A Pseudorandom Generator for Polynomial Threshold Functions of Gaussian with Subpolynomial Seed Length

Daniel M. Kane
Department of Mathematics
Stanford University
dankane@math.stanford.edu

November 4, 2013

Abstract

We present a new pseudorandom generator for polynomial threshold functions of Gaussians that for fixed degree achieves a seed length that is subpolynomial in the desired error.

1 Introduction

We say that a function $f : \mathbb{R}^n \to \{+1, -1\}$ is a degree-d *polynomial threshold function* (PTF) if it is of the form $f(x) = \mathrm{sgn}(p(x))$ for p some (degree-d) polynomial in n variables. Polynomial threshold functions make up a natural class of Boolean functions and have applications to a number of fields of computer science such as circuit complexity [2], communication complexity [15] and learning theory [12].

In this paper we study the question of pseudorandom generators for polynomial threshold functions of Gaussians. In particular, we wish to find explicit functions $F : \{0,1\}^s \to \mathbb{R}^n$ so that for any degree-d polynomial threshold function f

$$\left|\mathbb{E}_{x\sim_u\{0,1\}^s}[f(F(x))] - \mathbb{E}_{X\sim\mathcal{G}^n}[f(X)]\right| < \epsilon.$$

We say that such an F is a pseudorandom generator of seed length s that fools degree-d polynomial threshold functions with respect to the Gaussian distribution to within ϵ. In this paper, we develop a new such generator whose seed length is $O(\log(n)\epsilon^{-o(1)})$ for any fixed d.

1.1 Previous Work

There have been a number of previous papers dealing with the question of finding pseudorandom generators for polynomial threshold functions with respect to the Gaussian distribution or the Bernoulli distribution (i.e. uniform over $\{-1,1\}^n$). Several early works in this area showed that polynomial threshold functions of various degrees could be fooled by arbitrary k-wise independent families of Gaussian or Bernoulli random variables. It should be noted that a k-wise independent family of Bernoulli random variables can be generated from a seed of length $O(k\log(n))$. Although, any k-wise independent family of Gaussians will necessarily have infinite entropy, it is not hard to show that a simple discretization of these random variables leads to a generator of comparable seed length. These results on fooling polynomial threshold functions with k-independence are summarized in Table 1.1 below.

Paper	Bernoulli/Gaussian	d	k
Diakonikolas, Gopalan, Jaiswal, Servedio, Viola [4]	Bernoulli	1	$O(\epsilon^{-2}\log^2(\epsilon^{-1}))$
Diakonikolas, Kane, Nelson [5]	Gaussian	1	$O(\epsilon^{-2})$
Diakonikolas, Kane, Nelson [5]	Both	2	$O(\epsilon^{-8})$[1]
Kane [10]	Both	d	$O_d\left(\epsilon^{-2^{O(d)}}\right)$

Unfortunately, it is not hard to exhibit k-wise independent families of Bernoulli or Gaussian random variables that fail to ϵ-fool the class of degree-d polynomial threshold functions for $k = \Omega(d^2\epsilon^{-2})$, putting a limit on what can be obtained through mere k-independence.

There have also been a number of attempts to produce pseudorandom generators by using more structure than limited independence. In [13], Meka and Zuckerman develop a couple of such generators in the Bernoulli case. Firstly, they make use of pseudorandom generators against space bounded computation to produce a generator of seed length $O(\log(n) + \log^2(\epsilon^{-1}))$ in the special case where $d = 1$. By piecing together several k-wise independent families, they produce a generator for arbitrary degree PTFs of seed length $2^{O(d)}\log(n)\epsilon^{-8d-3}$. In [11], the author develops an improved analysis of this generator allowing for a seed length as small as $O_{c,d}(\log(n)\epsilon^{-11-c})$. For the Gaussian case, the author developed a generator of seed length $2^{O_c(d)}\log(n)\epsilon^{-4-c}$ in [9]. This generator was given essentially as an average several random variables each picked independently from a k-wise independent family of Gaussians. The analysis of this generator was also improved in [11], obtaining a seed length of $O_{c,d}(\log(n)\epsilon^{-2-c})$. For a summary of these results, see Table 1.1.

Paper	Bernoulli/Gaussian	d	s
Meka, Zuckerman [13]	Bernoulli	1	$O(\log(n) + \log^2(1/\epsilon))$
Kane this work	Gaussian	1	$O(\log(n) + \log^{3/2}(1/\epsilon))$
Meka, Zuckerman [13]	Bernoulli	d	$\log(n)2^{O(d)}\epsilon^{-8d-3}$
Kane [9]	Gaussian	d	$\log(n)2^{O(d)}\epsilon^{-4.1}$
Kane [11]	Gaussian	d	$\log(n)O_d(\epsilon^{-2.1})$
Kane [11]	Bernoulli	d	$\log(n)O_d(\epsilon^{-11.1})$
Kane this work	Gaussian	2	$\log(n)\exp(O(\log(1/\epsilon)^{2/3}\log\log(1/\epsilon)^{1/3}))$
Kane this work	Gaussian	d	$\log(n)O_{c,d}(\epsilon^{-c})$

In this paper, we improve on this bound further. We make use of a slight modification of the above generator, by using unequal weights in our averaging process and obtain a seed length of $O_{c,d}(\log(n)\epsilon^{-c})$. Furthermore, for the case $d = 1$, we manage to show that our generators can be constructed to have the much more explicit seed length of

$$O(\log(n) + \log(1/\epsilon)^{3/2}),$$

and for $d = 2$, we obtain a seed length of

$$\log(n)\exp\left(O\left(\log(\epsilon^{-1})^{2/3}\log\log(\epsilon^{-1})^{1/3}\right)\right),$$

1.2 Outline of Paper

In Section 2, we will introduce some conventions that we will use throughout the paper, and review some basic results on polynomials of Gaussians.

The key idea in our analysis is that for p an approximately linear polynomial, that $\mathbb{E}[p(X)]$ is a smooth function in the coefficients of p. Thus, in this case, $\mathbb{E}[p(X)]$ can be well approximated by a polynomial in these coefficients. A precise statement of this idea is presented in Proposition 6, whose proof takes up most of Section 3.

Hence, by the above claim, if p is an approximately linear polynomial, then for $f = \operatorname{sgn} \circ p$, and Y a random variable whose low-degree moments are correct, we will have that $\mathbb{E}[f(\epsilon Y + \sqrt{1-\epsilon^2}X)]$ for X a random Gaussian will be approximately correct. This is because p can be thought of as a nearly linear polynomial in X whose coefficients are given by polynomials in Y. Proposition 6 will therefore imply that this expectation is approximated by the expectation of some polynomial in Y.

Unfortunately, a generic polynomial will not necessarily be approximately linear. We fix this by evaluating the polynomial near a random input. In particular, if we consider $p(\epsilon X_1 + \sqrt{1-\epsilon^2}X_2)$ for a fixed random Gaussian X_2, the resulting polynomial in X_1 is likely to be approximately linear. Such an analysis will work for a sufficiently non-singular polynomial (i.e. a polynomial whose derivative is unlikely to be small). Not

[1] The bound in [5] for the Bernoulli case is actually $\tilde{O}(\epsilon^{-9})$, but this can be easily improved to $O(\epsilon^{-8})$ using technology from [11].

all polynomials are non-singular, but as we will show in Section 4, any polynomial can be written in terms of non-singular polynomials.

In Section 5, we use this theory to develop a sequence of iteratively more detailed generators eventually leading to one that satisfies our requirements. Using the ideas above, we show in Proposition 10 that for X a true n-dimensional Gaussian and Y a k-wise independent family of Gaussians that $\epsilon Y + \sqrt{1-\epsilon^2}X$ produces a PRG that fools degree-d PTFs to within $O_{d,k}(\epsilon^k)$. Iteratively replacing the X involved by such a generator, we obtain a PRG (see Proposition 11) given by

$$\sum_{i=0}^{\ell-1} \epsilon(1-\epsilon^2)^{i/2} Y_i + (1-\epsilon^2)^{\ell/2} X.$$

It is easy to see that for ℓ large, that the X term may safely be removed introducing at most a small error (see Proposition 12). Finally, in Theorem 13, we put these results together to produce a PRG of seed length $O_{c,d}(\log(n)\epsilon^{-c})$.

2 Background

2.1 Notation

We will use the notation $O_a(N)$ to denote a quantity whose absolute value is bounded above by N times some constant depending only on a. Throughout this paper, the variables $X, X_1, \ldots$ will be used to denote multidimensional Gaussian random variables unless stated otherwise.

