arXiv:2103.07809v1 [cs.CC] 13 Mar 2021

Fooling Gaussian PTFs via Local Hyperconcentration*

Ryan O'Donnell
Carnegie Mellon University

Rocco A. Servedio
Columbia University

Li-Yang Tan
Stanford University

with an appendix by

Daniel Kane
University of California, San Diego

March 16, 2021

Abstract

We give a pseudorandom generator that fools degree-d polynomial threshold functions over n-dimensional Gaussian space with seed length $\text{poly}(d) \cdot \log n$. All previous generators had a seed length with at least a 2^d dependence on d.

The key new ingredient is a *Local Hyperconcentration Theorem*, which shows that every degree-d Gaussian polynomial is hyperconcentrated almost everywhere at scale $d^{-O(1)}$.

Contents

1 Introduction

This paper is about pseudorandom generators (PRGs) for polynomial threshold functions (PTFs) over Gaussian space. Let us explain what this means. Let $\mathscr{C}$ be a class of functions from $\mathbb{R}^n$ to $\mathbb{R}$. A distribution $\mathcal{G}$ over $\mathbb{R}^n$ is an ε-*PRG for $\mathscr{C}$ over Gaussian space* if for every function $f \in \mathscr{C}$,

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{z} \sim \mathcal{G}}[f(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[f(\boldsymbol{x})] \right| \leq \varepsilon,$$

where $\mathrm{N}(0,1)^n$ is the standard n-dimensional Gaussian distribution. We equivalently say that $\mathcal{G}$ ε-*fools $\mathscr{C}$ over Gaussian space*. If a draw $\boldsymbol{z} \sim \mathcal{G}$ can be deterministically generated from a source of s independent uniformly random bits, we say that the *seed length* of $\boldsymbol{z}$ is s. If furthermore the generation can be performed by a computationally efficient algorithm, we say the PRG is *explicit*.

A *degree-d polynomial threshold function* (PTF) is a function $f(x) = \mathrm{sign}(p(x))$ where $p : \mathbb{R}^n \to \mathbb{R}$ is a real polynomial of total degree at most d. Now we can state the main theorem of this paper:

Theorem 1. *For all $n, d \in \mathbb{N}$ and $\varepsilon \in (0,1)$, there is an explicit PRG with seed length $\mathrm{poly}(d/\varepsilon) \cdot \log n$ that ε-fools the class of all degree-d PTFs over n-dimensional Gaussian space.*

The polynomial dependence on d here is a substantial improvement over previous PRGs, all of which had at least $2^{O(d)}$ dependence or worse. We view this as notable, as there are few prior works concerning structural properties of n-dimensional Gaussian or Boolean PTFs that are nontrivial for $d \gg \log n$.

1.1 Prior work

There has been significant work on PRGs for PTFs. Their study was initiated by Meka and Zuckerman [MZ10, MZ13], who gave a PRG with seed length[1] $(d/\varepsilon)^{O(d)} \cdot \log n$ that fools degree-d PTFs over the more general setting of *Boolean space*, $\{\pm 1\}^n$. PRGs over Boolean space can be shown to also yield PRGs over Gaussian space, thanks to the fact that $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_m$ has a nearly Gaussian distribution when $\boldsymbol{x} \sim \{\pm 1\}^m$ is uniformly random (see the discussion in Section 3.1), and the fact that degree-d PTFs are closed under taking linear combinations of inputs. Since the work of [MZ10, MZ13], there have been several works that focus just on fooling PTFs over Gaussian space, which we now discuss.

First, Kane [Kan11a] showed that limited independence (see Definition 11) suffices to fool Gaussian PTFs. The amount of independence required was $O_d(\varepsilon^{-2^{O(d)}})$, which translates into $O_d(\varepsilon^{-2^{O(d)}}) \cdot \log n$ in seed length. Using a different generator (one that is not based only on limited independence), Kane [Kan11b] then gave a PRG for Gaussian PTFs with seed length $2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon) \cdot \log n$. Note that this seed length strictly improves upon that in [MZ13], albeit only in the Gaussian setting.

Towards further improving the seed length dependence on ε, Kane [Kan12] gave a PRG with seed length $A(d, 1/c) \cdot (1/\varepsilon)^{2+c} \cdot \log n$ for any $c > 0$, where $A(\cdot, \cdot)$ is a variant of the Ackermann function.[2] This was improved to $A(d, 1/c) \cdot (1/\varepsilon)^c \cdot \log n$ in [Kan14]; while the seed length now has subpolynomial dependence on $1/\varepsilon$, its dependence on d limits its applicability to PTFs of constant (or very slightly superconstant) degree.

---

[1] They state $O(1/\varepsilon)^{O(d)} \cdot \log n$ just after [MZ13, Thm. 5.18], but they have appear to have dropped a factor of d when citing their Thm. 5.2 at the end of Lem. 5.20's proof. Correcting this leads to the seed length $(d/\varepsilon)^{O(d)} \cdot \log n$.

[2] In fact, it seems that correcting a typo in [Kan11a, Proof of Prop. 12], where a "θ" factor should be "θ^2", already leads to seed length $2^{O_c(d)} \cdot (1/\varepsilon)^{2+c} \cdot \log n$.

For degree-2 PTFs, Kane gives a PRG with seed length $O(\log^6(1/\varepsilon) \log \log(n/\varepsilon) \log n)$ [Kan15]; Diakonikolas, Kane, and Nelson [DKN10] showed that $\tilde{O}(1/\varepsilon^9)$-wise independence suffices to fool degree-2 PTFs over both Boolean and Gaussian space. For degree-1 PTFs (i.e. halfspaces), the current best PRG is due to Kothari and Meka [KM15], who achieve a near-optimal seed length of $O(\log(1/\varepsilon) \log \log(1/\varepsilon) + \log n)$.

Summarizing the prior state of the art, previous PRGs were either specific to $d = 1, 2$, or else had seed length with at least an exponential dependence on d. Consequently, there were no PRGs that could fool PTFs of degree $d = \log n$, even just to constant accuracy ε. Theorem 1 therefore represents the first PRG that is able to fool PTFs of degree $d \geq \log n$; our seed length remains nontrivial for d as large as $n^{\Omega(1)}$. Please see Table 1.

Reference	Seed length	Allowable / nontrivial range of d's
[DKN10]	$\tilde{O}(1/\varepsilon^9) \cdot \log n$	$d \leq 2$
[MZ13, MZ10]	$(d/\varepsilon)^{O(d)} \cdot \log n$	$d \leq O(\log n / \log \log n)$
[Kan11a]	$O_d\big(\varepsilon^{-2^{O(d)}}\big) \cdot \log n$	$d \leq$ slightly superconstant
[Kan11b]	$2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon) \cdot \log n$	$d \leq O(\log n)$
[Kan12]	$A(d, \frac{1}{c}) \cdot (1/\varepsilon)^{2+c} \cdot \log n$ for any $c > 0$	$d \leq$ slightly superconstant
[Kan14]	$A(d, \frac{1}{c}) \cdot (1/\varepsilon)^{c} \cdot \log n$ for any $c > 0$	$d \leq$ slightly superconstant
[Kan15]	$O(\log^6(1/\varepsilon) \log \log(n/\varepsilon) \log n)$	$d \leq 2$
[KM15]	$O(\log(1/\varepsilon) \log \log(1/\varepsilon) + \log n)$	$d = 1$
This work	$\mathrm{poly}(d/\varepsilon) \cdot \log n$	$d \leq n^{\Omega(1)}$

Table 1: Our work and the prior results on fooling degree-d Gaussian PTFs. The last column indicates the range of values of d's for which the seed length of the corresponding PRG is nontrivial (i.e. $o(n)$). The generators of [MZ13, DKN10] work for the more general setting of Boolean space, and [DKN10, Kan11a]'s analyses show that limited independence suffices.

1.2 Motivations

Geometric content. We now give a geometric perspective on the problem of constructing PRGs for Gaussian PTFs. Suppose one is given a set $F \subseteq \mathbb{R}^n$ and one wishes to approximately compute its Gaussian volume, $\mathbf{Pr}_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[\boldsymbol{x} \in F]$. There is an obvious Monte Carlo approach: picking $O(1/\varepsilon^2)$ Gaussian vectors $\boldsymbol{x}$ at random and outputting the fraction that fall into F will, with high probability, give an ε-accurate estimate. Our question is to what extent randomness is necessary for this problem.

The extent to which derandomization is possible depends on the "complexity" of the sets F we allow. If F is only given via a black-box membership oracle then no derandomization is possible. So we need to assume an "explicit description" of F is given, and in this paper we focus on the case that F is the set of points satisfying a polynomial inequality of degree at most d (i.e., F is
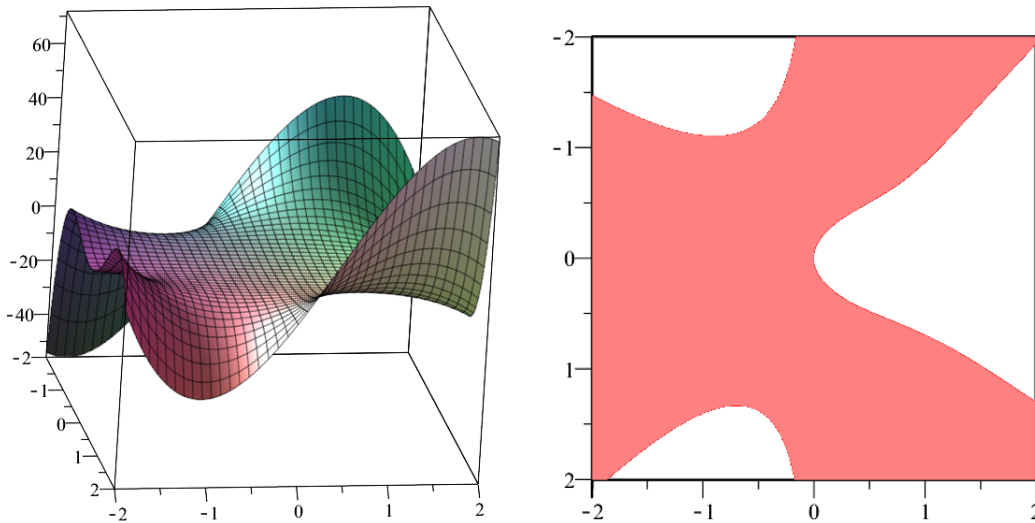
Figure 1: On the left, a plot of the degree-5 polynomial $p(x, y) = x^2 y^3 - 2x^3 y^2 + 3xy^4 - x + 2y^2$. On the right, the threshold set $F = \{(x, y) \in \mathbb{R}^2 : p(x, y) \geq 0\}$.

the set indicated by a degree-d PTF). Thus the $d = 1$ case allows halfspaces, the $d = 2$ case allows ellipsoids and hyperboloids, etc. For illustrative purposes, Figure 1 shows an example with $n = 2$ and $d = 5$, although we generally think of $n \gg d$.

One natural approach to this volume-approximation problem is the following: First, define some kind of explicit (nonrandom) finite "grid" of discrete points in $\mathbb{R}^n$; second, show that the Gaussian volume of any degree-d PTF set F is closely approximated by the fraction of grid points in F. A naive gridding scheme would use at least an exponential-in-n number of grid points (even for $d = O(1)$); the question is whether we can use a subexponential-in-n number of gridpoints, when $d \ll n$. Our Theorem 1 provides such a solution; by enumerating all seeds (essentially, taking the support of $\mathcal{G}$), we get an explicit set of just $n^{\text{poly}(d)}$ "grid points" that gives a high-quality volume approximation for any degree-d polynomial threshold set; this is nontrivial for d up to some $n^{\Omega(1)}$. Also note that this kind of "PRG solution" is stronger than just being an "volume-approximation" algorithm of the type "given F, approximate vol(F)"; as it is PRG-based, it gives *one* fixed, deterministic "grid" that simultaneously works to approximate the volume of all degree-d polynomial threshold sets F.

Boolean complexity theory. As mentioned earlier, the problem of PRGs (or deterministic volume approximation) for Gaussian polynomial threshold functions is a special case of the problem of PRGs (or approximate-counting) for *Boolean* polynomial threshold functions. This, in turn, is a very special case of the problem of derandomization for general Boolean circuits. Recall that the BPP vs. P problem is roughly equivalent to asking whether there is a deterministic polynomial-time algorithm that, given the explicit description of a subset $F \subseteq \{0, 1\}^n$ in the form of a poly(n)-gate Boolean circuit C computing the indicator function of F, computes a 0.1-accurate approximation to its "volume", $\mathbf{Pr}_{\boldsymbol{x} \sim \{0,1\}^n}[C(\boldsymbol{x}) = 1]$. Given how far we are from answering this question, the field of pseudorandomness has focused on special classes of circuits, of restricted depth and gate-types; the case of Boolean PTFs corresponds to depth-2 circuits C with a threshold gate on top and AND gates of width at most d at the bottom.

1.3 Our key new tool: the Local Hyperconcentration Theorem

For large d, the best prior PRG for degree-d Gaussian PTFs is Kane's [Kan11b], which has seed length $2^{O(d)}/\text{poly}(\varepsilon) \cdot \log n$. In this section we describe the most important new ingredient we introduce to Kane's framework, which lets us reduce the seed length's dependence on d down to $\text{poly}(d)$. In the next section we will give an overview of the constructions of [MZ10, MZ13, Kan11b], putting our new tool into context.

We call our main new tool the *Local Hyperconcentration Theorem*. To explain it, suppose $p : \mathbb{R}^n \to \mathbb{R}$ is a degree-d polynomial. Since p has high degree, it might fluctuate quite wildly near a given point $x \in \mathbb{R}^n$, causing $\text{sign}(p(x))$ to rapidly switch between ± 1 in small neighborhoods. However, we might hope that for *most* points x, the value of p in a local neighborhood of x is almost always within a $1 \pm \delta$ multiplicative factor of $p(x)$, and hence is almost always of constant sign.

The right definition of a "local neighborhood of x" is to choose a small scale parameter $\lambda > 0$, and then to consider a Gaussian $\widetilde{\boldsymbol{x}}$, centered at $\sqrt{1 - \lambda}x$, with variance λ in each coordinate.[3] Now if $\mathbf{Var}[p(\widetilde{\boldsymbol{x}})] \ll \mathbf{E}[p(\widetilde{\boldsymbol{x}})]^2$, we may say that p is *(multiplicatively) concentrated* in this λ-local neighborhood of x; and indeed, the second moment method (Chebyshev's inequality) tells us that $p(\widetilde{\boldsymbol{x}})$ almost always has the same sign (namely, the sign of $\mathbf{E}[p(\widetilde{\boldsymbol{x}})]$). The most important ingredient in Kane's work, [Kan11b, Cor. 10+Lem. 11], establishes this sort of result:

Theorem 2 (The key technical theorem of [Kan11b], simplified)**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial. Provided $\lambda \le 2^{-O(d)}$, with high probability over $\boldsymbol{x} \sim \mathrm{N}(0, 1)^n$ we have*

$$\mathbf{Var}[p(\widetilde{\boldsymbol{x}})] \ll \mathbf{E}[p(\widetilde{\boldsymbol{x}})]^2, \quad \text{where } \widetilde{\boldsymbol{x}} \sim \sqrt{1 - \lambda}\boldsymbol{x} + \mathrm{N}(0, \lambda)^n. \tag{1}$$

We may say that Kane shows degree-d polynomials have *local concentration at scale $\lambda = 2^{-O(d)}$*, almost everywhere. The value $L = 1/\lambda = 2^{O(d)}$ ends up becoming the dominant factor in Kane's PRG's seed length. At a high level, this is because the PRG has the form $\boldsymbol{z} = \boldsymbol{w}_1 + \boldsymbol{w}_2 + \cdots + \boldsymbol{w}_L$, where the $\boldsymbol{w}_i$'s are independent random vectors with $O(d)$-wise-independent $\mathrm{N}(0, \lambda)^n$ distributions.

By way of contrast, our new Local Hyperconcentration Theorem (stated in simplified form below) shows *local **hyper**concentration at scale $\lambda = d^{-O(1)}$*. For a high level sketch of the proof, see Section 5.1.

Theorem 3 (Simplified Local Hyperconcentration Theorem, see Theorem 47 and Theorem 85)**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial. Provided $\lambda \le d^{-O(1)}$, with high probability over $\boldsymbol{x} \sim \mathrm{N}(0, 1)^n$ we have*

$$\mathbf{HyperVar}_R[p(\widetilde{\boldsymbol{x}})] \ll \mathbf{E}[p(\widetilde{\boldsymbol{x}})]^2, \quad \text{where } \widetilde{\boldsymbol{x}} \sim \sqrt{1 - \lambda}\boldsymbol{x} + \mathrm{N}(0, \lambda)^n, \tag{2}$$

for any large constant R (indeed, for any $R \le \text{poly}(d)$).

Remark 4. The conference version of this paper [OST20] proved a quantitatively weaker version of the Local Hyperconcentration Theorem, showing local hyperconcentration at scale $\lambda = d^{-O(\log d)}$. This led to a seed length of $(d/\varepsilon)^{O(\log d)} \cdot \log n$. Subsequently, Kane improved the Local Hyperconcentration Theorem to show local hyperconcentration at scale $\lambda = d^{-O(1)}$, which yields the current seed length of $\text{poly}(d/\varepsilon) \cdot \log n$. Kane's proof is given in Appendix B and subsumes Section 5 of this paper.

---

[3]The $\sqrt{1 - \lambda}$ factor is included so that when we look at a *typical* $\boldsymbol{x}$ chosen from $\mathrm{N}(0, 1)^n$, the resulting "random point in the neighborhood" $\widetilde{\boldsymbol{x}}$ also has distribution $\mathrm{N}(0, 1)^n$.

4

We will define "hypervariance" $\mathbf{HyperVar}_R[\cdot]$ later (see Definition 27); here we only note that it is a stronger notion than variance, in the sense that $\mathbf{HyperVar}_R[p]$ is always at least as large as $\mathbf{Var}[p]$ for all $R \geq 1$. Whenever the theorem's conclusion holds for an outcome x of $\boldsymbol{x}$, the value of p in the λ-local neighborhood of x is "hyperconcentrated" (see Lemma 31), meaning that for any large constant q,

$$\mathbf{E}[|p(\widetilde{\boldsymbol{x}}) - \mu|^q] \ll |\mu|^q, \quad \text{where } \mu = \mathbf{E}[p(\widetilde{\boldsymbol{x}})].$$

The case $q = 2$ here is precisely the "concentration" conclusion in the theorem of [Kan11b]. Our hyperconcentration is a stronger conclusion: e.g., taking $q = 4$ lets us use the "fourth moment method", and in fact we'll eventually use $q = 8$.

To summarize, our theorem has two important improvements over [Kan11b]. First, it shows concentration at a much larger scale, $\lambda = d^{-O(1)}$, rather than $2^{-O(d)}$. This crucially gives us the *potential* to get our seed's dependence on d to be $1/\lambda = \text{poly}(d)$. This is far from automatic, though, because there are several other places in the [Kan11b] construction that "lose" a factor of $2^{O(d)}$. In all but one of these cases[4], it's because in [Kan11b] the variance bound Inequality (1) is bootstrapped using the *hypercontractivity inequality* in order to get control over p's behavior in various local neighborhoods. This hypercontractive inequality for degree-d polynomials inherently loses $2^{O(d)}$ factors (see Theorem 20). By contrast, since our theorem already establishes the stronger *hyperconcentration* conclusion Inequality (2) (this is the second key improvement, bounding hypervariance rather than variance), we are able to provide argumentation that eliminates all of these $2^{O(d)}$ factors.

1.4 Overview of the PRG framework we use

We use the same PRG for Gaussian PTFs as in the prior works of Meka–Zuckerman PRG [MZ10, MZ13] and Kane [Kan11a], namely

$$\boldsymbol{z} = \sqrt{\lambda}\boldsymbol{z}_1 + \sqrt{\lambda}\boldsymbol{z}_2 + \cdots + \sqrt{\lambda}\boldsymbol{z}_L, \tag{3}$$

where the key parameter λ is a small function of d and ε, where $L = 1/\lambda$, and where $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L$ are independent random vectors, each having an $O(d)$-wise independent n-dimensional Gaussian distribution. This leads to a seed length of essentially $O(d^2 L \cdot \log n)$ (see Theorem 10), and hence all the effort goes into finding the *largest* $\lambda = \lambda(d, \varepsilon)$ such that Equation (3) ε-fools degree-d Gaussian PTFs.

Here we review the Meka–Zuckerman and Kane works; our own analysis is heavily based on Kane's framework.

Meka–Zuckerman. The work of Meka and Zuckerman [MZ10] gave PRGs for degree-d Gaussian PTFs with seed length $(d/\varepsilon)^{O(d)} \cdot \log n$. In fact, they also extended their results to Boolean PTFs, but we do not review that extension here. At a high level, their construction followed a basic two-part paradigm used both in the proof of Central Limit Theorems and in PRG construction: *mollification + local low-degree behavior*. To explain this, recall that we are trying to design a PRG $\mathcal{G}$ with

$$\left| \underset{\boldsymbol{z} \sim \mathcal{G}}{\mathbf{E}}[f(\boldsymbol{z})] - \underset{\boldsymbol{x} \sim \text{N}(0,1)^n}{\mathbf{E}}[f(\boldsymbol{x})] \right| \leq \varepsilon,$$

where $f = \text{sign}(p)$, with p a degree-d polynomial. Suppose first that we did not have the discontinuous "sign" function, but rather we just wanted the above inequality for $f = p$. In that

---

[4]Namely, our "noise insensitivity extension lemma" Lemma 72, where we eliminate a factor of $2^{O(d)}$ from the analogous result of Kane [Kan11b, Cor. 16].

case, it would suffice for the components of the random vector $\boldsymbol{z} \sim \mathcal{G}$ to be "d-wise independent," and in fact this would achieve $\varepsilon = 0$. Furthermore, there are standard techniques to produce an appropriate "d-wise independent" $\mathcal{G}$ with seed length $\mathrm{poly}(d) \cdot \log n$, which would be an excellent bound for us.

Of course, when we return to the actual scenario of $f = \mathrm{sign}(p)$, the function f is not even a polynomial, let alone a low-degree one. The *mollification* portion of Meka and Zuckerman's work is to replace the sign function with a smooth approximator Φ, which is equal to sign outside some interval $[-\lambda, \lambda]$. Because the sign function is scale-invariant ($\mathrm{sign}(ty) = \mathrm{sign}(y)$ for $t > 0$), we may normalize p so that its variance $\mathbf{Var}[p(\boldsymbol{x})]$ is 1. Then one chooses the parameter $\lambda = \Theta(\varepsilon/d)^d$. The smooth mollifier Φ will have derivatives of all orders, with Tth derivative $\Phi^{(T)}$ bounded in magnitude by $O(1/\lambda^T)$. The replacement of sign by Φ leads to a *mollication error* of $O(\varepsilon)$, essentially due to the well-known anticoncentration bound for degree-d Gaussian polynomials due to Carbery and Wright [CW01]: $\mathbf{Pr}[|p(\boldsymbol{x})| \leq (\varepsilon/d)^d] \leq O(\varepsilon)$. (Note also that thanks to a trick, this only needs to hold for $\boldsymbol{x}$, and not the pseudorandom $\boldsymbol{z}$.) With the mollifier in place, Meka and Zuckerman can try to bound

$$\left| \mathbf{E}_{\boldsymbol{z} \sim \mathcal{G}}[\Phi(p(\boldsymbol{z}))] - \mathbf{E}_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[\Phi(f(\boldsymbol{x}))] \right| \leq O(\varepsilon).$$

Now although Φ is not a polynomial, it is "locally a low-degree polynomial" (say, of degree 4), thanks to Taylor's theorem. The error in this statement scales like the 4th derivative bound $\|\Phi^{(4)}\|_\infty \leq \mathrm{poly}(1/\lambda)$, times the "locality scale". Thus as long as we substitute $O(1)$-wise independent Gaussians for true Gaussians at a "scale" of $\lambda^{\Theta(1)}$, we will not incur more than $O(\varepsilon)$ error. This sort of argumentation allows Meka and Zuckerman to show that the PRG in Equation (3) ε-fools degree-d PTFs with $\lambda = \Theta(\varepsilon/d)^d$, which leads to their seed length of $(d/\varepsilon)^{O(d)} \cdot \log n$.

Kane. To repeat, our PRG analysis closely follows the structure of Kane's, which we now describe. Kane [Kan11b] shows that the PRG in Equation (3) succeeds with the improved (larger) value of $\lambda = 2^{-O(d)} \cdot \mathrm{poly}(\varepsilon)$, leading to his seed length of $2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon) \cdot \log n$. His "local concentration theorem" (Theorem 2) plays a central role in this, but he still needs to develop a complex framework (which we also employ) in order to complete the analysis.

Kane's Theorem 2 allows him to begin a new strategy for designing $\mathcal{G}$; rather than mollifying the sign function and taking $p(\boldsymbol{x})$ as a "black box" random variable, Kane instead mollifies the polynomial p itself. Roughly speaking, Kane's strategy begins by replacing p with $p \cdot \mathrm{Check}_1$, where $\mathrm{Check}_1(\widetilde{x})$ is a smoothed indicator function for the event that Inequality (1) holds at $\widetilde{x}$. The "with high probability over $\boldsymbol{x}$" in Kane's Theorem 2 is in fact probability $1 - \varepsilon$ provided $\lambda \leq 2^{-O(d)} \cdot \mathrm{poly}(\varepsilon)$, and this implies that the replacement of p by $p \cdot \mathrm{Check}_1$ only incurs error ε. Now we may hope that the construction from Equation (3) will work; roughly, this requires that in a λ-scale neighborhood of *every* point x, say $\widetilde{\boldsymbol{x}} = \sqrt{1-\lambda}x + \sqrt{\lambda}\boldsymbol{w}$, the function $\mathrm{sign}(p) \cdot \mathrm{Check}_1$ is essentially determined by low-degree moments of $\boldsymbol{w}$. There are two cases. If x is well into the region where $\mathrm{Check}_1(\widetilde{x})$ is 0, then $p \cdot \mathrm{Check}_1$ is essentially 0 and $\mathrm{sign}(p) \cdot \mathrm{Check}_1$ is essentially constant. Otherwise, if x is near the region where $\mathrm{Check}_1(\widetilde{x})$ is 1, then by definition $\mathbf{Var}[p(\widetilde{x})]$ is very small. Thus p is not varying very much in a neighborhood of x, and Taylor's theorem will tell us that low-degree moments suffice to essentially determine $p \cdot \mathrm{Check}_1$ in this neighborhood of x.

There are two catches here. First, the use of Taylor's theorem out to, say, degree 4 forces one to bound not just the expected *squared* deviation of p from $|p(x)|$ in the λ-neighborhood of $\widetilde{x}$; it requires one to control, say, the 4*th-power* deviation. This is where Kane uses the standard hypercontractivity-based fact that higher-power deviations can be controlled by the 2nd-power deviation (i.e., $\mathbf{Var}[p(\widetilde{\boldsymbol{x}})]$) at the expense of $2^{O(d)}$ losses. Kane is losing such factor anyway, since

6

he takes $\lambda = 2^{-O(d)} \cdot (1/\varepsilon)$. (This is one place where our analysis takes advantage of the local **hyper**concentration we prove in Theorem 3.)

The second catch is that Taylor's theorem needs to be applied not just to p but to Check_1 itself. Now Check_1 is concerned with the variance of p in a $\sqrt{\lambda}$-neighborhood of x. In order to control the Taylor error here, one needs to control the *variance of the variance*! Kane handles this by further mollifying p. He uses a *generalization* of Theorem 2 to show that at most points x, the variance of the variance in the neighborhood of x is small. (We must prove a similar generalization of our Local Hyperconcentration Theorem; see Theorem 49.) Thus p can be further mollified to $p \cdot \mathrm{Check}_1 \cdot \mathrm{Check}_2$ at only small loss. Now we have three cases to consider when analyzing $p \cdot \mathrm{Check}_1 \cdot \mathrm{Check}_2$; if x is well into the region where Check_2 is 0, then the mollified function is essentially 0 on the λ-neighborhood. Else, the variance of the variance of p in the neighborhood is suitably small. Next, if x is well into the region where Check_1 is 0, then the mollified function is again essentially 0 on the neighborhood; otherwise, the variance of p in the neighborhood is suitably small. In this third case, we are again in good shape to apply Taylor to p, and Check_1... but to handle Taylor error for Check_2, we need to introduce *another* check that the variance of the variance of the variance is small. Indeed, Kane's final mollifier needs not only this "descending" sequence of checks (that we will picture "vertically"), but for technical reasons needs additional "horizontally proliferating" checks (which, to avoid further lengthening this description, we will not discuss here).

