Random Restrictions and PRGs for PTFs in Gaussian Space

Zander Kelley
Department of Computer Science, University of Illinois at Urbana-Champaign, IL, USA

Raghu Meka
Department of Computer Science, University of California, Los Angeles, CA, USA

Abstract

A polynomial threshold function (PTF) $f : \mathbb{R}^n \to \mathbb{R}$ is a function of the form $f(x) = \mathsf{sign}(p(x))$ where p is a polynomial of degree at most d. PTFs are a classical and well-studied complexity class with applications across complexity theory, learning theory, approximation theory, quantum complexity and more. We address the question of designing pseudorandom generators (PRGs) for polynomial threshold functions (PTFs) in the gaussian space: design a PRG that takes a seed of few bits of randomness and outputs a n-dimensional vector whose distribution is indistinguishable from a standard multivariate gaussian by a degree d PTF.

Our main result is a PRG that takes a seed of $d^{O(1)} \log(n/\varepsilon) \log(1/\varepsilon)/\varepsilon^2$ random bits with output that cannot be distinguished from an n-dimensional gaussian distribution with advantage better than ε by degree d PTFs. The best previous generator due to O'Donnell, Servedio, and Tan (STOC'20) had a quasi-polynomial dependence (i.e., seed length of $d^{O(\log d)}$) in the degree d. Along the way we prove a few nearly-tight structural properties of *restrictions* of PTFs that may be of independent interest.

Similar results were obtained in [15] (independently and concurrently).[1]

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases polynomial threshold function, pseudorandom generator, multivariate gaussian

Digital Object Identifier 10.4230/LIPIcs.CCC.2022.21

Related Version *Previous Version*: https://arxiv.org/abs/2103.14134

Funding *Zander Kelley*: Supported by NSF grants CCF-1755921 and CCF-1814788.
Raghu Meka: Supported by NSF Career Award 1553605 and NSF AF 2007682.

1 Introduction

Polynomial threshold functions (PTFs) are a classical and well-studied class of functions with several applications in complexity theory, learning theory, theory of approximation, and more. Here we study the question of designing *pseudorandom generators* (PRGs) that fool test functions that are PTFs. We first start with some standard definitions. Let $\mathsf{sign} : \mathbb{R} \to \mathbb{R}$ be defined as $\mathsf{sign}(z) = 1$ if $z \geq 0$ and 0 otherwise.

▶ **Definition 1.** *For an integer $d > 0$, a degree d PTF $f : \mathbb{R}^n \to \{0, 1\}$ is a function of the form $f(x) = \mathsf{sign}(p(x))$, where $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree at most d.*

Our goal is to design a PRG that takes few random bits and outputs a high-dimensional vector whose distribution is indistinguishable from a standard multivariate gaussian by any low-degree PTF. Specifically:

[1] A version of our manuscript with the same proofs and results as in this submission appeared on arxiv on March 25, 2021. There are, however, differences in the technical overview and exposition (among other things, we've incorporated feedback from others within this submission).

▶ **Definition 2.** *A function $G : \{0,1\}^r \to \mathbb{R}^n$ is a pseudorandom generator for degree d PTFs with error ε if for every degree at most d PTF $f : \mathbb{R}^n \to \{0,1\}$,*

$$\left| \mathop{\mathbb{P}}_{y \in_u \{0,1\}^r} (f(G(y)) = 1) - \mathop{\mathbb{P}}_{x \sim N(0,1)^n} (f(x) = 1) \right| \leq \varepsilon.$$

We call r the seed length of the generator and say G ε-fools degree d PTFs with respect to the gaussian distribution [2]*. We say G is explicit if its output can be computed in time polynomial in n.*

Of particular interest is the *boolean case* where the target distribution is not gaussian but the uniform distribution on the hypercube $\{+1,-1\}^n$. The gaussian case is interesting by itself both from a complexity-theoretic view as well as a geometric one. For instance, a PRG as above can be used to get deterministic algorithms for approximating the gaussian volumes of polynomial surfaces. Further, the gaussian case is a necessary stepping-stone to obtaining PRGs in the Boolean case: a PRG for the latter implies a PRG for the gaussian case. Achieving similar parameters as we do for the boolean case would be a significant achievement: we do not even have non-trivial correlation lower bounds for NP against PTFs of degree $\omega(\log n)$ over the hypercube. A PRG would at the very least imply correlation lower bounds against a function in NP, resolving a longstanding bottleneck in circuit complexity.

Over the last several years, the question of designing PRGs for PTFs has received much attention. Non-explicitly (i.e., the generator is not necessarily efficiently computable), by the probabilistic method, it is known that there exists PRGs that ε-fool degree d PTFs with seed-length $O(d \log n + \log(1/\varepsilon))$. Meka and Zuckerman [12] gave the first non-trivial PRG for bounded degree PTFs with a seed length of $d^{O(d)} \log(n)/\varepsilon^2$ for the boolean and gaussian cases. Independent of [12], [2] showed that bounded independence fools degree-2 PTFs leading to seed length $O(\log(n)/\varepsilon^2)$. Since then, there have been several other works that make progress on the gaussian case [7, 6, 8, 10, 11]. The seed length in all of these works had an exponential dependence on the degree d of the PTF. In particular, until recently no non-trivial PRGs (i.e., seed length $o(n)$) were known for PTFs of degree $\omega(\log n)$. In a remarkable recent work, O'Donnell, Servedio, and Tan [14] got around this exponential dependence on the degree d, achieving a seed length of $(d/\varepsilon)^{O(\log d)} \log(n)$. Our work builds on their work (which in turn builds on a framework of [7]).

1.1 Main Results

Our main result is a PRG with seed length $(d/\varepsilon)^{O(1)} \log(n)$ that ε-fools n-variate degree-d PTFs:

▶ **Theorem 3** (PRG for PTFs)**.** *There exist constants c, C such that for all $\varepsilon > 0$ and $d \geq 1$, there exists an explicit PRG that ε-fools n-variate degree d PTFs with respect to the gaussian distribution with seed length $r(n,d,\varepsilon) = Cd^c \log(n/\varepsilon) \log(1/\varepsilon)/\varepsilon^2$.*

As remarked above, this is the first result with polynomial dependence on the degree for fooling PTFs against any distribution and gives the first non-trivial PRGs against PTFs of degree $n^{\Omega(1)}$. Previously, we could only handle degree at most $2^{O(\sqrt{\log n})}$.

Towards proving the above result, we develop several structural results on PTFs in the gaussian space that might be of independent interest. We expand on these later on. Briefly:

[2] Here, and henceforth, $y \in_u S$ denotes a uniformly random element from a multi-set S, and $N(0,1)$ represents the standard univariate gaussian distribution of variance 1.

- We show that the derivatives of a low-degree polynomial p, taken at a random point $x \sim N(0,1)^n$, are likely to have magnitudes $\|\nabla^k p(x)\|$ which grow slowly as k increases.
- We apply this fact to the study of random "gaussian restrictions" of a polynomial p,

$$p_{x,\lambda}(y) := p\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right),$$

  and conclude that for small enough λ, with high probability over $x \sim N(0,1)^n$, $p_{x,\lambda}(Y)$ becomes highly concentrated around its mean value μ when $Y \sim N(0,1)^n$, as quantified by a bound on the higher-moments $\mathbb{E}(p_{x,\lambda}(Y) - \mu)^R$.
- As this concentration result relies only on moment bounds, it extends easily to pseudorandom distributions Y over $\mathbb{R}^n$ which are k-moment-matching with $N(0,1)^n$, when $k \geq R \cdot \deg(p)$.

Note that the magnitudes of the derivatives $\nabla^k p_{x,\lambda}(0)$ (with respect to x) are the same as the magnitudes of the degree-k coefficients of $p_{x,\lambda}(y)$ (as a polynomial in y), up to a scaling factor of roughly $\lambda^{k/2}$. However, to obtain the moment bound, we must translate to the basis of Hermite polynomials and bound the degree-k coefficients with respect to this basis (rather than the standard basis). In contrast with our work, [14] derive coefficient-size bounds for the Hermite basis directly and work with it exclusively. However, there are some significant advantages in having the flexibility to work also within the standard basis which will become relevant later – mainly they are due to the fact that standard basis representations (or equivalently: derivatives) behave nicely under the scaling operator $p(t) \mapsto p(\gamma t)$. The Hermite-basis representation behaves poorly under scaling[3].

For an arbitrary fixed polynomial $p(t)$, a bound on the coefficient-sizes in one basis translates only to a fairly crude bound in the other basis[4]. Therefore, we come to the following rather technical contribution of our work which we would like to highlight: we find that, although it is rather painful to convert between bases while studying an arbitrary fixed polynomial, it is actually quite possible to do so when studying certain *average-case* behaviors of polynomials; for instance, to study the typical behavior of $p(x)$ in the neighborhood around a random point $x \sim N(0,1)^n$, or the typical moments of $p(\sqrt{1-\lambda}x + \sqrt{\lambda}Y)$, it is possible to pass freely between either polynomial basis, and we develop some simple tools for doing so. These tools appear to be new (at least with respect to the body of works on PTFs) and it seems likely that they could be helpful in future works.

Besides these structural results and technical contributions, we also manage to introduce some substantial simplifications to the analysis of the main PRG as compared to [14]. This is in part due to the flexibility we have to measure the *well-behavedness* of a polynomial p in the neighborhood around a point x directly via the derivatives at x, rather than indirectly by taking several Hermite expansions of p and other auxiliary polynomials (cf. *horizontal, diagonal mollifier checks* in [14]). We will expand on this in Section 2 when discussing our analysis, but we briefly summarize a few key points here.

- Following [7] and [14], the construction we analyze has the form $Z := \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Y_i$, where each Y_i is a k-moment-matching gaussian. This can be thought of as the gaussian analogue of the boolean construction from [12], which pseudorandomly partitions the n input bits into L buckets, and then assigns the bits in each bucket using k-wise independence. This

[3] In contrast, the Hermite basis representation behaves nicely under the *noise operator*, $p(t) \mapsto \mathbb{E}_{x \sim N(0,1)^n} p(\sqrt{1-\lambda}x + \sqrt{\lambda}t)$.

[4] This is especially true in the current setting where we must control the *relative* sizes of the magnitudes of coefficients at degree k vs. $k+1$.

construction and its variants are by now the most widely-applied pseudorandom tool for fooling various "geometric" function classes including linear threshold functions and their generalizations (such as PTFs and intersections of halfspaces).

