Pr_{\mathbf{z} \in \mathcal{N}^n}[\|g_{\mathbf{z}}\|_2 \leq \sqrt{\varepsilon}]. \end{aligned} \tag{6}$$

Thus, to finish the proof we simply need to bound the probability that $\mathrm{Var}[g_{\mathbf{z}}] \leq \varepsilon$. Note that $\mathrm{Var}[g_{\mathbf{z}}]$ is a quadratic polynomial in $\mathbf{z}$ which additionally is positive semi-definite. Using standard techniques we show (Claim 4.2) that for every $\delta > 0$,

$$\Pr_{\mathbf{z} \in \mathcal{N}^n}[\|g_{\mathbf{z}}\|_2 \leq \delta] \leq 2\delta, \tag{7}$$

which concludes the proof.

# 2 Preliminaries

**Notations** We denote by $\mathbb{N} := \{0, 1, 2, \ldots\}$ the set of nonnegative numbers. Let $[n] := \{1, \ldots, n\}$ and $[n]^d = \{(i_1, \ldots, i_d) : i_1, \ldots, i_d \in [n]\}$.

**Normal distribution** Let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution with mean $\mu$ and variance $\sigma^2$, and let $\mathcal{N} := \mathcal{N}(0, 1)$ denote a standard normal variable. We denote by $X \sim \mathcal{N}$ a normally distributed variable, and by $X = (X_1, \ldots, X_n) \sim \mathcal{N}^n$ a random variable where $X_1, \ldots, X_n$ are i.i.d normally distributed.

The (normalized) *Hermite polynomials* are univariate polynomials which form an orthogonal polynomial sequence under the normal distribution. That is, $H_k(x)$ is a degree $k$ polynomial such that $\mathbb{E}_{X \sim \mathcal{N}}[H_k(X)^2] = 1$ and $\mathbb{E}_{X \sim \mathcal{N}}[H_k(X) H_\ell(X)] = 0$ for any $k \neq \ell$. The first Hermite polynomials are $H_0(x) = 1, H_1(x) = x, H_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1), H_3(x) = \frac{1}{\sqrt{6}}(x^3 - x), \ldots$. The coefficient of $x^k$ in $H_k(x)$ is $1/\sqrt{k!}$.

**Multivariate polynomials** A function $f : \mathbb{R}^n \to \mathbb{R}$ is a degree $d$ polynomial if it can be represented as the sum of monomials of total degree at most $d$. It will be convenient to us to represent a polynomial $f$ in two basis: the usual monomial basis, and the Hermite polynomials basis. Let $e \in \mathbb{N}^n$. We denote by $|e| = \sum_i e_i$ the hamming weight of $e$. We represent $f$ in the monomial basis as

$$f(\mathbf{x}) = \sum_{e \in \mathbb{N}^n : |e| \leq d} f^M_e \prod_{i=1}^n x_i^{e_i},$$

where the superscript $M$ denotes that coefficient are in the monomial basis. We will denote by $f^{M;k}$ the part of $f$ which is homogeneous of degree $k$. That is, $f = \sum_{k=0}^d f^{M;k}$ where

$$f^{M;k}(\mathbf{x}) := \sum_{e \in \mathbb{N}^n : |e| = k} f^M_e \prod_{i=1}^n x_i^{e_i}.$$

We also represent $f$ in the basis of the Hermite polynomials,

$$f(\mathbf{x}) = \sum_{e \in \mathbb{N}^n : |e| \le d} f^H_e \prod_{i=1}^n H_{e_i}(x_i).$$

We denote by $f^{H;k}$ the homogeneous Hermite part of degree $k$. That is, $f = \sum_{k=0}^d f^{H;k}$ where

$$f^{H;k}(\mathbf{x}) := \sum_{e \in \mathbb{N}^n : |e| = k} f^H_e \prod_{i=1}^n H_{e_i}(x_i).$$

We note that the coefficients of $f$ in the monomial basis $\{f^M_e\}$ and in the Hermite basis $\{f^H_e\}$ are related by an invertible linear transformation. In particular, for $|e| = d$ this relation is particularly simple.