We recall here the definition of a polynomial threshold function:

Definition. *A function $f : \mathbb{R}^n \to \{\pm 1\}$ is a (degree-d)* polynomial threshold function *(or PTF) if it is of the form $f(x) = sgn(p(x))$ for some (degree-d) polynomial p.*

Another important definition will be the following:

Definition. *We say that a random variable Y taking values in $\mathbb{R}^n$ is* k-moment-matching*, if all of the moments of Y of order at most k agree with the corresponding moments of a standard n-dimensional Gaussian.*

Note that any k-wise independent family of Gaussians is k-moment-matching. Also note that applying any orthogonal transformation to a k-moment-matching random variable yields another k-moment-matching random variable. Throughout this paper we will use the variables $Y, Y_1, Y_i, \ldots$ to denote k-moment-matching random variables for some k unless otherwise specified.

2.2 Polynomials of Gaussians

We recall some basic facts about polynomials of Gaussians. We begin by recalling the L^t-norm of a function.

Definition. *For a function $p : \mathbb{R}^n \to \mathbb{R}$, we let*

$$|p|_t = \left(\mathbb{E}_X[|p(X)|^t]\right)^{1/t}.$$

We now recall some basic distributional results about polynomials evaluated at random Gaussians.

Lemma 1 (Carbery and Wright)**.** *If p is a degree-d polynomial then*

$$Pr(|p(X)| \leq \epsilon |p|_2) = O(d\epsilon^{1/d}).$$

Where the probability is over X, a standard n-dimensional Gaussian.

We will make use of the hypercontractive inequality. The proof follows from Theorem 2 of [14].

Lemma 2. *If p is a degree-d polynomial and $t > 2$, then*

$$|p|_t \leq \sqrt{t-1}^d |p|_2.$$

This bound on higher moments allows us to prove a concentration bound on the distribution of $p(X)$. The following result is a well-known consequence that can be found, for example, in [7].

Corollary 3. *If p is a degree-d polynomial and $N > 0$, then*

$$Pr_X(|p(X)| > N|p|_2) = O\left(2^{-(N/2)^{2/d}}\right).$$

Proof. Apply the Markov inequality and Lemma 2 with $t = (N/2)^{2/d}$. □

2.3 Orthogonal Polynomials

We recall that the Hermite polynomials form an orthonormal basis of the set of polynomials with respect to the Gaussian inner product. Thus any polynomial can be written uniquely as a linear combination of orthogonal polynomials

$$p(x) = \sum_{a \in \mathbb{Z}^n_{\geq 0}} c_a(p) h_a(x).$$

We let

$$p^{[k]}(x) := \sum_{|a|_1 = k} c_a(p) h_a(x)$$

be the sum of the terms in the above decomposition consisting of orthogonal polynomials of degree exactly k. Furthermore, we let

$$p^{[\geq k]} := \sum_{m \geq k} p^{[m]}.$$

We recall from [11] that

$$\mathbb{E}\left[|\partial_{X_1} \cdots \partial_{X_\ell} p(X)|^2\right] = \sum_k k(k-1)\cdots(k-\ell+1)\left|p^{[k]}\right|_2^2. \tag{1}$$

Where ∂_{X_i} above denotes the directional derivative in the X_i direction for X_i a random Gaussian.

We can use Hermite polynomials to prove a relationship between the size of a polynomial at a point and it's L^2 norm.

Lemma 4. *Let q be a degree-d polynomial in n variables, and let $x \in \mathbb{C}^n$. Then*

$$|q(x)| \leq ((n+1)(d+1)(|x|+1))^d |q|_2.$$

Proof. Let

$$q(x) = \sum_{|a|_1 \leq d} c_a h_a(x).$$

By Cauchy-Schwartz,

$$\begin{aligned}
|q(x)| &\leq \left(\sum_{|a|_1 \leq d} c_a^2\right)^{1/2} \left(\sum_{|a|_1 \leq d} h_a^2(x)\right)^{1/2} \\
&= |q|_2 \left(\sum_{|a|_1 \leq d} h_a^2(x)\right)^{1/2} \\
&\leq |q|_2 (n+1)^d \max_{|a|_1 \leq d} |h_a(x)|.
\end{aligned}$$

Let $a = (a_1, \ldots, a_n)$. Recall that the Hermite polynomial of one variable are defined by

$$h_a(y) = \frac{1}{\sqrt{n!}} \left(y - \frac{\partial}{\partial y}\right)^n 1.$$

From this it is easy to see by induction that the sum of the coefficients of $h_a(y)$ is at most $(a+1)^a$. Thus,

$$|h_a(x)| = \prod_{i=1}^{n} |h_{a_i}(x_i)| \leq \prod_{i=1}^{n} (d+1)^{a_i} |x|^{a_i} \leq ((d+1)(|x|+1))^d.$$

This completes the proof. □

2.4 Contraction Mappings

We will also make use of some standard results about contraction mappings. Recall that a function f from a metric space X to itself is a contraction mapping if there is a constant $c < 1$ so that for any $x, y \in X$, $d(f(x), f(y)) \leq cd(x, y)$. We recall the following standard result about contraction mappings:

Lemma 5. *For any contraction mapping $f : X \to X$, on a complete metric space X, f has a unique fixed point.*

Proof. Let $f^n(x)$ denote the n^{th} iterate of f on x. Pick some $x \in X$. It is easy to prove by induction that $d(f^n(x), f^{n+1}(x)) \leq c^n d(x, f(x))$. Thus $d(f^n(x), f^m(x)) \leq \frac{c^n}{1-c} d(x, f(x))$ for any $m > n$. Thus the sequence $\{f^n(x)\}$ is a Cauchy sequence and, thus has some limit, y. It is clear that $y = f(y)$, and is thus is a fixed point. To show that y is the only fixed point note that if x is another fixed point then $d(x, y) = d(f(x), f(y)) \leq cd(x, y)$, and thus that $d(x, y) = 0$, and therefore $x = y$. □

3 Polynomial Approximation of Expectations

In this Section, we prove the following Proposition, which says that the expectation of a threshold function of a polynomial p, that is approximately linear can be approximated by a polynomial in the coefficients of p.

Proposition 6. *Let d, m and k be positive integers. Let $p : \mathbb{R}^m \to \mathbb{R}^m$ be a degree-d polynomial given by $p(x) = x + q(x)$. Let $\epsilon, N > 0$ be real numbers and let $f : \mathbb{R}^m \to [-1, 1]$ be any function. Let $M = dmkN$. Then there exists a polynomial R in the coefficients of q of degree less than k, dependent only on d, m, k, ϵ and f so that for all q*

$$|\mathbb{E}[f(p(X))] - R(q)| \leq M^{O(M)}(\epsilon^N + \epsilon^{-1}|q|_2^k).$$

And so that $|R| \leq (M \log(\epsilon^{-1}))^{O(M)}$, where $|R|$ denotes the sum of the absolute values of the coefficients of R when written in terms of the $c_a(q)$.

In order to expand upon the intuition behind Proposition 6, we begin by sketching the proof in the case that $m = d = 1$. In this case we may write $q(x) = ax + b$. It is then the case that

$$\mathbb{E}[f(p(X))] = \mathbb{E}[f((1+a)X + b)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f((1+a)x + b) e^{-x^2/2} dx.$$

The key idea is to evaluate the above by making the change of variables $y = (1+a)x + b$. The above is then equal to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{-((y-b)/(1+a))^2/2} (1+a)^{-1} dy.$$

For small a and b, we may approximate the integrand above by a degree $k-1$ Taylor polynomial in a and b introducing an error on the order of $|q|^k$ in the process. Integrating then yields a polynomial in a and b plus a small error. The proof of Proposition 6 is a straightforward generalization of this idea, though we will see some technical difficulties arising from the more complicated change of variables, and the necessity of keeping better track of errors.

Proof. Note that $|\mathbb{E}[f(p(X))]| \leq 1$, therefore we may assume that $\epsilon \ll M^{-Cdmk}$, for any constant C or there is nothing to prove. Similarly, we may assume that $|q|_2 \ll \epsilon^{1/(3k)} M^{-CdmN}$ or else $M^{CM} \epsilon^{-1} |q|^k \gg |R(q)| + 1$ and there is again nothing to prove.