- A tempting first idea for analyzing Z is to apply a hybrid argument – this seems promising in light of the fact that for a low-degree polynomial, we know that $p\left(\sqrt{1-\frac{1}{L}}x+\sqrt{\frac{1}{L}}Y_i\right)$ should be highly-concentrated around its mean for typical x[5]. However, this naive idea fails quantitatively: The probability that we have good behavior at x is in general not smaller than $\sqrt{1/L}$, so we cannot afford a union-bound over L events as required by the standard hybrid argument. Remarkably in [7], Kane shows how to address this obstacle with a clever sandwiching argument which in some sense mimics the hybrid argument but manages to pay for the error caused by "bad points" x only once rather than L times.
- However, one drawback of Kane's analysis is that its implementation is highly elaborate. After the framework was extended by [14] to break the $\log(n)$-degree barrier, the complexity only increased and the details of the argument became only more specialized and technical[6]. Given the wide applicability of the aforementioned pseudorandom construction and its variants, it would be highly desirable to have a lean and more transparent analysis which might better serve as a flexible starting point for future adaptations. We propose that in this work, we do obtain such an analysis.

PTFs simplify under restrictions

As a byproduct of our analysis, we obtain a structural result on PTFs that is similar in spirit to the celebrated *switching lemmas* that show that certain classes of functions simplify significantly under random restrictions. Switching lemmas and random restrictions are a cornerstone in complexity theory, and are one of the main methods we have for proving lower bounds. We prove analogous results with nearly optimal parameters for the important class of PTFs in the continuous space.

In the *boolean case*, i.e., when studying distributions on the hypercube $\{+1,-1\}^n$, a *restriction* is a partial assignment of the form $\rho\in\{+1,-1,*\}^n$ with the understanding that the $*$-variables are left free. Typically, restrictions ρ as above are parametrized by some $\lambda>0$, the fraction of $*$'s.

Here, we study analogues of the above results in the continuous world, where the inputs are coming from the standard gaussian distribution. The first question however is what should the analogue of random restrictions be in the continuous space? As it turns out, adopting the usual interpretation (where some coordinates are fixed and some are free) is not a natural one to study in the continuous space especially for PTFs[7].

The answer comes from the work of [7] (further developed in [14]) who introduced the notion of a *zoom* of a polynomial. To draw a clearer parallel with random restrictions, we term these *gaussian restrictions*:

▶ **Definition 4.** *Given a function $p:\mathbb{R}^n\to\mathbb{R}$ and $x\in\mathbb{R}^n$, and a* restriction *parameter $\lambda\in(0,1)$, let $p_{x,\lambda}:\mathbb{R}^n\to\mathbb{R}$ be*[8] *the function $p_{x,\lambda}(y)=p(\sqrt{1-\lambda}x+\sqrt{\lambda}y)$.*

Intuitively, we can view $p_{x,\lambda}$ as a restriction where $(1-\lambda)$-fraction of the *variance* is already *fixed*. (Note that for independent $x,y\sim N(0,1)^n$, $\sqrt{1-\lambda}x+\sqrt{\lambda}y$ is distributed as $N(0,1)^n$.)

---

[5] For a quantitative version of this statement, see Lemma 13.

[6] Refer to [15], which fills in several details absent in [14], to see the full scope of the argument.

[7] One reason is that the class of PTFs is invariant under linear transformations, so it would be nice to have our notion of restrictions also have some symmetry under linear transformations.

[8] As the value of λ will often be clear, we will often in fact just use p_x for brevity.

We show that PTFs simplify significantly, i.e., become essentially constant, under *gaussian restrictions* for $\lambda \ll 1/d^6$.

▶ **Theorem 5.** *There is a constant $C > 0$ such that the following holds. For any $\delta, \varepsilon > 0$, if*
- *$f : \mathbb{R}^n \to \{0,1\}$ is a PTF of degree d, and*
- *$\lambda \leq C\frac{\delta^2}{d^6 \log(1/\varepsilon)}$,*

then with probability at least $1-\delta$ over $x \sim N(0,1)^n$, the gaussian restriction of the PTF $(f_{x,\lambda})$ is nearly fixed to a constant*: for some $b \in \{0,1\}$ we have*

$$\mathbb{P}_{y \sim N(0,1)^n}[f_{x,\lambda}(y) = b] > 1 - \varepsilon.$$

The work of [14] achieves a similar conclusion but when the restriction parameter is $\lambda = d^{-O(\log d)}$ as opposed to being polynomially small as above. This improved significantly on the work of [7] that implicitly shows a similar claim for $\lambda = 2^{-O(d)}$.

We remark that in a related line of work, [1, 4, 3, 5] study random restrictions of PTFs over the hypercube. Our focus here is on gaussian restrictions and obtaining stronger bounds quantitatively: these works had exponential dependence on the degree d.

Slow-growth of derivatives

The analysis of the PRG (Theorem 21) and the random restriction statement above (Theorem 5) rely crucially on a claim about the magnitude of the derivatives of a polynomial evaluated at random gaussian input which may itself be of independent interest (and can be stated in a self-contained way).

For a function $p : \mathbb{R}^n \to \mathbb{R}$, let $\|\nabla^k p(x)\|^2$ denote the sum of squares of all partial derivatives of p of order k at x. That is, $\|\nabla^k p(x)\|$ is the Frobenius norm of the tensor of k'th order partial derivatives of p. We show that for any degree d polynomial p, the Frobenius-norm of the k'th order derivatives are comparable to the $(k-1)$'th order derivatives on a random gaussian input with high probability:

▶ **Lemma 6.** *For any degree-d polynomial $p : \mathbb{R}^d \to \mathbb{R}$, and $x \sim N(0,1)^n$, the following holds with probability at least $1-\delta$:*

$$\|\nabla^k p(x)\| \leq O(d^3/\delta)\|\nabla^{k-1} p(x)\|, \quad \text{for all } 1 \leq k \leq d. \tag{1}$$

Note that the above lemma is tight up to the factor of $O(d^2)$: consider the example $p(x) = x_1^d$.

Independent and concurrent work

Independently and concurrent to our work, [15] (following up on [14]) also obtained similar results to Theorem 3. They first obtained an analogue of *hypervariance reduction* (cf., Lemma 11) as studied in [14] with better parameters and combined the improved hypervariance reduction lemma with the framework of [14] to yield a PRG with $d^{O(1)}$ dependence on the degree d.

Our approach differs in that we critically use our new bounds on the growth of derivatives of polynomials as in Lemma 6 (instead of Lemma 11 which follows from Lemma 6). Working with the derivatives directly allows us to get a substantially simpler analysis of the main PRG construction compared to [14, 15].

2 Proof Overview

We first describe the high-level ideas underlying our main PRG construction - the proof of Theorem 21. We then describe the main idea behind the proof of Lemma 6 which is critical in being able to handle PTFs of polynomially large degree. The proof of Lemma 6 is quite different from the approach taken in [7, 14] to prove analogous results in their analysis.

2.1 Analysis of the PRG

We will use the same generator as in [7], and the high-level strategy is similar in spirit to that of [7, 14]. However, we introduce several additional ingredients that exploit Lemma 6 and significantly simplify the analysis.

As in the works of [7] and [14], the PRG output will be

$$Z := \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Y_i,$$

where each Y_i is an independent k-moment-matching gaussian vector with $k = d^{\Theta(1)}$. For the time being let us work under the idealized assumption that each Y_i is exactly k-moment-matching with a standard gaussian: i.e., for any polynomial $h : \mathbb{R}^n \to \mathbb{R}$ of degree at most k, $\mathbb{E}[h(Y_i)] = \mathbb{E}_{z \sim N(0,1)^n}[h(z)]$. We will later relax this condition without too much additional work as is now standard (see Section 3 for details), and ultimately output a discrete approximation to Z with finite support. For now, it is appropriate to imagine that the seed length required for generating each Y_i will be roughly $O(k \log n)$; the total seed length will thus be $L \cdot O(k \log n)$. We improve prior works by showing that it suffices to let $L = d^{\Theta(1)}$, rather than $L = 2^{\Theta(d)}$ as in [7] or $L = d^{\Theta(\log d)}$ as in [14].

For the rest of this section, fix a degree d polynomial $p : \mathbb{R}^n \to \mathbb{R}$ and let $f : \mathbb{R}^n \to \{0, 1\}$ defined as $f(x) = \mathsf{sign}(p(x))$ be the corresponding PTF we are trying to fool. For simplicity in this introduction, we consider the case where p is multi-linear. The general case is similar but is slightly more nuanced.

We wish to compare $\mathbb{E}_Z[f(Z)]$ to $\mathbb{E}_z[f(z)]$ where $z \sim N(0,1)^n$. Note that we can rewrite $z \sim N(0,1)^n$ as $z := \frac{1}{\sqrt{L}} \sum_{i=1}^{L} y_i$ where each y_i is an independent standard gaussian.

First attempt: A hybrid argument

A natural approach to analyze the PRG is to use a hybrid argument by replacing each y_i with a k-moment matching Gaussian vector Y_i as in our PRG output. That is, show the following sequence of inequalities:

$$\begin{aligned}\mathbb{E}\left[f\left(\frac{y_1}{\sqrt{L}} + \frac{y_2}{\sqrt{L}} + \cdots + \frac{y_L}{\sqrt{L}}\right)\right] &\approx \mathbb{E}\left[f\left(\frac{Y_1}{\sqrt{L}} + \frac{y_2}{\sqrt{L}} + \cdots + \frac{y_L}{\sqrt{L}}\right)\right] \\ &\approx \mathbb{E}\left[f\left(\frac{Y_1}{\sqrt{L}} + \frac{Y_2}{\sqrt{L}} + \cdots + \frac{y_L}{\sqrt{L}}\right)\right] \approx \cdots \approx \mathbb{E}\left[f\left(\frac{Y_1}{\sqrt{L}} + \frac{Y_2}{\sqrt{L}} + \cdots + \frac{Y_L}{\sqrt{L}}\right)\right]. \quad (2)\end{aligned}$$

Let $\lambda = 1/L$ and $y' = \sqrt{\lambda}(y_2 + \cdots + y_L)$. Note that $y' \sim N(0, 1-\lambda)^n$. The first inequality in the sequence above, corresponding to a single-step of the hybrid argument is, equivalent to showing:

$$\mathbb{E}\left[f(\sqrt{\lambda} y_1 + y')\right] \approx \mathbb{E}\left[f(\sqrt{\lambda} Y_1 + y')\right].$$

In other words, the above inequality is asking to show that

$$\mathbb{E}[f_{y'/\sqrt{1-\lambda}}(y_1)] \approx \mathbb{E}[f_{y'/\sqrt{1-\lambda}}(Y_1)].$$

Intuitively, this is equivalent to showing that k-moment matching gaussians fool gaussian restrictions of a PTF with high probability over the restriction. Indeed, such a claim follows from our bounds on the derivatives of polynomials at random evaluation points (Lemma 6).

