A PRG for Boolean PTF of Degree 2 with Seed Length Subpolynomial in ϵ and Logarithmic in n

Daniel Kane[1]
UC San Diego
dakane@ucsd.edu

Sankeerth Rao
UC San Diego
skaringu@ucsd.edu

Abstract

We construct and analyze a *pseudorandom generator* for degree 2 boolean *polynomial threshold functions*. Random constructions achieve the optimal seed length of $O(\log n + \log \frac{1}{\epsilon})$, however the best known explicit construction of [8] uses a seed length of $O(\log n \cdot \epsilon^{-8})$. In this work we give an *explicit* construction that uses a seed length of $O(\log n + (\frac{1}{\epsilon})^{o(1)})$. Note that this improves the seed length substantially and that the dependence on the error ϵ is *additive* and only grows *subpolynomially* as opposed to the previously known multiplicative polynomial dependence.

Our generator uses *dimensionality reduction* on a *Nisan-Wigderson* based pseudorandom generator given by Lu, Kabanets [18].

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Pseudorandomness, Polynomial Threshold Functions

Digital Object Identifier 10.4230/LIPIcs.CCC.2018.2

Acknowledgements The authors would like to thank Zhenjian Lu and Valentine Kabanets for all their spontaneous help and collaboration.

1 Introduction

1.1 Background and importance

We say that a function $f : \mathbb{R}^n \to \{+1, -1\}$ is a (degree-d) *polynomial threshold function*(PTF) if it is of the form $f(x) = sgn(p(x))$ for p some (degree-d) polynomial in n variables. Polynomial threshold functions make up a natural class of Boolean functions and have applications to a number of fields of computer science such as circuit complexity [2], communication complexity [17] and learning theory [14].

In this paper, we study the question of pseudorandom generators (PRGs) for polynomial threshold functions of Bernoulli inputs(and in particular for $d=2$). In other words, we wish to find explicit functions $F : \{\pm 1\}^s \to \{\pm 1\}^n$ so that for any degree-2 polynomial threshold function f, we have

$$|\mathop{\mathbb{E}}_{x \sim_u \{\pm 1\}^s}[f(F(x))] - \mathop{\mathbb{E}}_{X \sim \{\pm 1\}^n}[f(X)]| < \epsilon.$$

We say that such an F is a pseudorandom generator of seed length s that fools degree-2 polynomial threshold functions with respect to the Bernoulli distribution to within ϵ. In

[1] Supported by NSF Award CCF-1553288(CAREER) and a Sloan Research Fellowship.

33rd Computational Complexity Conference (CCC 2018).
Editor: Rocco A. Servedio; Article No. 2; pp. 2:1–2:24
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Table 1 Pseudorandom Generators

Paper	Bernoulli/Gaussian	d	Seedlength s
Diakonikolas, Gopalan, etal [5]	Bernoulli	1	$\log n \cdot O(\epsilon^{-2}\log^2(1/\epsilon))$
Meka, Zuckerman [15]	Bernoulli	1	$O(\log n + \log^2(1/\epsilon))$
Gopalan, Kane, Meka [7]	Bernoulli	1	$O(\log(n/\epsilon)\cdot[\log\log(n/\epsilon)]^2)$
Diakonikolas, Kane, Nelson [8]	Gaussian	1	$\log n \cdot O(\epsilon^{-2})$
Kane [12]	Gaussian	1	$O(\log n + \log^{3/2}(1/\epsilon))$
Diakonikolas, Kane, Nelson [8]	Both	2	$\log n \cdot O(\epsilon^{-8})^\dagger$
Kane [12]	Gaussian	2	$\log n \cdot \exp[\tilde{O}(\log(1/\epsilon)^{2/3})]$
Kane [13]	Gaussian	2	$O(\log^6(1/\epsilon)\cdot\log n\cdot\log\log(n/\epsilon))$
Kane, Sankeerth This paper	**Bernoulli**	$\mathbf{2}$	$O(\log n + e^{\sqrt{\log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}}})$
Kane [9]	Both	d	$\log n \cdot O_d\left(\epsilon^{-2^{O(d)}}\right)$
Meka, Zuckerman [15]	Bernoulli	d	$\log n \cdot 2^{O(d)}\epsilon^{-8d-3}$
Kane [11]	Bernoulli	d	$\log n \cdot O_d(\epsilon^{-11.1})$
Kabanets, Lu [18]	Bernoulli	d	$e^{O(\sqrt{d\log n\log\log(n/\epsilon)})}$
Kane [10]	Gaussian	d	$\log n \cdot 2^{O(d)}\epsilon^{-4.1}$
Kane [11]	Gaussian	d	$\log n \cdot O_d(\epsilon^{-2.1})$
Kane [12]	Gaussian	d	$\log n \cdot O_{c,d}(\epsilon^{-c})$