Note that

$$\mathbb{E}[f(p(X))] = \int_{\mathbb{R}^m} f(p(x))\phi(x)dx$$

where $\phi(x) = (2\pi)^{-m/2}e^{-\frac{\sum_{i=1}^m x_i^2}{2}}$. Up to an error of $O(m\epsilon^N)$, we may ignore the integral outside of the range where $|x|_2 \leq M\log(\epsilon^{-1})$. Note furthermore, that in this range, for ϵ and $|q|$ sufficiently small, we have

$$|p(x)| \leq |x| + |q(x)| \leq M\log(\epsilon^{-1}) + O(M^d|q|_2\log(\epsilon^{-1})^d) \leq 2M\log(\epsilon^{-1}).$$

Claim. *If y is an element of $\mathbb{R}^m$, and if q is a degree-d polynomial with complex coefficients so that $|q|_2 \leq (CM^2(|y|_2+1))^{-d}$ for C a sufficiently large constant C, then there exists a unique $z \in \mathbb{C}^m$ with $|z-y| \leq 1$ so that $p(z) = y$.*

Proof. Consider the map a given by $a(z) = y - q(z)$. We note that for any z that $|a'(z)| = |q'(z)| = \sqrt{\sum_{i=1}^m |q_i(z)|^2}$. By Equation 1, $\sum_{i=1}^m |q_i|_2^2 \leq d|q|_2$. Thus by Lemma 4, $\sum_{i=1}^m |q_i(z)|^2 \leq d|q|_2 O(dm(|z|+1))^d$. Thus as long as $|z| \leq 2(|y|+1)$, we have that $|a'(z)| \leq 1/2$. Thus, a is a contraction from the ball of radius $2(|y|+1)$ to itself with constant $1/2$, and thus by Lemma 5 has a fixed point, z so that $a(z) = z$. But then $z = y - q(z)$, so $y = z + q(z) = p(z)$. This completes the proof. □

By this claim, if $y \in \mathbb{R}^m$ with $|y|_2 \leq 2M\log(\epsilon^{-1})$, we claim that there is a unique x with $|x|_2 \leq 3M\log(\epsilon^{-1})$ so that $p(x) = y$. We may now write our expectation as

$$\mathbb{E}[f(p(X))] = \int_{\substack{|x|_2 \leq 3M\log(\epsilon^{-1}) \\ |p(x)|_2 \leq 2M\log(\epsilon^{-1})}} f(p(x))\phi(x)dx + O(m\epsilon^N).$$

Our plan is now to compute this integral by making the change of variables $y = p(x)$. We know from the above that in the domain of interest there is a function p^{-1}, which by the Inverse Function Theorem is necessarily smooth. Thus,

$$\mathbb{E}[f(p(X))] = \int_{|y|_2 \leq 2M\log(\epsilon^{-1})} f(y)\left(\frac{\phi(p^{-1}(y))}{|\mathrm{Jac}(p(x))|_{x=p^{-1}(y)}}\right)dy + O(m\epsilon^N).$$

The fundamental idea of our proof will be to approximate $\left(\frac{\phi(p^{-1}(y))}{|\mathrm{Jac}(p(x))|_{x=p^{-1}(y)}}\right)$ by a polynomial in q with coefficients depending on y. Integrating the above formula for $\mathbb{E}[f(p(X))]$, will then yield our result. We note by the Implicit Function Theorem that for $|q|_2 \leq (CM^3(|y|_2+1))^{-d}$ that $p^{-1}(y)$ is a complex analytic function of q. In this range, $|p^{-1}(y) - y| \leq 1$, and the coefficients of the Jacobian of q at $p^{-1}(y)$ are all at most $d|q|_2 O(dm(|y|+1))^d < 1/(2m^2)$. Thus the eigenvalues of $\mathrm{Jac}(q)$ are all at most $1/(2m)$. The Jacobian of p (which is the Jacobian of q plus the identity) is the product of 1 plus these eigenvalues, and is therefore $\Theta(1)$. It is also easy to see that $\phi(p^{-1}(y)) = O(1)$ since the imaginary part of $p^{-1}(y)$ has size $O(1)$.

Thus, in the region where $|q|_2 \leq (CM^3(|y|_2+1))^{-d}$, we have that $I_y(q) := \frac{\phi(p^{-1}(y))}{|\mathrm{Jac}(p(x))|_{x=p^{-1}(y)}}$ is a complex analytic function of q with size $O(1)$. Letting $R_y(q)$ be the Taylor polynomial of $I_y(q)$ consisting of all terms of degree less than k, we have by standard results in complex analysis, that as long as $|q|_2 \leq (CM^3(|y|_2+1))^{-d}/2$ that

$$I_y(q) = R_y(q) + O(|q|_2^k(CM^3(|y|_2+1))^{dk}) = R_y(q) + O(M)^{O(M)}\epsilon^{-1/2}|q|_2^k.$$

Furthermore the coefficients of $R_y(q)$ are at most $k^k O(M^3(|y|_2+1))^{dk}$. Therefore the sum of the absolute values of the coefficients is at most $(M\log(\epsilon^{-1}))^{O(M)}$.

Thus we have that

$$\mathbb{E}[f(p(X))] = \int_{|y|\leq 2\log(\epsilon^{-1})} f(y)(R_y(q) + O(M)^{O(M)}\epsilon^{-1/2}|q|_2^k)dy + O(m\epsilon^N). \tag{2}$$

Letting,

$$R(q) := \int_{|y|\leq 2\log(\epsilon^{-1})} f(y)R_y(q)dy,$$

we have by Equation (2) that (noting that the domain of integration has volume at most $(4\log(\epsilon^{-1}))^m$)

$$\mathbb{E}[f(p(X))] = R(q) + O(M)^{O(M)}(\epsilon^{-1}|q|^k + \epsilon^N).$$

It is easy to see that $|R|$ is bounded appropriately, and thus this completes our proof.

□

We can use Proposition 6 to analyze a simple form of our generator.

Proposition 7. *Let p be a degree-d polynomial that can be written in the form $p(x) = h(q_1(x), \ldots, q_m(x))$ for some function h and some polynomials q_i of degree at most d. Let $f(x) = sgn(p(x))$ be the corresponding polynomial threshold function. Suppose that for each i that $q_i(x) = x_i + r_i(x)$ for some polynomial r_i. Let $\epsilon > 0$ be a real number and k be an even integer. Let X be a random Gaussian and Y is kd-moment-matching that is independent of X. Then for $M = dmk$,*

$$\left|\mathbb{E}[f(X)] - \mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right]\right| \leq M^{O(M)}\left(\epsilon^{k-1} + \epsilon^{-1}\sum_{i=1}^{m}|r_i|_2^k\right).$$

Proof. Note that X can be written as the sum $\epsilon X_1 + \sqrt{1-\epsilon^2}X_2$ for X_1 and X_2 independent Gaussians. Hence it suffices to show that $\mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right]$ is determined to within $M^{O(M)}\left(\epsilon^{k-1} + \epsilon^{-1}\sum_{i=1}^{m}|r_i|_2^k\right)$ simply by the low degree moments of Y.

We may rewrite X as (X_0, X_1), where X_0 is the Gaussian given by the first m coordinates of X and X_1 consists of the remaining coordinates. We let $Q(x_0, x_1, y)$ be the vector-valued polynomial given by

$$\begin{aligned} Q(X_0, X_1, Y)_i &= \frac{q_i\left(\epsilon Y + \sqrt{1-\epsilon^2}(X_0, X_1)\right)}{\sqrt{1-\epsilon^2}} \\ &= (X_0)_i + \left(\frac{\epsilon Y_i + r_i\left(\epsilon Y + \sqrt{1-\epsilon^2}(X_0, X_1)\right)}{\sqrt{1-\epsilon^2}}\right). \end{aligned}$$

Upon fixing values for Y and X_1 we let $q^{X_1,Y}(X_0)$ be the vector valued polynomial given by

$$q_i^{X_1,Y}(X_0) := \left(\frac{\epsilon Y_i + r_i\left(\epsilon Y + \sqrt{1-\epsilon^2}(X_0, X_1)\right)}{\sqrt{1-\epsilon^2}}\right).$$

We have that

$$Q(X_0, X_1, Y) = X_0 + q^{X_1,Y}(X_0).$$

We have that

$$\begin{aligned} \mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right] &= \mathbb{E}\left[\operatorname{sgn}\left(h\left(\sqrt{1-\epsilon^2}Q(X_0, X_1, Y)\right)\right)\right] \\ &= \mathbb{E}\left[g(Q(X_0, X_1, Y))\right] \\ &= \mathbb{E}_{X_1,Y}[\mathbb{E}_{X_0}[g(Q(X_0, X_1, Y))]] \\ &= \mathbb{E}_{X_1,Y}[R(q^{X_1,Y}) + O(M)^{O(M)}(\epsilon^{-1}|q^{X_1,Y}|^k + \epsilon^k)]. \end{aligned}$$