We say that a polynomial p is *well-behaved* at a point x if

$$\|\nabla^{k+1} p(x)\| \leq (1/\varepsilon)\|\nabla^k p(x)\| \text{ for all } k = 0, 1, \ldots, d-1,$$

where ε is a parameter that will be set to be slightly larger than $\sqrt{\lambda}$. We say p is poorly-behaved at x if the above condition does not hold.

The starting point of the analysis is that if p is well-behaved at x, then $\mathsf{sign}(p(x+\sqrt{\lambda}Y))$ is fooled by a moment-matching Y with *very good* error:

▶ **Proposition 7** (Direct Corollary of Lemma 13)**.** *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a degree d multi-linear polynomial and suppose that q is well-behaved at a point x. Let $R = \varepsilon^2/\lambda$. Then, for $y \sim N(0,1)^n$ and Y a dR-moment matching gaussian,*

$$\mathop{\mathbb{E}}_{y\sim N(0,1)^n}[\mathsf{sign}(q(x+\sqrt{\lambda}y)] - \mathop{\mathbb{E}}_{Y}[\mathsf{sign}(q(x+\sqrt{\lambda}Y)] \leq 2^{-\Omega(R)}.$$

This fact follows from the following argument. Since q is well-behaved at x, this in particular implies a non-negligible lower bound on the size of the constant term c of $h(t) := q(x+\sqrt{\lambda}t)$, relative to its other coefficients. In particular, $\mathsf{sign}(h(t))$ is *nearly fixed to a constant* in the sense of Theorem 5. Indeed, writing $h(t) = c + (h(t) - c)$, we see that $\mathsf{sign}(h(t))$ can only differ from $\mathsf{sign}(c)$ if we have a deviation with magnitude at least $|h(t) - c| \geq |c|$. We can use a concentration inequality to bound the probability that either $|h(y) - c| \geq |c|$ or $|h(Y) - c| \geq |c|$. In light of the bounds on $\|\nabla^k q(x)\|$, such a concentration inequality follows from moment bounds obtained from *hypercontractivity*.

The above lemma shows the first step of the hybrid argument and suggests the following strategy for analyzing the PRG. Define $Z_{-i} = Z - \sqrt{\lambda}Y_i$. We can now aim to show that the polynomial p is well-behaved at Z_{-i} with high probability. This indeed seems plausible as our Lemma 6 indeed shows that when Z is standard gaussian, the polynomial p is well-behaved at Z with high probability.

Immediately, there are two obstacles for this approach:

- First, Lemma 6 works only for truly random gaussian and not for our pseudorandom Z_{-i}.
- Second, even if we argue that p is likely to be well-behaved at Z_{-i}, we cannot apply a union bound over i. The error guarantee in Lemma 6, is $\gg \sqrt{\lambda}$; whereas, we have $L = 1/\lambda$ choices of i, so we cannot use such a straightforward union-bound argument to replace each Y_i with a y_i.

The second issue is especially problematic as the error probability in Lemma 6 cannot be improved, at least in that variant; the probability that the derivatives don't grow too fast is not small compared to $L = 1/\lambda$.

Beating the union bound

Roughly speaking, the main insight in going beyond the *union bound* obstacle mentioned above is as follows. There are two sources of error in the naive hybrid argument outlined above: (1) The probability of failure coming from p being poorly-behaved at the points Z_{-i}. (2) The error coming from applying Proposition 7 to replace a Y_i with y_i when p is well-behaved at Z_{-i}.

Note that we have very good control on the error of type (2) above: we could make it be much smaller than $1/L$ by increasing the amount of independence k. We will exploit this critically. We will complement this by showing that even though a naive union bound would be bad for errors of type (1) above, it turns out that we don't have to incur this loss: we (implicitly) show that $\mathbb{P}(\forall i,\ p \text{ is well-behaved at } Z_{-i}) \approx 1 - O(\varepsilon d^3)$. We do so by checking only that p is well-behaved at the single point Z (in a slightly stronger sense) and then we conclude that p is also highly-likely to be well-behaved at each of the "nearby" points Z_{-i}. Intuitively, this is what allows us to circumvent the union bound in the hybrid argument. However, it would be difficult to actually carry out the analysis as stated this way – we use a sandwiching argument to sidestep the complicated conditionings which would arise in this argument as stated.

We proceed to describe the sandwiching argument. We wish to lower-bound the PTF $\mathsf{sign}(p(x))$ by $\mathsf{sign}(p(x)) \cdot g(x)$, where $g(x)$ is some "mollifier" function taking values in $[0,1]$. The role of $g(x)$ is roughly to "test" whether p is well-behaved at x; we ideally want $g(x) = 1$ at points x where p is well-behaved and $g(x) = 0$ at points x where p is poorly-behaved. However, we also need $g(x)$ to be smooth[9], so there will be some intermediate region of points for which $g(x)$ yields a non-informative, non-boolean value.

We set $g(x)$ to be a smoothed version of the indicator function

$$g(x) \approx \prod_{k=0}^{d-1} \mathbb{1}\left(\|\nabla^{k+1} p(x)\| \le \frac{1}{\varepsilon} \|\nabla^k p(x)\| \right),$$

which tests whether the derivatives of p at x have controlled growth in the sense of Lemma 6. More specifically, we set

$$g(x) := \prod_{k=0}^{d-1} \rho\left(\log\left(\frac{1}{16\varepsilon^2} \frac{\|\nabla^k p(x)\|^2}{\|\nabla^{k+1} p(x)\|^2} \right) \right),$$

where $\rho(t) : \mathbb{R} \to [0,1]$ is some smooth univariate function with $\rho(t) = 0$ for $t \le 0$ and $\rho(t) = 1$ for $t \ge 1$.

Now, for every point $x \in \mathbb{R}^n$ we have

$$\mathsf{sign}(p(x)) \ge \mathsf{sign}(p(x)) g(x).$$

Furthermore, under truly-random gaussian inputs $z \sim N(0,1)^n$ we have

$$\mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) g(z) \ge \mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) - \mathop{\mathbb{E}}_z |g(z) - 1| \ge \mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) - O(\varepsilon d^3),$$

where the final inequality here follows from Lemma 6. Combining these, we get that

$$\mathop{\mathbb{E}}_Z \mathsf{sign}(p(Z)) \ge \mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) - O(\varepsilon d^3) - \left| \mathop{\mathbb{E}}_Z \mathsf{sign}(p(Z)) g(Z) - \mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) g(z) \right|.$$

Note that we can similarly obtain an upper-bound for $\mathbb{E}_Z \mathsf{sign}(p(Z))$ by repeating this argument on the polynomial $-p(x)$.

Thus, it suffices to bound $|\mathbb{E}_Z \mathsf{sign}(p(Z)) g(Z) - \mathbb{E}_z \mathsf{sign}(p(z)) g(z)|$. Having introduced the mollifier, we can now afford to do so by a standard hybrid argument. We represent z as $z := \frac{1}{\sqrt{L}} \sum_{i=1}^{L} y_i$ and recall that Z is of the form $Z = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Y_i$. We can replace each Y_i with y_i and get

$$\left| \mathop{\mathbb{E}}_Z \mathsf{sign}(p(Z)) g(Z) - \mathop{\mathbb{E}}_z \mathsf{sign}(p(z)) g(z) \right| \le \gamma L,$$

where γ is the (quite small) error coming from the following lemma.

[9] Ultimately we need g to be well approximated by some polynomial so that it can be fooled by limited independence. Smoothness will allow us to get such an approximation by truncating the Taylor series.

▶ **Lemma 8.** *There exists a constant c such that the following holds for $\lambda \le \varepsilon^2/Rd^c$. For any fixed vector $x \in \mathbb{R}^n$, Y a dR-moment-matching gaussian vector, and $y \sim N(0,1)^n$,*

$$|\mathop{\mathbb{E}}_{Y} \mathsf{sign}(p(x+\sqrt{\lambda}Y))g(x+\sqrt{\lambda}Y) - \mathop{\mathbb{E}}_{y} \mathsf{sign}(p(x+\sqrt{\lambda}y))g(x+\sqrt{\lambda}y)| \le \gamma = 2^{-\Omega(R)}.$$

Technically speaking, the above lemma is where our intuition on going around the union bound is quantified, allowing us to use the hybrid argument. We briefly outline our proof of this lemma, where for the purpose of illustration we continue with the simplifying assumption that the polynomial p is multilinear.

The proof is by a case analysis on the behavior of p at the the fixed point x. In the multilinear case it suffices to consider the derivatives $\nabla^k p(x)$; in the general case we need to consider something slightly different.

- Case 1: p is well-behaved at x, i.e., $\|\nabla^{k+1}p(x)\| \le (1/\varepsilon)\|\nabla^k p(x)\|$ for all k.
  - We can use Lemma 13 in this case to conclude that $\mathsf{sign}(p(x+\sqrt{\lambda}y))$, $\mathsf{sign}(p(x+\sqrt{\lambda}Y))$ are both almost constant with error $2^{-\Omega(R)}$.
  - So, it remains to show that Y fools $g(x+\sqrt{\lambda}y)$. We approximate g by a low-degree polynomial in y using a Taylor-truncation argument. Our assumption on the controlled growth of derivatives $\|\nabla^k p(x)\|$ allows us to bound the truncation error by bounding the higher-moments of the deviations $\|\nabla^k p(x+\sqrt{\lambda}Y)\| - \|\nabla^k p(x)\|$.
- Case 2: p is not well-behaved at x; let k_0 be the largest k such that $\|\nabla^{k_0+1}p(x)\| > (1/\varepsilon)\|\nabla^{k_0}p(x)\|$.
  - Intuitively, this says that the polynomial p is well behaved at degree above k_0, but not at degree k_0. This allows us to show, via an R-th moment bound, that both
    - $\|\nabla^{k_0}p(x+\sqrt{\lambda}Y)\| \le 2\varepsilon\|\nabla^{k_0+1}p(x)\|$
    - $\|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)\| \ge \frac{1}{2}\|\nabla^{k_0+1}p(x)\|$

    are highly likely. Thus, it is highly likely that

    $$\|\nabla^{k_0}p(x+\sqrt{\lambda}Y)\| \le 4\varepsilon\|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)\|.$$

    The latter means p is still sufficiently poorly-behaved at the point $x+\sqrt{\lambda}Y$ that the mollifier classifies it correctly as $g(x+\sqrt{\lambda}Y) = 0$.

2.2 Slow-growth of derivatives and simplification under restrictions

The proof of Lemma 6 is iterative and is relatively simple given Kane's *relative anti-concentration inequality* for degree d polynomials [9] developed in the context of studying the *Gotsman-Linial* conjecture for PTFs.