Luckily, all of these proliferating checks eventually "bottom out". The vertically descending checks bottom out because the "i-fold variance" is a polynomial of degree $d - i$, and hence the $(d + 1)$-fold variance is constantly 0. The horizontally proliferating checks may eventually be terminated due to the fact that a degree-d polynomial is determined by its values at $d + 1$ points. (Actually, one needs a quantitative version of this fact. Kane provides one involving another factor of $2^{O(d)}$; we eliminate this factor in Lemma 72.)

Ultimately, Kane's mollifier multiplies p by $\mathrm{poly}(d)$ "$\mathrm{Check}_{i,j}$" functions: one needs a generalization of Theorem 2 and another theorem to show that the mollification is close to p at almost all points; and, when using Taylor's theorem at x, one needs a $\mathrm{poly}(d)$-case analysis looking at the "deepest" check (if any) that "fails." If any check "fails," then the mollified function is essentially 0; otherwise, if they all pass, then in the λ-neighborhood of z, the variance of p, and the variance of the variance, and the variance of the variance of the variance, etc., are all suitably small for use in Taylor's theorem.

2 The high-level structure of our proof

Throughout this paper $p : \mathbb{R}^n \to \mathbb{R}$ is a nonzero polynomial of degree at most d, and we are interested in the degree-d polynomial threshold function $f(x) = \mathrm{sign}(p(x))$. For a given $0 < \varepsilon_{\mathrm{PRG}} < 1$, we determine a small value

$$\underline{\lambda} := \left(\frac{\varepsilon_{\mathrm{PRG}}}{d} \right)^{O(1)} \tag{4}$$

and we also let

$$L := 1/\underline{\lambda}, \qquad k_{\mathrm{indep}} := \Theta(d). \tag{5}$$

Our main goal is:

Theorem 5 (Main result: sum of k_{indep}-wise independent Gaussians fools degree-d PTFs)**.** *Let $\boldsymbol{x} \sim$ $\mathrm{N}(0, 1)^n$ be a standard n-dimensional Gaussian random vector, and let $\boldsymbol{z}_1, \dots, \boldsymbol{z}_L$ be independent*

k_{indep}-*wise independent n-dimensional Gaussian random vectors. Write*

$$\boldsymbol{Z} := \sqrt{\underline{\lambda}}\boldsymbol{z}_1 + \sqrt{\underline{\lambda}}\boldsymbol{z}_2 + \cdots + \sqrt{\underline{\lambda}}\boldsymbol{z}_L.$$

Then

$$|\mathbf{E}[\mathrm{sign}(f(\boldsymbol{Z}))] - \mathbf{E}[\mathrm{sign}(f(\boldsymbol{x}))]| \le \varepsilon_{\mathrm{PRG}}.$$

To prove Theorem 5, we will construct a certain function

$$\mathrm{Mollifier}_p : \mathbb{R}^n \to [0, 1],$$

which is a smoothed indicator function for a collection of events (related to local hyperconcentration of p) that are expected to almost always occur. We then show the following:

Theorem 6 (Mollification error theorem, analogue of Lemma 17 of [Kan11b])**.**

$$\Pr_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[\mathrm{Mollifier}_p(\boldsymbol{x}) \ne 1] \le \varepsilon_{\mathrm{PRG}}/4.$$

We then extend the mollifier to take into account the sign of p:

Definition 7. Define $I_+ : \mathbb{R}^n \to [0,1]$ by

$$I_+(x) = \mathrm{Mollifier}_p(x) \cdot \mathbf{1}[\mathrm{sign}(p(x)) = 1],$$

and define I_- similarly as $I_-(x) = \mathrm{Mollifier}_p(x) \cdot \mathbf{1}[\mathrm{sign}(p(x)) = -1]$.

The main thing we prove about $I_\pm$ is the following:

Theorem 8 (One step of the Replacement Method, analogue of Lemma 19 of [Kan11b])**.** *Fix any $x \in \mathbb{R}^n$, and assume the $\mathbb{R}^n$-valued random vectors $\boldsymbol{z}, \boldsymbol{z}'$ are each k_{indep}-wise independent n-dimensional Gaussian vectors. Then we have*

$$\left| \mathbf{E}\left[I_+\left(\sqrt{1 - \underline{\lambda}}x + \sqrt{\underline{\lambda}}\boldsymbol{z} \right) \right] - \mathbf{E}\left[I_+\left(\sqrt{1 - \underline{\lambda}}x + \sqrt{\underline{\lambda}}\boldsymbol{z}' \right) \right] \right| \le \frac{\varepsilon_{\mathrm{PRG}}}{4L}.$$

The analogous statement for I_- also holds.

From this, a "Replacement Method" argument easily yields the following:

Corollary 9. *For $\boldsymbol{Z}$ as in Theorem 5 and $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ we have*

$$|\mathbf{E}[I_+(\boldsymbol{Z})] - \mathbf{E}[I_+(\boldsymbol{x})]| \le \varepsilon_{\mathrm{PRG}}/4,$$

and similarly for I_-

Proof. We may view $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ as

$$\boldsymbol{x} = \sqrt{\underline{\lambda}}\boldsymbol{x}_1 + \sqrt{\underline{\lambda}}\boldsymbol{x}_2 + \cdots + \sqrt{\underline{\lambda}}\boldsymbol{x}_L,$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L \sim \mathrm{N}(0,1)^n$ are independent. For $0 \le t \le L$, write

$$\boldsymbol{w}_t = \sqrt{\underline{\lambda}}\boldsymbol{x}_1 + \cdots + \sqrt{\underline{\lambda}}\boldsymbol{x}_t + \sqrt{\underline{\lambda}}\boldsymbol{z}_{t+1} + \cdots + \sqrt{\underline{\lambda}}\boldsymbol{z}_L,$$

so $\boldsymbol{w}_0 = \boldsymbol{Z}$ and $\boldsymbol{w}_L = \boldsymbol{x}$. Thus by telescoping,

$$|\mathbf{E}[I_+(\boldsymbol{Z}))] - \mathbf{E}[I_+(\boldsymbol{x})]| \leq \sum_{t=1}^{L} |\mathbf{E}[I_+(\boldsymbol{w}_{t-1}))] - \mathbf{E}[I_+(\boldsymbol{w}_t))]|. \tag{6}$$

For a fixed $1 \leq t \leq L$, if we write

$$\boldsymbol{v} = \sqrt{\tfrac{\lambda}{1-\lambda}}\boldsymbol{x}_1 + \cdots + \sqrt{\tfrac{\lambda}{1-\lambda}}\boldsymbol{x}_{t-1} + \sqrt{\tfrac{\lambda}{1-\lambda}}\boldsymbol{z}_{t+1} + \cdots + \sqrt{\tfrac{\lambda}{1-\lambda}}\boldsymbol{z}_L,$$

then

$$\left|\mathbf{E}[I_+(\boldsymbol{w}_{t-1}))] - \mathbf{E}[I_+(\boldsymbol{w}_t))]\right| = \left|\mathbf{E}\left[I_+\left(\sqrt{1-\lambda} \cdot \boldsymbol{v} + \sqrt{\lambda}\boldsymbol{z}_t\right)\right] - \mathbf{E}\left[I_+\left(\sqrt{1-\lambda} \cdot \boldsymbol{v} + \sqrt{\lambda}\boldsymbol{x}_t\right)\right]\right|. \tag{7}$$

Since $\boldsymbol{z}_t$ and $\boldsymbol{x}_t$ are each k_{indep}-wise independent n-dimensional Gaussian vectors, Theorem 8 implies that $(7) \leq \varepsilon_{\text{PRG}}/(4L)$. Putting this into Inequality (6) completes the proof. $\square$

With the above ingredients in place, Theorem 5 follows almost immediately:

Proof of Theorem 5. Since $\text{sign}(p) \leq 1 - 2I_-$ pointwise,

$$\mathbf{E}[\text{sign}(p(\boldsymbol{Z}))] \leq \mathbf{E}[1 - 2I_-(\boldsymbol{Z})] \leq \mathbf{E}[1 - 2I_-(\boldsymbol{x})] + \varepsilon_{\text{PRG}}/2 \leq \mathbf{E}[\text{sign}(p(\boldsymbol{x}))] + \varepsilon_{\text{PRG}},$$

where the second inequality is thanks to Corollary 9 and the third is thanks to Theorem 6. The reverse direction, which lower bounds $\mathbf{E}[\text{sign}(p(\boldsymbol{Z}))]$ by $\mathbf{E}[\text{sign}(p(\boldsymbol{x}))] - \varepsilon_{\text{PRG}}$ using I_+, is similar. $\square$

Theorem 5 shows that a scaled sum of k_{indep}-wise independent Gaussians fools degree-d PTFs, but such a random variable is not quite the desired PRG since perfectly generating even a single Gaussian random variable formally requires infinitely many random bits. However, the following construction of Kane tells us that for fooling degree-d Gaussian PTFs, it essentially suffices to find the least L such that they are fooled by sums of L independent k-wise Gaussians; then, one gets an explicit PRG with seed length $O(kL \cdot d \log n)$.

Theorem 10 (Section 6 of [Kan11b]). *Let $n, d \in \mathbb{N}$, $0 < \varepsilon < 1$. Suppose that for some $k, L \in \mathbb{N}$, degree-d Gaussian PTFs are $(\varepsilon/2)$-fooled by $\sqrt{\lambda}\boldsymbol{z}_1 + \cdots + \sqrt{\lambda}\boldsymbol{z}_L$, where $\lambda = 1/L$ and $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L$ are k-wise independent n-dimensional Gaussians. Then there is an explicit PRG for ε-fooling degree-d n-dimensional Gaussian PTFs with seed length*

$$O(kL \cdot d \log(dLn/\varepsilon)),$$

which is simply $O(kL \cdot d \log n)$ under the reasonable assumptions that $d, 1/\varepsilon, L \leq \text{poly}(n)$.

As [Kan11b] does not quite explicitly state Theorem 10, we outline a proof in Appendix A for completeness. Theorem 1 follows immediately from Theorem 5 and Theorem 10.

The remaining tasks are to define Mollifier$_p$ and prove Theorems 6 and 8. We define Mollifier$_p$ in Section 4 and prove Theorem 6 and Theorem 8 in Sections 4.4 and 7 respectively.

3 Probabilistic preliminaries

In this section we introduce notation and collect several probabilistic facts we will use. Throughout, **boldface** is used to indicate random variables, $N(0,1)$ denotes the standard Gaussian (normal) distribution, and $N(0,1)^n$ is the associated n-dimensional product distribution.

3.1 Bits, Gaussians, and k-wise independence

Although this work is mainly concerned with Gaussian random variables, many (but not all) of the tools in it "generalize" to Boolean ± 1 random variables. In order to illustrate this, we will provide some definitions and notations in this section that work in both cases. However the Boolean results are never strictly needed in this work, and the reader may prefer to ignore them and focus only on the Gaussian case.

The fact that PTFs over Boolean space generalize PTFs over Gaussian space holds because, for large M and $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \sim \{\pm 1\}^M$ uniform and independent,

$$\boldsymbol{y} = \left(\frac{1}{\sqrt{M}} \sum_{i=1}^{M} \boldsymbol{x}_i^{(1)}, \ldots, \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \boldsymbol{x}_i^{(n)} \right) \tag{8}$$

is "close" to having an $N(0,1)^n$ distribution, and because a degree-d polynomial $p(\boldsymbol{y})$ is also a degree-d polynomial in the $\boldsymbol{x}_j^{(i)}$'s. One sense of "closeness" here is that each $\boldsymbol{y}_i$ may be coupled with a true Gaussian $\boldsymbol{z}_i \sim N(0,1)$ in such a way that $|\boldsymbol{y}_i - \boldsymbol{z}_i| \leq \frac{1}{\sqrt{M}}$ except with probability at most $O(\frac{1}{\sqrt{M}})$.

Definition 11. Let $\mathcal{D}$ be a probability distribution on $\mathbb{R}$. We say that a random vector $\boldsymbol{z}$ on $\mathbb{R}^n$ has a *k-wise independent $\mathcal{D}$ distribution* if each $\boldsymbol{z}_i$ has distribution $\mathcal{D}$, and for all choices of k indices $1 \leq i_1, \ldots, i_k \leq n$, the random variables $\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_k}$ are independent. Examples include $\mathcal{D}$ being the uniform distribution on $\{\pm 1\}$ ("k-wise independent bits") and the main concern in this paper, $\mathcal{D}$ being $N(0,1)$ ("k-wise independent Gaussians").

Remark 12. The main way we use k-wise independence is to say that if $\boldsymbol{x}$ is n-wise independent, $\boldsymbol{z}$ is k-wise independent, and $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree at most k, then $\mathbf{E}[p(\boldsymbol{z})] = \mathbf{E}[p(\boldsymbol{x})]$.

3.2 Polynomial expansions

We recall standard facts and notation from analysis of Boolean functions and Hermite polynomials; see, e.g., [O'D14] for a reference, and in particular [O'D14, Ch. 11.2] for Hermite analysis.

Every function $g : \{\pm 1\}^n \to \mathbb{R}$ can be represented by a multilinear polynomial,

$$g(x) = \sum_{\alpha \in \{0,1\}^n} \widehat{g}(\alpha) x^\alpha,$$

where each $\widehat{g}(\alpha) \in \mathbb{R}$ and we use the standard multi-index notation $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$ and $|\alpha| = \sum_i \alpha_i$. In "Gaussian space" the only functions we will ever analyze are polynomials; every degree-d polynomial $g : \mathbb{R}^n \to \mathbb{R}$ can be written in *Hermite polynomial decomposition* as

$$g(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq d}} \widehat{g}(\alpha) h_\alpha(x),$$

where each $\widehat{g}(\alpha) \in \mathbb{R}$, and the *multivariate Hermite polynomial* polynomial h_α is given by $h_\alpha(x) = h_{\alpha_1}(x_1) \cdots h_{\alpha_n}(x_n)$, where $h_k = \frac{1}{\sqrt{k!}} H_k$ is a normalized version of the univariate degree-k "probabilists' Hermite polynomial" H_k. The multivariate Hermite polynomials h_α are orthonormal under $N(0,1)^n$. Also, in the notation of Equation (8),

$$\sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| = k}} \left(\frac{1}{\sqrt{M}} \boldsymbol{x}^{(i)} \right)^\alpha \xrightarrow{M \to \infty} h_k(\boldsymbol{z}_i), \quad \boldsymbol{z} \sim N(0,1). \tag{9}$$

Let g denote either an n-variate Boolean or Gaussian polynomial. We use standard notation $\mathbf{E}[g]$ for its mean (that is, $\mathbf{E}[g(\boldsymbol{x})]$ for $\boldsymbol{x} \sim \{\pm 1\}^n$ in the former case, $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ in the latter), $\|g\|_r = \mathbf{E}[|g(\boldsymbol{x})|^r]^{1/r}$ for its r-norm ($r \geq 1$), and $\mathbf{Var}[g] = \mathbf{E}[g^2] - \mathbf{E}[g]^2$ for its variance. It holds that

$$\mathbf{E}[g] = \widehat{g}(0), \quad \mathbf{E}[g^2] = \sum_{\alpha} \widehat{g}(\alpha)^2, \quad \text{hence } \mathbf{Var}[g] = \sum_{\alpha \neq 0^n} \widehat{g}(\alpha)^2.$$

We write $g^{<k} = \sum_{|\alpha|<k} \widehat{g}(\alpha) h_\alpha$ for $k \in \mathbb{N}$, and similarly write $g^{=k}$ and $g^{\geq k}$. We also write $\mathrm{W}^{<k}[g] = \mathbf{E}[(g^{<k})^2] = \sum_{|\alpha|<k} \widehat{g}(\alpha)^2$ for the "weight of g below level k", and similarly write $\mathrm{W}^{=k}[g]$ and $\mathrm{W}^{\geq k}[g]$.

3.3 Noise and zooms

A basic fact about Gaussians is that if $\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{N}(0,1)^n$ are independent and $0 \leq \lambda \leq 1$, then $\sqrt{1-\lambda}\boldsymbol{x} + \sqrt{\lambda}\boldsymbol{y}$ is also distributed as $\mathrm{N}(0,1)$. In this work, λ typically denotes a "small" quantity; for fixed $x \in \mathbb{R}^n$ we view $\sqrt{1-\lambda}x + \sqrt{\lambda}\boldsymbol{y}$ as a "λ-noisy" version of x, and we view changing a polynomial g's input from $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ to $\sqrt{1-\lambda}x + \sqrt{\lambda}\boldsymbol{y}$ as "zooming into g at x with scale λ". We make a precise definition:

Definition 13. For g an n-variate Gaussian polynomial, $0 \leq \lambda \leq 1$, and $x \in \mathbb{R}^n$, we define the function $g_{\lambda|x}$ by

$$g_{\lambda|x}(y) = g\left(\sqrt{1-\lambda}x + \sqrt{\lambda}y\right).$$

The function $g_{\lambda|x}(y)$ is a polynomial in y of the same degree as g, and we (nonstandardly) refer to it as the λ-*zoom of g at x*.

Remark 14. Referring again to Equation (8), one may verify that a λ-zoom of g at a random x is the Gaussian analogue of a standard Boolean concept: a *random restriction* of a function $g : \{\pm 1\}^n \to \mathbb{R}$ at $x \in \{\pm 1\}^n$, meaning a subfunction obtained by proceeding through each coordinate i, and either fixing the ith input to be x_i with probability $1 - \lambda$, or else leaving it unfixed ("free") with probability λ.

The fact that random restrictions of a Boolean function interact well with its polynomial expansion is well known; e.g. [O'D14, Prop. 4.17] gives a formula for the expected square of any Fourier coefficient of a Boolean function under a random restriction. Carefully taking the "Gaussian special case" of this (using Equation (9)) yields the below analogue for random zooms. For completeness, we give a self-contained proof of this analogue in Appendix A.

Proposition 15. *For $g : \mathbb{R}^n \to \mathbb{R}$ a polynomial, $0 \leq \lambda \leq 1$, and $\beta \in \mathbb{N}^n$,*

$$\underset{\boldsymbol{x}\sim\mathrm{N}(0,1)^n}{\mathbf{E}} \left[\widehat{g_{\lambda|\boldsymbol{x}}}(\beta)^2\right] = \sum_{\gamma \geq \beta} \mathbf{Pr}[\mathrm{Bin}(\gamma,\lambda) = \beta]\, \widehat{g}(\gamma)^2,$$

where $\mathrm{Bin}(\gamma,\lambda)$ denotes an n-dimensional random vector with independent components, the jth of which is distributed as the binomial random variable $\mathrm{Bin}(\gamma_j,\lambda)$.

Summing the above proposition over all multi-indices β of a given weight $|\beta| = m$ immediately yields the following useful corollary:

Corollary 16. *For $g : \mathbb{R}^n \to \mathbb{R}$ a polynomial, $0 \leq \lambda \leq 1$, and $m \in \mathbb{N}$,*

$$\underset{\boldsymbol{x}\sim\mathrm{N}(0,1)^n}{\mathbf{E}} \left[\mathrm{W}^{=m}[g_{\lambda|\boldsymbol{x}}]\right] = \sum_{M} \mathbf{Pr}[\mathrm{Bin}(M,\lambda) = m] \cdot \mathrm{W}^{=M}[g].$$

3.4 Noise operator and hypercontractivity

Considering the mean of the zoom of a polynomial leads to the "Gaussian noise" (or "Ornstein–Uhlenbeck") operator (see, e.g., [O'D14, Def. 11.12]):

Definition 17. Given $0 < \rho \leq 1$, the operator U_ρ acts on Gaussian polynomials $g : \mathbb{R}^n \to \mathbb{R}$ via

$$(\mathrm{U}_\rho g)(x) = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n} \left[g\left(\rho x + \sqrt{1 - \rho^2} \boldsymbol{y} \right) \right] = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n} \left[g_{(1-\rho^2)|x}(\boldsymbol{y}) \right].$$

It is well known that U_ρ acts diagonally in the Hermite polynomial basis $(h_\alpha)_{\alpha \in \mathbb{N}^n}$:

$$\mathrm{U}_\rho g = \sum_{\alpha \in \mathbb{N}^N} \rho^{|\alpha|} \widehat{g}(\alpha) h_\alpha. \tag{10}$$

In particular, if g is a degree-d polynomial, so too is $\mathrm{U}_\rho g$.

We may also write U_ρ for the analogous Boolean noise operator (more usually denoted T_ρ, see [O'D14, Def. 2.46]), definable for $g : \{\pm 1\}^n \to \mathbb{R}$ either through Equation (10), or by stipulating that $\mathrm{U}_\rho g(x)$ is the mean of a random restriction of g at x with ρ-probability of fixing a coordinate.

Finally, somewhat unusually, we will need to extend the definition of U_ρ to $\rho > 1$, which we can do via the formula Equation (10); equivalently, by stipulating that $\mathrm{U}_{\rho^{-1}} = \mathrm{U}_\rho^{-1}$. For $\rho > 1$ this operator no longer has a "probabilistic interpretation", but it still maps degree-d polynomials to degree-d polynomials.

Remark 18. We will several times use the "semi-group property", $\mathrm{U}_{\rho_1} \mathrm{U}_{\rho_2} = \mathrm{U}_{\rho_1 \rho_2}$, which is immediate from Equation (10).

At one point in our analysis we will also need the notion of *Gaussian noise stability*:

Definition 19. For $g : \mathbb{R}^n \to \mathbb{R}$ and $\rho > 0$,

$$\mathrm{Stab}_\rho[g] = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n} [g(\boldsymbol{y}) \cdot \mathrm{U}_\rho g(\boldsymbol{y})] = \sum_\alpha \rho^{|\alpha|} \widehat{g}(\alpha)^2,$$

where the last equality is by Equation (10) and orthonormality of Hermite polynomials.

Hypercontractivity. A nontrivial and highly useful property of the Boolean/Gaussian noise operator U_ρ is *hypercontractivity* (see, e.g., [O'D14, Secs. 9.2, 11.1]):

Theorem 20 ((2, q)-hypercontractive inequality). *Let g be a Gaussian or Boolean polynomial. Then $\|\mathrm{U}_{1/\sqrt{q-1}} g\|_q \leq \|g\|_2$ holds for any $q > 2$.*

Hypercontractivity has the following consequences (see [O'D14, Thms. 9.22, 9.23]):

Theorem 21. *Let g be a Gaussian or Boolean polynomial of degree at most k. Then $\|g\|_2 \leq e^k \|g\|_1$.*

Theorem 22. *Let g be a Gaussian or Boolean polynomial of degree at most k. Then for any $t \geq \sqrt{2e}^k$,*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}} [|g(\boldsymbol{x})| \geq t \|g\|_2] \leq \exp\left(-\frac{k}{2e} t^{2/k} \right).$$

3.5 Hyperconcentration: our key tool

The ideas in this section, though technically standard, are part of the conceptual contribution of this work.

Very often we will need to show that a random variable is tightly concentrated around its mean in a *multiplicative* sense. Let us start with some notation.

Notation 23. We use the following notation to denote that two reals $a, b > 0$ are multiplicatively close: For $\nu \geq 0$,

$$a \approx_\nu b \quad \Longleftrightarrow \quad e^{-\nu} \leq a/b \leq e^\nu.$$

Note that this condition is indeed symmetric in a and b. We extend the notation to all $a, b \in \mathbb{R}$ by stipulating that $a \approx_\nu b$ if: $ab > 0$ and the above condition holds; or, $a = b = 0$.

Given a real random variable $\boldsymbol{w}$ with mean μ, a standard way to show that $\boldsymbol{w} \approx \mu$ with high probability is to first establish $\mathbf{stddev}[\boldsymbol{w}] \leq \eta|\mu|$ and then use Chebyshev's inequality. When this holds we informally say that $\boldsymbol{w}$ *concentrates* around its mean. In this work, a crucial concept will be improving this concentration using higher norms.

Definition 24. Let $q > 2$ and $\eta \geq 0$ be real numbers. We say a real random variable $\boldsymbol{w}$ with mean μ is (q, η)-*hyperconcentrated* if

$$\|\boldsymbol{w} - \mu\|_q = \mathbf{E}[|\boldsymbol{w} - \mu|^q]^{1/q} \leq \eta|\mu|.$$

The utility of this definition is that it gives an improvement to the Chebyshev inequality:

Proposition 25. *Suppose $\boldsymbol{w}$ with mean μ is (q, η)-hyperconcentrated. Then for any $t > 0$, except with probability at most $(\eta/t)^q$ we have $|\boldsymbol{w} - \mu| \leq t|\mu|$ (and in particular $\boldsymbol{w} \approx_{2t} \mu$ if $t \leq 1/2$).*

Proof. Apply Markov's inequality to the random variable $|\boldsymbol{w} - \mu|^q$. □

We'll also need the following simple consequence of hyperconcentration:

Lemma 26. *Suppose $\boldsymbol{z}$ is an $\mathbb{R}^m$-valued random vector with (q, η)-hyperconcentrated components, and write $\mu = \mathbf{E}[\boldsymbol{z}]$. Then for any multi-index $\alpha \in \mathbb{N}^m$ with $|\alpha| \leq q$,*

$$\mathbf{E}[|\boldsymbol{z} - \mu|^\alpha] \leq \eta^{|\alpha|}|\mu|^\alpha.$$

Proof. We have

$$\mathbf{E}[|\boldsymbol{z} - \mu|^\alpha] = \mathbf{E}\left[\prod_{i=1}^n |\boldsymbol{z}_i - \mu_i|^{\alpha_i}\right] \leq \prod_{i=1}^n \mathbf{E}\left[|\boldsymbol{z}_i - \mu_i|^{\alpha_i \cdot \frac{|\alpha|}{\alpha_i}}\right]^{\frac{\alpha_i}{|\alpha|}} \leq \prod_{i=1}^n (\eta\mu_i)^{\alpha_i} = \eta^{|\alpha|}|\mu|^\alpha,$$

where the first inequality is from Hölder's inequality and the second is from Definition 24. □

The random variables we'll show hyperconcentration for will be Gaussian polynomials. We will do this by bounding a quantity that we term their "hypervariance", and that plays a central role in our work:

Definition 27. Let g be a Gaussian or Boolean polynomial. Then for $R > 1$, we define the R-*hypervariance* of g to be

$$\mathbf{HyperVar}_R[g] \coloneqq \mathbf{Var}[\mathrm{U}_R g] = \sum_{\alpha \neq 0} R^{2|\alpha|} \widehat{g}(\alpha)^2.$$

(For $R = 1$, this reduces to the usual variance of g.)

Lemma 28. *Let g be a Gaussian or Boolean polynomial. Write $\mu = \mathbf{E}[g]$ and assume $\mathbf{HyperVar}_R[g] \leq \theta \mu^2$. Then the random variable $g(\boldsymbol{x})$ is $(1 + R^2, \sqrt{\theta})$-hyperconcentrated.*

Proof. Writing $\overline{g} = g - \mu$, our hypothesis is that $\|\mathrm{U}_R \overline{g}\|_2^2 \leq \theta \mu^2$. By hypercontractivity, we have $\|\overline{g}\|_{1+R^2} = \|\mathrm{U}_{1/R} \mathrm{U}_R \overline{g}\|_{1+R^2} \leq \|\mathrm{U}_R \overline{g}\|_2$. Thus $\|\overline{g}\|_{1+R^2} \leq \sqrt{\theta} |\mu|$, as needed. $\qquad\square$

The hypothesis in Lemma 28, that g's hypervariance is small compared to its squared-mean, will be an important one for us. It is essentially the same as the hypothesis that g's hypervariance is small compared to its squared-2-norm (since squared-2-norm equals squared-mean plus variance, and hypervariance is at least variance for all $R > 1$). It will be slightly more convenient in our Local Hypervariance Theorem to work with the latter hypothesis, so we codify it here and establish the analogue of Lemma 28.