*Claim* 2.1. Let $f$ be a degree $d$ polynomial. Then for every $e \in \mathbb{N}^n$ with $|e| = d$ we have

$$f^M_e = \left( \prod_{i=1}^n \frac{1}{\sqrt{e_i!}} \right) f^H_e.$$

The importance of the Hermite basis is that in this basis, the expected value and variance of $f$ under normal variables have a simple expression: $\mathbb{E}[f] := f^H_{0^n}$ and $\mathbb{E}[f^2] = \sum_e |f^H_e|^2$. We further denote $\|f\|_2 = \sqrt{\mathbb{E}[f^2]}$ and $\mathrm{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2$. We denote by $\mathrm{Poly}_{n,d}$ the family of polynomials $f(x_1, \ldots, x_n)$ of degree $d$ with $\mathrm{Var}[f] = 1$.

**Set-multilinear polynomials** A function $g : (\mathbb{R}^n)^d \to \mathbb{R}$ is a *set-multilinear polynomial* of degree $d$ if it has the following form. Let $\mathbf{x}^1, \ldots, \mathbf{x}^d \in \mathbb{R}^n$ be variables, where $\mathbf{x}^j = (x^j_1, \ldots, x^j_n)$. Then,

$$g(\mathbf{x}^1, \ldots, \mathbf{x}^d) = \sum_{I = (i_1, \ldots, i_d) \in [n]^d} g_I \prod_{j=1}^d x^j_{i_j}.$$

We have $\mathbb{E}[g] = 0$ and $\mathbb{E}[g^2] = \sum_{I \in [n]^d} |g_I|^2$ under the normal distribution. Analogously, let $\|g\|_2 = \sqrt{\mathbb{E}[g^2]}$ and $\mathrm{Var}[g] = \mathbb{E}[g^2] - \mathbb{E}[g]^2$. We denote by $\mathrm{Poly}^{ML}_{n,d}$ the family of set multilinear polynomials $g$ of degree $d$ with $\mathrm{Var}[g] = 1$.

# 3 Reduction to set-multilinear polynomials

Fix $n \in \mathbb{N}$. Let $\mathrm{B}_d(\varepsilon)$ denote the maximal probability that a degree $d$ multivariate polynomial, evaluated over normal variables, lies in some interval $(t - \varepsilon, t + \varepsilon)$, that is

$$\mathrm{B}_d(\varepsilon) := \sup \left\{ \Pr_{X \sim \mathcal{N}^n} [|f(X) - t| \le \varepsilon] : f \in \mathrm{Poly}_{n,d}, t \in \mathbb{R} \right\}.$$

The goal of this work is to provide bounds on $\mathrm{B}_d(\varepsilon)$. We do so in two steps: we first relate it to a related quantity for set-multilinear polynomials, which then follow to prove bounds for set-multilinear polynomials.

We denote by $\mathrm{B}_d^{ML}(\varepsilon)$ the maximal probability that a set-multilinear degree $d$ polynomial, evaluated over normal variables, lies in some interval $(t-\varepsilon, t+\varepsilon)$, that is

$$\mathrm{B}_d^{ML}(\varepsilon) := \sup \left\{ \Pr_{\mathbf{x}_1,\ldots,\mathbf{x}_d \sim \mathcal{N}^n} [|g(\mathbf{x}_1,\ldots,\mathbf{x}_d) - t| \le \varepsilon] : g \in \mathrm{Poly}_{n,d}^{ML}, t \in \mathbb{R} \right\}.$$

The main result we prove in this section is the following lemma.

**Lemma 3.1.** $B_d(\varepsilon) \le \left(B_d^{ML}(\varepsilon^{1/4d})\right)^{1/2^d} + 16(2d)^d \cdot \varepsilon^{1/2d}$.

The first step in the proof is to reduce polynomials to set-multilinear polynomials via iterated directional derivatives. The directional derivative of $f(\mathbf{x})$ in direction $\mathbf{y} \in \mathbb{R}^n$ is defined as

$$(\Delta_{\mathbf{y}} f)(\mathbf{x}) := f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}),$$

and iterated derivatives as

$$(\Delta_{\mathbf{y}^1,\ldots,\mathbf{y}^k} f)(\mathbf{x}) := (\Delta_{\mathbf{y}^1} \ldots \Delta_{\mathbf{y}^k} f)(\mathbf{x}) = \sum_{I \subseteq [k]} (-1)^{k-|I|} f(\mathbf{x} + \sum_{i \in I} \mathbf{y}^i).$$