Where g above is given by $g(x) = \operatorname{sgn}(h(\sqrt{1-\epsilon^2}x))$, and R is the appropriate polynomial given by Proposition 6. Since the expectation of $R(q^{X_1,Y})$ is determined by the moments of Y up to degree kd, this expectation is determined up to an error of

$$O(M)^{O(M)}\left(\epsilon^k + \epsilon^{-1}\mathbb{E}[|q^{X_1,Y}|^k]\right).$$

We note that $|q^{X_1,Y}|_2 \leq (|q^{X_1,Y}|_k)$. Therefore the error above is

$$\begin{aligned}
&O(M)^{O(M)}\left(\epsilon^k + \epsilon^{-1}\mathbb{E}[|q^{X_1,Y}(X_0)|^k]\right)\\
=&O(M)^{O(M)}\left(\epsilon^k + \epsilon^{-1}\sum_{i=1}^{m}\mathbb{E}[|\epsilon Y_i + r_i\left(\epsilon Y + \sqrt{1-\epsilon^2}(X_0,X_1)\right)|^k]\right)\\
=&O(M)^{O(M)}\left(\epsilon^k + \epsilon^{-1}\left(\epsilon^k + \sum_{i=1}^{m}|r_i|_2^k\right)\right)\\
=&O(M)^{O(M)}\left(\epsilon^{k-1} + \epsilon^{-1}\sum_{i=1}^{m}|r_i|_2^k\right).
\end{aligned}$$

Where the second to last line above is by Lemma 2 and the fact that Y is kd-moment-matching. □

4 Non-Singular Sets

Our basic plan will be to use Proposition 7 to show that the generator $\epsilon Y + \sqrt{1-\epsilon^2}X$ fools all polynomial threshold functions. The idea will be to let $\sqrt{1-\epsilon^2}X = \sqrt{\epsilon}X_1 + \sqrt{1-\epsilon-\epsilon^2}X_2$ for X_1 and X_2 independent Gaussians. Upon fixing a random value for X_2, it is not hard to show that the resulting polynomial of $\epsilon Y + \sqrt{\epsilon}X_1$ will likely have its quadratic terms of size $\tilde{O}(\epsilon)$. Were it the case that the linear term of this polynomial were $\Theta(\sqrt{\epsilon})$, (as seems likely) we could apply Proposition 7 almost immediately. Unfortunately, if this polynomial has essentially no linear terms, this technique may fail. The possibility of this failure is closely related to our original polynomial having small derivatives near X_2. We will want to consider polynomials for which this does not happen with non-negligible probability.

Definition. *Given a sequence of polynomials $(q_1,\ldots,q_m)$, we say that they form an* (ϵ,c,N)-non-singular set *if*

$$Pr_X\left(\left|\bigwedge_j \partial q_j(X)\right|_2 < \epsilon^c\right) < \epsilon^N.$$

Recall that the $\bigwedge$ above denotes the wedge product and that $|\bigwedge v_i|_2$ is the product of the singular values of the matrix $[v_1\cdots v_n]$. We recall the definition from [11] that for a degree-d polynomial $p:\mathbb{R}^n\to\mathbb{R}$, we say that a set of polynomials $(h,q_1,\ldots,q_m)$ is a decomposition of p of size m *if $q_i:\mathbb{R}^n\to\mathbb{R}$, and $h:\mathbb{R}^m\to\mathbb{R}$ are polynomials so that*

- $p(x) = h(q_1(x),\ldots,q_m(x))$
- *For every monomial $\prod x_i^{a_i}$ appearing in h, we have that $\sum a_1\deg(q_i)\leq d$*

Furthermore, we say that a polynomial p has an (ϵ,c,N)-non-singular decomposition of size m *if p has a decomposition $(h,q_1,\ldots,q_m)$ with $|q_i|_2\leq 1$ for all i and so that $(q_1,\ldots,q_m)$ is an (ϵ,c,N)-non-singular set.*

The key fact about these decompositions that we will need is the following structure theorem.

Theorem 8. *Let p be a degree-d polynomial, and let $\epsilon,c,N>0$ with $1/2>\epsilon$. Then there exists a degree-d polynomial p_0 with $|p-p_0|_2 = O_{c,d,N}(\epsilon^N)|p|_2$ so that p_0 has an (ϵ,c,N)-non-singular decomposition of size $O_{c,d,N}(1)$. Furthermore for $d=1$, the size may be taken to be 1 with no error, and for $d=2$, the size may be taken to be $O(N^2/c^2)$ with error at most $\epsilon^N|p|_2$ so long as $N>c$ and $\epsilon^c N/c$ is less than some sufficiently small constant.*

Proof. This arbitrary degree case follows from the proof of the Diffuse Decomposition Theorem of [11], essentially by replacing "(ϵ,ϵ^{-c})-diffuse" by "(ϵ,c,N)-non-singular" wherever it occurs and modifying some exponents slightly. In particular, one can begin with the trivial decomposition of p as $Id\circ p$ and refine until one has an (ϵ,c,N)-non-singular decomposition. If at some stage p approximately decomposes into polynomials $q_1,\ldots,q_m$, which are not (ϵ,c,N)-non-singular, we claim that we can decompose further. In

particular, it must be the case that a random linear combination of the q_i's evaluated at a random point, has a decent probability of having derivative less than ϵ^c. Then by [11] Proposition 10, this linear combination of the q_i must decompose in terms of lower degree polynomials plus a small error. Replacing one of the q's in our decomposition by a linear combination of the others plus a small error plus a function of lower degree polynomials yields a more refined decomposition of p. An ordinal monovariant can then be used as in [11] to show that this process will eventually terminate with an (ϵ, c, N)-non-singular decomposition.

For $d = 1$, we note that $p(X)$ can be written as $a+bL(X)$ for L a linear function with mean 0 and variance 1. $|\partial L(X)| = \Omega(1)$ for all X, thus letting $q_1(X) = L(X)$ and $h(q_1) = a+bq_1$, we have an (ϵ, c, N)-non-singular decomposition.

For $d = 2$, we note that any degree-2 polynomial $p(X)$ can be written as $Q(X) + L(X) + C$ with $Q(X)$ homogenous degree-2, $L(X)$ linear and C a constant. Diagonalizing the quadratic form, we may, after an orthogonal change of variables, write $Q(X)$ as $\sum c_i X_i^2$ for some constants c_i. Rewriting $p(X)$ further, we find that after an appropriate change of variables we may write $p(X)$ as $\sum c_i p_i(X_i)+C$ where p_i is a degree-2 polynomial in one variable with mean 0 and variance 1.

We may replace p by $p - C = \sum c_i p_i(X_i)$, and renormalize so that $|p|_2^2 = \sum c_i^2 = 1$.

We note that $|\partial p(X)|_2^2 = \sum c_i^2 (p_i'(X_i))^2$. We claim that (p) is already an (ϵ, c, N)-non-singular set if either:

1. There are more than $3N/c$ values of i with $|c_i| > \epsilon^{2c/3}$.
2. All of the $|c_i|$ are at most $\epsilon^{c/3}$.

In the first of these cases, for each such i there is a probability of at most $\epsilon^{c/3}$ that $|c_i p_i'(X_i)| < \epsilon^c$. Since all of these must hold for $|\partial p(X)|_2 < \epsilon^c$, and since these events are independent, the probability that $|\partial p(X)|_2 < \epsilon^c$ is less than ϵ^N.

In the second of these cases, we note that the mean value of $|\partial p(X)|_2^2$ is

$$\sum c_i^2 |p_i'|_2^2 = \Omega\left(\sum c_i^2\right) = \Omega(1).$$

On the other hand, the variance of $|\partial p(X)|_2^2$ is

$$\sum c_i^4 \mathrm{Var}((p_i')^2) = O\left(\sum c_i^4\right) = O(\epsilon^{2c/3}).$$

Letting $q(x) = |\partial p(X)|_2^2$ be a degree-2 polynomial with mean μ, we have $|q - \mu|_2 = O(\epsilon^{c/3})$. Thus, by the hypercontractive inequality and the fact that $\epsilon^c N/c$ is sufficiently small, we have that $|q - \mu|_{6N/c} \leq \epsilon^{c/6}|\mu/2|^{c/(6N)}$. Thus the probability that $|q(X) - \mu| < |\mu|/2$ is at most ϵ^N. Thus with probability at least $1 - \epsilon^N$, we have that $|\partial p(X)|_2 > \sqrt{|\mu|/2} > \epsilon^c$.