[9] shows that for any degree d polynomial, and $x, y \sim N(0,1)^n$ with probability at least $1-\delta$, we have $|\langle y, \nabla p(x)\rangle| \le (d^2/\delta)|p(x)|$. As y in the above statement is independent of x, for any x, $\langle y, \nabla p(x)\rangle$ is distributed as $N(0, \|\nabla p(x)\|^2)$. This says that the inequality is essentially equivalent to saying that with probability at least $1-\delta$ over x, we have $\|\nabla p(x)\|^2 \le O(d^2/\delta)|p(x)|$. The latter can be seen as the inequality corresponding to $k=1$ in the statement of Lemma 6. The full proof of the lemma is via iteratively applying the above argument using a vector-valued generalization of Kane's inequality.

Next, it is not too hard to prove Theorem 5 given Lemma 6. For illustration, suppose that we have a degree d multi-linear polynomial p, and write $f(t) := p(\sqrt{1-\lambda}t)$. Then, by elementary algebra[10], we have the identity

$$p_x(y) = p\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = \sum_{\alpha} \partial^\alpha f(x) \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} y^\alpha. \tag{3}$$

[10] If p is multi-linear, then the Hermite expansion (see Section 3) is just $p(x) = \sum_{\alpha\in\{0,1\}^n} \hat{p}(\alpha)h_\alpha(x) = \sum_{I\subseteq[n]} \hat{p}(I)\prod_{i\in I} x_i$. We can prove the identity for each monomial and use additivity.

Now, by Lemma 6, with probability $1-\delta$ over x, we have $\|\nabla^k f(x)\| \le O(d^3/\delta)\|\nabla^{k-1} f(x)\|$, for all k. Thus, if we take $\lambda \ll \delta^2/(R^2 d^6)$, the factor of λ will kill the growing derivatives leading to a bound on the higher-order moments of $p_x(y) - p_x(0)$ via hypercontractivity. These moment bounds in turn imply that $|p_x(y) - p_x(0)| < |p_x(0)|$ with high probability over y, and hence that $\mathsf{sign}(p_x(y)) = \mathsf{sign}(p_x(0))$ with high probability over y.

Notice that Equation (3) is essentially a Taylor expansion of p at $\sqrt{1-\lambda}x$: it expresses the function $p_x(y)$ as a polynomial in y in the standard basis, whose coefficients are determined by the derivatives of p at $\sqrt{1-\lambda}x$. We want to do something similar in the general case, but in the Hermite basis; for non-multi-linear polynomials these two bases no longer coincide. So, in the general case, we rely on the following identity, which we regard as an analogue of the Taylor expansion for the Hermite basis.

▶ **Lemma 9** (See Section 3). *Let $f(y) = \sum_\alpha \hat{f}(\alpha) h_\alpha(y)$. Then*

$$f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = \sum_\alpha \frac{\partial^\alpha g(x)}{\sqrt{\alpha!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} h_\alpha(y),$$

where $g(x) := U_{\sqrt{1-\lambda}} f(x) = \sum_\alpha \hat{f}(\alpha)(1-\lambda)^{|\alpha|/2} h_\alpha(x)$.

Hermite polynomials are such a ubiquitous tool used in such a wide range of fields that it seems unlikely that such an identity is new. However, we are not aware of any previous appearance of such an identity in the literature (at least in the body of work on PTFs) and we provide a proof.

Hypervariance reduction

We next remark on the relation between *slow-growth of derivatives* (as in Lemma 6) and *hypervariance reduction* as studied and introduced in [14]. The latter plays a similar role in their paper as the former does in this work. However, Lemma 6 importantly has only polynomial dependence on the degree d and is also much more conducive to our analysis of the PRG.

Recall the Hermite expansion (see Section 3) of polynomials: A degree d polynomial $p : \mathbb{R}^n \to \mathbb{R}$ can be uniquely expressed as

$$p(y) := \sum_{|\alpha| \le d} \hat{p}(\alpha) h_\alpha(y),$$

where $\alpha \in \mathbb{N}^n$ denotes a multi-index and $h_\alpha(y)$ is the α'th Hermite polynomial. The *hypervariance* and *normalized hypervariance* of a polynomial introduced in [14] are defined as follows:

▶ **Definition 10.** *For a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of the form $p(y) := \sum_\alpha \hat{p}(\alpha) h_\alpha(y)$, define its* hypervariance*,* $\mathrm{HyperVar}_R(\,)$*, and* normalized hypervariance*, $H_R(\,)$, as*

$$\mathrm{HyperVar}_R(p) := \sum_{\alpha \ne 0} \hat{p}(\alpha)^2 R^{2|\alpha|}, \quad H_R(p) := \frac{\mathrm{HyperVar}_R(p)}{\hat{p}(0)^2}.$$

Intuitively, if the normalized hypervariance $H_R(p)$ of a polynomial is small for a large R, then it means that the *weights* of the higher-order Hermite coefficients of p have a geometric decay.

[14] showed that for any polynomial p, for a suitable $\lambda > 0$, a gaussian restriction of p will have small normalized hypervariance with high probability. Specifically, they showed that if $\lambda = d^{-O(\log d)}$, then $H_R(p_{x,\lambda})$ is bounded with high probability over $x \sim N(0,1)^n$. They also asked whether this property holds when $\lambda = d^{-O(1)}$ instead of being quasi-polynomially small in d. Lemma 6 implies this conjecture without too much difficulty:

▶ **Lemma 11.** *For any degree d polynomial p and $\lambda, \delta > 0$, the following holds. Except with probability δ over $x \sim N(0,1)^n$, the normalized hypervariance $H_R(p_{x,\lambda}) = O(\lambda d^6 R^2/\delta^2)$.*

The proof of the analogue of Lemma 11 for quasi-polynomially small λ (i.e. $\lambda = d^{-O(\log d)}$) in [14] was by an iterative process: Intuitively, if one sets $\lambda_0 = d^{-O(1)}$, and $\lambda = \lambda_0^{\log d}$, then the random restriction $p_{\lambda,x}$ is equivalent to $(\log d)$ independent random restrictions with restriction parameter λ_0. The authors in [14] show that each such λ_0-restriction (essentially) decreases the degree by a factor of 2. We instead take a different approach by drawing a connection between norms of derivatives and to *relative anti-concentration* as developed in the context of studying the *Gotsman-Linial* conjecture for PTFs.

3 Preliminaries

The pseudorandom generator construction: idealization vs. discretization. Following [7] and [14], we analyze the idealized pseudorandom distribution

$$Z = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Y_i,$$

where each $Y_i \in \mathbb{R}^n$ is a k-moment-matching gaussian (that is, $\mathbb{E}[p(Y_i)] = \mathbb{E}_{x \sim N(0,1)^n}[p(x)]$ for all polynomials $p : \mathbb{R}^n \to \mathbb{R}$ of degree at most k).

Suppose that, for any such Z with parameters (L, k), it is the case that Z fools degree-d PTFs with error $\varepsilon = \varepsilon(L, k, d)$. Then, it is shown in [7] how to obtain a small-seed length PRG (in the sense of Definition 2) by providing a specific instantiation and discretization of this construction.

▶ **Theorem 12** ([7], implicit in Section 6)**.** *Suppose a Z as above with parameters (L, k) fools degree d-PTFs with error $\varepsilon = \varepsilon(L, k, d)$. Then, there is an explicit, efficiently computable PRG with seed length $O(dkL \log(ndL/\varepsilon)$ that (2ε)-fools degree d PTFs.*

Hermite polynomials. To argue about polynomials which are not necessarily multilinear, we need some simple facts concerning Hermite polynomials. For our purposes, Hermite polynomials are simply a convenient choice of polynomial basis which have nice properties (in particular being *orthonormal*) with respect to gaussian inputs. For a more detailed background on Hermite polynomials and their use for analyzing functions over gaussian space, see [13, Ch. 11].

One concrete way to define the Hermite polynomials is the following:

- For the univariate polynomials, the degree-m "Probabilist's" Hermite polynomial is the m-th coefficient of the generating function

$$e^{st - \frac{1}{2}s^2} = \sum_{m \geq 0} H_m(t) s^m.$$

- We define the degree-m univariate Hermite polynomial by the normalization

$$h_m(t) := \frac{1}{\sqrt{m!}} H_m(t).$$

- For a multi-index $\alpha \in \mathbb{N}^n$, we define the multivariate Hermite polynomial $h_\alpha : \mathbb{R}^n \to \mathbb{R}$ via the product

$$h_\alpha(x) := \prod_{i=1}^{n} h_{a_i}(x_i).$$

We record some basic properties of this particular choice of polynomial basis. The final two properties say that the Hermite basis is orthonormal with respect to correlation under the standard gaussian distribution – this is the reason for our choice of normalization.

- The set $\{h_\alpha(x) : |\alpha| \leq d\}$ is a basis for real polynomials in n variables of degree $\leq d$.
- h_0 is the constant polynomial $h_0 \equiv 1$.
- For multi-indicies $\alpha \in \{0,1\}^n$, $h_\alpha(x)$ is simply the monomial $\prod_{i:a_i=1} x_i$.
- For $x \sim N(0,1)^n$, and distinct multi-indices $\alpha \neq \beta$, $\mathbb{E}_x\, h_\alpha(x)h_\beta(x) = 0$.
- For $x \sim N(0,1)^n$, and any multi-index α, $\mathbb{E}_x\, h_\alpha(x)^2 = 1$.