Definition 29. Let g be a Gaussian or Boolean polynomial, and let $R > 1$. We say that g is (R, ε)-*attenuated* if $\mathbf{HyperVar}_R[g] \leq \varepsilon \|g\|_2^2$.

Remark 30. Intuitively, a polynomial g is attenuated if for each $i \geq 1$, the amount of Hermite weight it has at level i is "very small" compared with the total Hermite weight (squared 2-norm) of g. Crucially, the precise quantitative definition of "very small" in the preceding sentence depends on the weight level i, and gets exponentially stronger (smaller) as i gets larger. An intuition which may possibly be helpful is to think of an attenuated polynomial as a polynomial which is "morally constant" over Gaussian space.

Returning to hyperconcentration, we have the following:

Lemma 31. *Let g be a Gaussian or Boolean polynomial that is (R, θ)-attenuated, with $R \geq \sqrt{2}$ and $\theta \leq 1$. Then the random variable $g(\boldsymbol{x})$ is $(1 + \frac{1}{2} R^2, \sqrt{\theta})$-hyperconcentrated.*

Proof. Using the notation μ and $\overline{g}$ again, and starting with the (R, θ)-attenuation assumption, we have

$$\sum_{j \geq 1} R^{2j} \|g^{=j}\|_2^2 \leq \theta \left(\mu^2 + \sum_{j \geq 1} \|g^{=j}\|_2^2 \right) \implies \sum_{j \geq 1} (R^{2j} - \theta) \|g^{=j}\|_2^2 \leq \theta \mu^2 \implies \sum_{j \geq 1} \left(\tfrac{R}{\sqrt{2}} \right)^{2j} \|g^{=j}\|_2^2 \leq \theta \mu^2,$$

where the last step used $R \geq \sqrt{2}$ and $\theta \leq 1$. But this is equivalent to $\mathbf{HyperVar}_{\frac{R}{\sqrt{2}}}[g(\boldsymbol{x})] \leq \theta \mu^2$, so the result follows from Lemma 28. $\qquad\square$

Combining this with Lemma 31 and Proposition 25 yields the following useful result, which informally says that "attenuated polynomials are very likely to take values multiplicatively close to their means":

Proposition 32. *Let g be a Gaussian or Boolean polynomial that is (R, θ)-attenuated, with $R \geq \sqrt{2}$ and $\theta \leq 1$. Write $\mu = \mathbf{E}[g]$. Then assuming $0 < \gamma \leq 1$, we have $g(\boldsymbol{x}) \approx_\gamma \mu$ except with probability at most $(2\sqrt{\theta}/\gamma)^{\frac{1}{2} R^2 + 1}$.*

3.6 Special properties of Gaussian random variables

All of the results in this section so far have applied equally well to Gaussian or Boolean polynomials. We now give the two results we will use that are specific just to Gaussian polynomials. The first is a well known result of Carbery and Wright [CW01] on anticoncentration (see e.g. [Kan11b, Lem. 23], [O'D14, Sec. 11.6]):

Theorem 33 (Gaussian Carbery–Wright)**.** *There is a univeral constant C such that for any degree-d polynomial $g : \mathbb{R}^n \to \mathbb{R}$ and any $0 \le \delta \le 1$,*

$$\Pr_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n} \left[|g(\boldsymbol{x})| < \left(\tfrac{\delta}{C \cdot d} \right)^d \cdot \|g\|_2 \right] \le \delta.$$

The second Gaussian-specific result we use is a key lemma from Kane's work [Kan11b, Lemma 9]. This lemma was the essential ingredient he used to prove his "local concentration" result Theorem 2. At first glance, it may look much stronger than Theorem 2, because it gives a nontrivial kind of concentration result even for λ as large as $1/\mathrm{poly}(d)$. However the concentration one gets in (almost all) local neighborhoods is somewhat weak: one gets that $g_{\lambda|\boldsymbol{x}}$'s values are with high probability near a specific value, but this is not enough to even conclude that $\mathbf{stddev}[g_{\lambda|\boldsymbol{x}}]$ is small compared to that value. Kane uses hypercontractivity to bootstrap this to control over the variance when he obtains Theorem 2, and this loses a $2^{O(d)}$ factor. When we employ Lemma 34 below, we will already be working with hyperconcentrated functions, which means we will not lose much when similarly bootstrapping.

Lemma 34. *([Kan11b, Lemma 9] with parameters renamed.) Let $g : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial and let $0 < \beta < 1$. Then for $\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{N}(0,1)^n$ independent, except with probability β we have*

$$g_{\lambda|\boldsymbol{x}}(\boldsymbol{y}) \approx_\nu g(\boldsymbol{x}) \quad \text{for } \nu = O(d^2/\beta) \cdot \sqrt{\lambda}$$

(provided λ is small enough that $\nu \le 1$).

Kane's proof of this lemma (seemingly) crucially relies on the rotational invariance of n-dimensional Gaussians.

4 Defining $\mathrm{Mollifier}_p$

The definition of $\mathrm{Mollifier}_p$ involves a collection $\mathfrak{S}$ of "statistics" of the polynomial p. Each statistic $s \in \mathfrak{S}$ will be a certain nonnegative polynomial $s : \mathbb{R}^n \to \mathbb{R}^{\ge 0}$, defined in terms of p, of degree at most $2d$.

The definition also involves a collection MOLLIFIERCHECKS of "mollifier checks". Each mollifier check CHECK $\in$ MOLLIFIERCHECKS will consist of two ingredients:

$$\text{CHECK} = (\text{INEQ}, \delta), \quad \text{where INEQ} = (\gamma, s_u, s_v) \text{ means "} s_u \ge \gamma s_v \text{"} \tag{11}$$

for some statistics $s_u, s_v \in \mathfrak{S}$ and some nonnegative value γ, and where $\delta \ge 0$ is a "softness" parameter. The intuitive meaning of CHECK applied at a point $x \in \mathbb{R}^n$ is that it is "softly" checking that $s_u(x) \ge \gamma s_v(x)$, up to a multiplicative factor of roughly $1 \pm \delta$. More precisely:

Definition 35. Let $\sigma : \mathbb{R} \to [0,1]$ be a smooth function satisfying

$$\sigma(t) = \begin{cases} 0 & \text{if } t \le -1, \\ 1 & \text{if } t \ge +1, \end{cases}$$

and which is such that for all $j \le d$, the magnitude of σ's j-th derivative is everywhere bounded by $j^{O(j)}$. (This is easily achieved by standard constructions such as taking σ to be a suitable polynomial of degree $O(d)$ on the interval $[-1, 1]$.) Also, given a mollifier check CHECK $\in$ MOLLIFIERCHECKS as in (11), define

$$\mathrm{SoftCheck}_{\text{CHECK}} : \mathbb{R}^n \to [0, 1]$$

15

by

$$\text{SoftCheck}_{\text{CHECK}}(x) = \sigma\left(\delta^{-1}\ln\left(\frac{s_u(x)}{\gamma s_v(x)}\right)\right), \tag{12}$$

where we take $0/0 = +\infty$. We remark that

$$\text{SoftCheck}_{\text{CHECK}}(x) = 1 \text{ if } s_u(x) \geq \exp(\delta) \cdot \gamma s_v(x),$$
$$\text{SoftCheck}_{\text{CHECK}}(x) = 0 \text{ if } s_u(x) \leq \exp(-\delta) \cdot \gamma s_v(x).$$

The function Mollifier_p is the product of all the mollifier checks:

Definition 36. $\text{Mollifier}_p(x) = \displaystyle\prod_{\text{CHECK}\in\text{MOLLIFIERCHECKS}} \text{SoftCheck}_{\text{CHECK}}(x).$

To complete the definition of Mollifier_p we need to: (i) define the statistics in $\mathfrak{S}$; and, (ii) define the collection MOLLIFIERCHECKS of mollifier checks. We do each of these in turn below.

4.1 The statistics in $\mathfrak{S}$

4.1.1 Noisy derivatives of amplified polynomials

Now we arrive at a novel definition in this work which plays a key role in our results. We will make extensive use of the following notion, which can be thought of as a sort of "noisy derivative of the R-amplified version of g at x in directions y and y'." This is a variant of one of the key definitions of [Kan11b] (the second definition in Section 3 of that paper) but with the crucial difference that now we consider the "R-amplified" version of g in place of just g itself as was the case in the corresponding definition in [Kan11b]:

Definition 37. Given two vectors $y, y' \in \mathbb{R}^n$, define the operator $\triangle_{y,y'}^{(R,\lambda)}$ on polynomials $g : \mathbb{R}^n \to \mathbb{R}$ via

$$(\triangle_{y,y'}^{(R,\lambda)} g)(x) := \frac{\mathrm{U}_{\underline{R}}g_{\underline{\lambda}|x}(y) - \mathrm{U}_{\underline{R}}g_{\underline{\lambda}|x}(y')}{\sqrt{2}}.$$

We remark that $\mathrm{U}_{\underline{R}}g_{\underline{\lambda}|x}(y)$ is parenthesized as $(\mathrm{U}_{\underline{R}}(g_{\underline{\lambda}|x}))(y)$.

The following is easily verified:

Fact 38. *If $g : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree at most d, then for every $y, y' \in \mathbb{R}^n$, the function $\triangle_{y,y'}^{(R,\lambda)} g$ is a polynomial of degree at most $d - 1$.*

The following simple but crucial fact connects the derivative notion from Definition 37 to the hypervariance notion from Definition 27:

Fact 39. *For fixed $x \in \mathbb{R}^n$ and independent n-dimensional Gaussians $\boldsymbol{y}, \boldsymbol{y}' \sim \mathrm{N}(0,1)^n$, we have that*

$$\mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{y}'}\left[\left(\triangle_{\boldsymbol{y},\boldsymbol{y}'}^{(R,\lambda)} g(x)\right)^2\right] = \mathbf{HyperVar}_{\underline{R}}\left[g_{\underline{\lambda}|x}\right].$$

16

Proof. The left-hand side is equal to

$$\frac{1}{2}\,\mathbf{E}\left[\left(\sum_\alpha \underline{R}^{|\alpha|}\widehat{g_{\underline{\lambda}|x}}(\alpha)\left(h_\alpha(\boldsymbol{y})-h_\alpha(\boldsymbol{y}')\right)\right)^2\right]$$

$$=\frac{1}{2}\,\mathbf{E}\left[\left(\sum_{\alpha\neq0} \underline{R}^{|\alpha|}\widehat{g_{\underline{\lambda}|x}}(\alpha)\left(h_\alpha(\boldsymbol{y})-h_\alpha(\boldsymbol{y}')\right)\right)^2\right] \qquad \text{(since } h_\alpha(y)\equiv1 \text{ for } \alpha=0\text{)}$$

$$=\frac{1}{2}\sum_{\alpha,\beta\neq0} \underline{R}^{|\alpha|+|\beta|}\widehat{g_{\underline{\lambda}|x}}(\alpha)\widehat{g_{\underline{\lambda}|x}}(\beta)\,\mathbf{E}\left[(h_\alpha(\boldsymbol{y})-h_\alpha(\boldsymbol{y}'))(h_\beta(\boldsymbol{y})-h_\beta(\boldsymbol{y}'))\right]$$

$$=\sum_{|\alpha|>0} \underline{R}^{2|\alpha|}\widehat{g_{\underline{\lambda}|x}}(\alpha)^2=\mathbf{HyperVar}_{\underline{R}}[g_{\underline{\lambda}|x}]$$

where the penultimate equality is by orthonormality of the Hermite polynomials and independence of $\boldsymbol{y},\boldsymbol{y}'$. $\qquad\square$

4.1.2 The statistics in $\mathfrak{S}$

Fix a parameter

$$\underline{R}:=\Theta(T) \tag{13}$$

where T is a parameter we will set later. (See Equation (33); it will in fact be an absolute constant.) We can now define the set of statistics, $\mathfrak{S}$. Each statistic is doubly indexed by a pair of natural numbers; there are $(d+1)(D+1)$ many statistics, $\{s_{i,j}\}_{i\in\{0,\dots,d\},j\in\{0,\dots,D\}}$, where

$$D:=(2d+1)^2. \tag{14}$$

It is convenient for us to view the elements of $\mathfrak{S}$ as being arranged in a grid where the (i,j)-th statistic $s_{i,j}$ is in row i and column j (we will often use terminology of this sort). We remark that our statistic $s_{i,j}$ will closely correspond to the functions called $|p_\theta^{(\ell),m}(X)|_2^2$ in [Kan11b] ($i=\ell,j=m$), except that as mentioned we use *amplified* noisy derivatives where Kane just had noisy derivatives.

All statistics are defined in terms of the underlying degree-d polynomial p. We first define the 0th column of statistics:

Definition 40 (0th column of statistics)**.** For $i\in\{0,\dots,d\}$, we define

$$s_{i,0}(x):=\mathbf{E}\left[\left(\triangle_{\boldsymbol{y}_i,\boldsymbol{y}_i'}^{(R,\lambda)}\triangle_{\boldsymbol{y}_{i-1},\boldsymbol{y}_{i-1}'}^{(R,\lambda)}\cdots\triangle_{\boldsymbol{y}_1,\boldsymbol{y}_1'}^{(R,\lambda)}p(x)\right)^2\right],$$

where $\boldsymbol{y}_1,\boldsymbol{y}_1',\dots,\boldsymbol{y}_i,\boldsymbol{y}_i'\sim\mathrm{N}(0,1)^n$ are independent.

Remark 41. By Fact 38, $s_{d,0}$ is a constant function and $s_{d+1,0}$ is identically zero; this is why we consider $s_{i,0}$ only for $i\leq d$.

For technical reasons, we will also need to use slight variants of the statistics $s_{i,0}$, which correspond to taking an average over mildly noisy versions of the input:

Definition 42 (The remaining statistics)**.** For $i\in\{0,\dots,d\}$ and $j\in\{1,\dots,D\}$ we define $s_{i,j}$ by

$$s_{i,j}(x):=\mathrm{U}_{\sqrt{1-\underline{\lambda}}}s_{i,j-1}(x)=\mathop{\mathbf{E}}_{\boldsymbol{y}\sim\mathrm{N}(0,1)^n}\left[(s_{i,j-1})_{\underline{\lambda}|x}(\boldsymbol{y})\right].$$

17

Remark 43. Using the semigroup property, we have that

$$s_{i,j}(x) = \mathrm{U}_{(1-\underline{\lambda})^{j/2}} s_{i,0}(x) = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n} \left[(s_{i,0})_{1-(1-\underline{\lambda})^j | x}(\boldsymbol{y}) \right].$$

This completes the formal definition of the statistics in $\mathfrak{S}$; however, it will be very useful for us to view the statistics from a different perspective based on distributions of polynomials. We introduce this perspective in the next subsection.

4.2 A distributional view on the statistics

For the sake of probabilistic technicalities, we will need to define a notion of a distribution of polynomials being "nice":

Definition 44. We say that a distribution $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ of n-variable polynomials of degree at most d is *nice* if

- Υ is the normal distribution $\mathrm{N}(0,1)^I$ for some natural number I, and

- for each v in the support of Υ, the coefficients of p_v are polynomials of degree at most d in v.

It will be convenient for us to view the statistics $s_{i,j}$ as averages of squares of polynomials drawn from various nice distributions. To do this, we inductively define a grid of nice distributions of polynomials $\mathcal{F}_{i,j}$ as follows. The base distribution, $\mathcal{F}_{0,0}$ is just the probability distribution with a single outcome, namely the polynomial p. (Note that this corresponds to a nice distribution with $I = 0$.) Next, we inductively define the distributions $\mathcal{F}_{i,0}$ as follows:

- **To make a draw from $\mathcal{F}_{i,0}$, where $i > 0$:** First draw $\boldsymbol{f} \sim \mathcal{F}_{i-1,0}$. Then draw $\boldsymbol{y}, \boldsymbol{y}' \sim \mathrm{N}(0,1)^n$. Then output the polynomial $\triangle_{\boldsymbol{y},\boldsymbol{y}'}^{(R,\underline{\lambda})} \boldsymbol{f}$.

The following "operator notation" for zooms (local to this section and [Section 8.3](#)) will be convenient: for a polynomial p and a vector $y \in \mathbb{R}^n$, we define the notation

$$\mathrm{N}_y p := p_{1-\underline{\lambda}|y}. \tag{15}$$

With the distributions of polynomials $\mathcal{F}_{i,0}$ defined as above, we inductively define the distributions $\mathcal{F}_{i,j}$, where $j > 0$ as follows:

- **To make a draw from $\mathcal{F}_{i,j}$, where $j > 0$:** First draw $\boldsymbol{f} \sim \mathcal{F}_{i,j-1}$. Then draw $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$. Then output the polynomial $\mathrm{N}_{\boldsymbol{y}} \boldsymbol{f}$.

It is immediate from these definitions that each $\mathcal{F}_{i,j}$ is a nice distribution of polynomials of degree at most d. It is also immediate, comparing the above definition against [Definition 40](#) and [Definition 42](#), that for each $i \in \{0, \dots, d\}$ and each $j \in \{0, \dots, D\}$ we have that

$$s_{i,j} = \mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}} [\boldsymbol{f}^2]. \tag{16}$$

Finally, it is straightforward to check from these definitions (using also [Fact 39](#)) that

$$s_{i+1,0}(x) = \mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,0}} [\mathbf{HyperVar}_{\underline{R}}[\boldsymbol{f}_{\underline{\lambda}|x}]], \qquad s_{i,j+1}(x) = \mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}} [\|\boldsymbol{f}_{\underline{\lambda}|x}\|_2^2]. \tag{17}$$

These characterizations will be useful when we analyze the statistics later.

4.3 Defining the mollifier checks

Intuition. In this subsection we define the collection MOLLIFIERCHECKS of mollifier checks. Before formally defining these checks, we give some useful intuition concerning them. We will show in Section 4.4 that except with very small failure probability over $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, the statistics $s_{i,j}(\boldsymbol{x})$ satisfy the following properties, where $\hat{\lambda}, \delta_{\mathrm{horz}} > 0$ are suitable small parameters:

1. **Local hyperconcentration:** For each $i \in \{0, \ldots, d-1\}$,

$$s_{i+1,0}(\boldsymbol{x}) \le \hat{\lambda} s_{i,1}(\boldsymbol{x}). \tag{18}$$

2. **Insensitivity under noise:** For each $i \in \{0, \ldots, d\}, j \in \{0, \ldots, D-2\}$,

$$s_{i,j}(\boldsymbol{x}) \approx_{\delta_{\mathrm{horz}}} s_{i,j+1}(\boldsymbol{x}). \tag{19}$$

The parameter settings we require will turn out to be the following:

$$\hat{\lambda} \text{ satisfying } \mathrm{poly}(Td)^T \cdot \hat{\lambda}^{T/2} = \underline{\lambda}\varepsilon_{\mathrm{PRG}}, \qquad \delta_{\mathrm{horz}} := \frac{1}{KdD}, \tag{20}$$

where K is a suitably large absolute constant and $T > 2$ is a constant that will be set later in Equation (33) (for now, the most important thing to notice is that since $T > 2$, the exponent on $\hat{\lambda}$ above is strictly greater than 1).

The mollifier checks are designed precisely to check that the above properties Inequalities (18) and (19) actually hold at x, and thus Theorem 6 corresponds to the fact that these properties hold with high probability for a random $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$.

With the above intuition in place, we proceed to define the mollifier checks that check for each of the above types of properties (1) and (2). Please see below for a figure depicting the mollifier checks and the grid of statistics.
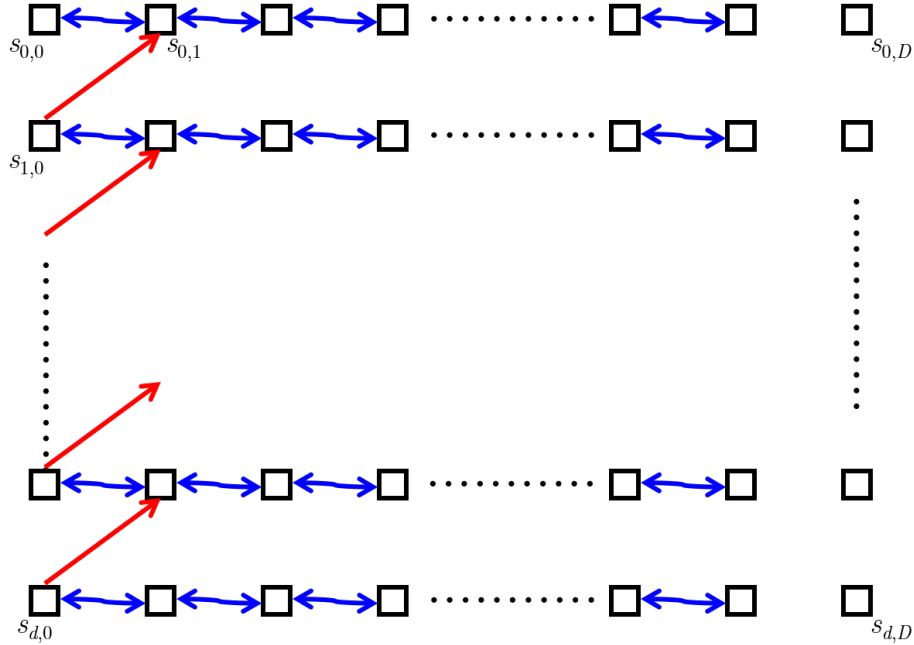

Figure 2: The grid of statistics and the mollifier checks. Local hyperconcentration checks are depicted in red and noise insensitivity checks are depicted in blue.

4.3.1 Checking local hyperconcentration

For each $i \in \{0, \ldots, d-1\}$, MOLLIFIERCHECKS contains a check corresponding to Inequality (18). The "inequality" portion of the check is

$$s_{i+1,0} \leq e\hat{\lambda}s_{i,1}$$

and the "softness" parameter of the check is 1; so for this check CHECK in MOLLIFIERCHECKS, the associated "soft check" is

$$\text{SoftCheck}_{\text{CHECK}}(x) = \sigma\left(\ln\left(\frac{e\hat{\lambda}s_{i,1}}{s_{i+1,0}}\right)\right). \tag{21}$$

We refer to these d elements of MOLLIFIERCHECKS as "anticoncentration checks" or (recalling the grid) as "diagonal checks."

4.3.2 Checking insensitivity under noise

For each $i \in \{0, \ldots, d\}, j \in \{0, D-2\}$, MOLLIFIERCHECKS contains a pair of checks corresponding to Inequality (19). The "inequality" portion of the first (respectively, second) check of the pair is

$$s_{i,j}(x) \geq \exp(2\delta_{\text{horz}}) \cdot s_{i,j+1}(x) \quad (\text{respectively, } s_{i,j+1}(x) \geq \exp(2\delta_{\text{horz}}) \cdot s_{i,j}(x)),$$

and the "softness" parameter of each of these checks is δ_{horz}. So for these two elements CHECK, CHECK$'$ of MOLLIFIERCHECKS the associated "soft checks" are

$$\text{SoftCheck}_{\text{CHECK}}(x) = \sigma\left(\delta_{\text{horz}}^{-1}\ln\left(\frac{\exp(2\delta_{\text{horz}}) \cdot s_{i,j}(x)}{s_{i,j+1}(x)}\right)\right) \quad \text{and} \tag{22}$$

$$\text{SoftCheck}_{\text{CHECK}'}(x) = \sigma\left(\delta_{\text{horz}}^{-1}\ln\left(\frac{\exp(2\delta_{\text{horz}}) \cdot s_{i,j+1}}{s_{i,j}}\right)\right). \tag{23}$$

We refer to these $2dD$ checks as "noise-insensitivity checks" or as "horizontal checks."

This concludes the definition of MOLLIFIERCHECKS, so recalling Definition 36 the definition of Mollifier$_p$ is now complete. We turn to proving Theorem 6.

4.4 Breaking down the mollification error for the proof of Theorem 6

Recalling the definition of MOLLIFIERCHECKS from Section 4.3, the approach to proving Theorem 6 is clear. We will show that each of the local hyperconcentration (diagonal) checks passes "with room to spare" with high probability over $\boldsymbol{x}$, and that likewise each of the noise-insensitivity (horizontal) checks passes with room to spare with high probability over $\boldsymbol{x}$. The two theorems stated below give the desired bounds:

Theorem 45 (Local hyperconcentration, rough analogue of Corollary 10 of [Kan11b])**.** *For each* $i \in \{0, \ldots, d-1\}$, *except with probability at most* $\varepsilon_{\text{PRG}}/(8d)$ *over* $\boldsymbol{x} \sim \text{N}(0,1)^n$, *Inequality (18) holds, i.e.*

$$s_{i+1,0}(\boldsymbol{x}) \leq \hat{\lambda}s_{i,1}(\boldsymbol{x}).$$

Theorem 46 (Noise-insensitivity, analogue of Lemma 11 of [Kan11b])**.** *For each* $i \in \{0, \ldots, d\}$ *and* $j \in \{0, \ldots, D-1\}$, *except with probability at most* $\varepsilon_{\text{PRG}}/(8(d+1)D)$ *over* $\boldsymbol{x} \sim \text{N}(0,1)^n$, *Inequality (19) holds, i.e.*

$$s_{i,j}(\boldsymbol{x}) \approx_{\delta_{\text{horz}}} s_{i,j+1}(\boldsymbol{x}).$$

Proof of Theorem 6 using Theorem 45 and Theorem 46. By a union bound over failure probabilities, we have that with probability at least $1 - \varepsilon_{\mathrm{PRG}}/4$ Inequality (18) holds for all i and Inequality (19) holds for all i, j. If Inequality (18) holds for a given i then $\ln \frac{e\hat{\lambda} s_{i,1}}{s_{i+1,0}} \geq 1$ and the diagonal check Equation (21) evaluates to 1. If Inequality (19) holds for a given i, j then $\delta_{\mathrm{horz}}^{-1} \ln\left(\frac{\exp(2\delta_{\mathrm{horz}}) \cdot s_{i,j}}{s_{i,j+1}} \right) \geq 1$ and the horizontal check Equation (22) evaluates to 1, and similarly $\delta_{\mathrm{horz}}^{-1} \ln\left(\frac{(\exp(2\delta_{\mathrm{horz}}) \cdot s_{i,j+1}}{s_{i,j}} \right) \geq 1$ and the horizontal check Equation (23) evaluates to 1. $\qquad \square$

It remains to prove Theorems 45 and 46.

5 Local hyperconcentration: Proof of Theorem 45

In this section we present the key new ingredient underlying our main result, the *Local Hyperconcentration Theorem* for degree-d polynomials. As alluded to in the Introduction, this result says that with high probability over a Gaussian $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, the λ-zoom of a degree-d polynomial p at $\boldsymbol{x}$ (i.e. the polynomial $p_{\lambda|\boldsymbol{x}}$) is attenuated — intuitively, it is "very close to a constant polynomial". We refer to this result as a "local hyperconcentration theorem" since by Lemma 31 attenuation of $p_{\lambda|\boldsymbol{x}}$ implies that the random variable $p_{\lambda|\boldsymbol{x}}(\boldsymbol{y})$ (for $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$) is hyperconcentrated; this property will play a crucial role in our later technical arguments.

For technical reasons related to the definition of our statistics (essentially because each statistic $s_{i,j}(x)$ is an average of polynomials — recall Section 4.2), the actual statement we will need is one that is about a distribution of polynomials rather than a single polynomial. However, for clarity of exposition we first state the "one-polynomial" version of the original local hyperconcentration theorem from [OST20] below:

Theorem 47 (Local hyperconcentration theorem for a single polynomial). *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most d. Fix parameters $R \geq 1, 0 < \varepsilon \leq 1$, $0 < \beta < 1$, and assume*

$$\lambda \leq \frac{\varepsilon}{R^2} \cdot \left(\frac{\beta}{d} \right)^{C \log d}$$

(where C is a certain universal constant). Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, except with probability at most β we have that the randomly zoomed polynomial $g_{\lambda|\boldsymbol{x}}$ is (R, ε)-attenuated; i.e.,

$$\mathbf{HyperVar}_R[g_{\lambda|\boldsymbol{x}}] \leq \varepsilon \cdot \|g_{\lambda|\boldsymbol{x}}\|_2^2.$$

Another way to phrase the conclusion is that for $\boldsymbol{h} = g_{\lambda|\boldsymbol{x}}$, except with probability β we have that $\boldsymbol{h}$ is such that

$$\mathbf{HyperVar}_R[\boldsymbol{h}] \leq (d/\beta)^{O(\log d)} \cdot R^2 \lambda \cdot \|\boldsymbol{h}\|_2^2.$$

Notice that the dependencies here on R and λ are "correct" in the sense that if one intuitively thinks of λ as "infinitesimal", we expect that $\boldsymbol{h}$ will have $\Theta(\lambda)$ weight at level 1, negligible weight above level 1, and the definition of $\mathbf{HyperVar}_R$ multiplies this $\Theta(\lambda)$ level-1 weight by R^2. The "error factor" in this result, $(d/\beta)^{O(\log d)}$ with $\beta \sim \varepsilon_{\mathrm{PRG}}$, essentially becomes our final seed length (divided by $\log n$).