this paper, we develop a generator with $s = O(\log n + e^{\sqrt{\log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}}})$. The main idea is to apply a Johnson-Lindenstrauss like dimensionality reduction on the Nisan-Wigderson based pseudorandom generator by Lu, Kabanets [18]. A random construction shows the existence of a PRG that uses a seed length of $s = O(\log n + \log\frac{1}{\epsilon})$, however there are no known constructions that achieve this. The best known constructions for Boolean degree 2 PTFs use a seed length of $s = \log n \cdot poly(\frac{1}{\epsilon})$, the current work improves this especially the error dependence to $s = O(\log n + subpoly(\frac{1}{\epsilon}))$.The Meka-Zuckerman PRG for LTFs in [15] uses a similar type of dimensionality reduction idea to reduce the seed length from $O(\log^2(\frac{n}{\epsilon}))$ to $O(\log n + \log^2\frac{1}{\epsilon})$.

1.2 Prior Work

An existential argument shows that there are optimal pseudo random generators of seed length $O(d\log n + \log\frac{1}{\epsilon})$. There has been a lot of research towards giving explicit constructions that approach this seed length. The following are the past results of pseudorandom generators constructed for PTFs of degree d.

1.3 Our results and merits of the paper

The main goal for degree 2 PRG constructions has been to achieve the optimal seed length of $O(\log n + \log(\frac{1}{\epsilon}))$ via explicit constructions. Random constructions do achieve this optimal seed length, however the best known explicit construction of [8] uses a seed length of $O(\log n \cdot \epsilon^{-8})$. In this paper we give an *explicit* construction that uses a seed length of $O(\log n + (\frac{1}{\epsilon})^{o(1)})$. Note that this improves the seed length substantially and that the dependence on the error ϵ is *additive* and only grows *subpolynomially* as opposed to the

[†] The original analysis only got $\log n \cdot \tilde{O}(\epsilon^{-9})$ until [11] led to an improved analysis using the same ideas.

previously known multiplicative polynomial dependence. In particular we give a construction for a seed length of $O(\log n + e^{\sqrt{\log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}}})$. The major improvement of this work is in separating out the n-dependence from the ϵ-dependence. It would be very interesting to improve this further to the optimal logarithmic dependence on ϵ.

The main theorem of this paper is:

▶ **Theorem 1.** *Given $\epsilon > 0, n \in \mathbb{N}$, we construct a function $F : \{\pm1\}^s \to \{\pm1\}^n$ such that for any degree 2 polynomial $p : \{\pm1\}^n \to \mathbb{R}$, the probability that $p(x) \geq 0$ at a uniformly random point in $\{\pm1\}^n$ is approximately (within ϵ) equal to the probability that $p(F[z]) \geq 0$ at a uniformly random point in $\{\pm1\}^s$. That is,*

$$|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^s} sgn(p(F[z]))| \leq \epsilon.$$

Here s is called the seed length of F and it is given by $s = O(\log n + e^{\sqrt{\log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}}})$.

We construct F by doing a dimensionality reduction like argument on a Nisan-Wigderson based pseudorandom generator for Boolean PTFs constructed by Kabanets, Lu in [18]. Their construction uses a seed length of $O(e^{\sqrt{\log n \log\log\frac{n}{\epsilon}}})$.