Gaussian noise operator. We recall the definition of the noise operator U_ρ, which here we regard as an operator on real polynomials in n variables (see [13, Ch. 11] for background and a more general viewpoint). For a polynomial $f : \mathbb{R}^n \to \mathbb{R}$ and a parameter $\rho \in [0,1]$, the action of U_ρ on f is specified by

$$(U_\rho f)(x) := \mathop{\mathbb{E}}_{Z \sim N(0,1)^n} f\left(\rho x + \sqrt{1-\rho^2} Z\right).$$

An important feature of the Hermite basis is that the noise operator acts on it *diagonally* (see [13, Ch. 11]):

$$U_\rho h_\alpha(x) = \rho^{|\alpha|} h_\alpha(x).$$

Thus, if f is a degree-d polynomial given in the Hermite basis as

$$f(x) = \sum_{|\alpha| \leq d} \hat{f}(\alpha) h_\alpha(x),$$

then we can express the result of the noise operator applied to f explicitly as

$$U_\rho f(x) = \sum_{|\alpha| \leq d} \hat{f}(\alpha) \rho^{|\alpha|} h_\alpha(x).$$

Higher moments and hypercontractivity. Fix a polynomial

$$f(x) := \sum_{|\alpha| \leq d} \hat{f}(\alpha) h_\alpha(x).$$

For an even natural number $q \geq 2$, we write the gaussian q-norm of f as

$$\|f\|_q := \left(\mathop{\mathbb{E}}_{x \sim N(0,1)^n} f(x)^q \right)^{1/q}.$$

We wish to be able to bound this quantity in terms of the magnitudes of the Hermite coefficients of f, $\hat{f}(\alpha)$. For this purpose, we extend the definition of U_ρ also to $\rho > 1$ by its action on the Hermite basis: $U_\rho h_\alpha(x) = \rho^{|\alpha|} h_\alpha(x)$. With this notation, we can express the well-known $(q,2)$-hypercontractive inequality [13, Ch. 9,11] as

$$\|f\|_q \leq \|U_{\sqrt{q-1}} f\|_2,$$

which is quite convenient for us, as we can use orthonormality of the Hermite basis to explicitly compute

$$\|U_{\sqrt{q-1}} f\|_2^2 = \sum_{|\alpha| \leq d} (q-1)^{|\alpha|} \hat{f}(\alpha)^2 \leq \sum_{|\alpha| \leq d} q^{|\alpha|} \hat{f}(\alpha)^2.$$

To get a feel for the utility of this bound, let's see how it can be used to prove the following concentration bound:

▶ **Lemma 13.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a degree d polynomial with normalized hypervariance $H_{\sqrt{q}}(f) \le \frac{1}{4}$, where q is an even natural number. Then,*

$$\mathop{\mathbb{P}}_{y\sim N(0,1)^n}\Big(\mathsf{sign}(f(y)) \neq \mathsf{sign}(\hat{f}(0))\Big) \le 2^{-q}.$$

Further, the same holds more generally for $y \sim Y$, as long as the distribution Y is dq-moment-matching.

Proof. Suppose that $f(y)$ is normalized so that

$$\mathop{\mathbb{E}}_{y\sim N(0,1)^n} f(y) = \hat{f}(0) = \pm 1.$$

We have the q-th moment bound

$$\|f(x) - \hat{f}(0)\|_q \le \|U_{\sqrt{q}}\left(f(y) - \hat{f}(0)\right)\|_2 \le \tfrac{1}{2}.$$

From the generic concentration inequality

$$\mathbb{P}\left(|X| \ge t\|X\|_q\right) \le t^{-q}$$

we obtain

$$\mathbb{P}\Big(\mathsf{sign}(f(y)) \neq \mathsf{sign}(\hat{f}(0))\Big) \le 2^{-q}.$$

Thus, we find that the PTF $\mathsf{sign}(f)$ almost always yields the value $\mathsf{sign}(\hat{f}(0))$ under random gaussian inputs. Crucially for us, this argument is also *easy to derandomize*: since the argument merely relies on a bound on the q-th moment $\mathbb{E}_{y\sim N(0,1)^n}(f(y) - \hat{f}(0))^q$, and for Y which is k-moment-matching for $k \ge dq$ we have

$$\mathop{\mathbb{E}}_{Y}(f(Y) - \hat{f}(0))^q = \mathop{\mathbb{E}}_{y\sim N(0,1)^n}(f(y) - \hat{f}(0))^q,$$

we conclude also that $\mathsf{sign}(f(Y))$ is typically equal to $\mathsf{sign}(\hat{f}(0))$. ◀

We remark that this lemma further implies that Y fools $\mathsf{sign}(f)$ when $H_{\sqrt{q}}(f)$ is small:

$$\mathop{\mathbb{E}}_{Y}\mathsf{sign}(f(Y)) = \mathop{\mathbb{E}}_{y\sim N(0,1)^n}\mathsf{sign}(f(y)) \pm O(2^{-q}).$$

Gaussian restrictions and derivatives on the Hermite basis. Besides the effect of the noise operator, it will also be important to understand the effect of two further operations on polynomials:

- The derivative map, $f(y) \mapsto \partial^\alpha f(y)$.
- The gaussian restriction at x, $f(y) \mapsto f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right)$.

In particular, we are concerned with how these operations affect the Hermite coefficients of a polynomial; ultimately, our goal will be to develop a "Hermite-basis analogue" of the Taylor expansion which can be applied to expand $f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right)$ as a function of y. We start by computing the effect of these two operations on univariate Hermite polynomials, and then on the full multivariate Hermite basis, and finally on a general polynomial $f(x)$ expressed in the Hermite basis.

▶ **Proposition 14.** *For univariate Hermite polynomials, we have the identities*

- $\frac{\partial^k}{\partial t^k} h_m(t) = \sqrt{\frac{m!}{(m-k)!}} h_{m-k}(t)$,
- $h_m\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = \sum_{k=0}^{m} \sqrt{\binom{m}{k}} (1-\lambda)^{(m-k)/2} \lambda^{k/2} h_{m-k}(x) h_k(y)$.

Proof. The first of these identities is standard (see e.g. [13, Ex. 11.10]); we provide a proof of the second.

The second identity can be proved by considering the generating function

$$e^{st - \frac{1}{2}s^2} = \sum_m \sqrt{m!} h_m(t) s^m,$$

and comparing the coefficient of s^m on both sides of

$$e^{s(\sqrt{1-\lambda}x + \sqrt{\lambda}y) - \frac{1}{2}s^2} = e^{(s\sqrt{1-\lambda})x - \frac{1}{2}(s\sqrt{1-\lambda})^2} \cdot e^{(s\sqrt{\lambda})y - \frac{1}{2}(s\sqrt{\lambda})^2}$$ ◀

The corresponding identities for multivariate Hermite polynomials follow easily from above.

▶ **Proposition 15.** *We have*

- $\partial^\alpha h_\beta(y) = \sqrt{\frac{\alpha!}{\gamma!}} h_\gamma(y)$, *where* $\gamma = \beta - \alpha$,
- $h_\beta\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = (1-\lambda)^{|\beta|/2} \sum_{\alpha \le \beta} \frac{\partial^\alpha h_\beta(x)}{\sqrt{\alpha!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} h_\alpha(y)$,
- $\partial^\alpha h_\beta\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = (1-\lambda)^{|\beta-\alpha|/2} \sum_{\gamma \le \beta-\alpha} \frac{\partial^{\alpha+\gamma} h_\beta(x)}{\sqrt{\gamma!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\gamma|/2} h_\gamma(y)$.

We conclude with a Taylor-like expansion in the Hermite basis that we use repeatedly.

▶ **Lemma 16.** *Let* $f(y) = \sum_\alpha \hat{f}(\alpha) h_\alpha(y)$. *Then*

$$f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = \sum_\alpha \frac{\partial^\alpha g(x)}{\sqrt{\alpha!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} h_\alpha(y),$$

where $g(x) := U_{\sqrt{1-\lambda}} f(x) = \sum_\alpha \hat{f}(\alpha)(1-\lambda)^{|\alpha|/2} h_\alpha(x)$.

Proof. We express

$$\begin{aligned} f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) &= \sum_\alpha \hat{f}(\alpha) h_\alpha\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) \\ &= \sum_\alpha \frac{h_\alpha(y)}{\sqrt{\alpha!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} \sum_{\beta \ge \alpha} \hat{f}(\beta)(1-\lambda)^{|\beta|/2} \partial^\alpha h_\beta(x) \\ &= \sum_\alpha \frac{h_\alpha(y)}{\sqrt{\alpha!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2} \partial^\alpha g(x). \end{aligned}$$ ◀

Lastly, we will also need an extension of this theorem which expresses $\partial^\alpha f$, at the point

$$\sqrt{1-\lambda}x + \sqrt{\lambda}y,$$

as a polynomial in y in the Hermite basis.

▶ **Theorem 17.** *Let* $f(y) = \sum_\alpha \hat{f}(\alpha) h_\alpha(y)$. *Then*

$$\partial^\alpha f\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right) = (1-\lambda)^{-|\alpha|/2} \sum_{\beta \ge \alpha} \partial^\beta g(x) \sqrt{\frac{\alpha!}{\beta!}} \left(\frac{\lambda}{1-\lambda}\right)^{|\beta-\alpha|/2} h_{\beta-\alpha}(y),$$

where $g(x) := U_{\sqrt{1-\lambda}} f(x)$.

Proof. We express

$$\begin{aligned}\partial^\alpha f\left(\sqrt{1-\lambda}x+\sqrt{\lambda}y\right) &= \sum_\beta \hat{f}(\beta)\partial^\alpha h_\beta\left(\sqrt{1-\lambda}x+\sqrt{\lambda}y\right)\\ &= \sum_\gamma \frac{h_\gamma(y)}{\sqrt{\gamma!}}\left(\frac{\lambda}{1-\lambda}\right)^{|\gamma|/2}\sum_{\beta\geq\gamma+\alpha}(1-\lambda)^{|\beta-\alpha|/2}\partial^{\alpha+\gamma}h_\beta(x)\\ &= (1-\lambda)^{-|\alpha|/2}\sum_\gamma \frac{h_\gamma(y)}{\sqrt{\gamma!}}\left(\frac{\lambda}{1-\lambda}\right)^{|\gamma|/2}\partial^{\alpha+\gamma}g(x).\end{aligned}$$

◀

4 Gaussian restrictions of polynomials

Here we prove the structural properties of gaussian restrictions of polynomials: Theorem 5, Lemma 6, Lemma 11. Note that Theorem 5 follows immediately from Lemma 11 and Lemma 13. We next prove Lemma 11 from Lemma 6.

Proof of Lemma 11 from Lemma 6. Define $f(x) := U_{\sqrt{1-\lambda}}p(x)$. Then, by Lemma 16,

$$p_x(y) = f(x) + \sum_{\alpha\neq 0}\frac{\partial^\alpha f(x)}{\sqrt{\alpha!}}\left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|/2}h_\alpha(y).$$

Thus,

$$\begin{aligned}\mathrm{HyperVar}_R(p_x) = \sum_{\alpha\neq 0}\left(\frac{\partial^\alpha f(x)}{\sqrt{\alpha!}}\right)^2\left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|}R^{2|\alpha|} &\leq \sum_{\alpha\neq 0}\left(\partial^\alpha f(x)\right)^2\left(\frac{\lambda}{1-\lambda}\right)^{|\alpha|}R^{2|\alpha|}\\ &= \sum_{k=1}^d R^{2k}\left(\frac{\lambda}{1-\lambda}\right)^k\|\nabla^k f(x)\|^2,\end{aligned}$$

where the first inequality follows as $\sqrt{\alpha!}\geq 1$.

We now conclude by applying Lemma 6 to f. We have

$$H_R(p_x) = \frac{\sum_{k=1}^d R^{2k}\left(\frac{\lambda}{1-\lambda}\right)^k\|\nabla^k f(x)\|^2}{f(x)^2}.$$

Except with probability δ over $x\sim N(0,1)^n$, we can bound this by

$$\sum_{k=1}^d R^{2k}\left(\frac{\lambda}{1-\lambda}\right)^k\left(\frac{Cd^3}{\delta}\right)^{2k}\leq O\left(\frac{\lambda d^6R^2}{\delta^2}\right).$$

◀

4.1 Proof of Lemma 6

Our main tool will be Kane's relative-anticoncentration lemma for gaussian polynomials.