Because of the need to analyze the statistics introduced in Section 4.1.2, we will often need to work with a distribution over polynomials rather than a single polynomial. We therefore introduce the following generalization of Definition 50, which captures the notion of a distribution over polynomials being attenuated on average:

Definition 48 (Nice distribution of polynomials is attenuated on average)**.** Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice (in the sense of Section 4.2) distribution of polynomials over $\mathbb{R}^n$. For $R \geq 1$ and $0 < \varepsilon \leq 1$, we say that the distribution $(g_{\boldsymbol{v}})_{\boldsymbol{v} \in \Upsilon}$ is (R, ε)-*attenuated on average* if

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R[g_{\boldsymbol{v}}]] \leq \varepsilon \cdot \mathbf{E}_{\boldsymbol{v}}\big[\|g_{\boldsymbol{v}}\|_2^2\big].$$

The actual main result we prove in this section is Theorem 49, which generalizes Theorem 47 to a nice distribution of polynomials and is the original local hyperconcentration theorem from [OST20]:

Theorem 49 (Local hyperconcentration theorem for a nice distribution of polynomials)**.** *Let* $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ *be a nice (in the sense of Section 4.2) distribution of degree-d polynomials. Fix parameters* $R \geq 1$, $0 < \varepsilon \leq 1$, $0 < \beta < 1$, *and assume*

$$\lambda \leq \frac{\varepsilon}{R^2} \cdot \left(\frac{\beta}{d}\right)^{C \log d} \tag{24}$$

(where C is a certain universal constant). Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, except with probability at most β we have that the distribution $((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})_{\boldsymbol{v} \in \Upsilon}$ is (R, ε)-attenuated on average; i.e.,

$$\mathbf{E}_{\boldsymbol{v}}\big[\mathbf{HyperVar}_R[(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}]\big] \leq \varepsilon \cdot \mathbf{E}_{\boldsymbol{v}}\big[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2\big].$$

In Appendix B an improved version of Theorem 49, namely Theorem 85, is proved, which only requires an upper bound on λ of $\varepsilon\beta/(Rd^{O(1)})$. Theorem 45 follows from Theorem 85 directly by setting parameters as follows:

Proof of Theorem 45 using Theorem 85. We instantiate Theorem 85 with its nice distribution "$(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$" being $\mathcal{F}_{i,0}$, its "R" parameter being set to $\underline{R}$ defined in Equation (13), its "λ" parameter being $\underline{\lambda}$, its "ε" parameter being $\hat{\lambda}$, and its "β" parameter being $\varepsilon_{\mathrm{PRG}}/(8d)$. Recalling Equation (16) and Equation (17) we have

$$s_{i,0}(x) = \mathbf{E}_{\boldsymbol{v}}[g_{\boldsymbol{v}}(x)^2], \quad s_{i,1}(x) = \mathbf{E}_{\boldsymbol{v}}\big[\|(g_{\boldsymbol{v}})_{\underline{\lambda}|x}\|_2^2\big], \quad \text{and } s_{i+1,0}(x) = \mathbf{E}_{\boldsymbol{v}}\big[\mathbf{HyperVar}_{\underline{R}}[(g_{\boldsymbol{v}})_{\underline{\lambda}|x}]\big].$$

Recalling the settings of $\underline{\lambda}$ and $\hat{\lambda}$ from Equation (4) and Equation (20), we see that the bound required in Equation (24) indeed holds, and so we can apply Theorem 85, and its conclusion gives precisely the desired conclusion of Theorem 45. □

In the rest of this section we prove Theorem 49. We first explain the high-level structure of the argument in Section 5.1 and then give the formal proof in the rest of the section.

5.1   A useful definition, and the high-level argument underlying Theorem 49

Before we can give the high level idea of the proof of Theorem 49 we need a refined notion of a polynomial being attenuated:

Definition 50 (Attenuated polynomial, refined notion)**.** Let $g : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most d. For $k \geq 0$, $R \geq 1$, and $0 < \varepsilon \leq 1$, we say that the polynomial g is (k, R, ε)-*attenuated* if

$$\sum_{|\beta| > k} R^{2|\beta|} \widehat{g}(\beta)^2 = \mathbf{HyperVar}_R[g^{>k}] \leq \varepsilon \cdot \|g\|_2^2. \tag{25}$$

Similarly, if $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ is a nice (in the sense of Section 4.2) distribution of polynomials over $\mathbb{R}^n$, we say that the distribution $(g_{\boldsymbol{v}})_{\boldsymbol{v}\in\Upsilon}$ is (k, R, ε)-*attenuated on average* if

$$\mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\mathbf{HyperVar}_R[g_{\boldsymbol{v}}^{>k}]\Big] \leq \varepsilon \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|g_{\boldsymbol{v}}\|_2^2\big].$$

Note that g being $(0, R, \varepsilon)$-attenuated is the same as g being (R, ε)-attenuated as defined earlier (see Definition 29), and likewise for $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ (see Definition 48).

With this refined notion of attenuation in hand we can explain the high level idea of our local hyperconcentration theorem. For ease of exposition, below we sketch the underlying ideas in the "one-polynomial" setting of Theorem 47 (the same ideas drive the proof of Theorem 49).

So, we are given a degree-d polynomial g and the goal is to argue that with high probability over a random point $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, the polynomial $g_{\lambda|\boldsymbol{x}}$ is (R, ε)-attenuated, i.e. $(0, R, \varepsilon)$-attenuated. A simple but crucial insight pointing the way is that *random zooms compose*: in more detail, if $0 < \lambda, \lambda' < 1$ are two noise rates and $\boldsymbol{x}, \boldsymbol{x}'$ are two independent $\mathrm{N}(0,1)^n$ random variables, then the distribution of the composed random zoom $(g_{\lambda|\boldsymbol{x}})_{\lambda'|\boldsymbol{x}'}$ is identical to the distribution of $g_{\lambda\lambda'|\boldsymbol{z}}$ where $\boldsymbol{z} \sim \mathrm{N}(0,1)^n$. With this in mind, it is natural to view a random zoom at the small noise rate λ as a "strong" random zoom which is obtained by composing a sequence of $\log d$ many "weaker" random zooms at larger noise rates.[5] If we can prove that a "weak" random zoom with high probability causes a (k, R, ε)-attenuated polynomial to become $(k/2, R, \varepsilon)$-attenuated, then since any degree-d polynomial is trivially (d, R, ε)-attenuated, a simple union bound over $\log d$ many applications of this "one-stage" result yields the desired random zoom lemma for g. This is precisely the high-level structure of our argument; see Theorem 54 for a formal statement of the one-stage result in the more general setting of a nice distribution of polynomials.

We proceed to give intuition for the proof of the one-stage result. In this setting we are now given g which is a (k, R, ε)-attenuated polynomial; intuitively this means that the amount of Hermite weight it has at levels $k+1, k+2, \dots$ is very small compared to the total Hermite weight of g at all levels $0, 1, \dots$. We must argue that with high probability over $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, after a random zoom at $\boldsymbol{x}$ the polynomial $\boldsymbol{q} := g_{\lambda|\boldsymbol{x}}$ is $(k/2, R, \varepsilon)$-attenuated, i.e. the amount of Hermite weight $\boldsymbol{q}$ has at levels $k/2 + 1, k/2 + 2, \dots$ is very small relative to the total Hermite weight of $\boldsymbol{q}$ at levels $0, 1, \dots$. This is naturally done via a two part argument. The first part is to argue *two-norm retention*: this amounts to showing that with high probability over $\boldsymbol{x}$, the squared two-norm of $\boldsymbol{q}$ does not become too small relative to the squared two-norm of p. The argument for this is based on the Carbery–Wright anticoncentration bound (Theorem 33) and the tail bound for Gaussian polynomials (Theorem 22); see Section 5.2 for a precise statement and proof of this part. The second part is to argue *attrition* of the high-degree Hermite weight: this amounts to showing that with high probability after a random zoom, the amount of Hermite weight at levels $k/2 + 1, k/2 + 2, \dots$ becomes very small relative to the squared two-norm of p. The argument for this is based on Corollary 16 and Markov's inequality; see Section 5.3 for a precise statement and proof.

5.2 First part of the proof of the one-stage local hyperconcentration theorem: Retention

The main result of this section is Lemma 52. Its proof uses the following proposition:

---

[5]The idea of decomposing a "strong" random zoom into multiple "weak" random zooms is due to Avi Wigderson.

Proposition 51. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution of polynomials of degree at most k. Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, except with probability at most β' we have*

$$\mathop{\mathbf{E}}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2] \geq \left(\frac{\beta'}{O(k)}\right)^{2k} \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

Proof. Let $r(x) \coloneqq \mathbf{E}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|x}\|_2^2]$. We observe that r is a nonnegative degree-$2k$ polynomial with mean

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[r(\boldsymbol{x})] = \mathop{\mathbf{E}}_{\boldsymbol{v}} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{N}(0,1)^n}[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2] = \mathop{\mathbf{E}}_{\boldsymbol{v}} \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{x}' \sim \mathrm{N}(0,1)^n}[g_{\boldsymbol{v}}(\sqrt{1-\lambda}\boldsymbol{x} + \sqrt{\lambda}\boldsymbol{x}')]$$
$$= \mathop{\mathbf{E}}_{\boldsymbol{v}} \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{N}(0,1)^n}[g_{\boldsymbol{v}}(\boldsymbol{g})^2] = \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

The claimed result now follows immediately from the Carbery–Wright anticoncentration bound Theorem 33 applied to $p = r$, since $\|r\|_2 \geq \|r\|_1 = \mathbf{E}[r] = \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]$. $\qquad\square$

One way to think of the nice distribution of degree-k polynomials in Proposition 51 is that it is "$(k, +\infty, 1)$-attenuated on average." Lemma 52 relaxes this requirement and shows that a similar result holds for a nice distribution that is $(k, S, 1)$-attenuated on average for a modestly large S.

Lemma 52. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution of polynomials that is $(k, S, 1)$-attenuated on average for some $k \in \mathbb{N}^+$. Fix a parameter $0 < \beta' < 1$. Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, except with probability at most β' we have*

$$\mathop{\mathbf{E}}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2] \geq \left(\frac{\beta'}{O(k)}\right)^{2k} \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2],$$

provided that (for a certain universal constant C)

$$S \geq Ck \log(3/\beta')/\beta'.$$

Proof. We introduce the notation $\ell_{\boldsymbol{v}} \coloneqq g_{\boldsymbol{v}}^{\leq k}$ and $h_{\boldsymbol{v}}^j \coloneqq g_{\boldsymbol{v}}^{=k+j}$ for $j = 1, 2, 3, \ldots$, so $g_{\boldsymbol{v}} = \ell_{\boldsymbol{v}} + \sum_{j \geq 1} h_{\boldsymbol{v}}^j$. We may assume without loss of generality that $\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2] = 1$, or equivalently, $\mathbf{E}_{\boldsymbol{v}}[\|\ell_{\boldsymbol{v}}\|_2^2] + \sum_{j \geq 1} \mathbf{E}_{\boldsymbol{v}}[\|h_{\boldsymbol{v}}^j\|_2^2] = 1$. Since $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ is $(k, S, 1)$-attenuated on average, we have that

$$\mathop{\mathbf{E}}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2] = 1 \geq \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathbf{HyperVar}_S[g_{\boldsymbol{v}}^{>k}]\right] = \sum_{i>k} S^{2i} \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}^{=i}\|_2^2] \tag{26}$$
$$\geq S^{2k} \sum_{j \geq 1} \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|h_{\boldsymbol{v}}^j\|_2^2] = S^{2k}\left(1 - \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|\ell_{\boldsymbol{v}}\|_2^2]\right).$$

We may deduce that

$$\mathop{\mathbf{E}}_{\boldsymbol{v}}[\|\ell_{\boldsymbol{v}}\|_2^2] \geq 1 - 1/S^{2k} \geq .99,$$

where the latter inequality holds assuming C is large enough. From Proposition 51, we conclude that

$$\text{except with probability at most } \beta'/2 \text{ over } \boldsymbol{x}, \qquad \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2] \geq \left(\frac{\beta'}{C_1 k}\right)^{2k}, \tag{27}$$

where C_1 is a universal (large) constant. Our goal will be to establish the following: for all $j \geq 1$,

$$\text{except with probability at most } (\beta'/3)/10^j \text{ over } \boldsymbol{x}, \qquad \mathop{\mathbf{E}}_{\boldsymbol{v}}[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2] \leq \left(\frac{\beta'}{C_1 k}\right)^{2(k+j)}. \tag{28}$$

24

Before establishing Inequality (28), we show how it yields the conclusion of the lemma. Given Inequalities (27) and (28), summing over j and taking a union bound, we get that except with probability at most β' over $\boldsymbol{x}$,

$$\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2\big]} - \sum_{j \geq 1}\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2\Big]} \geq \frac{1}{2}\left(\frac{\beta'}{C_1 k}\right)^k. \tag{29}$$

The triangle inequality easily gives that for functions $a(v), b_1(v), b_2(v), \ldots$, we have that

$$\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[(a(\boldsymbol{v}) - \sum_j b_j(\boldsymbol{v}))^2\Big]} \geq \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}[a(\boldsymbol{v})^2]} - \sum_j\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}[b(\boldsymbol{v})^2]}.$$

Applying this (for each outcome of $\boldsymbol{x}$) with $a(\boldsymbol{v}) = \|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2$ and $b(\boldsymbol{v}) = \|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2$, by Equation (29) we get that

$$\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[(\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2 - \sum_{j\geq 1}\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2)^2\Big]} \geq \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2\big]} - \sum_{j\geq 1}\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2\Big]} \geq \frac{1}{2}\left(\frac{\beta'}{C_1 k}\right)^k,$$

from which we get (using $(\|f\| - \|g_1\| - \|g_2\| - \|g_3\| - \cdots)^2 \leq \|f + g_1 + g_2 + g_3 + \cdots\|^2$) that

$$\frac{1}{4}\left(\frac{\beta'}{C_1 k}\right)^{2k} \leq \mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[(\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2 - \sum_{j\geq 1}\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2)^2\Big] \leq \mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\|(\ell_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}} + \sum_{j\geq 1}(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2\Big] = \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2\big],$$

which is the conclusion of the lemma since $\mathbf{E}_{\boldsymbol{v}}\big[\|g_{\boldsymbol{v}}\|_2^2\big] = 1$.

Thus it remains to establish Inequality (28). To do this, write $s_j(\boldsymbol{x}) = \mathbf{E}_{\boldsymbol{v}}\Big[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2\Big]$ for brevity, and note that s_j is a nonnegative polynomial of degree at most $2(k + j)$. Similar to the proof of Proposition 51, we have that

$$\|s_j\|_1 = \mathop{\mathbf{E}}_{\boldsymbol{x}}[s_j(\boldsymbol{x})] = \mathop{\mathbf{E}}_{\boldsymbol{x}}\Big[\mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2\big]\Big] = \mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\mathop{\mathbf{E}}_{\boldsymbol{x}}[\|(h_{\boldsymbol{v}}^j)_{\lambda|\boldsymbol{x}}\|_2^2]\Big] = \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|h_{\boldsymbol{v}}^j\|_2^2\big],$$

and by Equation (26) we have that $\mathbf{E}_{\boldsymbol{v}}\Big[\|h_{\boldsymbol{v}}^j\|_2^2\Big] \leq 1/S^{2(k+j)}$. Now using Theorem 21 we get that $\|s_j\|_2 \leq (e/S)^{2(k+j)}$ and therefore (using Theorem 22, with its "t" set to $u^{2(k+j)}$) we get that

$$\mathbf{Pr}\Big[s_j(\boldsymbol{x}) \geq (ue/S)^{2(k+j)}\Big] \leq \exp(-(k+j)u^2/e) = \exp(-u^2 k/e)\exp(-u^2/e)^j$$

for any choice of $u \geq \sqrt{2e}$. We will select $u = \frac{\beta' S}{C_1 ek}$, so that the preceding inequality aligns with Inequality (28); recalling the bound on R, this choice of u is indeed at least $\sqrt{2e}$ provided C is taken at least $\sqrt{2}e^{3/2}C_1$. Also taking C sufficiently large in our assumption on S, it is not hard to arrange for the error probability above to be at most $(\beta'/3)(.1)^j$. Thus Inequality (28) is established and the proof of Lemma 52 is complete. $\qquad\square$

It is interesting to observe that both the results of this section, Proposition 51 and Lemma 52, hold with no dependence on the value of $0 < \lambda < 1$.

5.3 Second part of the proof of the one-stage local hyperconcentration theorem: Attrition

The attrition result we establish in this subsection, Lemma 53, is a fairly direct consequence of Corollary 16. (Note that Lemma 53 does not require that the nice distribution $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be attenuated on average — it holds for any nice distribution of degree-d polynomials.)

Lemma 53. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution of polynomials of degree at most d. Fix parameters $0 < \beta' < 1$, $R' \geq 1$, $m \in \mathbb{N}^+$, $0 < \varepsilon \leq 1$, let $c > 0$ be a sufficiently small constant, and assume $\lambda \leq \frac{c^4 \varepsilon \beta'^5}{R'^2 m^3 d}$. Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$,*

$$\frac{1}{\varepsilon} \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathbf{HyperVar}_{R'}[((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})^{\geq m}]\right] \leq \left(\frac{O(c\beta')}{m}\right)^{4m} \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\|g_{\boldsymbol{v}}\|_2^2\right]$$

holds except with probability at most β'.

Proof. The expectation of the left-hand side is

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathbf{HyperVar}_{R'}[((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})^{\geq m}]\right]\right] = \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\sum_{i \geq m} R'^{2i} \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\mathrm{W}^{=i}[(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}]\right]\right]$$

$$= \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\sum_{i \geq m} R'^{2i} \sum_{j \leq d} \mathbf{Pr}[\mathrm{Bin}(j,\lambda) = i] \mathrm{W}^{=j}[g_{\boldsymbol{v}}]\right] \qquad \text{(Corollary 16)}$$

$$= \sum_{j \leq d} \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathrm{W}^{=j}[g_{\boldsymbol{v}}]\right] \sum_{i \geq m} \mathbf{Pr}[\mathrm{Bin}(j,\lambda) = i] R'^{2i}$$

$$\leq \sum_{j \leq d} \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathrm{W}^{=j}[g_{\boldsymbol{v}}]\right] \sum_{i \geq m} \left(\frac{ej}{i}\right)^i \lambda^i R'^{2i}$$

$$\leq \sum_{j \leq d} \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\mathrm{W}^{=j}[g_{\boldsymbol{v}}]\right] \left(\frac{2ejR'^2\lambda}{m}\right)^m \qquad \text{(if } c \text{ small enough)}$$

$$\leq \left(\frac{2edR'^2\lambda}{m}\right)^m \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\|g_{\boldsymbol{v}}\|_2^2\right]$$

$$\leq \left(\frac{2ec^4\varepsilon\beta'^5}{m^4}\right)^m \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\|g_{\boldsymbol{v}}\|_2^2\right] \qquad \text{(by the bound on } \lambda)$$

$$\leq \beta'\varepsilon\left(\frac{O(c\beta')}{m}\right)^{4m} \mathop{\mathbf{E}}_{\boldsymbol{v}}\left[\|g_{\boldsymbol{v}}\|_2^2\right].$$

The result now follows by Markov's inequality. $\qquad\square$

5.4 Putting the pieces together: Proof of the local hyperconcentration theorem

Combining Lemma 52 and Lemma 53 (with the "k" and "m" parameters satisfying $m = \lfloor k/2 \rfloor$, and adjusting constants), we may deduce the following, which is our "one-stage local hyperconcentration theorem:"

Theorem 54 (One-stage local hyperconcentration theorem)**.** *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution of polynomials of degree at most d, and assume the distribution is $(k, S, 1)$-attenuated on average for*

some $1 \leq k \leq d$. Fix parameters $R' \geq 1$, $0 < \beta' < 1$, $0 < \varepsilon' \leq 1$, and assume

$$S \geq Ck\log(3/\beta')/\beta', \qquad \lambda \leq \frac{\varepsilon'\beta'^5}{CR'^2 k^3 d}$$

for a suitably large universal constant C. Then except with probability at most β' over $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, the distribution $((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})_{\boldsymbol{v} \sim \Upsilon}$ is $(\lfloor k/2 \rfloor, R', \varepsilon')$-attenuated on average.

Note that a nice distribution of degree-d polynomials is $(d, S, 1)$-attenuated on average for any S. We can take $S = Cd\log(3/\beta')/\beta'$ and perform a first application of Theorem 54 on $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ with its R' parameter set to S and its ε' parameter set to 1, and infer that except with failure probability at most β' the distribution $((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{z}})_{\boldsymbol{v} \sim \Upsilon}$ is $(\lfloor d/2 \rfloor, S, 1)$-attenuated on average. Repeating this a total of $\lceil \log d \rceil$ times, with each repetition having its β' parameter set to $\beta/(\lceil \log d \rceil + 1)$, its R' parameter set to $S = Cd\log(3/\beta')/\beta'$ (for simplicity), and its ε' parameter set to 1, we get that except with probability $\beta \cdot (1 - \frac{1}{\lceil \log d \rceil + 1})$, the distribution $((g_v)_{\lambda'|\boldsymbol{z}})_{v \in \Upsilon}$ is $(1, S, 1)$-attenuated on average, where $\lambda' = (\frac{\beta'^6}{C^2 d^6 \log^2(3/\beta')})^{\lceil \log d \rceil}$. Finally, we perform one last application of Theorem 54 with its k parameter set to 1, its λ' parameter set to $\underline{\lambda}$, its ε' parameter set to the "ε" of Theorem 49, and its β' parameter set to $\beta/(\lceil \log d \rceil + 1)$ and its R' parameter set to R. We get the conclusion of Theorem 49 as stated at the beginning of this section, and the proof of the local hyperconcentration theorem is complete. $\qquad \square$

6 Noise insensitivity of the statistics: Proof of Theorem 46

Remark 55. Before entering into the proof, we note that Theorem 46 is analogous to Lemma 11 of [Kan11b], which shows that for every i, j, for most $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, the (i, j)-th statistic at $\boldsymbol{x}$ is multiplicatively close to the $(i, j+1)$-th statistic at $\boldsymbol{x}$. [Kan11b]'s proof of Lemma 11 uses his Lemma 34 (i.e., [Kan11b, Lem. 9]) together with hypercontractivity, but as we discussed earlier, this incurs a $2^{O(d)}$ factor.

Our arguments in this section also use Lemma 9 of [Kan11b], but they additionally use our Local Hyperconcentration Theorem and our notions of attenuation and hyperconcentration (specifically Proposition 32). These new ingredients let us avoid the $2^{O(d)}$ factor which is incurred at this point in the [Kan11b] argument.

We proceed with the proof of Theorem 46. We begin by recording a simple corollary of Lemma 34:

Corollary 56. *In the setting of Lemma 34, say that $x \in \mathbb{R}^n$ is "good" if*

except with probability at most 0.1 over $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$ we have $\quad g_{\lambda|x}(\boldsymbol{y}) \approx_\gamma g(x)$.

Then $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ is good except with probability 10β.

Next, combining Theorem 85 with Proposition 32, we derive the following:

Proposition 57. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial and let $R \geq \sqrt{2}$, $\theta \leq 1$, $\gamma \leq 1$. Say that $x \in \mathbb{R}^n$ is "well-behaved" if*

except with probability $(2\sqrt{\theta}/\gamma)^{\frac{1}{2}R^2+1}$ over $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$ we have $\quad g_{\lambda|x}(\boldsymbol{y}) \approx_\gamma \mathbf{E}[g_{\lambda|x}]$.

Then $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ is well-behaved except with probability β, provided $\lambda \leq \frac{c\theta}{R^2} \cdot \left(\frac{\beta}{d}\right)^{9/2}$.

Given some $\gamma \leq 1$, let us take
$$\theta = .01\gamma^2, \quad R = \sqrt{2}.$$
It follows that if $x \in \mathbb{R}^n$ is both good and well-behaved, then
$$\mathbf{Pr}[g_{\lambda|x}(\boldsymbol{y}) \not\approx_\nu g(x)] \leq .1, \quad \mathbf{Pr}[g_{\lambda|x}(\boldsymbol{y}) \not\approx_\gamma \mathbf{E}[g_{\lambda|x}]] \leq .04.$$
Since $.1 + .04 < 1$, the only way this can happens is that $g(x) \approx_{\nu+\gamma} \mathbf{E}[g_{\lambda|x}]$. Thus the above two propositions imply that except with probability at most $10\beta + \beta = 11\beta$ over $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$, we have
$$g(\boldsymbol{x}) \approx_{\nu+\gamma} \mathbf{E}[g_{\lambda|\boldsymbol{x}}],$$
provided $\lambda \leq \gamma^2 (\beta/d)^{O(1)}$. Selecting $\gamma = \delta_{\mathrm{horz}}/2$ and $\beta = \varepsilon_{\mathrm{PRG}}/(88(d+1)D)$, we conclude (recalling that $D = \mathrm{poly}(d)$) that
$$\lambda \leq \delta_{\mathrm{horz}}^2 (\varepsilon_{\mathrm{PRG}}/d)^{O(1)} \quad \implies \quad \mathbf{Pr}[g(\boldsymbol{x}) \approx_{\delta_{\mathrm{horz}}} \mathbf{E}[g_{\lambda|\boldsymbol{x}}]] \leq \varepsilon_{\mathrm{PRG}}/(8(d+1)D).$$
Applying this with $g = s_{i,j}$ completes the proof of Theorem 46. $\qquad\square$

7 Proof of Theorem 8: one step of the Replacement Method

In this section we define a collection of "analysis checks," which are inequalities among the statistics, and explain the high-level structure of the proof of Theorem 8. The analysis checks play a crucial role in the proof of Theorem 8: as we explain in Section 7.2, two very different arguments (corresponding to Lemma 59 and Lemma 60) are used to establish the conclusion of Theorem 8 at a given $x \in \mathbb{R}^n$, depending on whether or not all of the analysis checks hold at that x.

7.1 Analysis checks

In this subsection we define our set of "analysis checks," which we denote AnalysisChecks. They are related to, but somewhat different from, the mollifier checks MollifierChecks that were used to define the mollifier $\mathrm{Mollifier}_p$ in Section 4.

One difference between the analysis checks and the mollifier checks is that since the mollifier checks needed to be "actually encoded into the mollifier," each one needed to consist of both an inequality Ineq among the statistics and a "softness" parameter δ. In contrast, the analysis checks only play a role in our analysis and do not need to be encoded in the mollifier, and for this reason each analysis check consists only of an inequality Ineq among the statistics. Other than this, the difference between the analysis checks and mollifier checks is that the analysis checks essentially correspond to the mollifier checks "shifted right by one in the grid."

Below we describe the analysis checks in more detail and highlight the difference between them and the mollifier checks.

Definition 58. The set AnalysisChecks contains the following checks (inequalities among statistics):

- **The horizontal checks:** for every $0 \leq i \leq d$, for every $1 \leq j \leq D-1$, we check that
$$s_{i,j}(x) \approx_{\delta_a} s_{i,j+1}(x) \tag{30}$$
(so $2(d+1)(D-1)$ inequalities in total for the horizontal checks), where we set
$$\delta_a := \frac{1}{100dD}. \tag{31}$$

(Looking ahead, we note that this choice of δ_a is less than the upper bound on γ imposed by the noise insensitivity extension lemma, Lemma 72, which is the main result of Section 9.) Note also that while the mollifier checks defined in Section 4.3 check $s_{i,j}(x)$ against $s_{i,j+1}(x)$ for $j = 0, \ldots, D - 2$, here we are checking $s_{i,j}(x)$ against $s_{i,j+1}(x)$ for $j = 1, \ldots, D - 1$. Thus these checks correspond precisely to the mollifier's noise-insensitivity checks, but "shifted to the right by one."