Our generator is best thought of as a dimension reduction gadget. It reduces the problem of finding a PRG in n dimensions to that of finding a PRG in poly$(1/\epsilon)$ dimensions (with an additive loss of O$(\log n)$ in seed length). This means that if you combine it with a generator that has seed length $s(n, \epsilon)$, we get a new generator with seed length $O(\log n) + s(poly(1/\epsilon), \epsilon)$. This is particularly useful if the other generator is the Kabanets-Lu generator, since that generator has a great ϵ dependence at the expense of having a poor dependence on n. One could also use the trivial generator (i.e. the uniform distribution over the entire hypercube for which $s(n, \epsilon) = n$), and get a generator with seed length $O(\log n) + poly(1/\epsilon)$.

In particular we don't require the Kabanets-Lu generator, but since what we do only reduces the dimension of the problem, we do need *some* other generator. When we use the trivial PRG instead after using our technique, we can get $O(\log n) + poly(1/\epsilon)$. We believe that even this is new.

1.4 Proof overview with an outline of key technical ideas used

We construct our PRG F by composing a Johnson-Lindenstrauss matrix L^t with the following PRG H constructed by Kabanets, Lu in [18], that is $F = L \circ H$. They construct H by constructing a hard function that can't be computed by PTFs and using the Nisan-Wigderson hardness vs randomness template.

▶ **Theorem 2.** *Given $\epsilon > 0, n \in \mathbb{N}$, one can construct a function $H : \{\pm1\}^t \to \{\pm1\}^n$ such that for any degree 2 polynomial $q : \{\pm1\}^n \to \mathbb{R}$, the probability that $q(x) \geq 0$ at a uniformly random point in $\{\pm1\}^n$ is approximately (within ϵ) equal to the probability that $q(H[z]) \geq 0$ at a uniformly random point in $\{\pm1\}^t$. That is,*

$$|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(q(x)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^t} sgn(q(H[z]))| \leq \epsilon.$$

where the seed length of H is $t = O(e^{\sqrt{\log n \log\log\frac{n}{\epsilon}}})$.

Let's first understand the seed length needed for our PRG F.

Seed length

We use Kabanets PRG H to stretch from an initial seed of length t to dimension m. This is further stretched by L from m to n (think of m as $(\frac{1}{\epsilon})^{\Omega(1)}$). Thus the seed length t needed to make Kabanets PRG H work is $t = O(e^{\sqrt{\log m \log\log(\frac{m}{\epsilon})}})$, since $m = (\frac{1}{\epsilon})^{\Omega(1)}$ this would amount to a seed of $t = O(e^{\sqrt{\log(\frac{1}{\epsilon})\log\log(\frac{1}{\epsilon})}})$. L would further use randomness needing an extra seed of $O(\log n)$. Thus F would need a total seed length $s = O(\log n + e^{\sqrt{\log(\frac{1}{\epsilon})\log\log(\frac{1}{\epsilon})}})$.

Analysis

To analyse our PRG we split the error into two steps as follows:

- *Replace the n pure random bits input by m pure random bits* We replace the n pure random bits x by Ly, where y has only m purely random bits.
- *Replace the m pure random bits by t pseudorandom bits* We further replace the m pure random bits y by even fewer t purely random bits z. This is done via H, that is $y = H[z]$.

We depict this in the following equation:

$$\begin{aligned}
|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^t} sgn(p(F[z]))| &= |\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^t} sgn(pL(H[z]))| \\
&\leq |\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))| \\
&+ |\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^t} sgn(pL(H[z]))|.
\end{aligned}$$

Let's understand these steps:

1.4.1 Stretch t pure bits to m pure bits, $|\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y)) - \mathop{\mathbb{E}}_{z\sim\{\pm1\}^t} sgn(pL(H[z]))|$

As L is a linear operator, pL would still be a polynomial of degree 2. This error is small because H fools all degree 2 PTFs including pL. Thus we are using the PRG H to go from a space of dimension $t = O(e^{\sqrt{\log(\frac{1}{\epsilon})\log\log(\frac{1}{\epsilon})}})$ to a space of dimension $m = \frac{1}{\epsilon^{\Omega(1)}}$. The main technical idea used by [18] to achieve this is to give a hard function for PTFs and invoke the Nisan-Wigderson *hardness* vs *randomness* template.