Assuming that $N\epsilon^c/c$ is sufficiently small, we prove by induction on $\lfloor 3M/c \rfloor$ that p is within $\epsilon^M |p|_2$ of some p_0 with an (ϵ, c, N)-non-singular decomposition of size $9N(M + c/3)/c^2$. If $M < 0$, the result holds trivially since we may use $p_0 = 0$. Otherwise let $p = \sum c_i p_i(X_i)$ as above. Assume furthermore that $|c_1| \geq |c_2| \geq \ldots$. Let k be the largest i so that $|c_i| > \epsilon^{2c/3}$. If $i > 3N/c$, we are done, since (p) is already (ϵ, c, N)-non-singular. Otherwise, let $q_i(X) = X_i$ for $i = 1, \ldots, k$. Then $p(X) = \sum_{i=1}^k c_i p_i(q_i(X)) + ap'(X)$, where $a^2 = \sum_{i>k} c_i^2$ and $p'(X) = \sum_{i>k}(c_i/a)p_i(X_i)$. We claim that if $|a| > \epsilon^{c/3}$ that $(q_1, \ldots, q_k, p')$ is (ϵ, c, N)-non-singular. This is because under this assumption $|c_i/a| \leq \epsilon^{c/3}$ for all $i > k$. This implies by the above that p' is (ϵ, c, N)-non-singular. Since $\partial q_i(X)$ are always unit norm, orthogonal to each other and orthogonal to $\partial p'(X)$,

$$\left|\bigwedge \partial q_i(X) \wedge \partial p'(X)\right|_2 = |\partial p'(X)|_2.$$

Thus, we have an (ϵ, c, N)-non-singular decomposition of p. Otherwise, $|a| < \epsilon^{c/3}$. By the inductive hypothesis, we can find a p_0 within $\epsilon^{M-c/3}$ of p' so that p_0 has an (ϵ, c, N)-non-singular decomposition of size at most $9NM/c^2$. Furthermore, we may assume that the q's in this decomposition depend only on X_i for $i > k$. It is then clear that $\sum_{i=1}^k c_i p(q_i(X)) + ap_0(X)$ has an (ϵ, c, N)-non-singular decomposition of size at most $9NM/c^2 + 3N/c = 9N(M + c/3)/c^2$. This completes our inductive step and proves the Theorem. □

Definition. *For positive integers d and k we define $s(d,k)$ to be an integer m so that for any ϵ and for any degree-d polynomial p, there exists a degree-d polynomial p_0 with $|p-p_0|_2 \leq m^m \epsilon^{3dk}|p|_2$ and so that p_0 has an $(\epsilon, 1/10, k)$-non-singular decomposition of size at most m.*

By Theorem 8, $s(d,k)$ is finite for all d,k, and $s(1,k)=1$, $s(2,k)=O(k^2)$.

5 The PRG

In this Section, we will prove a sequence of increasingly more powerful results for PRGs. We begin by showing that if our polynomial has a non-singular decomposition that $\epsilon Y + \sqrt{1-\epsilon^2}X$ is an appropriate generator.

Proposition 9. *Let d,k be integers and $\epsilon > 0$. Let p be a degree-d polynomial with an $(\epsilon, 1/10, k)$-non-singular decomposition of size m. Let f be the corresponding polynomial threshold function. Let X be a Gaussian, and Y is $10kd$-moment-matching independent of X. Then letting $M = dmk$,*

$$\left|\mathbb{E}[f(X)] - \mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right]\right| = O(M)^{O(M)}(\epsilon^k).$$

Proof. First we assume that $\epsilon \ll M^{-Cdm}$ for a sufficiently large constant C, for otherwise there is nothing to prove.

It suffices to show that the expectation of $f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)$ is determined to within $O(M)^{O(M)}(\epsilon^k)$ by the low order moments of Y.

Let p have the $(\epsilon, 1/10, k)$-non-singular decomposition $(h, q_1, \ldots, q_m)$. Write $\sqrt{1-\epsilon^2}X$ as $\sqrt{\epsilon}X_1 + \sqrt{1-\epsilon-\epsilon^2}X_2$ for X_1 and X_2 independent Gaussians. Let $\epsilon X_0 + \sqrt{\epsilon}X_1 = \sqrt{\epsilon+\epsilon^2}Z$ for X_0 an independent Gaussian, and $W = \epsilon X_0 + \sqrt{1-\epsilon^2}X$. Consider each of the q_i as functions of Z and X_2. Thinking of X_2 as fixed let $q_i^{X_2}(Z) = q_i(X_2, Z) = q_i(\sqrt{1-\epsilon-\epsilon^2}X_2 + \sqrt{\epsilon+\epsilon^2}Z)$. Notice that

$$\begin{aligned}
\mathbb{E}_{X^2}\left[\left|\left(q_i^{X_2}\right)^{[\geq 2]}\right|_2^2\right] &\leq \mathbb{E}_{X_2,Z,X_3,X_4}[|\partial^Z_{X_3}\partial^Z_{X_4} q_i(X_2,Z)|_2^2] \\
&= (\epsilon+\epsilon^2)^2 \mathbb{E}[|\partial^W_{X_3}\partial^W_{X_4} q_i(W)|_2^2] \\
&= O(d^2\epsilon^2|q_i|_2^2) \\
&= O(d^2\epsilon^2).
\end{aligned}$$

Where $\partial^Z_{X_i}$ above denotes the directional derivative with respect to Z in the direction of X_i, and the first line is by Equation (1). Thus, since $\left|\left(q_i^{X_2}\right)^{[\geq 2]}\right|_2^2$ is given by a polynomial in X_2, we have by Corollary 3 that with probability $1 - O(M)^{O(M)}(\epsilon^k)$ that $\left|\left(q_i^{X_2}\right)^{[\geq 2]}\right|_2 \leq \epsilon \log(\epsilon^{-1})^d$ for all i. Similarly, we may show that with this same probability that $\left|\left(q_i^{X_2}\right)^{[1]}\right|_2 \leq \sqrt{\epsilon}\log(\epsilon^{-1})^d$ for all i. For X_2 fixed, let $L_i := \left(q_i^{X_2}\right)^{[1]}$.

Note that with high probability

$$\partial^Z q_i(X_2, Z) = \partial^Z L_i(Z) + O(\epsilon \log(\epsilon^{-1})^d).$$

on the other hand, we have that

$$\partial^Z q_i(Z, X_2) = \sqrt{\epsilon+\epsilon^2}\partial^W q_i(W).$$

By non-singularity this means that with probability $1 - O(M)^{O(M)}(\epsilon^k)$ over X_2 we have

$$\left|\bigwedge_i \left(\partial L_i(Z) + O(\epsilon \log(\epsilon^{-1})^d)\right)\right|_2^2 > \epsilon^{2m+1/5}.$$

On the other hand, the left hand side of the above is

$$\left|\bigwedge_i (\partial L_i(Z))\right|_2^2 + O(1)^m(\epsilon^{2m+1/2}\log(\epsilon^{-1})^{2dm}).$$

Thus for ϵ sufficiently small, we have with probability at least $1 - O(M)^{O(M)}(\epsilon^k)$ over the choice of X_2 that

$$\left|\bigwedge_i (\partial L_i(Z))\right|_2^2 > \frac{\epsilon^{2m+1/5}}{2}.$$

If this is the case, then the product of the singular values of the matrix with rows given by the gradients of the L_i is at least $\epsilon^{m/2+1/10}$. Since none of the singular values can be larger than $O_m(\epsilon^{1/2}\log(\epsilon^{-1})^d)$, this implies that all of the singular values of this matrix are at least $\epsilon^{1/4}$. Thus if we replace the q_i by appropriate linear combinations of themselves (with coefficients at most $\epsilon^{-3/4}$) we can ensure that the $\partial L_i(Z)$ are orthonormal. By making an appropriate change of variables for Z, we may assume that $L_i(Z) = Z_i$. Removing the degree-0 part of $q_i^{X_2}$, we may assume that $q_i^{X_2}(Z) = Z_i + r_i(Z)$ with $|r_i|_2 = O(\epsilon^{1/4}\log(\epsilon^{-1})^d)$.

To summarize, with probability at least $1 - O(M)^{O(M)}(\epsilon^k)$ over the choice of X_2, there is an orthogonal change of variables for Z, and a sequence of polynomials q_i', r_i with $q_i'(Z) = Z_i + r_i(Z)$ and $|r_i(Z)|_2 = O(\epsilon^{1/4}\log(\epsilon^{-1})^d)$ so that $p(Z, X_2)$ has a decomposition into the q_i'. Applying Proposition 7, we find that with probability $1 - O(M)^{O(M)}(\epsilon^k)$ over X_2 we have:

$$\begin{aligned}&\left|\mathbb{E}_Z\left[f\left(\sqrt{\epsilon+\epsilon^2}Z + \sqrt{1-\epsilon-\epsilon^2}X_2\right)\right] - \mathbb{E}_{Y,X_1}\left[f\left(\epsilon Y + \sqrt{\epsilon}X_1 + \sqrt{1-\epsilon-\epsilon^2}X_2\right)\right]\right| \\ &\quad = O(M)^{O(M)}(\epsilon^{5k-1} + (\epsilon^{1/4}\log(\epsilon^{-1})^d)^{10k}) = O(M)^{O(M)}(\epsilon^k).\end{aligned}$$

Taking an expectation over X_2 completes our proof. □

Next we use Theorem 8 to extend Proposition 9 to arbitrary polynomial threshold functions.