An important property of directional derivatives is that they reduce degrees. That is, if $f$ is a degree $d$ polynomial then $\Delta_{\mathbf{y}} f$ is a polynomial of degree at most $d-1$ for any $\mathbf{y} \in \mathbb{R}^n$, and

$$\deg(\Delta_{\mathbf{y}_1,\ldots,\mathbf{y}_k} f) \le d - k \tag{8}$$

for any $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^n$. In particular, if $f$ is a degree $d$ polynomial, then $f_{\mathbf{y}_1,\ldots,\mathbf{y}_d}(\mathbf{x})$ is a function of degree at most 0, i.e. a constant function, whose value depends only on $\mathbf{y}_1, \ldots, \mathbf{y}_d$ and not on $x$. The next claim establishes that this is in fact a set-multilinear polynomial in $\mathbf{y}_1, \ldots, \mathbf{y}_d$.

*Claim* 3.2. Let $f(\mathbf{x})$ be a degree $d$ polynomial. Consider the polynomial

$$g(\mathbf{x}, \mathbf{y}^1, \ldots, \mathbf{y}^d) := (\Delta_{\mathbf{y}^1,\ldots,\mathbf{y}^d} f)(\mathbf{x}).$$

For $I = (i_1, \ldots, i_d) \in [n]^d$ let $e(I) \in \mathbb{N}^n$ be defined as $e(I)_k = |\{j : i_j = k\}|$. Let $g_I := f^M_{e(I)} \cdot \prod_{k=1}^n (e(I)_k!)$. Then

$$g = g(\mathbf{y}^1, \ldots, \mathbf{y}^d) = \sum_{I \in [n]^d} g_I \prod_{j=1}^d y^j_{i_j}.$$

In particular, $g(\mathbf{y}^1, \ldots, \mathbf{y}^d)$ is set-multilinear of degree $d$ and

$$\|g\|_2 \ge \|f^{H;d}\|_2,$$

where $f^{H;d}$ is the homogeneous part of $f$ of degree $d$ in the Hermite basis.

*Proof.* We start by arguing about the structure of $g$. It will suffice to show it for each monomial of $f$ and then extend by linearity of the derivative operator. Let $m(\mathbf{x}) = x_{i_1} \dots x_{i_k}$ be a monomial where $k \le d$. If $k \le d-1$ we have $\Delta_{\mathbf{y}^1,\dots,\mathbf{y}^d} m \equiv 0$ as the $\mathbf{x}$ degree reduces below zero. Thus is suffices to study monomials of degree $d$. Let $m(x) = x_{i_1} \dots x_{i_d}$ where $i_1, \dots, i_d$ are not necessarily distinct. It is a routine calculation to verify that

$$\Delta_{\mathbf{y}^1,\dots,\mathbf{y}^d} m = \sum_{\sigma \in S_d} y_{i_1}^{\sigma(1)} \dots y_{i_d}^{\sigma(d)}, \tag{9}$$

where $S_d$ is the group of permutations on $[d]$. Let $I = (i_1, \dots, i_d)$. Each monomial $y_{i_1}^1 \dots y_{i_d}^d$ appears $\prod_{k=1}^n (e(I)_k!)$ times in (9). Hence we get the formula

$$g_I := f_{e(I)}^M \cdot \prod_{k=1}^n (e(I)_k!).$$

We next lower bound $\|g\|_2$. We have $\|g\|_2^2 = \sum_{e \in \mathbb{N}^n : |e|=d} |f_e^M|^2 \cdot (\prod_{i=1}^n e_i!)^2$. By Claim 2.1 we have for all $e \in \mathbb{N}^n$ with $|e| = d$ that

$$f_e^M = \left( \prod_{i=1}^n \frac{1}{\sqrt{e_i!}} \right) f_e^H.$$

Substituting we get the bound

$$\|g\|_2^2 \ge \sum_{e \in \mathbb{N}^n : |e|=d} |f_e^H|^2 = \|f^{H;d}\|_2^2.$$

□

The next claim bounds the probability that $f(\mathbf{x})$ is concentrated by the probability that $g(\mathbf{y}^1, \dots, \mathbf{y}^d) = \Delta_{\mathbf{y}^1,\dots,\mathbf{y}^d} f$ is concentrated.