1.4.2 Stretch m pure bits to n pure bits, $|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))|$

We show that this error is small in two steps:

- *Move from Boolean to Gaussian setting* We first move from the Boolean input to the Gaussian input setting. This can be done very easily for some special polynomials (*regular*). For a non regular polynomial we use technical ideas like the *regularity lemma* [6].
- *L is a good PRG for Gaussian inputs* When the input is Gaussian we have a lot of geometric structure. In particular using Central limit theorems(as done in [3]) any polynomial can be seen as a low dimensional very structured part and a lump mass that can be approximated by a single Gaussian. We show that L preserves this structure and thus we don't incur much error in changing x to Ly.

Move from Boolean to Gaussian setting

There are two technical ideas used here.

- **Regular polynomials** A polynomial is *regular* if no single input variable has a huge influence over the value of the polynomial. When a polynomial is regular one doesn't incur much loss when switching the input from boolean to gaussian as shown by the *Invariance principle* [16]. Think of this *replacement* as a telescope of replacing the variables one at a time and the error incurred when the ith variable is replaced is captured by its *influence*. In fact in this paper we show that L keeps *regular* polynomials regular. Thus if p is regular then so is pL. Thus we switch from boolean inputs to gaussian inputs for both p and pL.
- **Regularity Lemma** If a polynomial is not *regular*, then you could incur huge loss by directly switching the inputs from boolean to Gaussian. However there could be very few variables that have such a huge influence over the polynomial. So if these few variables are fixed the rest of the polynomial will either have negligible mass or be regular both of which are amenable to replacement from boolean to gaussian inputs. Thus the technical idea used here is the *Regularity* lemma of [6] which shows that every polynomial can be seen as a decision tree corresponding to the high influence variables that are fixed wherein the leaves are either regular or almost constant polynomials. We show that our JL matrix L interacts well with the Regularity Lemma. That is under the hash function of L we don't see any collision for the high influence variables whp. Also the low influence variables that do hash collide with these high influence variables contribute very little mass to pL.

L is a good PRG for Gaussian inputs

There are two technical ideas used here.

- **Central Limit Theorem** If all the eigenvalues of a polynomial are small relative to its variance then the polynomial can be well approximated by a single Gaussian as shown in [3] via a Central Limit Theorem. Since the variance of the polynomial is a constant, there can be only few large eigenvalues. Thus any polynomial can be seen as a structured polynomial consisting of the few large eigenvalues and an eigenregular polynomial that can be replaced by a single Gaussian. Thus the only essential information is in the top eigenstructure and the lump mass of the rest of the eigenvalues.
- **Structure preservation by L** We show that our JL matrix preserves the structure of these top few eigenvalues and also maps polynomials with small eigenvalues to polynomials with small eigenvalues with high probability. Thus L keeps this top eigenstructure+lump mass structure intact. It also approximately preserves the L^2 norms and covariance of polynomials and thus we see that L is a good PRG in the Gaussian setting.

The Meka-Zuckerman PRG for LTFs in [15] uses a similar type of dimensionality reduction idea to reduce the seed length from $O(\log^2(\frac{n}{\epsilon}))$ to $O(\log n + \log^2 \frac{1}{\epsilon})$.

1.5 Overview of the paper

We present the mathematical preliminaries required in section 2 and show that the PRG construction works under the assumption that p is regular in section 3. Then we prove a reduction from the general case to the special case of regular polynomials in Section 4. We present the conclusions in section 5.

Note:

All through the paper we will be bounding errors whp as $\frac{1}{m^{\Omega(1)}}$. Note that these errors are less than ϵ if m is chosen to be a sufficiently a large polynomial of $\frac{1}{\epsilon}$. Think of whp to mean with probability $1-\frac{1}{m^{\Omega(1)}}$.

All through the paper we leave the errors in terms of m, think of adding up all the errors and union bounding the probabilities and fixing all the parameters in terms of ϵ and then we choose a sufficiently large $m = \frac{1}{\epsilon^{\Omega(1)}}$ to make the sum of all the errors $O(\epsilon)$ and the union of all the error probabilities $O(\epsilon)$.