- **The diagonal checks:** for *just* the 1st column (note, *not* the 0th column), for all $i = 0, \ldots, d - 1$, we check that

$$s_{i+1,1}(x) \leq 100\hat{\lambda}s_{i,2}(x) \tag{32}$$

(so d diagonal checks in total). Note that while the strong anticoncentration checks defined in Section 4.3 check $s_{i+1,0}(x)$ against $s_{i,1}(x)$, here we are checking $s_{i+1,1}$ against $s_{i,2}(x)$. So similar to the previous bullet, these checks correspond precisely to the mollifier's strong anticoncentration checks, but again "shifted to the right by one."

Below we give an illustration of the analysis checks.
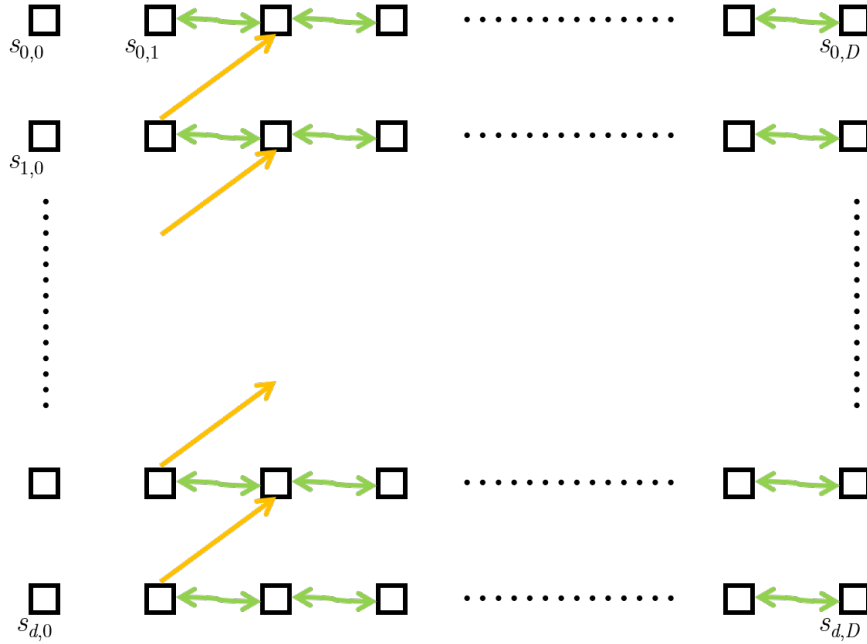
Figure 3: The analysis checks. Horizontal checks are depicted in green and diagonal checks are depicted in yellow.

7.2 High-level structure of the proof of Theorem 8

Let us define the following small integer parameter,

$$T = 4, \tag{33}$$

which will be the degree out to which we use Taylor's theorem.

Theorem 8 will be an immediate consequence of the below two lemmas, since by Equation (20), Equation (5) and Equation (4) we have that $\text{poly}((Td)^T) \cdot \hat{\lambda}^{T/2} \ll \varepsilon_{\text{PRG}}/(8L)$:

Lemma 59 (If analysis checks all pass, k_{indep}-wise moments determine mollifier's value)**.** *Suppose that $x \in \mathbb{R}^n$ is such that all of the checks in* ANALYSISCHECKS *hold at x, and that the random vector $\boldsymbol{z}$ is a k_{indep}-wise independent n-dimensional Gaussian. Then $\mathbf{E}\big[I_+\big(\sqrt{1-\underline{\lambda}} \cdot x + \sqrt{\underline{\lambda}} \cdot \boldsymbol{z}\big)\big]$ is determined up to an additive $\pm \mathrm{poly}((Td)^T) \cdot \hat{\lambda}^{T/2}$.*

Lemma 60 (If an analysis check fails, mollifier is close to zero)**.** *Suppose that $x \in \mathbb{R}^n$ is such that some check in* ANALYSISCHECKS *does not hold at x, and that the random vector $\boldsymbol{z}$ is a k_{indep}-wise independent n-dimensional Gaussian. Then $\mathbf{E}\big[I_+\big(\sqrt{1-\underline{\lambda}} \cdot x + \sqrt{\underline{\lambda}} \cdot \boldsymbol{z}\big)\big] \in [0, \mathrm{poly}((Td)^T) \cdot \hat{\lambda}^{T/2}]$.*

As we will see in the following sections, two very different arguments are used to prove Lemma 59 and Lemma 60. Lemma 59, which corresponds to the case in which x is such that all of the analysis checks pass, is based on a Taylor's theorem argument. In contrast, Lemma 60, which corresponds to the case in which x is such that some analysis check fails, employs a delicate argument, which takes advantage of the careful way that the analysis checks are structured vis-a-vis the mollifier checks, to argue that in this case almost all outcomes of $\boldsymbol{z}$ result in $I_+\big(\sqrt{1-\underline{\lambda}} \cdot x + \sqrt{\underline{\lambda}} \cdot \boldsymbol{z}\big) = 0$.

Before we can enter into the proofs of Lemma 59 and Lemma 60, there are several intermediate technical results which will be used in both proofs which we need to establish. We state and prove these technical results in Section 8 and Section 9, and prove Lemma 60 and Lemma 59 in Section 10 and Section 11 respectively.

8 Bounding the hypervariance of a statistic by its "neighbors"

The main goal of this section is to prove the following technical result which will be needed for our analysis. For every $x \in \mathbb{R}^n$, it gives an upper bound on the hypervariance of the zoom-at-x of our (i,j)-th statistic in terms of the values of some "nearby" statistics:

Theorem 61 (Bounding hypervariance of zooms (analogue of Proposition 12 of [Kan11b]))**.** *For all $i \in \{0, \dots, d-1\}$ all $j \in \{0, \dots, D-1\}$, and all $x \in \mathbb{R}^n$, it holds that*

$$\mathbf{HyperVar}_{\frac{\sqrt{\underline{R}}}{13}}\big[(s_{i,j})_{\underline{\lambda}|x}\big] \le 8(s_{i,j+1}(x) + s_{i+1,j}(x)) \cdot s_{i+1,j}(x)$$

$$\le 16 \max\{s_{i,j+1}(x)s_{i+1,j}(x), s_{i+1,j}(x)^2\}.$$

We stress that Theorem 61 holds for *every* input $x \in \mathbb{R}^n$. This is important because Theorem 61 will be used to prove Theorem 8, and in that setting we are dealing with an arbitrary $x \in \mathbb{R}^n$.

Remark 62. Theorem 61 is analogous to Proposition 12 of [Kan11b], which upper bounds the *variance* of the zoom-at-x of [Kan11b]'s (i,j)-th statistic in terms of the values of the $(i+1,j)$-th and $(i, j+1)$-th statistics. However, there is a factor of 2^d present in the bound of [Kan11b] (again because of hypercontractivity) which as always is incompatible with our goal of achieving an overall quasipolynomial rather than exponential dependence on d.

Before entering into the proof of Theorem 61, we record some corollaries and related results which we will use in Sections 10 and 11. Fix any $x \in \mathbb{R}^n$, let $\boldsymbol{z}$ denote a k_{indep}-wise independent n-dimensional Gaussian random vector, and let us write $\boldsymbol{w}$ to denote $\sqrt{1-\underline{\lambda}}x + \sqrt{\underline{\lambda}}\boldsymbol{z}$. Let us also introduce the notation

$$\underline{s}_{i,j} = s_{i,j}(x), \qquad \boldsymbol{s}_{i,j} = s_{i,j}(\boldsymbol{w}) = (s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{z}).$$

We first record the following:

Fact 63. *For all $0 \le i \le d$ and all $0 \le j \le D-1$, we have that*

$$\underline{s}_{i,j+1} = \mathbf{E}[\boldsymbol{s}_{i,j}] = |\mathbf{E}[\boldsymbol{s}_{i,j}]|. \tag{34}$$

Proof. We have

$$\underline{s}_{i,j+1} = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n}[(s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{y})] = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n}[(s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{z})] = \mathbf{E}[\boldsymbol{s}_{i,j}] = |\mathbf{E}[\boldsymbol{s}_{i,j}]|,$$

where the first equality is Definition 42, the second is because $s_{i,j}$ is a polynomial of degree at most $2d$ and $k_{\mathrm{indep}} \ge 2d$, and the last is by the non-negativity of $s_{i,j}$. $\qquad\square$

Next, as a corollary of Theorem 61 we have the following:

Corollary 64. *For all $0 \le i \le d-1$ and all $0 \le j \le D-1$,*

$$\boldsymbol{s}_{i,j} \text{ is } (2T, 4\max\{\sqrt{\zeta_{i,j}}, \zeta_{i,j}\})\text{-hyperconcentrated}, \qquad \text{where } \zeta_{i,j} := \frac{\underline{s}_{i+1,j}}{\underline{s}_{i,j+1}} \tag{35}$$

(where we interpret $0/0 = 0$).

Proof. We begin by noting that if $\underline{s}_{i,j+1} = 0$, then (recalling Remark 43 and Definition 40) it must be the case that $s_{i,0}$ is the constant-0 polynomial and hence $s_{i+1,j}$ is also zero; in this case $\boldsymbol{s}_{i,j}$ is the identically-0 random variable, which is certainly $(2T, 0)$-hyperconcentrated. Hence we subsequently assume that $\underline{s}_{i,j+1} > 0$.

By Theorem 61, we have that

$$\mathbf{HyperVar}_{\frac{\sqrt{\underline{R}}}{13}}\big[(s_{i,j})_{\underline{\lambda}|x}\big] \le 16\max\{\underline{s}_{i,j+1}\underline{s}_{i+1,j}, \underline{s}_{i+1,j}^2\}. \tag{36}$$

We apply Lemma 28 to the function $(s_{i,j})_{\underline{\lambda}|x}$, observing that the "$\mu$" of Lemma 28 is $\mathbf{E}_{\boldsymbol{y} \sim \mathrm{N}(0,1)^n}[(s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{y})] = \underline{s}_{i,j+1}$ and hence that Equation (36) lets us take the "θ" of Lemma 28 to be $16\max\{\zeta_{i,j}, \zeta_{i,j}^2\}$. Since $1 + \underline{R}/13 \ge 2T$, Lemma 28 thus gives that (for $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$) the random variable $(s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{y})$ is $(2T, 4\max\{\sqrt{\zeta_{i,j}}, \zeta_{i,j}\})$-hyperconcentrated. Since $(s_{i,j})_{\underline{\lambda}|x}(y)$ is a polynomial of degree at most $2d$ and $4dT = k_{\mathrm{indep}}$, the $2T$-th moments of $(s_{i,j})_{\underline{\lambda}|x}(\boldsymbol{y})$ and of $\boldsymbol{s}_{i,j}$ are identical (recall Remark 12). Now by the definition of hyperconcentration of a random variable we get that Inequality (35) holds as desired. $\qquad\square$

8.1   Proof of Theorem 61

Theorem 61 is proved using the following two results. We note that each of these results holds in a fairly general setting: in Lemma 65 $(q_v)_{v \in \Upsilon}$ can be any nice distribution of polynomials, and in fact both lemmas hold for polynomials over either Boolean space or Gaussian space (it will be clear from the proofs that they go through essentially unchanged in the Boolean context).

Lemma 65. *Let $(q_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution over polynomials (as defined in Section 4.2) over Gaussian space. Then for $R_0 := \frac{1}{13}\underline{R}^{1/4}$, we have*

$$\mathbf{HyperVar}_{R_0}\Big[\mathop{\mathbf{E}}_{\boldsymbol{v}}\big[q_{\boldsymbol{v}}^2\big]\Big] \le 8\Big(\mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\|q_{\boldsymbol{v}}\|_2^2\big] + \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\mathbf{HyperVar}_{\underline{R}}[q_{\boldsymbol{v}}]\big]\Big) \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[\mathbf{HyperVar}_{\underline{R}}[q_{\boldsymbol{v}}]\big].$$

Lemma 66. [6] *For all $0 \le i \le d$ and all $0 \le j \le D-1$, and all $x \in \mathbb{R}^n$, we have*

$$\mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}}[\mathbf{HyperVar}_{\underline{R}}[\boldsymbol{f}_{\underline{\lambda}|x}]] \le s_{i+1,j}(x).$$

---

[6] We note at this point that [Kan11b] appears to have a gap in the proof of its Proposition 12. Specifically, it is not true that the second equality following "Notice that" in that proof holds true (roughly speaking, because the derivative operator D and the noise operator N of [Kan11b] do not in general commute, as can be verified by considering the polynomial $p(x) = x_1 x_2$). The raison d'être of our Lemma 66 is to fill this gap.

Proof of Theorem 61 using Lemma 65 and Lemma 66. We instantiate Lemma 65 by taking the nice distribution $(q_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ to be $(\boldsymbol{f}_{\underline{\lambda}|x})_{\boldsymbol{f} \sim \mathcal{F}_{i,j}}$. With this choice $\mathbf{HyperVar}_{R_0}\big[\mathbf{E}_{\boldsymbol{v}}\big[q_{\boldsymbol{v}}^2\big]\big]$ corresponds to $\mathbf{HyperVar}_{R_0}\big[(s_{i,j})_{\underline{\lambda}|x}\big]$ (by Equation (16)), and $\mathbf{E}_{\boldsymbol{v}}\big[\|q_{\boldsymbol{v}}\|_2^2\big]$ corresponds to $s_{i,j+1}(x)$ (by Equation (17)). The final quantity on the right-hand side of Lemma 65, $\mathbf{E}_{\boldsymbol{v}}\big[\mathbf{HyperVar}_{R}[q_{\boldsymbol{v}}]\big]$, corresponds to $\mathbf{E}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}}[\mathbf{HyperVar}_{R}[\boldsymbol{f}_{\underline{\lambda}|x}]]$, so from Lemma 65 we get that

$$
\mathbf{HyperVar}_{R_0}\big[(s_{i,j})_{\underline{\lambda}|x}\big] \leq 8\left(s_{i,j+1}(x) + \mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}}[\mathbf{HyperVar}_{R}[\boldsymbol{f}_{\underline{\lambda}|x}]]\right) \cdot \mathop{\mathbf{E}}_{\boldsymbol{f} \sim \mathcal{F}_{i,j}}[\mathbf{HyperVar}_{R}[\boldsymbol{f}_{\underline{\lambda}|x}]]
$$
$$
\leq 8(s_{i,j+1}(x) + s_{i+1,j}(x)) \cdot s_{i+1,j}(x) \qquad\qquad \text{(Lemma 66)}
$$

and the proof of Theorem 61 is complete. $\qquad\square$

8.2 Proof of Lemma 65

Our main goal in this section is to prove the following:

Lemma 67. *Let $(q_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution over polynomials (as defined in Section 4.2) over Gaussian space. Define the function*
$$
g = \mathop{\mathbf{E}}_{\boldsymbol{v}}\big[q_{\boldsymbol{v}}^2\big].
$$

Then if $R_0 > \frac{3^{1/8}}{13}$, we have
$$
\mathbf{HyperVar}_{R_0}[g] \leq 8 \mathop{\mathbf{E}}_{\boldsymbol{v}}\Big[\|\mathrm{U}_{\sqrt{3}}q_{\boldsymbol{v}}\|_2^2\Big] \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}}[\mathbf{HyperVar}_{R}[q_{\boldsymbol{v}}]]
$$

where $R = 3e^4 R_0^4 \leq 165 R_0^4$.

Lemma 65 follows from Lemma 67 since $13^2 > 165$, $\mathbf{HyperVar}_{S}[p]$ is an increasing function of S for all p, and

$$
\|\mathrm{U}_{\sqrt{3}}q_v\|_2^2 = \sum_{|\alpha| \geq 0} 3^{|\alpha|}\widehat{q}_v(\alpha)^2 \qquad\qquad \text{(Equation (10) and Plancherel)}
$$
$$
= \sum_{|\alpha| \geq 0}\widehat{q}_v(\alpha)^2 + \sum_{|\alpha| \geq 1}(3^{|\alpha|} - 1)\widehat{q}_v(\alpha)^2
$$
$$
= \|q_v\|_2^2 + \mathbf{HyperVar}_{R}[q_v] \qquad\qquad \text{(Definition 27 and choice of } \underline{R})
$$

We will use the following lemma in the proof of Lemma 67:

Lemma 68. *Given any polynomial q and any $k \in \mathbb{N}^+$, we have*

$$
\mathrm{W}^{\geq k}[q^2] \leq 4 \cdot \left\|\mathrm{U}_{\sqrt{3}}(q^{\geq k/2})\right\|_2^2 \cdot \left\|\mathrm{U}_{\sqrt{3}}q\right\|_2^2.
$$

Proof. Let us write $q = \ell + h$ where $\ell = q^{<k/2}$ and $h = q^{\geq k/2}$. Since $(\ell^2)^{\geq k} = 0$, we have that

$$
(q^2)^{\geq k} = (\ell^2 + 2\ell h + h^2)^{\geq k} = (\ell^2)^{\geq k} + (2\ell h)^{\geq k} + (h^2)^{\geq k} = (h(2\ell + h))^{\geq k},
$$

and thus

$$
\mathrm{W}^{\geq k}[q^2] = \left\|(q^2)^{\geq k}\right\|_2^2 = \left\|(h(2\ell + h))^{\geq k}\right\|_2^2 \leq \|h(2\ell + h)\|_2^2 = \mathbf{E}[h^2(2\ell + h)^2] \leq \left\|h^2\right\|_2 \cdot \left\|(2\ell + h)^2\right\|_2,
$$

where the latter inequality is by Cauchy–Schwarz. By $(2,4)$-hypercontractivity (recalling Theorem 20), we have that

$$\|h^2\|_2 \le \|U_{\sqrt{3}}h\|_2^2$$

and similarly

$$\|(2\ell + h)^2\|_2 \le \|U_{\sqrt{3}}(2\ell + h)\|_2^2 = 4\|U_{\sqrt{3}}\ell\|_2^2 + \|U_{\sqrt{3}}h\|_2^2,$$

where the last equality holds because ℓ and h are orthogonal. Thus we have shown that

$$
\begin{aligned}
W^{\ge k}[q^2] &\le \left\| U_{\sqrt{3}}(q^{\ge k/2}) \right\|_2^2 \cdot \left(4\|U_{\sqrt{3}}\ell\|_2^2 + \|U_{\sqrt{3}}h\|_2^2 \right) \\
&\le \left\| U_{\sqrt{3}}(q^{\ge k/2}) \right\|_2^2 \cdot \left(4\|U_{\sqrt{3}}\ell\|_2^2 + 4\|U_{\sqrt{3}}h\|_2^2 \right) \\
&= 4 \cdot \left\| U_{\sqrt{3}}(q^{\ge k/2}) \right\|_2^2 \cdot \left\| U_{\sqrt{3}}q \right\|_2^2.
\end{aligned}
$$

$\square$

Proof of Lemma 67. We have

$$
\begin{aligned}
\mathbf{HyperVar}_{R_0}[g] &= \sum_{k \ge 1} R_0^{2k} \cdot \left\| g^{=k} \right\|_2^2 && \text{(Definition 27)} \\
&\le \sum_{k \ge 1} (eR_0)^{2k} \cdot \left\| g^{=k} \right\|_1^2 && \text{(Theorem 21)} \\
&= \sum_{k \ge 1} (eR_0)^{2k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{x} \sim N(0,1)^n} \left[\left| g^{=k}(\boldsymbol{x}) \right| \right]^2. && \text{(definition of one-norm)}
\end{aligned}
$$

Next we observe that

$$\left| g^{=k}(x) \right| = \left| \left(\mathop{\mathbf{E}}_{\boldsymbol{v}}[q_{\boldsymbol{v}}^2] \right)^{=k}(x) \right| = \left| \mathop{\mathbf{E}}_{\boldsymbol{v}} \left[(q_{\boldsymbol{v}}^2)^{=k} \right](x) \right| \le \mathop{\mathbf{E}}_{\boldsymbol{v}} \left[\left| (q_{\boldsymbol{v}}^2)^{=k}(x) \right| \right],$$

where the second equality holds because the operator which maps p to $p^{=k}$ (i.e. projecting to the k-th Wiener chaos) is a linear operator, and the inequality is the triangle inequality. Continuing the above, we deduce that

$$
\begin{aligned}
\mathbf{HyperVar}_{R_0}[g] &\le \sum_{k \ge 1} (eR_0)^{2k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{x}} \left[\mathop{\mathbf{E}}_{\boldsymbol{v}} \left[\left| (q_{\boldsymbol{v}}^2)^{=k}(x) \right| \right] \right]^2 \\
&= \sum_{k \ge 1} (eR_0)^{2k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}} \left[\left\| (q_{\boldsymbol{v}}^2)^{=k} \right\|_1 \right]^2 \\
&= \sum_{k \ge 1} (eR_0)^{2k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}} \left[\left\| (q_{\boldsymbol{v}}^2)^{=k} \right\|_2 \right]^2 && \text{(monotonicity of norms)} \\
&\le \sum_{k \ge 1} (eR_0)^{2k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}} \left[\sqrt{W^{\ge k}[q_{\boldsymbol{v}}^2]} \right]^2. && \left(\text{as } \left\| h^{=k} \right\|_2^2 \le \left\| h^{\ge k} \right\|_2^2 \text{ for any } h \right)
\end{aligned}
$$

Now we apply Lemma 68 to each q_v, which lets us continue as follows:

$$\mathbf{HyperVar}_{R_0}[g] \leq \sum_{k \geq 1} (eR_0)^{2k} \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[2\sqrt{\left\| \mathrm{U}_{\sqrt{3}}(q_{\overline{\boldsymbol{v}}}^{\geq k/2}) \right\|_2^2 \cdot \left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2} \right]^2$$

$$\leq 4 \sum_{k \geq 1} (eR_0)^{2k} \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}}(q_{\overline{\boldsymbol{v}}}^{\geq k/2}) \right\|_2^2 \right] \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2 \right] \qquad \text{(Cauchy–Schwarz)}$$

$$= 4 \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2 \right] \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\sum_{k \geq 1} (eR_0)^{2k} \cdot \left\| \mathrm{U}_{\sqrt{3}}(q_{\overline{\boldsymbol{v}}}^{\geq k/2}) \right\|_2^2 \right]$$

$$= 4 \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2 \right] \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\sum_{k \geq 1} (eR_0)^{2k} \cdot \sum_{j \geq k/2} 3^j \mathrm{W}^{=j}[q_{\boldsymbol{v}}] \right]$$

$$\text{(definition of } \left\| \mathrm{U}_{\sqrt{3}}(q_{\overline{\boldsymbol{v}}}^{\geq k/2}) \right\|_2^2)$$

$$\leq 8 \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2 \right] \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\sum_{j \geq 1} (3e^4 R_0^4)^j \cdot \mathrm{W}^{=j}[q_{\boldsymbol{v}}] \right]$$

$$= 8 \underset{\boldsymbol{v}}{\mathbf{E}} \left[\left\| \mathrm{U}_{\sqrt{3}} q_{\boldsymbol{v}} \right\|_2^2 \right] \cdot \underset{\boldsymbol{v}}{\mathbf{E}} \left[\mathbf{HyperVar}_{3e^4 R_0^4}[q_{\boldsymbol{v}}] \right],$$

completing the proof. $\qquad \square$

8.3 Proof of Lemma 66

We begin by re-expressing the right-hand side of Lemma 66:

$$s_{i+1,j}(x) = \underset{\boldsymbol{f} \sim \mathcal{F}_{i+1,j}}{\mathbf{E}} [\boldsymbol{f}(x)^2] \qquad \text{(Equation (16))}$$

$$= \underset{\boldsymbol{f} \sim \mathcal{F}_{i+1,0}}{\mathbf{E}} \, \underset{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j}{\mathbf{E}} (\mathrm{N}_{\boldsymbol{y}_1} \cdots \mathrm{N}_{\boldsymbol{y}_j} \boldsymbol{f})^2 \qquad \text{(definition of } \mathcal{F}_{i+1,j})$$

$$= \underset{\boldsymbol{f} \sim \mathcal{F}_{i+1,0}}{\mathbf{E}} \left[\underset{\boldsymbol{z}}{\mathbf{E}} [\boldsymbol{f}_{(1-\underline{\lambda})^j | \boldsymbol{z}}(x)^2] \right] \qquad \text{(semigroup property / Remark 43)}$$

$$= \underset{\boldsymbol{f} \sim \mathcal{F}_{i,0}}{\mathbf{E}} \, \underset{\boldsymbol{y}, \boldsymbol{y}'}{\mathbf{E}} \left[\underset{\boldsymbol{z}}{\mathbf{E}} \left[\left((\triangle_{\boldsymbol{y}, \boldsymbol{y}'}^{(R, \underline{\lambda})} \boldsymbol{f})_{(1-\underline{\lambda})^j | \boldsymbol{z}}(x) \right)^2 \right] \right], \qquad (37)$$

where the last equality is by the definition of $\mathcal{F}_{i+1,0}$ in terms of $\mathcal{F}_{i,0}$.

We similarly re-express the left-hand side of Lemma 66:

$$\underset{\boldsymbol{f} \sim \mathcal{F}_{i,j}}{\mathbf{E}} [\mathbf{HyperVar}_{\underline{R}}[\boldsymbol{f}_{\underline{\lambda} | x}]] = \underset{\boldsymbol{f} \sim \mathcal{F}_{i,j}}{\mathbf{E}} \, \underset{\boldsymbol{y}, \boldsymbol{y}'}{\mathbf{E}} \left[\left(\triangle_{\boldsymbol{y}, \boldsymbol{y}'}^{(R, \underline{\lambda})} \boldsymbol{f}(x) \right)^2 \right] \qquad \text{(Fact 39)}$$

$$= \underset{\boldsymbol{f} \sim \mathcal{F}_{i,0}}{\mathbf{E}} \, \underset{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j}{\mathbf{E}} \, \underset{\boldsymbol{y}, \boldsymbol{y}'}{\mathbf{E}} \left[\left((\triangle_{\boldsymbol{y}, \boldsymbol{y}'}^{(R, \underline{\lambda})} \mathrm{N}_{\boldsymbol{y}_1} \cdots \mathrm{N}_{\boldsymbol{y}_j} \boldsymbol{f})(x) \right)^2 \right] \qquad \text{(definition of } \mathcal{F}_{i,j})$$

$$= \underset{\boldsymbol{f} \sim \mathcal{F}_{i,0}}{\mathbf{E}} \, \underset{\boldsymbol{y}, \boldsymbol{y}'}{\mathbf{E}} \, \underset{\boldsymbol{z}}{\mathbf{E}} \left[\left((\triangle_{\boldsymbol{y}, \boldsymbol{y}'}^{(R, \underline{\lambda})} \boldsymbol{f}_{(1-\underline{\lambda})^j | \boldsymbol{z}})(x) \right)^2 \right], \qquad (38)$$

where the last equality is by the semigroup property / Remark 43. Comparing Equation (37) and Equation (38), Lemma 66 is an immediate consequence of Proposition 69, stated and proved below (setting its ρ parameter to be $(1 - \underline{\lambda})^j$), which states that the desired inequality holds "outcome by outcome" for outcomes of $\boldsymbol{f} \sim \mathcal{F}_{i,0}$.

Proposition 69. *For all $R \geq 1$ (and in particular $R = \underline{R}$), for every polynomial $g : \mathbb{R}^n \to \mathbb{R}$, and every $0 < \rho < 1$, we have that*

$$\mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{y'}} \mathop{\mathbf{E}}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R,\underline{\lambda})} g_{\rho|\boldsymbol{z}})(x)^2 \Big] \leq \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{y'}} \mathop{\mathbf{E}}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R,\underline{\lambda})} g)_{\rho|\boldsymbol{z}}(x)^2 \Big]. \tag{39}$$

Since the proof of Proposition 69 is somewhat involved we explain the high-level idea underlying it before entering into the technical details. When $R > 1$ the quantities in Equation (39) are somewhat difficult to work with since the Gaussian noise operator U_R, which is involved in the definition of the $\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R,\underline{\lambda})}$ operator, does not admit a convenient probabilistic interpretation (recall that for $R > 1$ the definition of U_R is through Equation (10)). The proof of Proposition 69 takes advantage of the fact that for $0 \leq R' \leq 1$, the quantity $\mathbf{E}_{\boldsymbol{y},\boldsymbol{y'}} \mathbf{E}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R',\underline{\lambda})} g_{\rho|\boldsymbol{z}})(x)^2 \Big]$ corresponding to the left-hand side of Equation (39) does have a natural probabilistic interpretation, and likewise for the quantity $\mathbf{E}_{\boldsymbol{y},\boldsymbol{y'}} \mathbf{E}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R',\underline{\lambda})} g)_{\rho|\boldsymbol{z}}(x)^2 \Big]$ corresponding to the right-hand side. These probabilistic interpretations let us give tractable expressions for each of the two quantities, and as we will see, it is evident from these expressions that the corresponding quantities correspond to polynomials in R' of degree at most $2d$. These polynomials can then be analyzed to show that the left-hand side is indeed at most the right-hand side for all $R \geq 1$, as asserted by the proposition.