▶ **Lemma 18** ([9]). *For a degree d polynomial p, and independent standard gaussian vectors $x,y\in\mathbb{R}^n$,*

$$\mathbb{P}(|p(x)|\leq\varepsilon|\left\langle y,\nabla p(x)\right\rangle|)\leq O(\varepsilon d^2).$$

In fact, we will actually work with the following corollary which is essentially the first of the d inequalities in Lemma 6.

▶ **Corollary 19.** *For a degree d polynomial p, and independent standard gaussian vector $x \in \mathbb{R}^n$,*

$$\mathbb{P}(|p(x)| \le \varepsilon \|\nabla p(x)\|) \le O(\varepsilon d^2).$$

Proof. We note that for any fixed x, $\langle y, \nabla p(x)\rangle$ is identical in distribution to $Z\|\nabla p(x)\|$, where $Z \sim N(0,1)$ is a standard gaussian. So, we express

$$\begin{aligned}\mathbb{P}(|p(x)| \le \varepsilon |\langle y, \nabla p(x)\rangle|) &= \mathbb{P}(|p(x)| \le \varepsilon |Z| \|\nabla p(x)\|) \\ &\ge \mathbb{P}(|p(x)| \le \varepsilon \|\nabla p(x)\|) \cdot \mathbb{P}(|Z| \ge 1)\,.\end{aligned}$$

Since $\mathbb{P}(|Z| \ge 1) \ge \Omega(1)$, we conclude that

$$\mathbb{P}(|p(x)| \le \varepsilon \|\nabla p(x)\|) \le O(\varepsilon d^2).$$ ◀

The heart of the proof of Lemma 6 is a vector-valued variant of the above corollary:

▶ **Lemma 20.** *Let $\vec{f}(x) := (f_1(x), f_2(x), \ldots, f_m(x))$ be a collection of m degree-at-most d polynomials $f_j(x)$. If $x \in \mathbb{R}^n$ is a standard gaussian vector, then*

$$\mathbb{P}\left(\|\vec{f}(x)\|^2 \le \varepsilon^2 \sum_{j=1}^{m} \|\nabla f_j(x)\|^2 \right) \le O(\varepsilon d^2).$$

Proof of Lemma 6. We simply apply the above lemma d times and take a union bound. For $1 \le k \le d$, let $\vec{f}_k(x) := ((\partial^\alpha f(x) : |\alpha| = k))$. Note that $\|\vec{f}_k(x)\|^2 = \|\nabla^k f(x)\|^2$. Further, note that

$$\sum_{\alpha : |\alpha| = k} \|\nabla(\partial^\alpha f(x))\|^2 \ge \|\nabla^{k+1} f(x)\|^2,$$

where the inequality follows as each $(k+1)$'th order derivative would be counted at least once in the expression on the left hand side. Therefore, by the above lemma, for $x \sim N(0,1)^n$, we have

$$\mathbb{P}\big(\|\nabla^k f(x)\|^2 \le \varepsilon^2 \|\nabla^{k+1} f(x)\|^2\big) \le O(\varepsilon d^2)$$

Setting $\varepsilon = \delta/d^3$, and taking a union bound over all k, we get that for a constant $C > 0$,

$$\mathbb{P}\big(\forall k, \|\nabla^k f(x)\|^2 > C(\delta^2/d^6)\|\nabla^{k+1} f(x)\|^2\big) \ge 1 - \delta.$$

This proves Lemma 6. ◀

Proof of Lemma 20. Consider the auxiliary polynomial

$$h(x, y) := \sum_{j=1}^{m} f_j(x) y_j.$$

As a function of both x and y, we have

$$\nabla h(x, y) = \vec{f}(x) \circ M_x y,$$

where M_x is the matrix with columns $\nabla f_j(x)$ (that is, M_x has (i,j)-th entry $\frac{\partial}{\partial x_i} f_j(x)$). So, applying Corollary 19 to this auxiliary polynomial gives the probability bound

$$\begin{aligned} q &:= \mathbb{P}\big(h(x,y)^2 \le \varepsilon^2 \|\nabla g(x,y)\|^2\big) \\ &= \mathbb{P}\left(\left\langle y, \vec{f}(x)\right\rangle^2 \le \varepsilon^2 \left(\|\vec{f}(x)\|^2 + \|M_x y\|^2\right)\right) \\ &\le O(\varepsilon d^2). \end{aligned}$$

Now, for some constant $C \ge 2$ to be specified later, let E denote the event that

$$(C^2-1)\|\vec{f}(x)\|^2 \le \frac{\varepsilon^2}{2}\|M_x\|_F^2,$$

where $\|M_x\|_F$ is the Frobenius norm of M_x. We note that we can lower-bound the probability q by

$$q \ge \mathbb{P}(E) \cdot \mathbb{P}\left(\left|\left\langle y, \vec{f}(x)\right\rangle\right| \le C\|\vec{f}(x)\| \text{ and } \|M_x y\|^2 \ge \frac{1}{2}\|M_x\|_F^2 \mid E\right).$$

We claim that for large enough choice of constant C, this conditional probability can be lower-bounded by $\Omega(1)$. Indeed, we can argue for any fixed x:

- $\mathbb{P}\left(\left|\left\langle y, \vec{f}(x)\right\rangle\right| \ge C\|\vec{f}(x)\|\right) \le \frac{1}{C^2}$.
- $\mathbb{P}\big(\|M_x y\|^2 \ge \frac{1}{2}\|M_x\|_F^2\big) \ge \Omega(1)$.

The first item is just a Chebyshev inequality; the second item can be derived e.g. from the basic anticoncentration bound one obtains for degree-2 polynomials from the Paley-Zygmund bound together with hypercontractivity (since, for any fixed matrix M, the quadratic form $g(y) := \|My\|^2$ has second-moment $\mathbb{E}\, g(y)^2 \ge (\mathbb{E}\, g(y))^2 = \|M\|_F^2$).

Thus, by choosing C large enough, we can lower-bound this conditional probability by

$$\Omega(1) - \frac{1}{C^2} \ge \Omega(1).$$

We conclude that $\mathbb{P}(E) \le O(q) = O(\varepsilon d^2)$. This gives the desired conclusion

$$\mathbb{P}\left(\|\vec{f}(x)\| \le \Omega(\varepsilon)\|M_x\|_F\right) \le O(\varepsilon d^2). \qquad \blacktriangleleft$$

5 Pseudorandom Generator for PTFs

The following theorem gives quantitative bounds on the error of our main generator:

▶ **Theorem 21.** *Fix some parameters $\varepsilon > 0$ and $R \in \mathbb{N}$. Let z be a standard gaussian, and let $Z = \frac{1}{\sqrt{L}}\sum_{i=1}^{L} Y_i$, where each Y_i is dR-moment-matching. Then for some sufficiently large absolute constant c and any polynomial p of degree d,*

$$\mathop{\mathbb{E}}_{Z} \mathsf{sign}(p(Z)) \ge \mathop{\mathbb{E}}_{z\sim N(0,1)^n} \mathsf{sign}(p(z)) - O(\varepsilon d^3) - L \cdot 2^{-\Omega(R)},$$

as long as L is at least Rd^c/ε^2.

Combining the above with Theorem 12 immediately implies our main result Theorem 3.

Proof of Theorem 3. Given a target error ε', set $\varepsilon = \varepsilon'/Cd^3$, and $R = C\log(d/\varepsilon)$ for a sufficiently big constant so that the error in the above lemma is at most $\varepsilon'/2$ for $L = Rd^c/\varepsilon^2 = O(d^c \log(d/\varepsilon)/\varepsilon^2)$. While the above theorem only gives a lower bound, we can get an upper bound by applying the result to $-p$. Now, by applying Theorem 12 there exists an efficient PRG that fools degree d PTFs with error at most ε' and seed length $O(d^{O(1)} \log(nd/\varepsilon') \log(d/\varepsilon')/(\varepsilon')^2$ which can be simplified to the bound in the theorem. ◀

We now prove the above theorem by the lower-sandwiching argument outlined in Section 2.1. Fix a polynomial $p(x)$ of degree d. We remind the reader of our convention $\mathsf{sign}(t) := \mathbb{1}(t \geq 0)$.

We define the mollifier function

$$g(x) := \prod_{k=0}^{d-1} \rho\left(\log\left(\frac{1}{16\varepsilon^2}\frac{\|\nabla^k p(x)\|^2}{\|\nabla^{k+1} p(x)\|^2}\right)\right),$$

where $\rho : \mathbb{R} \to [0,1]$ is some smooth univariate function with $\rho(t) = 0$ for $t \leq 0$, $\rho(t) = 1$ for $t \geq 1$, and $\|\frac{\partial^k \rho}{\partial t^k}\|_\infty \leq k^{O(k)}$ for all k. [11]

Proof of Theorem 21. For every point $x \in \mathbb{R}^n$ we have

$$\mathsf{sign}(p(x)) \geq \mathsf{sign}(p(x))g(x).$$

Furthermore, under the truly-random gaussian inputs $z \sim N(0,1)^n$ we have

$$\mathop{\mathbb{E}}_{z} \mathsf{sign}(p(z))g(z) \geq \mathop{\mathbb{E}}_{z} \mathsf{sign}(p(z)) - \mathop{\mathbb{E}}_{z} |g(z) - 1| \geq \mathop{\mathbb{E}}_{z} \mathsf{sign}(p(z)) - O(\varepsilon d^3),$$

where the final inequality here follows from Lemma 6. Combining these, we get that

$$\mathop{\mathbb{E}}_{Z} \mathsf{sign}(p(Z)) \geq \mathop{\mathbb{E}}_{z\sim N(0,1)^n} \mathsf{sign}(p(z)) - O(\varepsilon d^3) - |\mathop{\mathbb{E}}_{Z} \mathsf{sign}(p(Z))g(Z) - \mathop{\mathbb{E}}_{z\sim N(0,1)^n} \mathsf{sign}(p(z))g(z)|.$$

Thus, it suffices to bound $|\mathbb{E}_Z \mathsf{sign}(p(Z))g(Z) - \mathbb{E}_{z\sim N(0,1)^n} \mathsf{sign}(p(z))g(z)|$, which we do by a hybrid argument. We first represent z as $z := \frac{1}{\sqrt{L}} \sum_{i=1}^{L} y_i$ where each y_i is an independent standard gaussian. We can replace each Y_i with y_i and get

$$|\mathop{\mathbb{E}}_{Z} \mathsf{sign}(p(Z))g(Z) - \mathop{\mathbb{E}}_{y} \mathsf{sign}(p(y))g(y)| \leq 2^{-\Omega(R)} L,$$

as a consequence of the following lemma (restatement of Lemma 8) that we prove in the next section. Theorem 21 now follows. ◀

▶ **Lemma 22** (Main hybrid-step). *There exists a constant c such that the following holds for $\lambda \leq \varepsilon^2/Rd^c$. For any fixed vector $x \in \mathbb{R}^n$, Y a dR-moment-matching gaussian vector, and $y \sim N(0,1)^n$,*

$$|\mathop{\mathbb{E}}_{Y} \mathsf{sign}(p(x + \sqrt{\lambda}Y))g(x + \sqrt{\lambda}Y) - \mathop{\mathbb{E}}_{y} \mathsf{sign}(p(x + \sqrt{\lambda}y))g(x + \sqrt{\lambda}y)| \leq \gamma = 2^{-\Omega(R)}.$$

[11] For example, it suffices to let $\rho(t)$ be the standard mollifier $\rho(t) := 0$ for $t \leq 0$, $\rho(t) := 1$ for $t \geq 1$, and $\rho(t) := e \cdot \exp\left(\frac{1}{(t-1)^2-1}\right)$ for $t \in (0,1)$.