2 Preliminaries

2.1 Basic results on polynomials, concentration, anticoncentration, invariance and regularity

Concentration

We begin with a standard concentration bound from [4] that says that Gaussian degree-2 polynomials are concentrated around their mean. We would need this multiple times in the paper to show concentration of Gaussian polynomials.

▶ **Lemma 3.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-2 polynomial. We have*

$$\Pr_{x\sim\mathcal{N}^n(0,1)}\left[|p(x) - \mathbb{E}[p(x)]| > t\sqrt{Var[p]}\right] \leq e^{-\Omega(t)}.$$

Anticoncentration

We will need the following standard Carbery-Wright anticoncentration bound from [1],[4] that proves a bound on the mass a Gaussian degree-2 polynomial could have around any point. This would be useful in many instances including when we change functions of Gaussians.

▶ **Lemma 4.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-2 polynomial that is not identically 0. Then for all $\epsilon > 0$ and all $\theta \in \mathbb{R}$, we have*

$$\Pr_{x\sim\mathcal{N}^n(0,1)}\left[|p(x) - \theta| < \epsilon\sqrt{Var[p]}\right] \leq O(\sqrt{\epsilon}).$$

The following lemma from [4] is very useful as it helps us bound the distributional distance between two Gaussian polynomials by just bounding the L^2 norm. The proof follows from an application of Lemmas 3,4.

▶ **Lemma 5.** *Let $a(x), b(x)$ be degree-2 polynomials over $\mathbb{R}^n$. For $x \sim \mathcal{N}^n(0,1)$, if $\mathbb{E}[a(x){-}b(x)] = 0$, $Var[a] = 1$ and $Var[a{-}b] \leq (\beta/2)^6$, then*

$$|\mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(a(x)) - \mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(b(x))| \leq O(\beta).$$

Invariance Principle

The Invariance principle bounds the change in $\mathbb{E}[sgn(p(x))]$ when the input is changed from Boolean to Gaussian. We use the following lemma based on [16].

▶ **Lemma 6.** *For any degree 2 multilinear polynomial* $p = \sum_{i,j\in[n]} a_{ij}x_ix_j + \sum_{l\in[n]} b_lx_l + C$, *we have the following bound:*

$$|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x))| \le O\left[\frac{\sum_{i=1}^{n} Inf_i^2(p)}{(Var[p])^2}\right]^{\frac{1}{9}}.$$

where the ith influence of p is defined as $Inf_i(p) = \mathbb{E}|\frac{\partial p}{\partial x_i}|^2 = 2\sum_{j\in[n]} a_{ij}^2 + b_i^2$.

Think of ith influence as the variance of p along the ith coordinate.

Observe that $Var[p] \le \sum_{i=1}^{n} Inf_i(p) \le 2Var[p]$. Now we define the notion of *regularity* for polynomials which essentially means that there is no single variable whose influence is very large as compared to the rest of the variables.

▶ **Definition 7.** We say that the polynomial p is τ-regular if $\max_{i\in[n]} Inf_i(p) \le \tau Var[p]$.

Thus for a τ-regular polynomial p we can bound the replacement error above as $O(\tau^{\frac{1}{9}})$ because

$$\frac{\sum_{i=1}^{n} Inf_i^2(p)}{(Var[p])^2} \le \frac{[\max_i Inf_i(p)]\sum_{i=1}^{n} Inf_i(p)}{(Var[p])^2} \le 2\tau.$$

Note that when we apply this, we pick $\tau = \epsilon^{O(1)}$.

Regularity Lemma

We will use the following Regularity Lemma from [6]:

▶ **Lemma 8.** *Every multilinear degree 2 polynomial* $p : \{\pm1\}^n \to \mathbb{R}$ *can be written as a decision tree of depth* $D = \frac{1}{\tau} \cdot O\left(\log\frac{1}{\tau\theta}\right)^{O(1)}$ *such that with probability* $(1-\theta)$ *over a random leaf the resulting polynomial* p_α *is either*

(i) τ *regular, OR*

(ii) $Var(p_\alpha) < \theta||p||_2^2$.

Note that when we apply this Regularity lemma we will choose $\theta = \frac{1}{\sqrt{m}}, \tau = \epsilon^{O(1)}$ so that $D = (\log m)^{O(1)}$. After all the parameters are fixed we finally pick $m = \frac{1}{\epsilon^{\Omega(1)}}$ large enough so that all the errors get bounded by $O(\epsilon)$.