Proposition 10. *Let f be a degree-d polynomial threshold function. Let $\epsilon > 0$ and k be an integer. Let X be a random Gaussian and Y is $10kd$-moment-matching and independent of X. Letting $M = dks(d,k)$,*

$$\left|\mathbb{E}[f(X)] - \mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right]\right| = O(M)^{O(M)}(\epsilon^k).$$

Proof. Let $f = \mathrm{sgn}(p(x))$ for some degree-d polynomial p with $|p|_2 = 1$. We know that there exists a degree-d polynomial p_0 so that $|p - p_0|_2 = s^s(\epsilon^{2kd+k})$ so that p_0 has an $(\epsilon, 1/10, k)$-non-singular decomposition of size $m \leq s(d,k)$. Since $\epsilon Y + \sqrt{1-\epsilon^2}X$ is $2d$-moment-matching, we have by the Markov bound that with probability $1 - O(M)^{O(M)}(\epsilon^k)$ that

$$\left|p\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right) - p_0\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right| \leq \epsilon^{kd}.$$

Note that the polynomials $p_0 \pm \epsilon^{kd}$ also have $(\epsilon, 1/10, k)$-non-singular decompositions of size m. Therefore, we have by the above, Proposition 9 and Lemma 1 that

$$\begin{aligned}\mathbb{E}\left[f\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right] &= \mathbb{E}\left[\mathrm{sgn}\left(p\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right)\right)\right] \\ &\leq \mathbb{E}\left[\mathrm{sgn}\left(p_0\left(\epsilon Y + \sqrt{1-\epsilon^2}X\right) + \epsilon^{kd}\right)\right] + O(M)^{O(M)}(\epsilon^k) \\ &= \mathbb{E}[\mathrm{sgn}(p_0(X) + \epsilon^{kd})] + O(M)^{O(M)}(\epsilon^k) \\ &= \mathbb{E}[\mathrm{sgn}(p_0(X) - \epsilon^{kd})] + O(M)^{O(M)}(\epsilon^k) \\ &\leq \mathbb{E}[\mathrm{sgn}(p(X))] + O(M)^{O(M)}(\epsilon^k) \\ &= \mathbb{E}[f(X)] + O(M)^{O(M)}(\epsilon^k).\end{aligned}$$

And the other direction of the inequality follows analogously. □

Iterating applying Proposition 10 yields the following:

Proposition 11. *Let f be a degree-d polynomial threshold function and $\epsilon > 0$. Let k and ℓ be integers. For $1 \le i \le \ell$ let Y_i be $10kd$-moment-matching and X a Gaussian so that X and the Y_i are independent. Letting $M = dks(d,k)$,*

$$\left|\mathbb{E}[f(X)] - \mathbb{E}\left[f\left(\sum_{i=1}^{\ell} \epsilon\left(\sqrt{1-\epsilon^2}\right)^{i-1} Y_i + \left(\sqrt{1-\epsilon^2}\right)^{\ell} X\right)\right]\right| = O(M)^{O(M)}(\ell\epsilon^k).$$

Proof. The proof is by induction on ℓ and noting that by fixing the values of $Y_1, \ldots, Y_{\ell-1}$ Proposition 10 implies that

$$\left|\mathbb{E}\left[f\left(\sum_{i=1}^{\ell-1} \epsilon\left(\sqrt{1-\epsilon^2}\right)^{i-1} Y_i + \left(\sqrt{1-\epsilon^2}\right)^{\ell-1} X\right)\right] - \mathbb{E}\left[f\left(\sum_{i=1}^{\ell} \epsilon\left(\sqrt{1-\epsilon^2}\right)^{i-1} Y_i + \left(\sqrt{1-\epsilon^2}\right)^{\ell} X\right)\right]\right| = O(M)^{O(M)}(\epsilon^k).$$

□

It is not hard to get rid of the X in the above generator

Proposition 12. *Let f be a degree-d polynomial threshold function and $\epsilon > 0$. Let k and ℓ be integers. For $1 \le i \le \ell$ let Y_i be independent $10kd$-moment-matching random variables and X a Gaussian. Letting $M = dks(d,k)$,*

$$\left|\mathbb{E}[f(X)] - \mathbb{E}\left[f\left(\frac{\sum_{i=1}^{\ell}\left(\sqrt{1-\epsilon^2}\right)^{i-1} Y_i}{\sqrt{\sum_{i=1}^{\ell}(1-\epsilon^2)^i}}\right)\right]\right| = O(M)^{O(M)}\left(\ell\epsilon^k + (1-\epsilon^2)^{\frac{\ell}{2d+1}}\right).$$

Proof. Let $f(x) = \mathrm{sgn}(p(x))$ for p a degree-d polynomial with $|p|_2 = 1$.

Let

$$Y := \frac{\sum_{i=1}^{\ell}\left(\sqrt{1-\epsilon^2}\right)^{i-1} Y_i}{\sqrt{\sum_{i=1}^{\ell}(1-\epsilon^2)^i}}.$$

Assume that X and Y are independent and let

$$Z := \sqrt{1-(1-\epsilon^2)^{\ell}}Y + \left(\sqrt{1-\epsilon^2}\right)^{\ell} X.$$

It is not hard to show that since Y is $2d$-moment-matching that

$$\mathbb{E}[|p(Y) - p(Z)|^2] \le O(d^2(1-\epsilon^2)^{\ell}).$$

Thus by the Markov inequality, with probability at least $1 - \left(d^2(1-\epsilon^2)^{\frac{\ell}{2d+1}}\right)$, we have that

$$|p(Y) - p(Z)| \le (1-\epsilon^2)^{\frac{d\ell}{2d+1}}.$$

Therefore, we have that

$$\begin{aligned}
\mathbb{E}[f(Y)] &= \mathbb{E}[\mathrm{sgn}(p(Y))] \\
&\le \mathbb{E}\left[\mathrm{sgn}\left(p(Z) + (1-\epsilon^2)^{\frac{d\ell}{2d+1}}\right)\right] + O\left(d^2(1-\epsilon^2)^{\frac{\ell}{2d+1}}\right) \\
&\le \mathbb{E}\left[\mathrm{sgn}\left(p(X) + (1-\epsilon^2)^{\frac{d\ell}{2d+1}}\right)\right] + O(M)^{O(M)}\left(\ell\epsilon^k + (1-\epsilon^2)^{\frac{\ell}{2d+1}}\right) \\
&\le \mathbb{E}\left[\mathrm{sgn}(p(X))\right] + O(M)^{O(M)}\left(\ell\epsilon^k + (1-\epsilon^2)^{\frac{\ell}{2d+1}}\right) \\
&= \mathbb{E}\left[f(X)\right] + O(M)^{O(M)}\left(\ell\epsilon^k + (1-\epsilon^2)^{\frac{\ell}{2d+1}}\right).
\end{aligned}$$

The other direction of the inequality holds analogously. □

We can finally prove our main result:

Theorem 13. *For d, k positive integers and $\epsilon > 0$, there exists an explicit pseudorandom generator, Y of seed length $O(d^2k^2\log(n)\epsilon^{-1})$ so that for X an n-dimensional Gaussian, and f any degree-d polynomial threshold function in n variables, and $M = dks(d, 3k)$*

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| = O(M)^{O(M)}(\epsilon^k).$$

In particular, such a generator is given by letting

$$Y = \frac{\sum_{i=1}^{[\epsilon^{-2/3}(2d+1)k]}(1-\epsilon^{2/3})^{i/2}Y_i}{\sqrt{\sum_{i=1}^{[\epsilon^{-1}dk]}(1-\epsilon^{2/3})^i}}$$

Where the Y_i are independent of each other and each within $O(\epsilon^2/(dk))$ statistical distance of a $10d(3k+3)$-moment-matching random variable.