Proof of Proposition 69. We define the function $p_{\mathrm{LHS}}(R')$ to be

$$p_{\mathrm{LHS}}(R') := \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{y'}} \mathop{\mathbf{E}}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R',\underline{\lambda})} g_{\rho|\boldsymbol{z}})(x)^2 \Big] \tag{40}$$

and the function $p_{\mathrm{RHS}}(R')$ to be

$$p_{\mathrm{RHS}}(R') := \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{y'}} \mathop{\mathbf{E}}_{\boldsymbol{z}} \Big[(\triangle_{\boldsymbol{y},\boldsymbol{y'}}^{(R',\underline{\lambda})} g)_{\rho|\boldsymbol{z}}(x)^2 \Big]. \tag{41}$$

Let

$$h(u) := g(\sqrt{\rho} R' \sqrt{1 - \underline{\lambda}} x + \sqrt{1 - \rho R'^2 (1 - \underline{\lambda})} u).$$

The following two claims provided the probabilistic interpretations alluded to earlier:

Claim 70. *For all $0 \leq R' \leq 1$, we have that*

$$p_{\mathrm{LHS}}(R') = \mathrm{Stab}_{\sigma_{\mathrm{LHS}}(R')}[h] - \mathrm{Stab}_{\tau_{\mathrm{LHS}}(R')}[h] = \sum_{\alpha} \widehat{h}(\alpha)^2 ((\sigma_{\mathrm{LHS}}(R'))^{|\alpha|} - (\tau_{\mathrm{LHS}}(R'))^{|\alpha|}), \text{ where}$$

$$\tag{42}$$

$$\sigma_{\mathrm{LHS}}(R') := \frac{1 - \rho + R'^2 \underline{\lambda} \rho}{1 - \rho R'^2 (1 - \underline{\lambda})}, \qquad \tau_{\mathrm{LHS}}(R') := \frac{1 - \rho}{1 - \rho R'^2 (1 - \underline{\lambda})}.$$

Claim 71. *For all $0 \leq R' \leq 1$, we have that*

$$p_{\mathrm{RHS}}(R') = \mathrm{Stab}_{\sigma_{\mathrm{RHS}}(R')}[h] - \mathrm{Stab}_{\tau_{\mathrm{RHS}}(R')}[h] = \sum_{\alpha} \widehat{h}(\alpha)^2 ((\sigma_{\mathrm{RHS}}(R'))^{|\alpha|} - (\tau_{\mathrm{RHS}}(R'))^{|\alpha|}), \text{ where}$$

$$\tag{43}$$

$$\sigma_{\mathrm{RHS}}(R') := \frac{(1 - \underline{\lambda})(1 - \rho) R'^2 + R'^2 \underline{\lambda}}{1 - \rho R'^2 (1 - \underline{\lambda})}, \qquad \tau_{\mathrm{RHS}}(R') := \frac{(1 - \underline{\lambda})(1 - \rho) R'^2}{1 - \rho R'^2 (1 - \underline{\lambda})}.$$

Proof of Claim 70. For all $0 \leq R' \leq 1$, we have that

$$p_{\text{LHS}}(R') = \frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}} \left[\left(\mathrm{U}_{R'} g_{\rho|\boldsymbol{z}}(\sqrt{1-\underline{\lambda}} x + \sqrt{\underline{\lambda}} \boldsymbol{y}) - \mathrm{U}_{R'} g_{\rho|\boldsymbol{z}}(\sqrt{1-\underline{\lambda}} x + \sqrt{\underline{\lambda}} \boldsymbol{y}') \right)^2 \right] \tag{44}$$

$$= \frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}} \left[\left(\mathop{\mathbf{E}}_{\boldsymbol{v}} \left[g_{\rho|\boldsymbol{z}} \left(R'(\sqrt{1-\underline{\lambda}} x + \sqrt{\underline{\lambda}} \boldsymbol{y}) + \sqrt{1-R'^2} \boldsymbol{v} \right) \right] \right. \right.$$

$$\left. \left. - \mathop{\mathbf{E}}_{\boldsymbol{v}'} \left[g_{\rho|\boldsymbol{z}} \left(R'(\sqrt{1-\underline{\lambda}} x + \sqrt{\underline{\lambda}} \boldsymbol{y}') + \sqrt{1-R'^2} \boldsymbol{v}' \right) \right] \right)^2 \right] \tag{45}$$

$$= \frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}} \left[\left(\mathop{\mathbf{E}}_{\boldsymbol{v}} \left[g(\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho} R' \sqrt{1-\underline{\lambda}} x + \sqrt{\rho} R' \sqrt{\underline{\lambda}} \boldsymbol{y} + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}) \right] \right. \right.$$

$$\left. \left. - \mathop{\mathbf{E}}_{\boldsymbol{v}'} \left[g(\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho} R' \sqrt{1-\underline{\lambda}} x + \sqrt{\rho} R' \sqrt{\underline{\lambda}} \boldsymbol{y}' + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}') \right] \right)^2 \right], \tag{46}$$

where Equation (44) is by definition of $\triangle_{\boldsymbol{y}, \boldsymbol{y}'}^{(R', \underline{\lambda})}$, Equation (45) is by Definition 17 (the probabilistic definition of $\mathrm{U}_{R'}$, valid when $0 \leq R' \leq 1$), and Equation (46) is by definition of the zoom. Let us define

$$h(u) \coloneqq g(\sqrt{\rho} R' \sqrt{1-\underline{\lambda}} x + \sqrt{1 - \rho R'^2 (1-\underline{\lambda})} u), \tag{47}$$

so expanding the square, we may re-express Equation (46) as

$$\frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}} \left[h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y} + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \cdot h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y} + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \right] \tag{48}$$

$$- \mathop{\mathbf{E}}_{\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{v}'} \left[h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y} + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \cdot h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y}' + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}'}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \right] \tag{49}$$

$$+ \frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{y}', \boldsymbol{z}, \boldsymbol{v}'} \left[h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y}' + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}'}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \cdot h\left(\frac{\sqrt{1-\rho} \boldsymbol{z} + \sqrt{\rho R'^2 \underline{\lambda}} \boldsymbol{y}' + \sqrt{\rho} \sqrt{1-R'^2} \boldsymbol{v}'}{\sqrt{1 - \rho R'^2 (1-\underline{\lambda})}} \right) \right], \tag{50}$$

where all the random variables above are distributed as $\mathrm{N}(0,1)^n$. It is easy to see that $(48) = (50)$, and inspection reveals that both quantities are equal to $\frac{1}{2} \mathrm{Stab}_{\sigma_{\text{LHS}}}[h]$. Inspection also reveals that $(49) = \mathrm{Stab}_{\tau_{\text{LHS}}}[h]$, giving the first equality of Equation (42). The second equality of Equation (42) follows from the Hermite formula for Stab given in Definition 19, and the proof of Claim 70 is complete. $\qquad \square$

The proof of Claim 71 is very similar to the above proof so we omit it.

To complete the proof of Proposition 69, we must show that $p_{\text{LHS}}(R) \leq p_{\text{RHS}}(R)$ for all $R \geq 1$. By Claim 70 and Claim 71, this would follow immediately from showing that

$$(\sigma_{\text{LHS}}(R))^{|\alpha|} - (\tau_{\text{LHS}}(R))^{|\alpha|} \leq (\sigma_{\text{RHS}}(R))^{|\alpha|} - (\tau_{\text{RHS}}(R))^{|\alpha|}.$$

Letting a plays the role of $|\alpha|$ and clearing the common denominator of $1 - \rho R^2 (1 - \underline{\lambda})$ that is present in all of $\sigma_{\text{LHS}}, \tau_{\text{LHS}}, \sigma_{\text{RHS}}, \tau_{\text{RHS}}$, it remains to show the following: for all natural numbers a and all real $R \geq 1$,

$$(R^2 \underline{\lambda} \rho - \rho + 1)^a - (1 - \rho)^a \leq (R^2 (1 - \underline{\lambda})(1 - \rho) + R^2 \underline{\lambda})^a - (R^2 (1 - \underline{\lambda})(1 - \rho))^a. \tag{51}$$

Equation (51) is a consequence of the following stronger inequality (obtained by replacing the quantity $(R^2 (1 - \underline{\lambda})(1 - \rho))^a$ in Equation (51) by the larger quantity $(R^2 (1 - \rho))^a$):

$$(R^2 \underline{\lambda} \rho - \rho + 1)^a - (1 - \rho)^a \leq (R^2 (1 - \underline{\lambda})(1 - \rho) + R^2 \underline{\lambda})^a - (R^2 (1 - \rho))^a. \tag{52}$$

Equation (52) can be rewritten as

$$(R^2\underline{\lambda}\rho + (1-\rho))^a - (1-\rho)^a \le (R^2\underline{\lambda}\rho + R^2(1-\rho))^a - (R^2(1-\rho))^a, \tag{53}$$

which is of the form

$$(x+y_0)^a - y_0^a \le (x+y_1)^a - y_1^a \tag{54}$$

where $x = R^2\underline{\lambda}\rho \ge 0$, $y_1 = R^2(1-\rho)$, and $y_0 = (1-\rho)$; recalling that $R \ge 1$, we have $y_0 \le y_1$. Expanding out both sides of Equation (54) using the binomial theorem, the right-hand side is at least as large as the left-hand side term by term, and the proof of Proposition 69, and hence also Lemma 66, is complete. □

9   Noise insensitivity extension lemma

For technical reasons our analysis will require a technical result which we state and prove below. Intuitively, this result says that if $x \in \mathbb{R}^n$ is an input to a degree-d polynomial $r(\cdot)$ at which poly(d) many successive "noisifications" of r, at increasing but all small noise rates, are all multiplicatively close to each other, then they are all multiplicatively close to the value $r(x)$.

Recall that $D = (2d+1)^2$ and that $\underline{\lambda} = \left(\frac{\varepsilon_{\mathrm{PRG}}}{d}\right)^{O(1)}$. The lemma is as follows (recall that the notation "$a \approx_\gamma b$" means that $\exp(-\gamma) \le \frac{a}{b} \le \exp(\gamma)$):

Lemma 72 (Noise insensitivity extension lemma (analogue of Corollary 16 of [Kan11b])). *Let $r_0 : \mathbb{R}^n \to \mathbb{R}$ be a non-negative degree-$(2d)$ polynomial. Let $a + b = 1$ and suppose $0 \le b \le \frac{1}{Cd^{10}}$ for a suitable large absolute constant C. For $1 \le j \le D$ write $r_j(x)$ to denote $\mathrm{U}_{a^{j/2}}r_0$.*

Suppose $x \in \mathbb{R}^n$ is a point such that for all $1 \le j \le D-1$ we have $r_j(x) \approx_\gamma r_{j+1}(x)$, where $\gamma \le \frac{1}{12D(2d+1)}$. Then $r_0(x) \approx_1 r_1(x)$, i.e. $e^{-1} \le \frac{r_0(x)}{r_1(x)} \le e$.

We note that later when we apply this lemma it will be with the polynomial r_0 instantiated to be a zeroth-column statistic $s_{i,0}$, of degree $2d$, and with $a = 1 - \underline{\lambda}$, so we will have that $b = \underline{\lambda} = \left(\frac{\varepsilon_{\mathrm{PRG}}}{d}\right)^{O(1)}$ satisfies $b \le \frac{1}{Cd^{10}}$ with room to spare. Recalling Remark 43, Lemma 72 implies that if the statistics $s_{i,1}(x),\ldots,s_{i,D}(x)$ are all multiplicatively close to each other then $s_{i,0}(x)$ is also multiplicatively (fairly) close to this common value.

It is interesting to contast Lemma 72 with Corollary 16 of [Kan11b]. That corollary gives a qualitatively similar result, also establishing constant-factor multiplicative closeness of $r_0(x)$ as its conclusion, but is quantitatively very different in the assumptions it uses to reach that conclusion. In Corollary 16 of [Kan11b] only $O(d)$ many noisifications $r_1,\ldots,r_{O(d)}$ are considered, but they are assumed to be much closer to each other, multiplicatively $(1 \pm \exp(-d))$-close (and it can be shown that such a strong assumption is required if only $O(d)$ many noisifications are considered). In contrast, Lemma 72 assumes closeness now of poly(d) rather than $O(d)$ many noisifications, but the closeness that we need to assume is much weaker, only multiplicative $(1 \pm \frac{1}{\mathrm{poly}(d)})$-closeness; this is crucial for our overarching goal of "getting rid of all factors of 2^d."

Proof of Lemma 72. Recall from Equation (10) that for any fixed x and varying ρ, the quantity

$$\mathrm{U}_\rho r_0(x) = \sum_{|\alpha| \le d} \rho^{|\alpha|} \widehat{r_0}(\alpha) h_\alpha(x) \coloneqq A(\rho)$$

is a polynomial in ρ of degree at most $2d$. Let $A_* \coloneqq A((1-q)^{1/2}) = r_1(x)$. The hypothesis of Lemma 72 tells us that for all $j = 1,\ldots,D$, we have $A((1-q)^{j/2}) \approx_{\gamma D} A_*$; defining the polynomial

37

$B(\rho) \coloneqq \frac{A(\rho)}{A_*} - 1$, we get that

$$-\gamma D \leq e^{-\gamma D} - 1 \leq B((1-q)^{j/2}) \leq e^{\gamma D} - 1 \leq 2\gamma D$$

for all $j = 1, \ldots, D$. Next, let us define the degree-$2d$ polynomial $C(\rho)$ by

$$C((2/q)\rho) \coloneqq B(1-\rho), \qquad \text{so} \qquad C((2/q)(1 - (1-q)^{j/2})) = B((1-q)^{j/2}) \in [-\gamma D, 2\gamma D]$$

for $j = 1, \ldots, D$. For notational convenience, for $j = 1, \ldots, D$ we write "j'" to denote the value $(2/q)(1 - (1-q)^{j/2})$, and we observe that

$$j\left(1 - \frac{1}{d^5}\right) \leq j' \leq j, \tag{55}$$

where the upper bound is immediate and the lower bound holds (with room to spare) since by assumption we have $q \leq \frac{1}{Cd^{10}}$. So intuitively, we have that $C(1'), \ldots, C(D')$ are all very close to zero — between $-\gamma D$ and $2\gamma D$ — and to prove the lemma it suffices to show that $C(0) \in [-1/2, 1/2]$.

We do this using Lagrange interpolation. Recall that the Lagrange interpolation formula tells us that for any degree-$2d$ polynomial C and any $2d+1$ points $x_1, \ldots, x_{d+1}$, we have

$$C(x) = \sum_{j=1}^{2d+1} C(x_j)\ell_j(x), \qquad \text{where} \qquad \ell_j(x) = \prod_{m \in [1,d+1]\setminus j} \frac{x - x_m}{x_j - x_m}. \tag{56}$$

We apply this formula at $x = 0$ where we take the $2d+1$ values x_i to be $x_i = (i^2)'$. Fix a $j \in [1, d]$ and let us consider $\ell_j(0)$; it is equal to

$$\ell_j(0) = \frac{(1^2)' \cdot (2^2)' \cdots ((j-1)^2)' \cdot ((j+1)^2)' \cdots ((2d+1)^2)'}{((1^2)' - (j^2)') \cdot ((2^2)' - (j^2)') \cdots (((j-1)^2)' - (j^2)') \cdot (((j+1)^2)' - (j^2)') \cdots (((2d+1)^2)' - (j^2)')}$$

Note that in the preceding expression, every multiplicand in the numerator is of the form a' for some integer $a \in [1, D]$ and every multiplicand in the denominator is of the form $(a' - b')$ for distinct integers $a, b \in [1, D]$. It follows straightforwardly from this and from Equation (55) that $\ell_j(0)$ is within a multiplicative $\left[1 - \frac{1}{d}, 1 + \frac{1}{d}\right]$ factor of the above expression "without the primes", i.e. of

$$\frac{1^2 \cdot 2^2 \cdots (j-1)^2 \cdot (j+1)^2 \cdots (2d+1)^2}{(1^2 - j^2) \cdot (2^2 - j^2) \cdots ((j-1)^2 - j^2) \cdot ((j+1)^2 - j^2) \cdots ((2d+1)^2 - j^2)}. \tag{57}$$

Now we require the following bound on the above fraction, which we prove after using it to finish the proof of Lemma 72:

Claim 73. *For all $j \in \{1, \ldots, 2d+1\}$ it holds that $|(57)| \leq 2$.*

It follows that for each j we have $|\ell_j(0)| \leq 2(1 + \frac{1}{d}) < 3$, and hence by Equation (56) we have that $|C(0)| \leq 6(d+1)\gamma D < 1/2$. This proves Lemma 72. $\qquad \square$

Proof of Claim 73. We have that

$$|(57)| = \left| \prod_{i=1}^{2d+1} \frac{i^2}{i^2 - j^2} \right| = \overbrace{\left(\prod_{i=1}^{j-1} \frac{i^2}{(j-i)(j+i)} \right)}^{=A} \cdot \overbrace{\left(\prod_{i=j+1}^{2d+1} \frac{i^2}{(i-j)(i+j)} \right)}^{=B}, \quad \text{where by inspection}$$

$$A = \frac{1 \cdot 2 \cdots \cdots (j-1)}{(j+1) \cdot (j+2) \cdots \cdots (2j-1)}, \quad \text{and}$$

$$B \le B' := \prod_{i=j+1}^{\infty} \frac{i^2}{(i-j)(i+j)} = \frac{(j+1)(j+2) \cdots \cdots (2j)}{1 \cdot 2 \cdots \cdots j}, \quad \text{so}$$

$$|(57)| = A \cdot B \le A \cdot B' = \frac{2j}{j} = 2. \qquad \qquad \Box$$

10 Proof of Lemma 60: if some analysis check fails, then with high probability some mollifier check fails

As per the assumptions of Lemma 60, in this section we completely fix an $x = \underline{x} \in \mathbb{R}^n$ which is such that some check in ANALYSISCHECKS does not hold at x, and we let $\boldsymbol{z}$ denote a k_{indep}-wise independent n-dimensional Gaussian random vector. We recall the notation from the start of Section 8,

$$\boldsymbol{w} = \sqrt{1 - \underline{\lambda}}\underline{x} + \sqrt{\underline{\lambda}}\boldsymbol{z}, \qquad \underline{s}_{i,j} = s_{i,j}(\underline{x}), \qquad \boldsymbol{s}_{i,j} = s_{i,j}(\boldsymbol{w}) = (s_{i,j})_{\underline{\lambda}|\underline{x}}(\boldsymbol{z}),$$

and we remark that we will be making extensive use of Corollary 64 in the arguments that follow.

Recalling the statement of Lemma 60, we assume through the rest of Section 10 that $\underline{x}$ causes some analysis check to fail, and our goal is to show that

$$0 \le \mathbf{E}[I_+(\sqrt{1 - \underline{\lambda}} \cdot \underline{x} + \sqrt{\underline{\lambda}} \cdot \boldsymbol{z})] \le \text{poly}(Td^T) \cdot \hat{\lambda}^{T/2}.$$

Recalling that I_+ is the product of functions bounded in $[0, 1]$ (namely the indicator of $\text{sign}(p) = 1$ and all of the $\text{SoftCheck}_{\text{CHECK}}$ functions as CHECK ranges over MOLLIFIERCHECKS), to prove Lemma 60 it suffices to establish the following:

There exists some CHECK $\in$ MOLLIFIERCHECKS "$s_u \ge c s_v$ with softness δ",

such that $\text{SoftCheck}_{\text{CHECK}}(\sqrt{1 - \underline{\lambda}} \cdot \underline{x} + \sqrt{\underline{\lambda}} \cdot \boldsymbol{z}) = 0$ — equivalently, $\boldsymbol{s}_u < \exp(-\delta) \cdot c\boldsymbol{s}_v$ —

except with probability at most $\text{poly}((Td)^T) \cdot (\hat{\lambda})^{T/2}$ over $\boldsymbol{z}$. (58)

To establish Inequality (58), we will consider the checks in ANALYSISCHECKS in a careful order, specifically, the order shown below. All subsequent references to the "first" analysis check that fails, "earlier" or "later" analysis checks, etc. are with respect to this ordering.
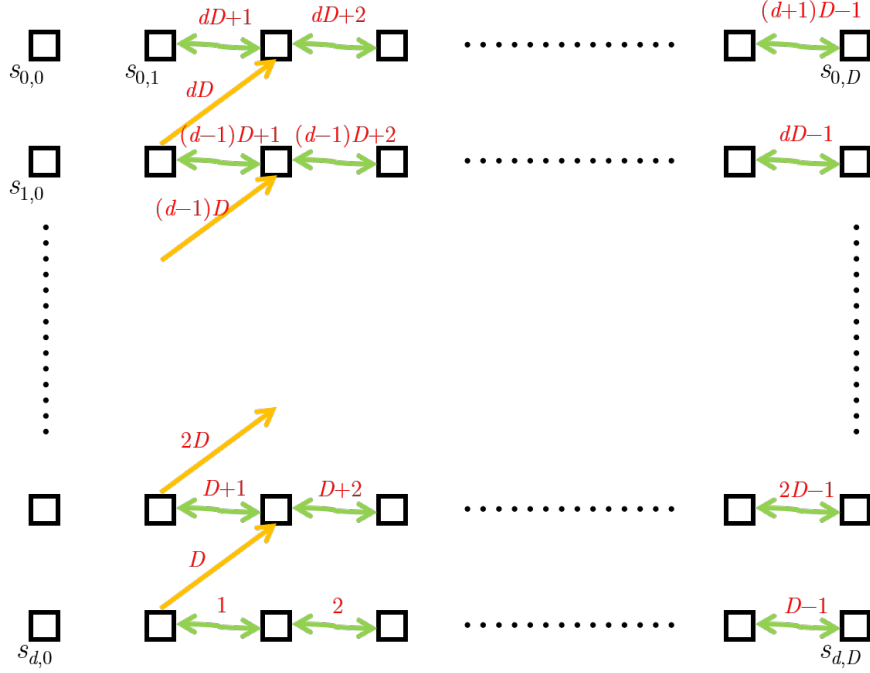
Figure 4: Ordering of analysis checks (indicated in red)

The argument has three cases depending on where is the first analysis check that fails for $\underline{x}$ (a horizontal check in the bottom row; a horizontal check in a higher row; or a diagonal check). Before entering into the case analysis, which involves detailed and careful arguments, we stress two high level points. First, the overall *qualitative* structure of the following arguments follows [Kan11b] quite closely (in particular page 13 of that paper). Second, to obtain our *quantitative* improvement over [Kan11b] (essentially, getting a $\text{poly}(d^T) = \text{poly}(d)$ factor in the failure probability of Inequality (58) rather than the $2^{(O(d)}$ factor that is present in [Kan11b]), crucially requires the technical tools that we developed in Section 8 and Section 9.

10.1 The first failing analysis check is horizontal and is in the bottom row $(i = d)$.

Recall that the horizontal analysis checks in the bottom row are $\underline{s}_{d,j} \approx_{\delta_a} \underline{s}_{d,j+1}$ for all $1 \leq j \leq D-1$. But Definition 40 and Fact 38 imply that $s_{d,0}$ is a constant function, and it follows from Definition 42 that $s_{d,j}$ is the same constant function for all j. Thus all numbers $\underline{s}_{d,j}$ are equal to the same constant, and hence the analysis checks in the bottom row cannot actually fail. So this case cannot occur.

10.2 The first failing analysis check is horizontal and in some row $0 \leq i < d$

Suppose that the first analysis check to fail is one of the two implicit in the statement

$$\underline{s}_{i^*,j^*} \not\approx_{\delta_a} \underline{s}_{i^*,j^*+1}, \tag{59}$$

for some $0 \leq i^* \leq d - 1$ and $1 \leq j^* \leq D - 1$. We first note that (similar to the beginning of the proof of Corollary 64) if one of the two quantities $\underline{s}_{i^*,j^*}, \underline{s}_{i^*,j^*+1}$ is zero then $s_{i,0}$ must be the constant-0 polynomial and hence the other quantity must be zero as well. But since Equation (59)

holds, it cannot be the case that $\underline{s}_{i^*,j^*}$ and $\underline{s}_{i^*,j^*}$ are both zero. Hence in the rest of the proof we assume that $\underline{s}_{i^*,j^*}, \underline{s}_{i^*,j^*+1} > 0$.

In this case we will analyze the random variables $\boldsymbol{s}_{i^*,j^*-1}$ and $\boldsymbol{s}_{i^*,j^*}$. By Fact 63 we have that

$$\mathbf{E}[\boldsymbol{s}_{i^*,j^*-1}] = \underline{s}_{i^*,j^*}, \qquad \mathbf{E}[\boldsymbol{s}_{i^*,j^*}] = \underline{s}_{i^*,j^*+1}, \qquad \text{and let us write } M \coloneqq \max\{\underline{s}_{i^*,j^*}, \underline{s}_{i^*,j^*+1}\}. \quad (60)$$

Since Equation (59) is the first analysis check to fail, it must be the case that all horizontal analysis checks in the (i^*+1)-th row passed, i.e.

$$\text{for all } 1 \le j \le D-1, \text{ we have } \underline{s}_{i^*+1,j} \approx_{\delta_a} \underline{s}_{i^*+1,j+1},$$

which immediately gives that

$$\text{for all } 1 \le j \le D, \text{ we have } \underline{s}_{i^*+1,j} \approx_{D\delta_a} \underline{s}_{i^*+1,1}.$$

Now we apply the noise insensitivity extension lemma, Lemma 72, to the degree-$(2d)$ polynomial $s_{i^*,0}$ in place of "r_0", with δ_a in place of "γ," and with $\underline{\lambda}$ in place of "b." Since δ_a is indeed less than $\frac{1}{12D(2d+1)}$ as required by that lemma, Lemma 72 gives us that we can extend the above closeness (with slightly weaker parameters) even to the $j = 0$ case:

$$\text{for all } 0 \le j \le D, \text{ we have } \underline{s}_{i^*+1,j} \approx_1 \underline{s}_{i^*+1,1}. \quad (61)$$

As another consequence of the fact that Equation (59) is the first analysis check to fail, we have that the diagonal analysis check relating the (i^*+1)-th to the i^*-th row passed, i.e. we have that

$$\underline{s}_{i^*+1,1} \le 100\hat{\lambda}\underline{s}_{i^*,2}. \quad (62)$$

Combining Equations (61) and (62), we conclude that

$$\text{for all } 0 \le j \le D, \text{ we have } \underline{s}_{i^*+1,j} \le 100e\hat{\lambda}\underline{s}_{i^*,2}. \quad (63)$$

Now, recalling Corollary 64, we have that

$$\boldsymbol{s}_{i^*,j^*-1} \text{ is } (T, 4\max\{\sqrt{\zeta_{i^*,j^*-1}}, \zeta_{i^*,j^*-1}\})\text{-hyperconcentrated, where } \zeta_{i^*,j^*-1} = \frac{\underline{s}_{i^*+1,j^*-1}}{\underline{s}_{i^*,j^*}} \le \frac{100e\hat{\lambda}\underline{s}_{i^*,2}}{\underline{s}_{i^*,j^*}},$$

$$\boldsymbol{s}_{i^*,j^*} \text{ is } (T, 4\max\{\sqrt{\zeta_{i^*,j^*}}, \zeta_{i^*,j^*}\})\text{-hyperconcentrated, where } \zeta_{i^*,j^*} = \frac{\underline{s}_{i^*+1,j^*}}{\underline{s}_{i^*,j^*+1}} \le \frac{100e\hat{\lambda}\underline{s}_{i^*,2}}{\underline{s}_{i^*,j^*+1}},$$

where both inequalities are by Equation (63). (Note that the above ratios are well-defined since $\underline{s}_{i^*,j^*}, \underline{s}_{i^*,j^*+1} > 0$.)