2.2 Eigenvalues of polynomials, Central Limit Theorem.

Eigenvalues

Let $p : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial of degree 2. Thus there exist a real symmetric matrix A, a vector B^t and a constant C such that

$$p(x) = x^tAx + B^tx + C.$$

The eigenvalues of p are defined to be the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the real symmetric matrix A. Since p is a multilinear polynomial we have $\sum_{i=1}^{n} \lambda_i = 0$.

We have the following expression for variance of the polynomial from [3]:

$$Var[p] = \sum_{i=1}^{n}(b_i^2 + 2a_{ii}^2) + \sum_{1 \le i < j \le n} a_{ij}^2.$$

The eigenvalues capture a lot of information about the polynomial. For instance if all the eigenvalues are small then the polynomial behaves like a single Gaussian. Let's define this notion of regular polynomials.

▶ **Definition 9.** If all the eigenvalues of a polynomial p are small relative to it's variance, that is $|\lambda_{max}(p)| \le \epsilon\sqrt{Var[p]}$, then it is called an ϵ-regular polynomial.

Central Limit theorem

We would need the following Central Limit Theorem from [3](Lemma 31 in their paper). It essentially says that if all the eigenvalues of a degree 2 polynomial p are small then the polynomial can be well approximated with a single Gaussian which has the same mean and variance. That is,

▶ **Lemma 10.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-2 polynomial over independent standard Gaussians. If $|\lambda_{max}(p)| \le \epsilon\sqrt{Var[p]}$, then p is $O(\epsilon)$-close to the Gaussian $\mathcal{N}(\mathbb{E}[p], Var[p])$ in total variation distance(hence also in Kolmogorov distance).*

2.3 Definition of L and basic facts.

We define L as follows: L is determined by a hash function $h : [n] \to [m]$ and a sign function $\sigma : [n] \to \{\pm 1\}$ as follows:

$$L(y)_i = \sigma(i) y_{h(i)}.$$

Note that for each $i \in [n]$, h is uniformly random on $[m]$ and σ is ± 1 uniformly at random. h, σ are chosen from 8-wise independent families. Thus L can be represented by a $n \times m$ matrix where the ith row of L is $c_i = \sigma(i) e_{h(i)}$ where e_j is the jth standard basis vector of $\mathbb{R}^m$. It is depicted in the following figure:

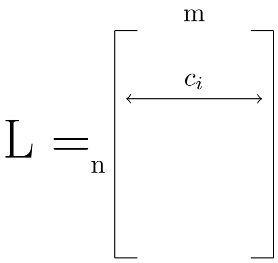

Figure 1 Construction of L

Note that the rows of L satisfy the following properties:

$$\mathbb{E}_L[\langle c_i, c_j \rangle^1] = \mathbb{E}_L[\langle c_i, c_j \rangle^3] = \delta_{ij}.$$

$$\mathbb{E}_L[\langle c_i, c_j \rangle^2] = \begin{cases} 1, & \textit{if i=j.} \\ \frac{1}{m}, & \textit{else.} \end{cases}$$

Note that this is a standard *Johnson-Lindenstrauss* matrix. In the following Lemma we show that they preserve L^2 norms and inner products of vectors to give a feel for the kind of computations we need. In fact L^t preserves a lot more structure as we shall see in the next section.

▶ **Lemma 11.** *For any $n, \epsilon > 0$, there exists an $m = poly(\frac{1}{\epsilon})$ and an explicit family of Linear transformations L^t(with seed length $O(\log n)$ from $\{\pm 1\}^m \to \{\pm 1\}^n$) so that for any two unit vectors $v_1, v_2 \in \mathbb{R}^n$ we have*

$$|\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle| < \epsilon \text{ wp } 1-2\epsilon \text{ over } L.$$

Proof. We know that $L^t v_1 = \sum_{i=1}^{n} v_1^i c_i, L^t v_2 = \sum_{j=1}^{n} v_2^j c_j$.