Proof. Let $\delta = \epsilon^{1/3}$. Let $\ell = \delta^{-2}\log(\epsilon^{-k(2d+1)})$. Let $Z_1, \ldots, Z_\ell$ be independent and $10d(3k+3)$-moment-matching. Let

$$Z := \frac{\sum_{i=1}^{\ell}\left(\sqrt{1-\delta^2}\right)^{i-1}Z_i}{\sqrt{\sum_{i=1}^{\ell}(1-\delta^2)^i}}.$$

By Proposition 12 we have that

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Z)]| = O(M)^{O(M)}\left(\ell\delta^{3k+3} + (1-\delta^2)^{\frac{\ell}{2d+1}}\right) = O(M)^{O(M)}(\epsilon^k).$$

By Gauss-Jacobi quadrature, there is a 1-dimensional $10d(3k+3)$-moment-matching random variable supported on a set of size $10d(3k+3)$. Therefore there is an explicit random variable with seed $O(dk\log(n/\epsilon))$ which differs from this by at most $\epsilon^k n^{-1}\ell^{-1}$ in statistical distance. A $10d(3k+3)$-wise-independent family of n of these variables, has seed length $O(d^2k^2\log(n/\epsilon))$ and is within a statistical distance of $O(\epsilon^k\ell^{-1})$ of some $10d(3k+3)$-moment-matching variable. If we take ℓ independent copies of such random variables, calling them Y_i and let

$$Y := \frac{\sum_{i=1}^{\ell}\left(\sqrt{1-\delta^2}\right)^{i-1}Y_i}{\sqrt{\sum_{i=1}^{\ell}(1-\delta^2)^i}}.$$

then Y can be generated from seed length

$$O(d^2k^2\log(n/\epsilon)\ell) = O(d^2k^2\log(n)\epsilon^{-1}),$$

and has statistical distance at most $O(\epsilon^k)$ from Z. Thus

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(Z)] + O(\epsilon^k) = \mathbb{E}[f(X)] + O(M)^{O(M)}(\epsilon^k).$$

□

Changing the value of ϵ appropriately, we have that

Corollary 14. *Let d be a positive integer and $c, \epsilon > 0$. Letting $k = \lceil c^{-1}\rceil$ and $M = dks(d, k)$, there exists an explicit pseudorandom generator Y with seed length $O(M)^{O(M)/k}(\log(n)\epsilon^{-c})$ so that for any degree-d polynomial threshold function in n variables, and X an n-dimensional Gaussian,*

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \epsilon.$$

For $d = 1$ we have that $M = O(k)$. Thus letting $k = \log(1/\epsilon)$, we have:

Corollary 15. *For $\epsilon > 0$, there exists an explicit pseudorandom generator Y with seed length*

$$\log(n)\log(1/\epsilon)^{O(1)}$$

so that for any degree-1 polynomial threshold function in n variables, and X an n-dimensional Gaussian,

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \epsilon.$$

For $d = 2$ we have that $M = O(k^3)$. Thus letting $k = \Theta\left(\log(\epsilon^{-1})/\log\log(\epsilon^{-1})\right)^{1/3}$, we have:

Corollary 16. *For $\epsilon > 0$, there exists an explicit pseudorandom generator Y with seed length*

$$\log(n)\exp\left(O\left(\log(\epsilon^{-1})^{2/3}\log\log(\epsilon^{-1})^{1/3}\right)\right)$$

so that for any degree-2 polynomial threshold function in n variables, and X an n-dimensional Gaussian,

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \epsilon.$$

6 PRG for LTFs

The result in Corollary 15 can actually be significantly improved upon. This comes in three steps. Firstly, a more careful analysis is needed to clarify the $O(1)$ in the exponent of the $\log(1/\epsilon)$ term. Secondly, the generator given here can be further compressed using ideas similar to the generator in [13]. Finally, standard dimension reduction techniques can be used to effectively reduce the value of n, providing a further improvement.

For the first of these, we will need a more explicit version of Theorem 13 for LTFs.

Proposition 17. *Let k be a positive, even integer, X an n-dimensional Gaussian, and Y and independent k-moment-matching random variable. Let f by a linear threshold function and $1/2 > \epsilon > 0$ a real number. Then*

$$\left|\mathbb{E}[f(x)] - \mathbb{E}[f(\sqrt{1-\epsilon^2}X + \epsilon Y)]\right| = O(\epsilon)^k.$$

Proof. Let $f(x) = \mathrm{sgn}(v \cdot x + \theta)$ for $|v| = 1$. Note that

$$f(\sqrt{1-\epsilon^2}X + \epsilon Y) = \mathrm{sgn}(\sqrt{1-\epsilon^2}v \cdot X + \epsilon v \cdot Y + \theta).$$

Thus, for fixed Y, the expectation of the above over X is

$$\mathrm{erf}\left(\frac{\theta + \epsilon v \cdot Y}{\sqrt{1-\epsilon^2}}\right).$$

Expanding the above as a Taylor series in $v \cdot Y$, we find that it is given by a polynomial in $v \cdot Y$, we find that it is a polynomial in $v \cdot Y$ plus an error of size at most $\frac{O(k)^{\sqrt{k}}}{k!}\left(\frac{\epsilon v \cdot Y}{\sqrt{1-\epsilon^2}}\right)^k$. The expectation of the polynomial is determined by the k-moment-matching of Y, and the expectation of the latter term is $O(\epsilon)^k$. Since the expectation for Y a Gaussian is exactly $\mathbb{E}[f(X)]$, this yields our result. □

Iterating the above we find:

Proposition 18. *Let k be a positive even integer, ℓ a positive integer and $1/2 > \epsilon > 0$. Let X be a random Gaussian. Let Y be given by*

$$Y = \sum_{i=0}^{\ell} \epsilon(1-\epsilon^2)^{i/2} Y_i$$

where Y_i are independent k-moment-matching random variables. Then for any LTF f,

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| = \ell O(\epsilon)^k + O(k\log(1/\epsilon)(1-\epsilon^2)^{\ell/2}).$$

Proof. Let $Y' = Y + (1-\epsilon^2)^{(\ell+1)/2}X$. Let

$$Z_j = \sum_{i=0}^{j} \epsilon(1-\epsilon^2)^{i/2}Y_i + (1-\epsilon^2)^{(1+j)/2}X.$$

Note that $Z_{-1} = X$ and $Z_\ell = Y'$. By Proposition 17,

$$|\mathbb{E}[f(Z_j)] - \mathbb{E}[f(Z_{j+1})]| = O(\epsilon)^k,$$

thus

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y')]| = \ell O(\epsilon)^k.$$

We have yet to consider the difference between $\mathbb{E}[f(Y)]$ and $\mathbb{E}[f(Y')]$. We will show that this is small by showing that $f(Y) = f(Y')$ with high probability. Let $f(x) = \mathrm{sgn}(v \cdot x + \theta)$ with $|v| = 1$. Note that $v \cdot X$ is normally distributed. With probability $1 - O(\epsilon)^k$ we have that $|v \cdot X| \leq k\log(1/\epsilon)$. If this is the case, then $f(Y) = f(Y')$, unless $|p(Y')| < k\log(1/\epsilon)(1-\epsilon^2)^{\ell/2}$. But since Y' fools linear threshold functions to within $\ell O(\epsilon)^k$, this happens with probability

$$\Pr(|p(X)| < k\log(1/\epsilon)(1-\epsilon^2)^{\ell/2}) + \ell O(\epsilon)^k = \ell O(\epsilon)^k + O(k\log(1/\epsilon)(1-\epsilon^2)^{\ell/2}).$$

This completes our proof. □

A modification of this generator yields the following theorem.

Theorem 19. *If Y is given by an appropriate weighted sum of approximate moment-matching random variables, which are seeded by a generator fooling read once branching programs with appropriate parameters, then it ϵ-fools Linear Threshold Functions of Gaussians and has seed length*

$$O(\log(n) + \sqrt{\log(n/\epsilon)}\log(1/\epsilon)).$$

Proof. By Proposition 18, to fool LTFs to within δ (for δ sufficiently small), it suffices to use Y as in Proposition 18 for any $k \leq \log(1/\delta)$, $\epsilon = \delta^{2/k}$, $\ell = \delta^{-5/k}$. In other words, changing around variables names some, in order to fool LTFs to within ϵ, it suffices to use an appropriate weighted sum of $\ell = \epsilon^{-5/k}$ random variables chosen from k-moment-matching families (or even families within ϵ^2 statistical distance of being k-moment-matching). Note that such a family can be generated from a seed of length $O(k\log(n/\epsilon))$. Thus, we can use a generator of the form

$$Y(s) = \sum_{i=1}^{\ell} f_i(s_i)$$

where the f_i are some explicit approximate k-moment-matching generators times weights, and the s_i are seeds of length $O(k\log(n/\epsilon))$. Our basic idea will be to show that it is sufficient to choose the s_i not independently, but instead with a generator that fools an appropriate family of read once branching programs.