To analyze the $\underline{s}_{i^*,2}$ factor which appears in both numerators above, we consider two cases. If $j^* = 1$, then $\underline{s}_{i^*,2} = \underline{s}_{i^*,j^*+1} \le M$ (recalling Equation (60)). Otherwise, by virtue of the fact that all preceding analysis checks in the i^*-th row passed, we conclude that $\underline{s}_{i^*,2} \approx_{D\delta_a} \underline{s}_{i^*,j^*}$ and hence $\underline{s}_{i^*,2} \le eM$ (since $D\delta_a \le e$). Either way, we conclude that

$$\zeta_{i^*,j^*-1} \le \frac{100e^2\hat{\lambda}M}{\underline{s}_{i^*,j^*}}, \quad \zeta_{i^*,j^*} \le \frac{100e^2\hat{\lambda}M}{\underline{s}_{i^*,j^*+1}}. \quad (64)$$

For the rest of the analysis of this case, we will reason the exact same way about $\boldsymbol{s}_{i^*,j^*-1}$ and about $\boldsymbol{s}_{i^*,j^*}$. Let us write $\boldsymbol{s}$ to denote either $\boldsymbol{s}_{i^*,j^*-1}$ or $\boldsymbol{s}_{i^*,j^*}$, and we similarly write just ζ for

either ζ_{i^*,j^*-1} or ζ_{i^*,j^*}, and write $\mu = \mathbf{E}[\mathbf{s}]$ (note that recalling Equation (60), we have that μ is either $\underline{s}_{i^*,j^*}$ or $\underline{s}_{i^*,j^*+1}$). Inequality (64) tells us that in either case we have

$$\zeta \leq 100e^2\hat{\lambda}(M/\mu). \tag{65}$$

Now we apply Proposition 25 to the random variable $\mathbf{s}$ (which was shown above to be $(T, 4\max\{\sqrt{\zeta}, \zeta\})$-hyperconcentrated), taking its "t" parameter to be $\frac{M/\mu}{d^C}$ for a large absolute constant C, to deduce that

$$|\mathbf{s} - \mu| \leq M/d^C \text{ except with probability at most } \left(\frac{4\max\{\sqrt{\zeta}, \zeta\} \cdot d^C}{M/\mu}\right)^T \leq \text{poly}(d^T) \cdot \hat{\lambda}^{T/2}. \tag{66}$$

(To justify the last inequality, we observe that if $\zeta \geq 1$ then $\max\{\sqrt{\zeta}, \zeta\}/(M/\mu) = \zeta/(M/\mu) \leq 100e^2\hat{\lambda}$ by Equation (65), which is at most $O(\sqrt{\hat{\lambda}})$ since $\hat{\lambda} \leq 1$. On the other hand, if $\zeta < 1$ then $\max\{\sqrt{\zeta}, \zeta\}/(M/\mu) = \sqrt{\zeta}/(M/\mu)$, which is at most $10e\sqrt{\hat{\lambda}}/\sqrt{M/\mu}$ by Equation (65), which is in turn at most $10e\sqrt{\hat{\lambda}}$ using $M = \max\{\underline{s}_{i^*,j^*}, \underline{s}_{i^*,j^*+1}\} \geq \mu$.)

Thus except with a $\text{poly}(d^T) \cdot \hat{\lambda}^{T/2}$ failure probability, each of $\mathbf{s}_{i^*,j^*-1}, \mathbf{s}_{i^*,j^*}$ is within an additive $\pm M/d^C$ of its mean. Without loss of generality (the other case is entirely similar), let us assume that $\mathbf{s}_{i^*,j^*}$ has the larger mean, so $\mu = M = \mathbf{E}[\mathbf{s}_{i^*,j^*}] = s_{i^*,j^*+1}$; given this, Equation (59) tells us that

$$\underline{s}_{i^*,j^*} \leq \exp(-\delta_a)\underline{s}_{i^*,j^*+1} = \exp(-\delta_a)M.$$

Now, even if $\underline{s}_{i^*,j^*} = \mathbf{E}[\mathbf{s}_{i^*,j^*-1}]$ were as large as possible (by the above, this largest possible value is $\exp(-\delta_a)M$), except with overall failure probability at most $\text{poly}(d^T) \cdot \hat{\lambda}^{T/2}$, we have that both

$$\mathbf{s}_{i^*,j^*-1} \leq s_{i^*,j^*} + \frac{M}{d^C} \leq \exp(-\delta_a)M + \frac{M}{d^C}$$

(by Equation (66) applied with its $\mathbf{s}$ being $\mathbf{s}_{i^*,j^*-1}$ and its μ being s_{i^*,j^*}) and

$$\mathbf{s}_{i^*,j^*} \geq M - \frac{M}{d^C}$$

(by Equation (66) applied with its $\mathbf{s}$ being $\mathbf{s}_{i^*,j^*}$ and its μ being M). Recalling the definitions of δ_{horz} and δ_a from Equation (20) and Equation (31), and that the C above is a large constant, the two preceding inequalities imply that Inequality (58) holds for the $(i^*, j^* - 1)$ vs. (i^*, j^*) noise insensitivity mollifier check (see Section 4.3.2), as desired.

10.3 The first failing analysis check is diagonal

Finally, the last case we must consider is that the first analysis check to fail is the diagonal check

$$\underline{s}_{i^*+1,1} \not\leq 100\hat{\lambda}\underline{s}_{i^*,2}$$

for some $0 \leq i^* \leq d - 1$. In this case we analyze the random variables $\mathbf{s}_{i^*+1,0}$ and $\mathbf{s}_{i^*,1}$.

We first observe that by Equation (34) and the above inequality we can lower bound the expectation of $\mathbf{s}_{i^*+1,0}$ by

$$\mathbf{E}[\mathbf{s}_{i^*+1,0}] = \underline{s}_{i^*+1,1} > 100\hat{\lambda}\underline{s}_{i^*,2} = 100\hat{\lambda}\,\mathbf{E}[\mathbf{s}_{i^*,1}], \tag{67}$$

which will be useful for us later. Next we give a high-probability lower bound on $\mathbf{s}_{i^*+1,0}$:

Claim 74. $s_{i^*+1,0} \geq \frac{1}{2}\underline{s}_{i^*+1,1}$ *except with probability at most* $(80e)^T \cdot \hat{\lambda}^{T/2}$.

Proof. The claim is immediate from Proposition 25, taking its "t" to be $1/2$, once we establish the following:
$$s_{i^*+1,0} \text{ is } (T, 40e\sqrt{\hat{\lambda}})\text{-hyperconcentrated.} \tag{68}$$

To establish Inequality (68), first suppose that $i^*+1 = d$. In this case $s_{i^*+1,0}$ is a constant function, so $s_{i^*+1,0}$ is a constant random variable, and Inequality (68) is clearly true. The other possibility is that $i^* + 1 < d$. In this case since $i^* < d - 1$, there must have been at least one earlier diagonal analysis check, and it succeeded, meaning that
$$\underline{s}_{i^*+2,1} \leq 100\hat{\lambda}\underline{s}_{i^*+1,2}.$$

We further have that the horizontal analysis checks in rows $i^* + 1$ and $i^* + 2$ all succeeded, and hence
$$\underline{s}_{i^*+1,2} \approx_{\delta_a} \underline{s}_{i^*+1,1}, \qquad \underline{s}_{i^*+2,1} \approx_1 \underline{s}_{i^*+2,0},$$

where the second inequality is by an application of Lemma 72 (which we may apply because $\delta_a \leq \frac{1}{12D(2d+1)}$) to the degree-$2d$ polynomial $s_{i^*+2,0}$. Upper-bounding e^{δ_a} by e for simplicity and combining these relations, we deduce that
$$\underline{s}_{i^*+2,0} \leq 100e^2\hat{\lambda}\underline{s}_{i^*+1,1}.$$

Now Inequality (68) follows from Corollary 64 applied to $s_{i,j} = s_{i^*+1,0}$. $\square$

Next we establish a high-probability upper bound on $s_{i^*,1}$:

Claim 75. $s_{i^*,1} \leq \underline{s}_{i^*+1,1}/(20\hat{\lambda})$ *except with probability at most* $16^T \cdot \hat{\lambda}^{T/2}$.

Proof. Applying Inequality (35) to $s_{i^*,1}$, we get that $s_{i^*,1}$ is $(T, 4\max\{\sqrt{\zeta_{i^*,1}}, \zeta_{i^*,1}\})$-hyperconcentrated, and Inequality (67) tells us that $\zeta_{i^*,1}/\hat{\lambda} > 100$. Applying Proposition 25 to $s_{i^*,1}$ with its "t" parameter set to $\zeta_{i^*,1}/(20\hat{\lambda}) - 1 > \zeta_{i^*,1}/(40\hat{\lambda})$, we get that except with failure probability at most
$$\left(\frac{\eta}{t}\right)^T \leq \left(\frac{4\max\{\sqrt{\zeta_{i^*,1}}, \zeta_{i^*,1}\}}{\zeta_{i^*,1}/(40\hat{\lambda})}\right)^T = (160\hat{\lambda}\max\{\zeta_{i^*,1}^{-1/2}, 1\})^T, \tag{69}$$

we have
$$s_{i^*,1} \leq (t+1)\mathbf{E}[s_{i^*,1}] = (\zeta_{i^*,1}/(20\hat{\lambda})) \cdot \mathbf{E}[s_{i^*,1}] = (\zeta_{i^*,1}/(20\hat{\lambda})) \cdot \underline{s}_{i^*,2} = \underline{s}_{i^*+1,1}/(20\hat{\lambda}),$$

where the second equality is by Fact 63 and the third is by the definition of $\zeta_{i^*,1}$ (recalling Inequality (35)). Using again $\zeta_{i^*,1} > 100\hat{\lambda}$, the failure probability bound Equation (69) is at most $16^T \cdot \hat{\lambda}^{T/2}$. $\square$

Putting Claim 74 and Claim 75 together, we conclude that
$$s_{i^*+1,0} \geq 10\hat{\lambda}s_{i^*,1} \quad \text{except with probability } O(\hat{\lambda})^{T/2},$$

which establishes Inequality (58) for the i^*-th local hyperconcentration mollifier check (see Section 4.3.1), as desired.

11 Proof of Lemma 59: Using a Taylor-based argument if all analysis checks pass

As per the assumptions of Lemma 59, in this section we completely fix an $x = \underline{x} \in \mathbb{R}^n$ such that all of the analysis checks pass, and we let $\boldsymbol{z}$ denote a k_{indep}-wise independent n-dimensional Gaussian random vector. We recall the notation from the start of Section 8,

$$\boldsymbol{w} = \sqrt{1 - \underline{\lambda}}\underline{x} + \sqrt{\underline{\lambda}}\boldsymbol{z}, \qquad \underline{s}_{i,j} = s_{i,j}(\underline{x}), \qquad \boldsymbol{s}_{i,j} = s_{i,j}(\boldsymbol{w}) = (s_{i,j})_{\underline{\lambda}|\underline{x}}(\boldsymbol{z}),$$

and we note that we will again be making use of Corollary 64 in the arguments that follow. We further introduce the notation

$$\mu_p := \mathbf{E}[p(\boldsymbol{w})].$$

To prove Lemma 59 we must show that the expectation of the random variable

$$I_+(\boldsymbol{w}) = \text{Mollifier}_p(\boldsymbol{w}) \cdot \mathbf{1}[\text{sign}(p(\boldsymbol{w}))], \tag{70}$$

is determined up to an additive $\pm\text{poly}((Td)^T)\cdot\hat{\lambda}^{T/2}$ just by virtue of $\boldsymbol{z}$ being k_{indep}-wise independent. To do this it is useful to observe that $\text{Mollifier}_p(\boldsymbol{w})$ depends only on the $\boldsymbol{s}_{i,j}$ random variables; more precisely, we may rewrite Equation (70) as

$$I_+(\boldsymbol{w}) = \widetilde{\text{Mollifier}}_p(\{\boldsymbol{s}_{i,j}\}_{0\leq i\leq d,0\leq j\leq D-1}) \cdot \mathbf{1}[\text{sign}(p(\boldsymbol{w}))], \tag{71}$$

where the function $\widetilde{\text{Mollifier}}_p : (\mathbb{R}^{\geq 0})^{(d+1)\times D} \to [0,1]$ is defined in the obvious way,

$$\widetilde{\text{Mollifier}}_p(\{s_{i,j}\}_{0\leq i\leq d,0\leq j\leq D-1}) := \prod_{(\text{INEQ}=(c,s_u,s_v),\delta)\in\text{MOLLIFIERCHECKS}} \sigma\left(\delta^{-1}\ln\left(\frac{s_u}{cs_v}\right)\right). \tag{72}$$

We begin to analyze Equation (71) by first analyzing the simpler random variable $\mathbf{1}[\text{sign}(p(\boldsymbol{w}))]$ which is a part of $I_+(\boldsymbol{w})$:

Claim 76. $\mathbf{Pr}[\text{sign}(p(\boldsymbol{w})) \neq \text{sign}(\mu_p)] \leq O(\hat{\lambda})^{T/2}.$

Proof. Recalling that $s_{0,0} = p^2$, by Definition 42 we have that

$$\underline{s}_{0,1} = \mathbf{E}_{\boldsymbol{y}\sim\text{N}(0,1)^n}[p_{\underline{\lambda}|\underline{x}}(\boldsymbol{y})^2],$$

and by Fact 39 and Definition 40 we have that

$$\underline{s}_{1,0} = \mathbf{HyperVar}_{\underline{R}}[p_{\underline{\lambda}|\underline{x}}].$$

Furthermore, since by the assumption of Lemma 59 we have that all the analysis checks pass at $\underline{x}$, we may draw the following conclusions:

$$\underline{s}_{1,1} \leq 100\hat{\lambda}\underline{s}_{0,2}, \quad \underline{s}_{0,1} \approx_{\delta_a} \underline{s}_{0,2}, \quad \underline{s}_{1,0} \approx_1 \underline{s}_{1,1},$$

where the first of these is by the $i = 1$ diagonal analysis check passing, the second is by the $(i,j) = (0,1)$ horizontal analysis check passing, and the last of these follows from the noise insensitivity extension lemma Lemma 72 and the passing of the horizontal analysis checks $s_{1,j}(\underline{x}) \approx_{\delta_a} s_{1,j+1}(\underline{x})$ for all $1 \leq j \leq D - 1$. Combining these bounds we deduce that

$$\mathbf{HyperVar}_{\underline{R}}[p_{\underline{\lambda}|\underline{x}}] = \underline{s}_{1,0} \leq 100e\hat{\lambda}\underline{s}_{0,2} = 100e^2\hat{\lambda}\mathbf{E}_{\boldsymbol{y}\sim\text{N}(0,1)^n}[p_{\underline{\lambda}|\underline{x}}(\boldsymbol{y})^2],$$

or in other words $p_{\underline{\lambda}|\underline{x}}$ is $(\underline{R}, 100e^2\hat{\lambda})$-attenuated. Since $T \leq 1 + \frac{1}{2}\underline{R}^2$, by Lemma 31 we may conclude that for $\boldsymbol{y} \sim \mathrm{N}(0,1)^n$, the random variable $p_{\underline{\lambda}|\underline{x}}(\boldsymbol{y})$ is $(T, 10e\sqrt{\hat{\lambda}})$-hyperconcentrated. Since the definition of $(T, 10e\sqrt{\hat{\lambda}})$-hyperconcentration only uses T-th moments, and $p_{\underline{\lambda}|\underline{x}}$ has degree at most d and $Td \leq k_{\mathrm{indep}}$, recalling that $\boldsymbol{z}$ is k_{indep}-wise independent we may conclude that the random variable $p(\boldsymbol{w})$ is also $(T, 10e\sqrt{\hat{\lambda}})$-hyperconcentrated. (Recall Remark 12.) Now applying Proposition 25 with its "t" parameter set to $1/2$, we get that

$$|p(\boldsymbol{w}) - \mu_p| \leq (1/2)|\mu_p|, \quad \text{except with probability at most } O(\hat{\lambda})^{T/2},$$

which is easily seen to imply the claim if $\mu_p \neq 0$. (If $\mu_p = 0$, then the above condition translates into $p(\boldsymbol{w}) = 0$ with probability at least (say) 0.9, which can only be the case if p is identically 0, in which case the claim holds trivially.) $\qquad\square$

Since $\widetilde{\mathrm{Mollifier}}_p$ is bounded in $[0,1]$, as an immediate consequence of Claim 76 we have that $\mathbf{E}[I_+(\boldsymbol{w})]$ and $\mathbf{E}[\widetilde{\mathrm{Mollifier}}_p(\{\boldsymbol{s}_{i,j}\}_{0 \leq i \leq d, 0 \leq j \leq D-1})]$ can differ by at most an additive $O(\hat{\lambda})^{T/2}$. Hence the remaining task, to prove Lemma 59, is to show that

$$\widetilde{\mathrm{Mollifier}}_p(\{\boldsymbol{s}_{i,j}\}_{0 \leq i \leq d, 0 \leq j \leq D-1}) \text{ is determined up to an additive } \pm \mathrm{poly}((Td)^T) \cdot \hat{\lambda}^{T/2}. \quad (73)$$

In the rest of this section we do this as follows: first, in Section 11.1, we use the assumption that all analysis checks pass at $\underline{x}$ to show that the statistics $\boldsymbol{s}_{i,j}$ are suitably hyperconcentrated. Next, in Section 11.2 we show that the $\widetilde{\mathrm{Mollifier}}_p$ function satisfies a certain technical "relaxedness" condition (essentially a bound on the magnitude of its derivatives). Finally, in Section 11.3 we use hyperconcentration of the $\boldsymbol{s}_{i,j}$ statistics and relaxedness of $\mathrm{Mollifier}_p$ in an argument based on Taylor's theorem to establish Equation (73).

11.1 Establishing hyperconcentration of the $\boldsymbol{s}_{i,j}$'s

The proof of the following lemma uses the assumption that all analysis checks pass at $\underline{x}$:

Lemma 77. *For $0 \leq i \leq d$, $0 \leq j \leq D-1$, the random variable $\boldsymbol{s}_{i,j}$ is $(2T, 40e\sqrt{\hat{\lambda}})$-hyperconcentrated.*

Proof. Fix any $0 \leq i \leq d-1$. Since the diagonal analysis check between the i-th and $(i+1)$-th row passes at $\underline{x}$, we have (recalling Equation (32)) that $\underline{s}_{i+1,1} \leq 100\hat{\lambda}\underline{s}_{i,2}$. Similar to the beginning of the proof of Corollary 64, if $\underline{s}_{i,2} = 0$ then $\boldsymbol{s}_{i,j}$ must be the constant-0 random variable and the lemma holds. So we assume that $\underline{s}_{i,2} > 0$ and we have that

$$\frac{\underline{s}_{i+1,1}}{\underline{s}_{i,2}} \leq 100\hat{\lambda}. \quad (74)$$

Further, since the horizontal analysis checks in rows i and $i+1$ all passed at $\underline{x}$, we can apply Lemma 72 to $s_{i,0}$ and $s_{i+1,0}$ to get that

$$\text{for all } 0 \leq j \leq D, \quad \underline{s}_{i,j} \approx_1 \underline{s}_{i,2}, \quad \underline{s}_{i+1,j} \approx_1 \underline{s}_{i+1,1}. \quad (75)$$

Combining Equation (74) and Equation (75), and recalling the notation $\zeta_{i,j} = \frac{\underline{s}_{i+1,j}}{\underline{s}_{i,j+1}}$ from Inequality (35), we may conclude that

$$\text{for all } 0 \leq i \leq d-1, 0 \leq j \leq D-1 \text{ we have that } \zeta_{i,j} \leq 100e^2\hat{\lambda}.$$

45

Putting this into Inequality (35), we get that

$$\text{for } 0 \le i \le d-1,\ 0 \le j \le D-1, \quad \boldsymbol{s}_{i,j} \text{ is } (2T, 40e\sqrt{\hat{\lambda}})\text{-hyperconcentrated.}$$

This hyperconcentration trivially extends to $\boldsymbol{s}_{d,j}$ because $\boldsymbol{s}_{d,j}$ is a constant random variable, and Lemma 77 is proved. $\qquad\square$

11.2 $\widetilde{\text{Mollifier}}_p$ is relaxed

For notational simplicity in the remainder of this section, rather than using the (i,j)-indexing for the elements of $\boldsymbol{s} = (\{\boldsymbol{s}_{i,j}\}_{0 \le i \le d, 0 \le j \le D-1})$, we will use generic indices $1 \le u \le m := (d+1)D = \text{poly}(d)$, so we write $\boldsymbol{s}$ as $\boldsymbol{s} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m)$.

Definition 78. For $a \in \mathbb{N}$, $B \ge 1$, we say a function $\psi : (\mathbb{R}^{\ge 0})^m \to [0,1]$ is (a,B)-*relaxed* if it is smooth and satisfies

$$\text{for all } \alpha \in \mathbb{N}^n \text{ with } 0 \le |\alpha| \le a \text{ and all } s \in (\mathbb{R}^{\ge 0})^m, \qquad |s|^\alpha \cdot |\partial_\alpha \psi(s)| \le B.$$

Lemma 79 ($\widetilde{\text{Mollifier}}_p$ is relaxed)**.** *The function $\psi = \widetilde{\text{Mollifier}}_p$ is $(2T, \text{poly}((Td)^T))$-relaxed.*

Proof. Let us write B_0 to denote the maximum of $1/\delta$ where δ ranges over all of the "softness parameters" involved in the definition of $\psi = \widetilde{\text{Mollifier}}_p$ (recall Equation (72)). Recalling Section 4.3.1 and Section 4.3.2 we have that $B_0 = 1/\delta_{\text{horz}} = \text{poly}(d)$.

Say that a function $\phi(s)$ is "ψ-like" if it takes the form of the right-hand side of Equation (72), except that some of the multiplicands may have derivatives σ', σ'', σ''', etc., in place of σ. We show by induction on $a = |\alpha|$ that $\partial_\alpha \psi(s)$ consists of a sum of at most $(m+1)^a$ terms, each being of the form $b\frac{\phi(s)}{s^\alpha}$, where $\phi(s)$ is ψ-like and b is a constant that is at most $(aB_0)^a$ in magnitude. The base case $a = 0$ is immediate. For the induction step, we consider differentiating a term $b\frac{\phi(s)}{s^\alpha}$ with respect to some s_u. We view this term as a product of up to $m+1$ factors involving s_u, namely the multiplicands in $\phi(s)$ involving s_u, and also any power s_u^j in the denominator. Now we use the calculus product rule. Differentiating a multiplicand of the form $\sigma^{(i)}(\pm \delta^{-1} \ln(s_u/s_v) + \text{const.})$ with respect to s_u gives a similar factor, but with a higher derivative $\sigma^{(i+1)}$ and picking up a factor of $\pm \delta^{-1}/s_u$. Thus we indeed get another ψ-like term, with an extra factor of s_u in the denominator and a constant factor increased in magnitude by at most $B_0 \le aB_0$, as is sufficient for the induction. Similarly, differentiating the factor of $1/s_u^j$ picks up a constant factor of $j \le a \le aB_0$ in magnitude, as well as an extra factor of s_u in the denominator. This completes the induction.

Next, we observe that in all the ψ-like terms that are present in $\partial_\alpha \psi(s)$, the maximum-order derivative on σ that arises is at most $a = |\alpha|$. Recalling Definition 35, for $a > 0$ all of these ψ-like terms are uniformly bounded in magnitude on $(\mathbb{R}^{\ge 0})^m$ by $a^{O(a)}$ (and when $a = 0$ we have that the desired inequality holds since $|s|^0 \cdot |\psi(s)| = |\psi(s)| \le 1$). Given the induction and these observations, it is clear that ψ is $(2T, B)$-relaxed, where $B = (m+1)^{2T} \cdot (2TB_0)^{2T} \cdot O(2T)^{2T} = \text{poly}((Td)^T)$. $\quad\square$

Remark 80. Lemma 79 is analogous to Lemma 18 of [Kan11b], which similarly gives an upper bound on the partial derivatives of the mollifier of [Kan11b]. The upper bound given in [Kan11b] is exponential in d because of $\exp(d)$-type factors which are involved in the definition of the mollifier in that work.

11.3 The core Taylor's theorem argument

In this subsection we prove the following lemma:

Lemma 81. *Let $\boldsymbol{s} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m)$ be a vector of nonnegative random variables, each of which is $(2T, \eta)$-hyperconcentrated. Let $\psi : (\mathbb{R}^{\geq 0})^m \to [0, 1]$ be $(2T, C)$-relaxed. Assume $\eta \leq \frac{1}{4m}$. Then up to an additive error of $C \cdot O(m)^T \cdot \eta^T$, the expectation $\mathbf{E}[\psi(\boldsymbol{s})]$ is determined by the moments of $\boldsymbol{s}$ of degree up to $2T$.*

Given Lemma 77 and Lemma 79, observing that for our mollfier $\psi = \text{Mollifier}_p$ we have $m = \text{poly}(d)$ and hence $40e\sqrt{\hat{\lambda}} \ll \frac{1}{4m}$, we can indeed combine these results with Lemma 81. Recalling that the moments of $\boldsymbol{s}$ are determined by our assumptions and parameter settings (since each statistic $s_{i,j}$ has degree at most $2d$, the random variable $\boldsymbol{z}$ is a k_{indep}-wise independent Gaussian, and $2d \cdot 2T \leq k_{\text{indep}}$), this establishes Equation (73) as desired.