*Claim* 3.3. Let $\mathcal{D}$ be a distribution over $\mathbb{R}^n$. Let $\{X^{i,j} \sim \mathcal{D}\}_{i \in [d], j \in \{0,1\}}$ be $2d$ independent random variables. For all $I \in \{0,1\}^d$ define random variables

$$\overline{X}^I := \sum_{i=1}^d X^{i,I_i}.$$

Let $W^1, \dots, W^d \sim \mathcal{D}$ be another collection of independent random variables. Then for any measurable set $S \subset \mathbb{R}^n$ we have

$$\Pr\left[\forall I \in \{0,1\}^d,\ \overline{X}^I \in S\right] \ge \Pr\left[W^1 + \ldots + W^d \in S\right]^{2^d}.$$

*Proof.* Let $f(\mathbf{x}) = \mathbf{1}_{\mathbf{x} \in S}$ be the indicator function of $S$. For $0 \le k \le d$ define

$$E_k := \mathbb{E}\left[ \prod_{I \in \{0,1\}^k} f\left( \sum_{i=1}^k X^{i,I_i} + \sum_{j=k+1}^d W^j \right) \right],$$

where $E_0 := \mathbb{E}[f(W^1 + \ldots + W^d)]$. We need to show that $E_d \geq (E_0)^{2^d}$. We will do so by showing that $E_k \geq E_{k-1}^2$ for all $1 \leq k \leq d$. To this end, we have

$$E_{k-1}^2 = \left(\mathbb{E}_{\{X^{i,j}\},W^{k+1},\ldots,W^d}\mathbb{E}_{W^k}\left[\prod_{I\in\{0,1\}^{k-1}} f\left(\sum_{i=1}^{k-1} X^{i,I_i} + \sum_{j=k}^{d} W^j\right)\right]\right)^2$$
$$\leq \mathbb{E}_{\{X^{i,j}\},W^{k+1},\ldots,W^d}\left(\mathbb{E}_{W^k}\left[\prod_{I\in\{0,1\}^{k-1}} f\left(\sum_{i=1}^{k-1} X^{i,I_i} + \sum_{j=k}^{d} W^j\right)\right]\right)^2,$$

where the inequality follows from the Cauchy-Schwarz inequality. Opening brackets, we have two identical copies $W^{k,0}, W^{k,1}$ for $W^k$, which gives

$$E_{k-1}^2 \leq \mathbb{E}_{\{X^{i,j}\},W^{k+1},\ldots,W^d}\mathbb{E}_{W^{k,0},W^{k,1}}\left[\prod_{\ell\in\{0,1\}}\prod_{I\in\{0,1\}^{k-1}} f\left(\sum_{i=1}^{k} X^{i,I_i} + W^{k,\ell} + \sum_{j=k+1}^{d} W^j\right)\right]$$
$$= E_k.$$

where the last equality follows by definition (simply rename $W^{k,0}, W^{k,1}$ to $X^{k,0}, X^{k,1}$). □

We next use Claim 3.3 to bound the probability that $|f(\mathbf{x}) - t| \leq \varepsilon$ by $B_d^{ML}(\cdot)$, as long as $\|f^{H;d}\|_2$ is not too small.

*Claim* 3.4. Let $f(\mathbf{x})$ be a degree $d$ polynomial. Then for any $t \in \mathbb{R}$ and any $\varepsilon > 0$,

$$\Pr_{X\in\mathcal{N}^n}[|f(X) - t| \leq \varepsilon] \leq B_d^{ML}\left(\varepsilon \cdot \frac{(2d)^{d/2}}{\|f^{H;d}\|_2}\right)^{1/2^d}.$$

*Proof.* Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a degree $d$ polynomial with $\mathrm{Var}[f] = 1$, and let $t \in \mathbb{R}$ and $\varepsilon > 0$. Define $S \subset \mathbb{R}^n$ by $S := \{\mathbf{x} \in \mathbb{R}^n : |f(\mathbf{x}) - t| \leq \varepsilon\}$. Our goal is to bound the measure of $S$ under the normal distribution. Let $\{X^{i,j} \sim \mathcal{N}(0, 1/d)^n\}_{i\in[d],1\leq j\leq\{0,1\}}$ be $2d$ independent random variables. For all $I \in \{0,1\}^d$ define new random variables $\overline{X}_I := \sum_{i=1}^d X^{i,I_i}$. Note that for any $I \in \{0,1\}^d$ we have $\overline{X}_I \sim \mathcal{N}^n$ since $X^{i,j} \sim \mathcal{N}(0, 1/d)$. By Claim 3.3 we have

$$\Pr_{X\in\mathcal{N}^n}[X \in S] \leq \left(\Pr\left[\forall I \in \{0,1\}^d,\ \overline{X}_I \in S\right]\right)^{1/2^d}. \tag{10}$$