5.1 Analysis of the main hybrid-step

The proof of Lemma 22 is by a case-analysis as outlined in the introduction. Consider the setting as in the lemma and define

$$\phi(z) := U_{\sqrt{1-\lambda}}p\left(\frac{z}{\sqrt{1-\lambda}}\right).$$

The core argument will be a case-analysis on the derivatives of ϕ at the fixed point x and whether these are slow-growing. Note that if p were multi-linear, then we would simply have $\phi \equiv p$. The starting point is the following re-scaling of Lemma 16:

$$p\left(x+\sqrt{\lambda}y\right) = \sum_{|\alpha|\le d}\frac{\partial^\alpha\phi(x)}{\sqrt{\alpha!}}\lambda^{|\alpha|/2}h_\alpha(y). \tag{4}$$

Further, by a re-scaling of Theorem 17, we get the following identity which gives a nice nearly self-referential expression relating the derivatives of p to those of ϕ:

$$\partial^\alpha p\left(x+\sqrt{\lambda}y\right) = \sum_{\beta\ge\alpha}\sqrt{\frac{\alpha!}{\beta!}}\partial^\beta\phi(x)\lambda^{|\beta-\alpha|/2}h_{\beta-\alpha}(y). \tag{5}$$

Now, note that for a truly random gaussian y we have $\partial^\alpha\phi(x) = \mathbb{E}_y\,\partial^\alpha p(x+\sqrt{\lambda}y)$. Thus, it is reasonable to expect that for typical points x and small enough λ, $\partial^\alpha p(x+\sqrt{\lambda}y)$ will be strongly concentrated around $\partial^\alpha\phi(x)$. The following lemma gives quantitative bounds on how much the derivatives $\partial^\alpha p(x+\sqrt{\lambda}y)$ deviate from their expectations $\partial^\alpha\phi(x)$ for a random $y\sim N(0,1)^n$. As we will need such bounds even for k-moment-matching Y, we state the deviation bound in terms of moments:

▶ **Lemma 23.** *Suppose f is a degree-d polynomial, and let $\phi(z) = U_{\sqrt{1-\lambda}}f(\frac{z}{\sqrt{1-\lambda}})$. Consider the polynomial*

$$D(y) := \|\nabla^k f(x+\sqrt{\lambda}y) - \nabla^k\phi(x)\|^2,$$

which measures the euclidean distance between the k-th order derivatives $\nabla^k f(x+\sqrt{\lambda}y)$ and their expectations $\nabla^k\phi(x)$.

For $y\sim N(0,1)^n$, we have the moment bound

$$\|D(y)\|_{q/2} \le \sum_{t=k+1}^{d}(\lambda dq)^{t-k}\|\nabla^t\phi(x)\|^2.$$

That is,

$$\left(\mathop{\mathbb{E}}_{y\sim N(0,1)^n}\|\nabla^k f(x+\sqrt{\lambda}y) - \nabla^k\phi(x)\|^q\right)^{1/q} \le \sqrt{\sum_{t=k+1}^{d}(\lambda dq)^{t-k}\|\nabla^t\phi(x)\|^2}.$$

Proof. We express

$$D(y) = \sum_\alpha\left(\partial^\alpha f(x+\sqrt{\lambda}y) - \partial^\alpha\phi(x)\right)^2 = \sum_\alpha\left(\sum_{\beta>\alpha}\sqrt{\frac{\alpha!}{\beta!}}\partial^\beta\phi(x)\lambda^{|\beta-\alpha|/2}h_{\beta-\alpha}(y)\right)^2.$$

First, by triangle-inequality, we get

$$\|D(y)\|_{q/2} \leq \sum_{\alpha} \left\| \left(\sum_{\beta > \alpha} \sqrt{\frac{\alpha!}{\beta!}} \partial^{\beta} \phi(x) \lambda^{|\beta-\alpha|/2} h_{\beta-\alpha}(y) \right)^2 \right\|_{q/2}$$
$$= \sum_{\alpha} \left\| \sum_{\beta > \alpha} \sqrt{\frac{\alpha!}{\beta!}} \partial^{\beta} \phi(x) \lambda^{|\beta-\alpha|/2} h_{\beta-\alpha}(y) \right\|_{q}.$$

Applying hypercontractivity, we now get

$$\|D(y)\|_{q/2} \leq \sum_{\alpha} \left\| U_{\sqrt{q}} \sum_{\beta > \alpha} \sqrt{\frac{\alpha!}{\beta!}} \partial^{\beta} \phi(x) \lambda^{|\beta-\alpha|/2} h_{\beta-\alpha}(y) \right\|_{2}$$
$$= \sum_{\alpha} \sum_{\beta > \alpha} \frac{\alpha!}{\beta!} \partial^{\beta} \phi(x)^2 \lambda^{|\beta-\alpha|} q^{|\beta-\alpha|}$$
$$\leq \sum_{\alpha} \sum_{\beta > \alpha} \partial^{\beta} \phi(x)^2 \lambda^{|\beta-\alpha|} q^{|\beta-\alpha|}$$
$$= \sum_{t=k+1}^{d} \binom{t}{t-k} (\lambda q)^{t-k} \|\nabla^t \phi(x)\|^2$$
$$\leq \sum_{t=k+1}^{d} (\lambda d q)^{t-k} \|\nabla^t \phi(x)\|^2. \qquad \blacktriangleleft$$

We are now ready to prove Lemma 22.

Proof of Lemma 22. We study two cases:

1. x is poorly-behaved for ϕ. In this case, we will show that $g(x+\sqrt{\lambda}Y) = 0$ with probability at least $1 - 2^{-\Omega(R)}$.
2. x is well-behaved for ϕ: In this case, we will exploit the fact that $\mathsf{sign}(p(x+\sqrt{\lambda Y}))$ will equal $\mathsf{sign}(\phi(x))$ with probability $1 - 2^{-\Omega(R)}$. We then have to show that Y fools the mollifier g which is a bit technically involved (hence we deal with this case second unlike in Section 2.1).

We begin with the first case.

Case 1: x is poorly-behaved for ϕ. Consider the case where the inequality $\|\nabla^k \phi(x)\| \geq \varepsilon \|\nabla^{k+1}\phi(x)\|$ is violated for some k, and indeed let k_0 be the largest k such that this inequality is violated. We will argue that with probability at least $1 - 2^{-\Omega(R)}$, over random choice of Y, that

$$\|\nabla^{k_0} p(x+\sqrt{\lambda}Y)\| \leq 4\varepsilon \|\nabla^{k_0+1} p(x+\sqrt{\lambda}Y)\|,$$

in which case $g(x+\sqrt{\lambda}Y) = 0$.

More specifically, we will show that it is highly likely that both

- $\|\nabla^{k_0} p(x+\sqrt{\lambda}Y)\| \leq 2\varepsilon \|\nabla^{k_0+1}\phi(x)\|$, and
- $\|\nabla^{k_0+1} p(x+\sqrt{\lambda}Y)\| \geq \frac{1}{2}\|\nabla^{k_0+1}\phi(x)\|$.

For this, we will use Equation (5) and Lemma 23. Supposing k_0 is the largest k such that

$$\|\nabla^k \phi(x)\| < \varepsilon \|\nabla^{k+1}\phi(x)\|,$$

we have

- $\|\nabla^{k_0}\phi(x)\| \le \varepsilon\|\nabla^{k_0+1}\phi(x)\|$ and
- $\|\nabla^{k_0+1}\phi(x)\| \ge \varepsilon^t\|\nabla^{k_0+1+t}\phi(x)\|$ for all $t \ge 0$.

Lemma 23 therefore gives the bounds

$$\left(\mathop{\mathbb{E}}_{Y}\|\nabla^{k_0}p(x+\sqrt{\lambda}Y)-\nabla^{k_0}\phi(x)\|^R\right)^{1/R} \le \varepsilon\|\nabla^{k_0+1}\phi(x)\|\sqrt{\sum_{t\ge1}(\lambda dR/\varepsilon^2)^t}$$

and

$$\left(\mathop{\mathbb{E}}_{Y}\|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)-\nabla^{k_0+1}\phi(x)\|^R\right)^{1/R} \le \|\nabla^{k_0+1}\phi(x)\|\sqrt{\sum_{t\ge1}(\lambda dR/\varepsilon^2)^t}.$$

So, as long as $\lambda dR/\varepsilon^2$ is at most a sufficiently small constant, we conclude that the following bounds hold with probability at least $1-2^{-R}$:

- $\|\nabla^{k_0}p(x+\sqrt{\lambda}Y)\| \le \|\nabla^{k_0}\phi(x)\| + \|\nabla^{k_0}p(x+\sqrt{\lambda}Y)-\nabla^{k_0}\phi(x)\| \le 2\varepsilon\|\nabla^{k_0+1}\phi(x)\|$, and
- $\|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)\| \ge \|\nabla^{k_0+1}\phi(x)\| - \|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)-\nabla^{k_0+1}\phi(x)\| \ge \frac{1}{2}\|\nabla^{k_0+1}\phi(x)\|$.