Let $f(x) = \mathrm{sgn}(v \cdot x + \theta)$ with $|v| = 1$. Then

$$f(Y(s)) = \mathrm{sgn}(\sum_{i=1}^{\ell} g_i(s_i) + \theta)$$

where $g_i(s_i) = v \cdot f_i(s_i)$. Let $h_i(s_i)$ be obtained from $g_i(s_i)$ by rounding to the nearest multiple of ϵ/ℓ and then truncating if its absolute value is more than $\ell\epsilon^{-2}$. Note that except with probability $O(\epsilon)$ that $h_i(s_i)$ is within ϵ/ℓ of $g_i(s_i)$, and thus $|\sum g_i(s_i) - \sum h_i(s_i)| < \epsilon|$. Since Y fools LTFs to within ϵ, and since $\Pr(|p(X)| < \epsilon) = O(\epsilon)$, this means that

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f(Y)] + O(\epsilon) \\
&= \mathbb{E}[\mathrm{sgn}(\sum_{i=1}^{\ell} g_i(s_i) + \theta)] + O(\epsilon) \\
&= \mathbb{E}[\mathrm{sgn}(\sum_{i=1}^{\ell} h_i(s_i) + \theta)] + O(\epsilon).
\end{aligned}$$

Notice that the partial sums of $\sum_{i=1}^{\ell} h_i(s_i)$ can be stored in $O(\log(\ell/\epsilon))$ bits. Thus, the sign of the sum plus θ can be computed by a read once branching program of width $O(\log(\ell/\epsilon))$ that runs for ℓ rounds on inputs of size $O(k\log(n/\epsilon))$. Let s' be the output of a generator that fools such read once branching programs, and let $Y' = Y(s')$. Note that by results of [6], such generators can be produced from seeds of length

$$O(k\log(n/\epsilon) + \log(\ell)\log(\ell/\epsilon)) = O(k\log(n/\epsilon) + \log^2(1/\epsilon)/k).$$

Note that

$$p(Y') = \sum_{i=1}^{\ell} g_i(s'_i) + \theta.$$

Again, except with probability $O(\epsilon)$, we have that

$$\left| p(Y') - \left(\sum_{i=1}^{\ell} h_i(s'_i) + \theta \right) \right| = O(\epsilon).$$

Thus we have that

$$\begin{aligned}\mathbb{E}[f(Y')] &= \mathbb{E}[\mathrm{sgn}(p(Y'))] \\ &\leq \mathbb{E}\left[\mathrm{sgn}\left(\sum_{i=1}^{\ell} h_i(s'_i) + \theta + O(\epsilon) \right) \right] + O(\epsilon).\end{aligned}$$

Note that by construction s' is enough to fool the functions

$$\mathrm{sgn}\left(\sum_{i=1}^{\ell} h_i(s'_i) + \theta + O(\epsilon) \right).$$

Therefore, we have that

$$\begin{aligned}\mathbb{E}[f(Y')] &\leq \mathbb{E}\left[\mathrm{sgn}\left(\sum_{i=1}^{\ell} h_i(s_i) + \theta + O(\epsilon) \right) \right] + O(\epsilon) \\ &= \mathbb{E}[f(X)] + O(\epsilon).\end{aligned}$$

Similarly,

$$\mathbb{E}[f(Y')] \geq \mathbb{E}[f(X)] + O(\epsilon).$$

Thus Y' provides a pseudorandom generator with error $O(\epsilon)$. Modifying ϵ by a constant, we achieve a generator of error ϵ and seed length $O(k\log(n/\epsilon)+\log^2(1/\epsilon)/k)$. Taking k to be the floor of $1+\log(1/\epsilon)/\sqrt{\log(n/\epsilon)}$ yields our result.

□

If n is larger than $1/\epsilon$, the second term above can be slightly improved via dimension reduction techniques. In particular, we show the following:

Theorem 20. *There exists a pseudorandom generator Y with seed length $O(\log(n)+\log(1/\epsilon)^{3/2})$ that ϵ-fools linear threshold functions with respect to the n-dimensional Gaussian distribution.*

Proof. Suppose that A is a probability distribution over $m \times n$ matrices so that for any $x \in \mathbb{R}^n$,

$$\Pr_A(|Ax| = |x|(1 \pm \epsilon)) \geq 1 - \epsilon.$$

Such families were shown to exist in [8], and many explicit constructions are known. Let Y be a generator that ϵ-fools LTFs in m dimensions. We claim that $A^T Y$ $O(\epsilon)$-fools LTFs in n dimensions.

To show this consider the LTF given by $f(x) = v \cdot x + \theta$ with $|v| = 1$. We know that $\mathbb{E}[f(X)] = \text{erf}(\theta)$. On the other hand, we have that

$$\begin{aligned}
\mathbb{E}[f(Y)] &= \mathbb{E}_A[\mathbb{E}_Y[\text{sgn}(v \cdot A^T Y + \theta)]] \\
&= \mathbb{E}_A[E_Y[\text{sgn}(Av \cdot Y + \theta)]] \\
&= \mathbb{E}_A[\text{erf}(\theta/|Av|) + O(\epsilon)] \\
&= \mathbb{E}_A[\text{erf}(\theta/|Av|)] + O(\epsilon) \\
&= \text{erf}(\theta(1 \pm \epsilon)) + O(\epsilon) \\
&= \text{erf}(\theta) + O(\epsilon).
\end{aligned}$$

Using a generator from [1], we have a family A for $m = O(\epsilon^{-3})$ and seed length $O(\log(n))$. By the above, we have a generator for Y of seed length $O(\log(1/\epsilon)^{3/2})$, thus combining these we have a generator of seed length $O(\log(n) + \log(1/\epsilon)^{3/2})$. □

7 Concluding Remarks

It should be noted that although our construction behaves very well asymptotically as ϵ goes to zero, the constants are potentially somewhat unwieldy. In particular, working through the details of our analysis the constant hidden by the $s(d, k)$ in our seed length would be $A(d + O(1), k)$, where A here is the Ackermann function. The size of this constant is largely due to the size of the non-singular decompositions coming from Theorem 8. Fortunately, the reality is probably much better than these bounds would suggest. In particular, it seems reasonable to conjecture that the such non-singular decompositions can be found of polynomial size. Were this the case, $s(d, k)$ would be only $(d/c)^{O(1)}$. Optimizing the parameter c in our construction would then lead to generators of seed length $\log(n) \exp\left(d^{O(1)} \log(\epsilon^{-1})^a\right)$ for some $a < 1$.

Acknowledgements

This research was done with the support of an NSF postdoctoral fellowship.

References

[1] Noga Alon, Yossi Matias, Mario Szegedy *The space complexity of approximating the frequency moments* Journal of Computer and System Sciences Vol. 58(1) (1999), p. 137-147.

[2] Richard Beigel *The polynomial method in circuit complexity*, Proc. of 8th Annual Structure in Complexity Theory Conference (1993), pp. 82-95.

[3] A. Carbery, J. Wright *Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$* Mathematical Research Letters, Vol. 8(3), pp. 233248, 2001.

[4] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, E. Viola, *Bounded Independence Fools Halfspaces* SIAM Journal on Computing, Vol. 39(8), p. 3441-3462, 2010.

[5] Ilias Diakonikolas, Daniel M. Kane, Jelani Nelson, *Bounded Independence Fools Degree-2 Threshold Functions*, Foundations of Computer Science (FOCS), 2010.

[6] Russell Impagliazzo, Noam Nisan, Avi Wigderson *Pseudorandomness for network algorithms*, STOC (1994), p. 356364.

[7] Svante Janson *Gaussian Hilbert Spaces*, Cambridge University Press, 1997.

[8] W.B. Johnson, J. Lindenstrauss *Extensions of Lipschitz mappings into a Hilbert space* Contemp Math Vol. 26 (1984), p. 189-206.

[9] Daniel M. Kane *A Small PRG for Polynomial Threshold Functions of Gaussians* Symposium on the Foundations Of Computer Science (FOCS), 2011.

[10] Daniel M. Kane *k-Independent Gaussians Fool Polynomial Threshold Functions*, Conference on Computational Complexity (CCC), 2011.

[11] Daniel M. Kane *A Structure Theorem for Poorly Anticoncentrated Gaussian Chaoses and Applications to the Study of Polynomial Threshold Functions*, manuscript http://arxiv.org/abs/1204.0543.

[12] Adam R. Klivans, Rocco A. Servedio *Learning DNF in time* $2^{O(n^{1/3})}$, J. Computer and System Sciences Vol. 68 (2004), p. 303-318.

[13] Raghu Meka, David Zuckerman *Pseudorandom generators for polynomial threshold functions*, Proceedings of the 42nd ACM Symposium on Theory Of Computing (STOC 2010).

[14] Nelson *The free Markov field*, J. Func. Anal. Vol. 12(2), p. 211-227, 1973.

[15] Alexander A. Sherstov *Separating AC0 from depth-2 majority circuits*, SIAM J. Computing Vol. 38 (2009), p. 2113-2129.