Proof of Lemma 81. We write μ to denote $\mathbf{E}[\boldsymbol{s}]$. For a generic $s \in (\mathbb{R}^{\geq 0})^m$, Taylor's theorem implies that:

$$\psi(s) = P(s) + \text{Err}(s), \quad \text{where } P(s) = \sum_{0 \leq |\alpha| < T} \tfrac{1}{\alpha!} \partial_\alpha \psi(\mu) \cdot (s - \mu)^\alpha, \quad \text{Err}(s) = \sum_{|\alpha| = T} \tfrac{1}{\alpha!} \partial_\alpha \psi(s^*) \cdot (s - \mu)^\alpha$$

for some s^* on the open line segment from μ to s. As P is a polynomial in s of degree at most $T - 1$, we have that $\mathbf{E}[P(\boldsymbol{s})]$ is exactly determined by the moments of $\boldsymbol{s}$ of degree up to $T - 1$. It therefore suffices to bound

$$\left|\mathbf{E}[\text{Err}(\boldsymbol{s})]\right| = \left|\mathbf{E}[\text{Err}(\boldsymbol{s})] \cdot \left(\mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}} + \mathbf{1}_{\{\boldsymbol{s} \napprox_1 \mu\}}\right)\right| \leq \mathbf{E}\left[|\text{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right] + \mathbf{E}\left[|\text{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \napprox_1 \mu\}}\right],$$
(76)

where $\boldsymbol{s} \approx_1 \mu$ means that $\boldsymbol{s}_i$ is within a multiplicative factor of e of μ_i (i.e., $\boldsymbol{s}_i \approx_1 \mu_i$) for each $1 \leq i \leq m$. Note that when this event occurs we also have $\boldsymbol{s}^* \approx_1 \mu$, and hence by the $(2T, C)$-relaxed property of ψ, for all α such that $|\alpha| = T$, we have

$$|\mu|^\alpha \cdot |\partial_\alpha \psi(\boldsymbol{s}^*)| \leq e^T C.$$
(77)

We proceed to analyze $\mathbf{E}\left[|\text{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right]$ as follows: we have

$$\mathbf{E}\left[|\text{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right] \leq \sum_{|\alpha| = T} |\partial_\alpha \psi(\boldsymbol{s}^*)| \, \mathbf{E}\left[|(\boldsymbol{s} - \mu)^\alpha| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right].$$
(78)

Now, if an α with $|\alpha| = T$ has $|\mu|^\alpha = 0$, this means that there is an index $\alpha_i > 0$ such that $\mu_i = 0$. Recalling the definition of hyperconcentration (Definition 24), it must be the case that $\boldsymbol{s}_i$ is zero with probability 1 and hence any such α contributes zero to the right-hand side of Equation (78). Thus we have that

$$\begin{aligned}
\mathbf{E}\left[|\text{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right] &\leq \sum_{|\alpha| \leq T, |\mu|^\alpha > 0} |\partial_\alpha \psi(\boldsymbol{s}^*)| \, \mathbf{E}\left[|(\boldsymbol{s} - \mu)^\alpha| \cdot \mathbf{1}_{\{\boldsymbol{s} \approx_1 \mu\}}\right] \\
&\leq e^T C \sum_{|\alpha| \leq T, |\mu|^\alpha > 0} |\mu|^{-\alpha} \, \mathbf{E}[|\boldsymbol{s} - \mu|^\alpha] \qquad \text{(using Equation (77))} \\
&\leq e^T C \cdot \#\{\alpha : |\alpha| = T\} \cdot \eta^T \qquad\qquad \text{(Lemma 26)} \\
&\leq C(em)^T \cdot \eta^T.
\end{aligned}$$
(79)

For the second term in Inequality (76) we use Cauchy–Schwarz:

$$\mathbf{E}\big[|\mathrm{Err}(\boldsymbol{s})| \cdot \mathbf{1}_{\{\boldsymbol{s} \not\approx_1 \mu\}}\big] \leq \sqrt{\mathbf{E}\Big[|\mathrm{Err}(\boldsymbol{s})|^2\Big]}\sqrt{\mathbf{E}\Big[\mathbf{1}_{\{\boldsymbol{s} \not\approx_1 \mu\}}^2\Big]} = \sqrt{\mathbf{E}\Big[|\psi(\boldsymbol{s}) - P(\boldsymbol{s})|^2\Big]}\sqrt{\mathbf{Pr}[\boldsymbol{s} \not\approx_1 \mu]}$$

$$\leq \left(1 + \sqrt{\mathbf{E}[P(\boldsymbol{s})^2]}\right) \cdot \sqrt{m} \cdot (2\eta)^T, \quad (80)$$

where in the last step we used $|\psi(\boldsymbol{s})| \leq 1$ for the first factor, and Proposition 25 (plus a union bound over the m coordinates of $\boldsymbol{s}$) for the second factor. Our handling of $\mathbf{E}\big[P(\boldsymbol{s})^2\big]$ will be similar to Inequality (79): we have that

$$\mathbf{E}\big[P(\boldsymbol{s})^2\big] \leq \sum_{0 \leq |\alpha|,|\beta| < T} |\partial_\alpha \psi(\mu)| \cdot |\partial_\beta \psi(\mu)| \cdot \mathbf{E}\Big[|\boldsymbol{s} - \mu|^{\alpha+\beta}\Big]$$

$$= \sum_{0 \leq |\alpha|,|\beta| < T : |\mu|^{\alpha+\beta} > 0} |\partial_\alpha \psi(\mu)| \cdot |\partial_\beta \psi(\mu)| \cdot \mathbf{E}\Big[|\boldsymbol{s} - \mu|^{\alpha+\beta}\Big]$$

$$\leq \sum_{0 \leq |\alpha|,|\beta| < T : |\mu|^{\alpha+\beta} > 0} C^2 |\mu|^{-\alpha-\beta} \cdot \mathbf{E}\Big[|\boldsymbol{s} - \mu|^{\alpha+\beta}\Big] \qquad (\psi \text{ is } (2T,C)\text{-relaxed})$$

$$\leq C^2 \sum_{0 \leq |\alpha|,|\beta| < T} \eta^{|\alpha+\beta|} \qquad\qquad\qquad (\text{Lemma 26})$$

$$\leq C^2 \sum_{k=0}^{T} (2m\eta)^k \leq 2C^2, \qquad\qquad\qquad (81)$$

where the last inequality used the assumption $\eta \leq \frac{1}{4m}$ and the equality uses reasoning similar to our earlier analysis of $|\alpha| = T$ such that $|\mu|^\alpha = 0$. Putting Inequalities (79) to (81) into Inequality (76) yields

$$|\mathbf{E}[\mathrm{Err}(\boldsymbol{s})]| \leq C(em)^T \cdot \eta^T + (1 + \sqrt{2}C) \cdot \sqrt{m} \cdot (2\eta)^T = C \cdot O(m)^T \cdot \eta^T,$$

as claimed.

This concludes the proof of Lemma 81 and thus also of Lemma 59. $\qquad\square$

Acknowledgments

References

[ABI86]   Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986.

[CW01]   Anthony Carbery and James Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, 8(3):233–248, 2001.

[DKN10] Ilias Diakonikolas, Daniel Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.

[Jof74] Anatole Joffe. On a set of almost deterministic k-independent random variables. *The Annals of Probability*, 2(1):161–162, 1974.

[Kan11a] Daniel Kane. k-independent Gaussians fool polynomial threshold functions. In *Proceedings of the 26th Conference on Computational Complexity (CCC)*, pages 252–261, 2011.

[Kan11b] Daniel Kane. A small PRG for polynomial threshold functions of Gaussians. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 257–266, 2011.

[Kan12] Daniel Kane. A structure theorem for poorly anticoncentrated Gaussian chaoses and applications to the study of polynomial threshold functions. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 91–100, 2012.

[Kan14] Daniel Kane. A pseudorandom generator for polynomial threshold functions of Gaussians with subpolynomial seed length. In *Proceedings of the 29th Annual Conference on Computational Complexity (CCC)*, pages 217–228, 2014.

[Kan15] Daniel Kane. A polylogarithmic PRG for degree 2 threshold functions in the Gaussian setting. In *Proceedings of the 30th Conference on Computational Complexity (CCC)*, pages 567–581, 2015.

[KM15] Pravesh Kothari and Raghu Meka. Almost optimal pseudorandom generators for spherical caps. In *Proceedings of the 47th Annual on Symposium on Theory of Computing (STOC)*, pages 247–256, 2015.

[MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 427–436, 2010.

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM Journal on Computing*, 42(3):1275–1301, 2013.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. Available at http://analysisofbooleanfunctions.net/.

[OST20] 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 (STOC)*, page 1170–1183, 2020.

A Omitted proofs

Proof of Proposition 15. Our goal will be to establish the following identity:

$$g_{\lambda|x}(y) = g(\sqrt{1-\lambda}x + \sqrt{\lambda}y) = \sum_{\alpha,\beta \in \mathbb{N}^n} \widehat{g}(\alpha + \beta)\sqrt{\mathbf{Pr}[\mathrm{Bin}(\alpha + \beta, \lambda) = \beta]}\, h_\alpha(x)\, h_\beta(y). \qquad (82)$$

From this we immediately deduce

$$\widehat{g_{\lambda|x}}(\beta) = \sum_{\gamma \geq \beta} \widehat{g}(\gamma) \sqrt{\mathbf{Pr}[\mathrm{Bin}(\gamma, \lambda) = \beta]} \, h_{\gamma - \beta}(x).$$

Now Parseval's identity (i.e., taking the expected square over $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ and using orthonormality of the $h_{\gamma-\beta}$'s) yields Proposition 15.

It remains to verify Equation (82). This identity is a direct consequence of the following univariate special case:

$$h_m(\sqrt{1-\lambda}x + \sqrt{\lambda}y) = \sum_{i+j=m} \sqrt{\mathbf{Pr}[\mathrm{Bin}(m, \lambda) = j]} \, h_i(x) \, h_j(y). \tag{83}$$

To obtain Equation (82), one simply substitutes Equation (83) into the multivariate Hermite expansion $g(z) = \sum_\gamma \widehat{g}(\gamma) h_\gamma(z)$.

Finally, Equation (83) is a variant of the standard identity [O'D14, Ex. 11.11] concerning $H_m(x + y)$. To prove Equation (83), we recall ([O'D14, Eq. (11.8)]) the generating function definition of h_m:

$$\exp(tz - \tfrac{1}{2}t^2) = \sum_{m=0}^{\infty} \tfrac{1}{\sqrt{m!}} h_m(z) t^m.$$

Substitute $z = \sqrt{1-\lambda}x + \sqrt{\lambda}y$, and use $-\tfrac{1}{2} = -\tfrac{1-\lambda}{2} - \tfrac{\lambda}{2}$ on the left-hand side. This yields

$$\exp(t\sqrt{1-\lambda}x - \tfrac{1-\lambda}{2}t^2)\exp(t\sqrt{\lambda}y - \tfrac{\lambda}{2}t^2) = \sum_{m=0}^{\infty} \tfrac{1}{\sqrt{m!}} h_m(\sqrt{1-\lambda}x + \sqrt{\lambda}y) t^m.$$

Now on the left we use the generating function twice again (with $\sqrt{1-\lambda}t$ and $\sqrt{\lambda}t$ replacing t), yielding

$$\left(\sum_{i=0}^{\infty} \tfrac{1}{\sqrt{i!}} h_i(y)(\sqrt{\lambda}t)^i\right)\left(\sum_{j=0}^{\infty} \tfrac{1}{\sqrt{j!}} h_j(x)(\sqrt{1-\lambda}t)^j\right) = \sum_{m=0}^{\infty} \tfrac{1}{\sqrt{m!}} h_m(\sqrt{\lambda}y + \sqrt{1-\lambda}x) t^m.$$

Equation (83) now follows by considering the coefficient on t^m on both sides. $\qquad \square$

Proof of Theorem 10. As mentioned, this result essentially appears in Section 6 of [Kan11b].

The following construction of k-wise independent tuples is well known [Jof74, ABI86]:

Theorem 82. *For any $k, n, M \in \mathbb{N}$, there is an efficient deterministic algorithm that takes in a uniformly random bit-string of length $O(k \max\{M, \log n\})$ and outputs a sequence $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in \{0, 1, \ldots, 2^M - 1\}^n$ that is k-wise independent with respect to the uniform distribution on $\{0, 1, \ldots, 2^M - 1\}$.*

We could convert each M-bit $\boldsymbol{X}_i$ to a near-Gaussian using Equation (8); the result would be a k-wise independent vector in with respect to an "$O(\tfrac{1}{\sqrt{M}})$-near-Gaussian" distribution. However one can convert M-bit strings to near-Gaussians with exponentially better accuracy, via the Box–Muller transform; as Kane [Kan11b, Proof of Cor. 2] shows, one can deterministically and efficiently convert $\boldsymbol{X}_i$ to a random variable $\boldsymbol{y}_i$ whose distribution can be coupled to a true Gaussian $\boldsymbol{z}_i \sim \mathrm{N}(0,1)]$ such that $\mathbf{Pr}[|\boldsymbol{y}_i - \boldsymbol{z}_i| \leq \delta] \geq 1 - \delta$ for $\delta = \Theta(2^{-M/2})$.

Given this, we may then use the following lemma explicitly proven by Kane [Kan11b], which is a relatively straightforward consequence of the Carbery–Wright theorem (Theorem 33):

Lemma 83. ([Kan11b, Lem. 21].) *Let $\boldsymbol{z}$ be a k-wise independent n-dimensional Gaussian random vector. Suppose that $\boldsymbol{y}$ is an n-dimensional random vector such that $\boldsymbol{y}$ and $\boldsymbol{z}$ may be coupled so that $\mathbf{Pr}[|\boldsymbol{y}_i - \boldsymbol{z}_i| \leq \delta] \geq 1 - \delta$. Finally, suppose one can show that for $L = 1/\lambda$ and $\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(L)}$ i.i.d. copies of $\boldsymbol{z}$, the sum $\sqrt{\lambda}\boldsymbol{z}^{(1)} + \cdots + \sqrt{\lambda}\boldsymbol{z}^{(L)}$ is $\varepsilon/2$-fooling for degree-d Gaussian PTFs. Then for $\boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(L)}$ being i.i.d. copies of $\boldsymbol{y}$, the sum $\sqrt{\lambda}\boldsymbol{y}^{(1)} + \cdots + \sqrt{\lambda}\boldsymbol{y}^{(L)}$ is ε'-fooling for degree-d Gaussian PTFs, where*

$$\varepsilon' = \varepsilon/2 + O(Ln\delta) + O(d\sqrt{Ln}\delta^{1/d}\log(1/\delta)).$$

Theorem 10 now follows by taking $M = O(d\log(dLn/\varepsilon))$ and thereby achieving $\delta = (dLn/\varepsilon)^{O(d)}$. $\qquad \square$

B   Appendix by Daniel Kane

Lemma 84. *Let p be a degree-d polynomial with $p(x) \geq 0$ for all x. Then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ and $\varepsilon > 0$ we have that*

$$\mathbf{Pr}[p(\boldsymbol{x}) < \varepsilon \mathbf{E}_{\boldsymbol{x}}[p(\boldsymbol{x})]] = O(d\varepsilon^{1/d}).$$

Proof. This follows from Carbery-Wright and the observation that $\|p\|_2 \geq \mathbf{E}_{\boldsymbol{x}}[p(\boldsymbol{x})]$. $\qquad \square$

Theorem 85 (Local Hyperconcentration Theorem, improved version)**.** *There exists a constant $c > 0$ so that for any $1 > \varepsilon, \beta > 0$ and $R \geq 1$ sufficiently small if $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ is a nice distribution over degree-d polynomials in n variables then for*

$$\lambda \leq \frac{c\varepsilon\beta}{Rd^{9/2}}$$

then for $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ with probability at least $1 - \beta$ we have that

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})] \leq \varepsilon^2 \mathbf{E}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2].$$

Proof. The proof of this theorem will depend on basic facts about the sequence of derivatives of $(g_{\boldsymbol{v}})_{\lambda|x}$. Firstly, we establish some notation, we let $D_y f(x)$ denote the directional derivative of f at x in the y direction. We begin with the following Lemma relating the size of functions and their derivatives:

Lemma 86. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v} \sim \Upsilon}$ be a nice distribution over degree-d polynomials. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be independent $\mathrm{N}(0,1)^n$ Gaussian random variables. Then for $\varepsilon > 0$ we have that*

$$\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}}[\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(\boldsymbol{x})|^2] \leq \varepsilon^2 \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}}(g_{\boldsymbol{v}}(\boldsymbol{x}))|^2]] = O(d^2\varepsilon).$$

Proof. We begin with the case where $g_{\boldsymbol{v}}$ is actually a constant family (i.e. is just a single function). This result then follows immediately from Lemma 9 of [D. Kane "The Correct Exponent for the Gotsman-Linial Conjecture"].

From here we generalize to the case where $g_{\boldsymbol{v}}$ is a linear polynomial in $\boldsymbol{v}$. By the previous case, we have that

$$\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{v}}[|g_{\boldsymbol{v}}(\boldsymbol{x})|^2 \leq \varepsilon^2 |D_{\boldsymbol{y}}(g_{\boldsymbol{v}}(\boldsymbol{x}))|^2] = O(d^2\varepsilon).$$

On the other hand since $g_{\boldsymbol{v}}(x)$ and $D_y(g_{\boldsymbol{v}}(x))$ are linear functions of $\boldsymbol{v}$. Therefore, for any x and y, with at least 50% probability over the choice of $\boldsymbol{v}$ we have that $\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(x)|^2] \ll |g_{\boldsymbol{v}}(x)|^2$ and $|D_y(g_{\boldsymbol{v}}(x))|^2 \ll \mathbf{E}_{\boldsymbol{v}}[|D_y(g_{\boldsymbol{v}}(x))|^2]$. Therefore, whenever

$$\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(x)|^2] \leq \varepsilon^2 \mathbf{E}_{\boldsymbol{v}}[|D_y(g_{\boldsymbol{v}}(x))|^2]$$

there is at least a 50% probability over the choice of $\boldsymbol{v}$ that

$$|g_{\boldsymbol{v}}(x)|^2 \leq \varepsilon^2 C |D_y(g_{\boldsymbol{v}}(x))|^2$$

for some positive constant C. However, the latter happens with probability $O(d^2\varepsilon)$ and is at least half of

$$\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}}[\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(\boldsymbol{x})|^2] \leq \varepsilon^2 \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}}(g_{\boldsymbol{v}}(\boldsymbol{x}))|^2]].$$

Therefore, this latter probability is $O(d\varepsilon^2)$.

Finally, we can handle the generic case. Let $h_a(v)$ be an orthonormal basis for the polynomials in v. We can write $g_v(x)$ as $\sum_a h_a(v)p_a(x)$ for some polynomials p_a. Define

$$h_{\boldsymbol{\mu}}(x) := \sum_a \boldsymbol{\mu}_a p_a(x)$$

where $\boldsymbol{\mu}$ is a Gaussian random variable with as many components as there are terms in the above decomposition of g_v. It is easy to see that for any x and y that

$$\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(x)|^2] = \sum_a p_a^2(x) = \mathbf{E}_{\boldsymbol{\mu}}[|h_{\boldsymbol{\mu}}(x)|^2]$$

and

$$\mathbf{E}_{\boldsymbol{v}}[|D_y g_{\boldsymbol{v}}(x)|^2] = \sum_a (D_y p_a(x))^2 = \mathbf{E}_{\mu}[|D_y h_\mu(x)|^2].$$

Since our Lemma holds for h, it must therefore also hold for g. This completes our proof. $\qquad\square$

Our theorem will rest upon the following notion of a random derivative sequence:

Definition 87. Given a nice distribution $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ of degree-d polynomials, a *random derivative sequence* for g is a sequence of the form

$$\mathbf{E}_{\boldsymbol{v}}[|g_{\boldsymbol{v}}(\boldsymbol{x})|^2], \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}(\boldsymbol{x})|^2], \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}_2}D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}(\boldsymbol{x})|^2], \dots \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}_d}D_{\boldsymbol{y}_{d-1}}\cdots D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}(\boldsymbol{x})|^2],$$

where $\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_d$ are independent Gaussian random variables.

We will often denote the kth term by

$$\mathcal{D}^k_{\boldsymbol{x},\boldsymbol{y}_i}(g_v) := \mathbf{E}_{\boldsymbol{v}}[|D_{\boldsymbol{y}_k}D_{\boldsymbol{y}_{k-1}}\cdots D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}(\boldsymbol{x})|^2].$$

Corollary 88. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ be a nice distribution over degree-d polynomials. Then with probability at least $1 - \varepsilon$ over the choice of $\boldsymbol{x}, \boldsymbol{y}_i$ we have that:*

$$\mathcal{D}^{k+1}_{\boldsymbol{x},\boldsymbol{y}_i}(g_v) = O(d^6/\varepsilon^2)\mathcal{D}^k_{\boldsymbol{x},\boldsymbol{y}_i}(g_v)$$

for all $0 \leq k \leq d$.

Proof. It follows immediately from Lemma 86 that for each k

$$\mathcal{D}^{k+1}_{\boldsymbol{x},\boldsymbol{y}_i}(g_v) = O(d^6/\varepsilon^2)\mathcal{D}^k_{\boldsymbol{x},\boldsymbol{y}_i}(g_v)$$

with probability $1 - \varepsilon/d$. The full result follows from a union bound over k. $\qquad\square$

This says that a typical random derivative sequence does not increase too rapidly. However, we show that there is a kind of converse for non-hyperconcentrated families:

Proposition 89. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ be a nice distribution of degree-d polynomials so that for some $R, \varepsilon > 0$ we have that*

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R((g_{\boldsymbol{v}}))] \geq \varepsilon^2 \, \mathbf{E}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})\|_2^2].$$

Then with at least 50% probability over a choice of random Gaussians $\boldsymbol{x}, \boldsymbol{y}_i$ we have that for some $0 \leq k \leq d$ that

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{k+1}(g_v) > \Omega(\varepsilon^2/(d^3 R^2))\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^k(g_v).$$

In order to prove this, we will need to talk about the Hermite parts of g. Recall that for g a polynomial, the notation $g^{=k}$ denotes the degree-k Hermite part of g.

We will make use of the following facts:

Fact 90. *For any polynomial g and $R > 0$,*

$$\mathbf{HyperVar}_R(g) = \sum_k kR^{2k}\|g^{=k}\|_2^2.$$

Fact 91. *For any polynomial g and vector v:*

$$(D_v g)^{=k} = D_v(g^{=k+1}).$$

Fact 92. *For any g with $g = g^{=k}$ and $\boldsymbol{x}$ a random Gaussian*

$$\mathbf{E}_{\boldsymbol{x}}[\|D_{\boldsymbol{x}}g\|_2^2] = k\|g\|_2^2.$$

We will also need to know that with reasonable probability that some element of a random derivative sequence is not too small.

Lemma 93. *Let $(g_{\boldsymbol{v}})_{\boldsymbol{v}\sim\Upsilon}$ be a nice distribution of degree-d polynomials and let $m \geq d$. Then with probability at least $1 - d/m$ over a choice of random Gaussians $\boldsymbol{x}, \boldsymbol{y}_i$ there exists a $0 \leq k \leq d$ so that*

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^k(g_v) \geq \Omega(1/m)^{3k} \, \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]/2.$$

Proof. We proceed by induction on d. If $d = 0$, then $\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^0(g_v) = \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]$ and we are done.

For the inductive step, we break into two cases. On the one hand, if

$$\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2] \geq 100m \, \mathbf{E}_{\boldsymbol{v}}[\mathbf{Var}_{\boldsymbol{x}}(g_{\boldsymbol{v}}(\boldsymbol{x}))],$$

then we have that

$$\mathbf{E}_{\boldsymbol{x}}[\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2 \mathbf{1}\{|g_{\boldsymbol{v}}(\boldsymbol{x})| < (2/3)\|g_{\boldsymbol{v}}\|_2\}]]$$

$$= \mathbf{E}_{\boldsymbol{v}}[\mathbf{E}_{\boldsymbol{x}}[\|g_{\boldsymbol{v}}\|_2^2 \mathbf{1}\{|g_{\boldsymbol{v}}(\boldsymbol{x})| < (2/3)\|g_{\boldsymbol{v}}\|_2\}]]$$

$$\leq \mathbf{E}_{\boldsymbol{v}}\left[9(\|g_{\boldsymbol{v}}\|_2^2)\left(\frac{\mathbf{Var}_{\boldsymbol{x}}(g_{\boldsymbol{v}}(\boldsymbol{x}))}{\|g_{\boldsymbol{v}}\|_2^2}\right)\right]$$

$$= 9 \, \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]/2$$

$$\leq (1/10m) \, \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

However, if

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^0(g_v) < \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]/2,$$

it must be the case that

$$\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2 \mathbf{1}\{|g_{\boldsymbol{v}}(x)| < (2/3)\|g_{\boldsymbol{v}}\|_2\}] \geq (1/6)\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2],$$

and by the Markov Inequality, this happens with probability at most $1/m$.

On the other hand, if

$$\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2] \leq 100m\,\mathbf{E}_{\boldsymbol{v}}[\mathbf{Var}_{\boldsymbol{x}}(g_{\boldsymbol{v}}(\boldsymbol{x}))],$$

then

$$\mathbf{E}_{\boldsymbol{y}_1}[\mathbf{E}_{\boldsymbol{v}}[\|D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}\|_2^2]] = \mathbf{E}_{\boldsymbol{v}}[\mathbf{Var}_{\boldsymbol{x}}(g_{\boldsymbol{v}}(\boldsymbol{x}))].$$

Since $\mathbf{E}_{\boldsymbol{v}}[\|D_{y_1}g_{\boldsymbol{v}}\|_2^2]$ is a non-negative quadratic function of y_1, Lemma 84 implies that with probability at least $1 - 1/m$ we have that

$$\mathbf{E}_{\boldsymbol{v}}[\|D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}\|_2^2] \geq (1/m^2)\mathbf{E}_{\boldsymbol{v}}[\mathbf{Var}_{\boldsymbol{x}}(g_{\boldsymbol{v}}(\boldsymbol{x}))] \geq \Omega(1/m^3)\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

For such an outcome of $\boldsymbol{y}_1$, we can apply our inductive hypothesis to $(D_{\boldsymbol{y}_1}g_v)$. $\qquad\square$

We are now prepared to prove Proposition 89.

Proof. Note that

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R(g_{\boldsymbol{v}})] = \sum_{k=1}^{d} kR^{2k}\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}^{=k}\|_2^2].$$

Therefore, under our hypothesis, there must be a k so that

$$\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}^{=k}\|_2^2] \geq (\varepsilon^2/(4R^2))^k \mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

Notice that

$$\mathbf{E}_{\boldsymbol{v}}[\|D_{y_k}D_{y_{k-1}}\cdots D_{y_1}g_{\boldsymbol{v}}^{=0}\|_2^2]$$

is a non-negative degree $2k$ polynomial in $y_1,\ldots,y_k$ with average value at least $(\varepsilon^2/(4R^2))^k\,\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2]$. Therefore, Lemma 84 implies that with probability at least $5/6$ we have that

$$\|D_{\boldsymbol{y}_k}D_{\boldsymbol{y}_{k-1}}\cdots D_{\boldsymbol{y}_1}g_{\boldsymbol{v}}^{=0}\|_2^2 \geq \Omega(\varepsilon^2/(d^2R^2))^k\,\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

By Lemma 93 with probability at least $5/6$ there is a $k' > k$ so that

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{k'}(g_v) \geq \Omega(1/d)^{3(k'-k)}\|D_{\boldsymbol{y}_k}D_{\boldsymbol{y}_{k-1}}\cdots D_{\boldsymbol{y}_1}g_v^{=0}\|_2^2/2 \geq \Omega(\varepsilon^2/(d^3R^2))^{k'}\,\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

Finally, with probability at least $5/6$ we have that

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{0}(g_v) = O(1)\,\mathbf{E}_{\boldsymbol{v}}[\|g_{\boldsymbol{v}}\|_2^2].$$

Thus, if all three of these events hold (which happens with probability at least $1/2$), there will be some $d \geq k > 1$ so that

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{k}(g_v) \geq \Omega(\varepsilon^2/(d^3R^2))^k \mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{0}(g_v).$$

Therefore, there must also be a k so that

$$\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{k+1}(g_v) \geq \Omega(\varepsilon^2/(d^3R^2))\mathcal{D}_{\boldsymbol{x},\boldsymbol{y}_i}^{k}(g_v). \quad \square$$

Our Theorem will now follow from the tension between Corollary 88 and Proposition 89 along with the observation that

$$\mathcal{D}^k_{\boldsymbol{x},\boldsymbol{y}_i}((g_\upsilon)_{\lambda|\boldsymbol{x}}) = \lambda^{2k}\mathcal{D}^k_{\boldsymbol{z},\boldsymbol{y}_i}(g_\upsilon)$$

where $\boldsymbol{z} = \sqrt{1-\lambda}\boldsymbol{x}' + \lambda\boldsymbol{x}$. Note also that $\boldsymbol{z}$ is a standard Gaussian if $\boldsymbol{x}'$ and $\boldsymbol{x}$ are. In particular, Corollary 88 tells us that with probability $1 - \beta/2$ that

$$\mathcal{D}^{k+1}_{\boldsymbol{z},\boldsymbol{y}_i}(g_\upsilon) = O(d^6/\beta^2)\mathcal{D}^k_{\boldsymbol{z},\boldsymbol{y}_i}(g_\upsilon) \tag{84}$$

for all $0 \le k \le d$. On the other hand, if

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})] > \varepsilon^2 \mathbf{E}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|x}\|_2^2]$$

then Proposition 89 implies that with at least 50% probability that there is a $0 \le k \le d$ so that

$$\mathcal{D}^{k+1}_{\boldsymbol{x},\boldsymbol{y}_i}((g_\upsilon)_{\lambda|x}) > \Omega(\varepsilon^2/(d^3R^2))\mathcal{D}^k_{\boldsymbol{x},\boldsymbol{y}_i}((g_\upsilon)_{\lambda|\boldsymbol{x}}).$$

But this is equivalent to saying that

$$\mathcal{D}^{k+1}_{\boldsymbol{z},\boldsymbol{y}_i}(g_\upsilon) > \Omega(\lambda^{-2}\varepsilon^2/(d^3R^2))\mathcal{D}^k_{\boldsymbol{z},\boldsymbol{y}_i}(g_\upsilon).$$

However, given our setting of λ this would contradict Equation (84). Therefore, the probability of Equation (84) being violated is at most $\beta/2$, but is at least half the probability that

$$\mathbf{E}_{\boldsymbol{v}}[\mathbf{HyperVar}_R((g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}})] > \varepsilon \mathbf{E}_{\boldsymbol{v}}[\|(g_{\boldsymbol{v}})_{\lambda|\boldsymbol{x}}\|_2^2].$$

Hence we conclude that the latter probability is at most β. $\qquad\square$