We now bound the latter term. Define $h : (\mathbb{R}^n)^{2d} \to \mathbb{R}$ by

$$h(\{X^{i,j}\}) := \sum_{I\in\{0,1\}^d} (-1)^{d-|I|} f(\overline{X}_I),$$

where $|I| = I_1 + \ldots + I_d$ is the hamming weight of $I$. Assume that indeed $\overline{X}_I \in S$ for all $I \in \{0,1\}^d$. That is, we have $|f(\overline{X}_I) - t| \leq \varepsilon$ for all $I \in \{0,1\}^d$. In particular, we get that $|h(\{X^{i,j}\})| \leq 2^d\varepsilon$. We thus have

$$\Pr[\forall I \in \{0,1\}^d,\ \overline{X}_I \in S] \leq \Pr[|h(\{X^{i,j}\})| \leq 2^d\varepsilon]. \tag{11}$$

We next turn to study the function $h$. Define $X' := \sum_{i=1}^{d} X^{i,0}$ and $Y^i := X^{1,i} - X^{0,i}$. It is simple to verify that

$$h(\{X^{i,j}\}) = \Delta_{Y^1,\dots,Y^d} f(X').$$

Thus, by claim 3.2, $h(\{X^{i,j}\}) = g(Y^1, \dots, Y^d)$ where $g$ is a set-multilinear polynomial of degree $d$ with $\|g\|_2 \geq \|f^{H;d}\|_2$. Moreover, $Y^1, \dots, Y^d \sim \mathcal{N}(0, 2/d)$ are independent. Recalling that $g$ is multilinear, we have

$$\begin{aligned}
&\Pr_{Y^1,\dots,Y^d \in \mathcal{N}(0,2/d)^n}[|g(Y^1, \dots, Y^d)| \leq 2^d \varepsilon] \\
&= \Pr_{Z^1,\dots,Z^d \in \mathcal{N}^n}[|g(\sqrt{2/d} \cdot Z^1, \dots, \sqrt{2/d} \cdot Z^d)| \leq 2^d \varepsilon] \\
&= \Pr_{Z^1,\dots,Z^d \in \mathcal{N}^n}[|g(Z^1, \dots, Z^d)| \leq \left(\sqrt{d/2}\right)^d 2^d \varepsilon] && (12) \\
&\leq B_d^{ML}\left(\varepsilon \cdot (2d)^{d/2} / \|g\|_2\right). && (13)
\end{aligned}$$

and the claim follows since $\|g\|_2 \geq \|f^{H;d}\|_2$. □

We are now ready to prove Lemma 3.1.

*Proof of Lemma 3.1.* Let $f(\mathbf{x})$ be a degree $d$ polynomial with $\mathrm{Var}[f] = 1$. Let $f^{H;k}$, $k \in [d]$ be the homogeneous parts of $f$ in the Hermite basis. By the orthogonality of the Hermite polynomials we have $\mathbb{E}[f^{(H;k)}] = 0$ and

$$\sum_{k=1}^{d} \|f^{(H;k)}\|_2^2 = \mathrm{Var}[f] = 1.$$

Let $0 < \eta \leq 1/2$ to be determined later, and let $\ell \in [d]$ be maximal such that $\|f^{H;\ell}\|_2 \geq \eta^\ell$ (note there must exist such $\ell$). Decompose $f = f_1 + f_2$ where $f_1 = \sum_{k=1}^{\ell} f^{H;k}$ and $f_2 = \sum_{k=\ell+1}^{d} f^{H;k}$. Let $c > 1$ be a parameter to be determined later. We will bound

$$\Pr_{X \in \mathcal{N}^n}[|f(X) - t| > \varepsilon] \leq \Pr_{X \in \mathcal{N}^n}[|f_1(X) - t| \leq (c+1)\varepsilon] + \Pr_{X \in \mathcal{N}^n}[|f_2(X)| \geq c\varepsilon].$$

We will establish the claim by bounding both terms for appropriate choices of $\eta, c$.