In the case that these bounds hold, we get

$$\|\nabla^{k_0}p(x+\sqrt{\lambda}Y)\| \le 4\varepsilon\|\nabla^{k_0+1}p(x+\sqrt{\lambda}Y)\|,$$

and so $g(x+\sqrt{\lambda}Y)=0$. As this holds with probability at least $1-2^{-\Omega(R)}$ for both $y\sim N(0,1)^n$ as well as Y, the conclusion of Lemma 22 follows. This finishes the proof of Case 1. ⌟

Case 2: x is well-behaved for ϕ. We now consider the complimentary case where

$$\|\nabla^k\phi(x)\| \ge \varepsilon\|\nabla^{k+1}\phi(x)\|$$

for all $k=0,1,\dots,d-1$. Consider the normalized polynomial

$$f(y) := \frac{p(x+\sqrt{\lambda}y)}{\phi(x)} = 1+\frac{1}{\phi(x)}\sum_{\alpha\neq0}\partial^\alpha\phi(x)\lambda^{|\alpha|/2}h_\alpha(y).$$

Using hypercontractivity, we bound the R-th moment of $f(y)-1$ by its $\sqrt{R}$-hypervariance:

$$\|f(y)-1\|_R \le \|U_{\sqrt{R}}(f(y)-1)\|_2 \le \sqrt{\sum_{k\ge1}\left(\frac{\lambda R}{\varepsilon^2}\right)^k} \le \frac{1}{2}.$$

So, by a Markov argument, we have

$$\mathbb{P}\Big(\mathsf{sign}(p(x+\sqrt{\lambda}Y)) \neq \mathsf{sign}(\phi(x))\Big) \le 2^{-R},$$

and this holds whenever Y is k-moment-matching for $k \ge dR$. So, $\mathsf{sign}(p(x+\sqrt{\lambda}Y))$ is nearly a constant for random Y; it remains to show that Y fools $g(x+\sqrt{\lambda}Y)$. We do this by (essentially) truncating the Taylor-series of g about x so that we are left with a degree dR polynomial, which is fooled by Y. The truncation-error will be small because our assumption,

$$\|\nabla^k\phi(x)\| \ge \varepsilon\|\nabla^{k+1}\phi(x)\| \text{ for all } k,$$

gives us good control on the R-th order moments of the deviations $\|\nabla^k\phi(x)\| - \|\nabla^k p(x+\sqrt{\lambda}Y)\|$. The exact calculations are somewhat cumbersome and are given below. We will show that Y fools the mollifier function

$$g(x+\sqrt{\lambda}y) = \prod_{k=0}^{d-1}\rho\left(\log\left(\frac{1}{16\varepsilon^2}\frac{\|\nabla^k p(x+\sqrt{\lambda}y)\|^2}{\|\nabla^{k+1}p(x+\sqrt{\lambda}y)\|^2}\right)\right).$$

To simplify notation we define the shifted function $\sigma(t) := \rho(t - \log(16\varepsilon^2))$, and express

$$g(x + \sqrt{\lambda}y) = \prod_{k=0}^{d-1} \sigma\left(\log \|\nabla^k p(x + \sqrt{\lambda}y)\|^2 - \log \|\nabla^{k+1} p(x + \sqrt{\lambda}y)\|^2\right).$$

It will be convenient to think of g (redundantly) as function of $2d$ auxiliary variables $s_1 \dots s_d$, $t_1, \dots t_d$, which we will eventually fix to

- $s_i := \|\nabla^{i-1} p(x + \sqrt{\lambda}y)\|^2$
- $t_i := \|\nabla^i p(x + \sqrt{\lambda}y)\|^2$,

so we write

$$g(s,t) := \prod_{i=1}^{d} \sigma\left(\log(s_i) - \log(t_i)\right).$$

We Taylor-expand $g(s,t)$ around the points

- $a_i := \|\nabla^{i-1}\phi(x)\|^2$
- $b_i := \|\nabla^i \phi(x)\|^2$,

which gives

$$g(s,t) = \ell(s,t) + h(s,t),$$

with low-degree part

$$\ell(s,t) := \sum_{\substack{\alpha,\beta \in \mathbb{N}^d \\ |\alpha|+|\beta|<R}} \frac{\partial_s^\alpha \partial_t^\beta g(a,b)}{\alpha!\beta!} (s-a)^\alpha (t-b)^\beta$$

and remainder

$$|h(s,t)| \le \sum_{\substack{\alpha,\beta \in \mathbb{N}^d \\ |\alpha|+|\beta|=R}} \frac{|\partial_s^\alpha \partial_t^\beta g(s^*,t^*)|}{\alpha!\beta!} |s-a|^\alpha |t-b|^\beta,$$

where "$|\partial_s^\alpha \partial_t^\beta g(s^*,t^*)|$" is notation for the maximum magnitude of $\partial_s^\alpha \partial_t^\beta g$ on any point on the line segment from (a,b) to (s,t). We need the following fact to bound the size of the derivatives of g,

▷ Claim 24. Suppose σ is a smooth univariate function with uniform derivative bounds

$$\|\sigma^{(n)}\|_\infty \le n^{O(n)}.$$

The bivariate function

$$r(u,v) := \sigma(\log(u) - \log(v))$$

has derivatives bounded in size by

$$\left|\frac{\partial^n}{\partial u^n}\frac{\partial^m}{\partial v^m} r(u,v)\right| \le \frac{n^{O(n)}}{|u|^n}\frac{m^{O(m)}}{|v|^m}.$$

This claim follows easily from the generalized chain rule (Faà di Bruno's formula). As a result, we get the derivative bounds

$$\left|\partial_s^\alpha \partial_t^\beta g(s,t)\right| \le \frac{|\alpha|^{O(|\alpha|)}}{|s^\alpha|}\frac{|\beta|^{O(|\beta|)}}{|t^\beta|}.$$

Using this, we bound the remainder

$$|h(s,t)| \le \sum_{\substack{\alpha,\beta\in\mathbb{N}^d \\ |\alpha|+|\beta|=R}} d^{O(R)} \prod_{i=1}^{d} \left(\frac{|1-\frac{s_i}{a_i}|}{1-|1-\frac{s_i}{a_i}|}\right)^{\alpha_i} \left(\frac{|1-\frac{t_i}{b_i}|}{1-|1-\frac{t_i}{b_i}|}\right)^{\beta_i}.$$

Now, consider the event E (which depends on y) that

$$(1-\delta)\|\nabla^i\phi(x)\|^2 \le \|\nabla^i p(x+\sqrt{\lambda}y)\|^2 \le (1+\delta)\|\nabla^i\phi(x)\|^2$$

holds for all i, where $\delta \le 1/2$ is a parameter we will set shortly. In the case that this indeed holds, we get

$$|h(s,t)| \le d^{O(R)}O(\delta)^R.$$

We set δ just small enough to ensure

$$|h(s,t)| \le 2^{-R}.$$

Now, we express g (which we now think of as a function of the underlying variable y) as

$$\begin{aligned} g &= g\cdot \mathbb{1}_E + g\cdot \mathbb{1}_{\bar{E}} \\ &= \ell\cdot \mathbb{1}_E + h\cdot \mathbb{1}_E + g\cdot \mathbb{1}_{\bar{E}} \\ &= \ell - \ell\cdot \mathbb{1}_{\bar{E}} + h\cdot \mathbb{1}_E + g\cdot \mathbb{1}_{\bar{E}}, \end{aligned}$$

and we obtain the pointwise bound

$$|g-\ell| \le 2^{-R} + \mathbb{1}_{\bar{E}} + |\ell|\cdot \mathbb{1}_{\bar{E}}.$$

On average over Y, we get truncation error

$$\begin{aligned} \mathop{\mathbb{E}}_{Y}\left|g(x+\sqrt{\lambda}Y)-\ell(Y)\right| &\le 2^{-R} + \mathop{\mathbb{E}}_{Y}\mathbb{1}_{\bar{E}}(Y) + \sqrt{\mathop{\mathbb{E}}_{Y}\ell^2(Y)}\sqrt{\mathop{\mathbb{E}}_{Y}\mathbb{1}_{\bar{E}}(Y)} \\ &\le 2^{-R} + O\left(\frac{d}{\delta}\right)^R \cdot \left(\frac{\lambda dR}{\varepsilon^2}\right)^{-\Omega(R)} \\ &\le 2^{-R} + d^{O(1)}\cdot \left(\frac{\lambda dR}{\varepsilon^2}\right)^{-\Omega(R)} \end{aligned}$$

where the second inequality here follows from the moment bounds in Lemma 23. As required by the conditions of Lemma 22, we insist that λ is small enough that this error is at most $2^{-\Omega(R)}$. Since this bound holds also for truly-random standard gaussian y, and $\mathbb{E}_Y \ell(Y) = \mathbb{E}_y \ell(y)$, we obtain the desired bound

$$|\mathop{\mathbb{E}}_{Y} g(x+\sqrt{\lambda}Y) - \mathop{\mathbb{E}}_{y} g(x+\sqrt{\lambda}y)| \le 2^{-\Omega(R)}.$$

This finishes the proof in Case 2 and hence of Lemma 22. ◀

References

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

2 Ilias Diakonikolas, Daniel M Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 11–20. IEEE, 2010.

3 Ilias Diakonikolas, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Average sensitivity and noise sensitivity of polynomial threshold functions. *SIAM J. Comput.*, 43(1):231–253, 2014. doi:10.1137/110855223.

4 Prahladh Harsha, Adam Klivans, and Raghu Meka. Bounding the sensitivity of polynomial threshold functions. *Theory of Computing*, 10(1):1–26, 2014. doi:10.4086/toc.2014.v010a001.

5 Valentine Kabanets, Daniel M Kane, and Zhenjian Lu. A polynomial restriction lemma with applications. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 615–628, 2017.

6 Daniel M Kane. k-independent gaussians fool polynomial threshold functions. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 252–261. IEEE Computer Society, 2011.

7 Daniel M Kane. A small PRG for polynomial threshold functions of gaussians. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 257–266. IEEE, 2011.

8 Daniel M Kane. A structure theorem for poorly anticoncentrated gaussian chaoses and applications to the study of polynomial threshold functions. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 91–100. IEEE, 2012.

9 Daniel M Kane. The correct exponent for the Gotsman-Linial conjecture. In *2013 IEEE Conference on Computational Complexity*, pages 56–64. IEEE, 2013.

10 Daniel M Kane. A pseudorandom generator for polynomial threshold functions of gaussian with subpolynomial seed length. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 217–228. IEEE, 2014.

11 Daniel M Kane. A polylogarithmic PRG for degree 2 threshold functions in the gaussian setting. In *Proceedings of the 30th Conference on Computational Complexity*, pages 567–581, 2015.

12 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM Journal on Computing*, 42(3):1275–1301, 2013.

13 Ryan O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014.

14 Ryan O'Donnell, Rocco A Servedio, and Li-Yang Tan. Fooling gaussian PTFs via local hyperconcentration. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1170–1183, 2020.

15 Ryan O'Donnell, Rocco A. Servedio, Li-Yang Tan, and Daniel Kane. Fooling gaussian PTFs via local hyperconcentration, 2021. arXiv:2103.07809. arXiv:2103.07809.