We start by bounding $\Pr[|f_1(X) - t| \geq (c+1)\varepsilon]$. Note that $f_1$ is a polynomial of degree $\ell$, and $f_1^{H;\ell} = f^{H;\ell}$. In particular $\|f_1^{H;\ell}\|_2 \geq \eta^\ell$. Applying Claim 3.4 we get that

$$\begin{aligned}
\Pr_{X \in \mathcal{N}^n}[|f_1(X) - t| \leq (c+1)\varepsilon] &\leq B_\ell^{ML}\left((c+1)\varepsilon \cdot \frac{(2\ell)^{\ell/2}}{\eta^\ell}\right)^{1/2^\ell} \\
&\leq B_d^{ML}\left(\varepsilon (2d)^{d/2} \cdot \frac{2c}{\eta^\ell}\right)^{1/2^d},
\end{aligned}$$

where the second inequality follows from the monotonicity of $B^{ML}$ and since $c > 1$.

We now turn to bound $\Pr[|f_2(X)| \le c\varepsilon]$. We have $\mathbb{E}[f_2] = 0$ and $\mathbb{E}[f_2^2] = \sum_{k=\ell+1}^{d} \eta^{2k} \le 2\eta^{2(\ell+1)}$. Applying Chebychev's inequality we get

$$\Pr_{X\in\mathcal{N}^n}[|f_2(X)| \ge c\varepsilon] \le \frac{\mathrm{Var}[f_2]}{(c\varepsilon)^2} \le \left(\frac{2\eta^{\ell+1}}{c\varepsilon}\right)^2.$$

We now set parameters. Set $\eta = \varepsilon^{1/2d}$ and $c := \frac{(2d)^{-d/2}\eta^{\ell+1/2}}{2\varepsilon}$. Assuming that $c \ge 1$, we have

$$\Pr_{X\in\mathcal{N}^n}[|f_1(X) - t| \le (c+1)\varepsilon] \le B_d^{ML}(\sqrt{\eta})^{1/2^d} = B_d^{ML}(\varepsilon^{1/4d})^{1/2^d}$$

and

$$\Pr_{X\in\mathcal{N}^n}[|f_2(X)| \ge c\varepsilon] \le 16\eta(2d)^d = 16\varepsilon^{1/2d}(2d)^d.$$

Note that if $16\varepsilon^{1/2d}(2d)^d \ge 1$ then the bound is trivial; we can thus assume that $16\varepsilon^{1/2d}(2d)^d \le 1$, in which case $c > 1$ as required. □

# 4 A bound for set-multilinear polynomials

We prove in this section the following result.

**Lemma 4.1.** *For every $d \ge 2$,*

$$B_d^{ML}(\varepsilon) \le B_{d-1}^{ML}(\sqrt{\varepsilon}) + 2\sqrt{\varepsilon}.$$

*In particular,*

$$B_d^{ML}(\varepsilon) \le 2d \cdot \varepsilon^{1/2^{d-1}}.$$

The conclusion of Lemma 4.1 follows immediately from the reduction from $B_d^{ML}$ to $B_{d-1}^{ML}$ and by standard estimates for normal variables for the case of $d = 1$. Let $f(\mathbf{x}^1, \dots, \mathbf{x}^d)$ be a set-multilinear polynomial of degree $d$ with $\mathrm{Var}[f] = 1$. Fix $t \in \mathbb{R}$ and $\varepsilon > 0$. We will derive bounds on $\Pr[|f(\mathbf{x}^1, \dots, \mathbf{x}^d) - t| \le \varepsilon]$. Consider fixings of $\mathbf{x}^d$. For $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$ denote $f_{\mathbf{z}}(\mathbf{x}^1, \dots, \mathbf{x}^{d-1}) := f(\mathbf{x}^1, \dots, \mathbf{x}^{d-1}, \mathbf{z})$. Note that $f_{\mathbf{z}} \in \mathrm{Poly}_{d-1}^{ML}$ and that

$$\|f_{\mathbf{z}}\|_2^2 = \sum_{i_1,\dots,i_{d-1}\in[n]} \left(\sum_{i_d} f_{i_1,\dots,i_d} z_{i_d}\right)^2. \tag{14}$$

We can bound

$$\begin{aligned}
&\Pr_{\mathbf{x}^1,\dots,\mathbf{x}^d\in\mathcal{N}^n}\left[|f(\mathbf{x}^1, \dots, \mathbf{x}^d) - t| \le \varepsilon\right] \\
&= \mathbb{E}_{\mathbf{z}\in\mathcal{N}^n} \Pr_{\mathbf{x}^1,\dots,\mathbf{x}^{d-1}\in\mathcal{N}^n}\left[|f_{\mathbf{z}}(\mathbf{x}^1, \dots, \mathbf{x}^{d-1}) - t| \le \varepsilon\right] \\
&\le \mathbb{E}_{\mathbf{z}\in\mathcal{N}^n}[\min(B_{d-1}^{ML}(\varepsilon/\|f_{\mathbf{z}}\|_2), 1)]. \qquad (15)
\end{aligned}$$

Set $\delta = \sqrt{\varepsilon}$. We can condition on whether $\|f_{\mathbf{z}}\|_2 \ge \delta$ or not. That it,

$$\mathbb{E}_{\mathbf{z}\in\mathcal{N}^n}[\min(B_{d-1}^{ML}(\varepsilon/\|f_{\mathbf{z}}\|_2), 1)] \le B_{d-1}^{ML}(\varepsilon/\delta) + \Pr_{\mathbf{z}\in\mathcal{N}^n}[\|f_{\mathbf{z}}\|_2 \le \delta].$$

We conclude the proof by showing that with high probability $\|f_{\mathbf{z}}\|_2$ is not too small.

*Claim* 4.2. For any $\delta > 0$,
$$\Pr_{\mathbf{z}\sim\mathcal{N}^n}[\|f_{\mathbf{z}}\|_2 \le \delta] \le 2\delta.$$

*Proof.* We first claim that there exists an $m \times n$ real matrix $A$ such that $\|f_{\mathbf{z}}\|_2 = \|A\mathbf{z}\|_2$ and such that $\|A\|_F := \sqrt{\sum_{i,j} A_{i,j}^2} = 1$. To see that, identify $[m] = [n]^{d-1}$ and for each $i_1, \dots, i_{d-1} \in [n]$ define the $(i_1, \dots, i_{d-1})$ row of $A$ as $A_{(i_1,\dots,i_{d-1}),j} = c_{i_1,\dots,i_{d-1},j}$.

Let $B := A^t A$. Note that $\|f_{\mathbf{z}}\|_2^2 = \mathbf{z}^t B \mathbf{z}$ and that $B$ is an $n \times n$ real symmetric matrix. Let $u_1, \dots, u_n \in \mathbb{R}^n$ be the eigenvectors of $B$ with corresponding real eigenvalues $\lambda_1, \dots, \lambda_n \ge 0$. We have $\sum \lambda_i = \mathrm{Tr}(B) = \|A\|_F^2 = 1$. As $B$ is symmetric, we can assume that $u_1, \dots, u_n$ form an orthonormal basis of $\mathbb{R}^n$. Define $y_i := \langle u_i, \mathbf{z}\rangle$, and note that $\mathbf{y} = (y_1, \dots, y_n) \sim \mathcal{N}^n$ since the normal distribution remains invariant under an orthogonal transformation. Thus we have
$$\|f_{\mathbf{z}}\|_2^2 = \sum_{i=1}^n \lambda_i \langle u_i, \mathbf{z}\rangle^2 = \sum_{i=1}^n \lambda_i y_i^2.$$
We thus need to bound $\Pr_{\mathbf{y}\in\mathcal{N}^n}[\sum_{i=1}^n \lambda_i y_i^2 \le \delta^2]$. By Markov's inequality we have
$$\Pr_{\mathbf{y}\sim\mathcal{N}^n}[\sum_{i=1}^n \lambda_i y_i^2 \le \delta^2] = \Pr_{\mathbf{y}\sim\mathcal{N}^n}[e^{-\sum_{i=1}^n (\lambda_i/\delta^2) y_i^2} \ge e^{-1}] \le \frac{\mathbb{E}[e^{-\sum_{i=1}^n (\lambda_i/\delta^2) y_i^2}]}{e^{-1}} = e \cdot \prod_{i=1}^n \mathbb{E}[e^{-(\lambda_i/\delta^2) y_i^2}].$$
Using the simple fact that $\mathbb{E}[e^{-\alpha y^2}] = 1/\sqrt{2\alpha+1}$ we get that
$$\Pr_{\mathbf{y}\sim\mathcal{N}^n}[\sum_{i=1}^n \lambda_i y_i^2 \le \delta^2] \le \frac{e}{\sqrt{\prod_{i=1}^n (1 + 2\lambda_i/\delta^2)}}.$$
We next apply the inequality $(1 + x\lambda) \ge (1+x)^\lambda$ which holds for any $x > 0$ and $0 \le \lambda \le 1$ (this follows from the fact that the function $(1+x)^{1/x}$ is monotone decreasing). We thus conclude as
$$\Pr_{\mathbf{y}\sim\mathcal{N}^n}[\sum_{i=1}^n \lambda_i y_i^2 \le \delta^2] \le \frac{e}{(\sqrt{1+2/\delta^2})^{\sum \lambda_i}} = \frac{e}{\sqrt{1+2/\delta^2}} \le (e/\sqrt{2})\delta \le 2\delta$$
as claimed. □

# 5 A bound for general distributions

We sketch in this section the proof of Theorem 1.2. The proof is identical to that of Theorem 1.1 except that we do not get explicit constants.

The reduction from general polynomials to set-multilinear polynomials is done in the same way as in Lemma 3.1. Let $\mathrm{B}_{d\mathcal{D},d}(\cdot)$ be a bound for for general degree $d$ polynomial under the distribution $d\mathcal{D}$, and $\mathrm{B}^{ML}_{\mathcal{D}-\mathcal{D},d}(\cdot)$ be a bound for set-multilinear degree $d$ polynomials under the distribution $\mathcal{D} - \mathcal{D}$. Following exactly the same proof of Lemma 3.1, we get
$$\mathrm{B}_{d\mathcal{D},d}(\varepsilon) \le B^{ML}_{\mathcal{D}-\mathcal{D},d}(C_d \cdot \varepsilon^{1/2d})^{1/2^d} + C_d \cdot \varepsilon^{1/2d}, \tag{16}$$

where the value of $C_d > 0$ depends on the distribution $\mathcal{D}$ (in particular, it depends on the coefficients of the orthogonal polynomials under $\mathcal{D}$).

The proof of Lemma 4.1 can similarly be extended to general distributions. Assume $\mathcal{D}-\mathcal{D}$ has PSD anti-concentration with constants $c, C$. Then the proof of Lemma 4.1 yields

$$\mathrm{B}^{ML}_{\mathcal{D}-\mathcal{D},d}(\varepsilon) \leq O(C\varepsilon^{(c/2)^d}). \tag{17}$$

Combining (16) and (17) yield Theorem 1.2.

# References


[BELY09] Ido Ben-Eliezer, Shachar Lovett, and Ariel Yadin. Polynomial threshold functions: Structure, approximation and pseudorandomness. *CoRR*, abs/0911.3473, 2009.

[CW01] A. Carbery and J. Wright. Distributional and $l^q$ norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Lett*, 8(3):233–248, 2001.

[DGJ+09] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. In *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 171–180, Washington, DC, USA, 2009. IEEE Computer Society.

[DHK+10] Ilias Diakonikolas, Prahladh Harsha, Adam Klivans, Raghu Meka, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Bounding the average sensitivity and noise sensitivity of polynomial threshold functions. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 533–542, New York, NY, USA, 2010. ACM.

[DRST09] Ilias Diakonikolas, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Average sensitivity and noise sensitivity of polynomial threshold functions. *CoRR*, abs/0909.5011, 2009.

[DSTW10] Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 211–222, Washington, DC, USA, 2010. IEEE Computer Society.

[HKM09] Prahladh Harsha, Adam Klivans, and Raghu Meka. Bounding the sensitivity of polynomial threshold functions. *CoRR*, abs/0909.5175, 2009.

[Kan10] Daniel M. Kane. The gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 205–210, Washington, DC, USA, 2010. IEEE Computer Society.

[MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 427–436, New York, NY, USA, 2010. ACM.



ECCC ISSN 1433-8092

http://eccc.hpi-web.de