Thus we have

$$\langle L^t v_1, L^t v_2\rangle = \langle \sum_{i=1}^{n} v_1^i c_i, \sum_{j=1}^{n} v_2^j c_j\rangle = \sum_{i,j\in[n]} v_1^i v_2^j \langle c_i, c_j\rangle$$

$$\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle = \sum_{i\neq j\in[n]} v_1^i v_2^j \langle c_i, c_j\rangle$$

$$(\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle)^2 = \sum_{\substack{i_1\neq j_1\\ i_2\neq j_2}} v_1^{i_1} v_1^{i_2} v_2^{j_1} v_2^{j_2} \langle c_{i_1}, c_{j_1}\rangle\langle c_{i_2}, c_{j_2}\rangle.$$

Note that when averaged wrt $\mathbb{E}_L$, the only terms that survive are those that are paired either as $(i_1 = i_2, j_1 = j_2)$ or $(i_1 = j_2, i_2 = j_1)$.

The rest of the terms average to 0 because of the sign σ, that is $\mathbb{E}_\sigma[\sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)]$ only survives if the indices are paired and we already have the constraints $i_1 \neq j_1, i_2 \neq j_2$.

Thus we have

$$\begin{aligned}\mathbb{E}_L(\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle)^2 &= \sum_{i\neq j}(v_1^i)^2(v_2^j)^2\mathbb{E}_L\langle c_i, c_j\rangle^2 + \sum_{i\neq j} v_1^i v_2^i v_1^j v_2^j \mathbb{E}_L\langle c_i, c_j\rangle^2\\ &= \frac{1}{m}\sum_{i\neq j}[(v_1^i)^2(v_2^j)^2 + v_1^i v_2^i v_1^j v_2^j]\\ &\leq \frac{1}{m}(|v_1|_2^2|v_2|_2^2 + \langle v_1, v_2\rangle^2) \leq \frac{2}{m}\end{aligned}$$

Thus using Chebyshev's inequality we have

$$|\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle| \leq \frac{1}{m^{1/3}} \text{ wp } \left(1 - \frac{2}{m^{1/3}}\right) \text{ over } L.$$

Now we choose $m = \frac{1}{\epsilon^3}$ to have

$$|\langle L^t v_1, L^t v_2\rangle - \langle v_1, v_2\rangle| \leq \epsilon \text{ wp } 1-2\epsilon \text{ over } L.$$

This completes the proof. ◀

To see that norms are preserved too just choose $v_1 = v_2$ above.

Note

All through the paper we will be computing such expected moments and bounding them by $\frac{1}{m^{\Omega(1)}}$ and then use Markov|Chebyshev's inequality (We can't use big moments because L has limited independence). Think of these errors as small because after all the parameters are fixed we pick $m = \frac{1}{\epsilon^{\Omega(1)}}$, to be a sufficiently large polynomial of $\frac{1}{\epsilon}$ to bound all the terms by $O(\epsilon)$. We showed the constants explicitly in the above Lemma but we would not be computing them exactly later on and just denote them with $O(1)$.

2.4 Technical Lemmas involving L

We show that the transformation $p \to pL$ doesn't change the variance by a lot. If $p(x) = x^tAx + B^tx + C$ then $pL(y) = y^t(L^tAL)y + (B^tL)y + C$. Note that this is just a basic moment computation and doesn't involve anything non trivial.

▶ **Lemma 12.** *If $p(x) = x^tAx + B^tx + C$ is a multilinear polynomial, $pL(y) = y^t(L^tAL)y + (B^tL)y + C$. Then,*

$$\mathbb{E}_L Var[pL] = \sum_{i=1}^{n} b_i^2 + \Big(1 + \frac{3}{m}\Big)|A|_F^2 = Var[p] + \frac{3}{m}|A|_F^2$$

Proof. We know that

$$Var[p] = \sum_{i=1}^{n}(b_i^2 + a_{ii}^2) + ||A||_F^2 = \sum_{i=1}^{n} b_i^2 + ||A||_F^2.$$

Let's compute the same for pL. Note that $L^tAL = \sum_{i,j\in[n]} a_{ij}c_i \otimes c_j$.

Thus,

$$\begin{aligned}|L^tAL|_F^2 &= \sum_{i_1,j_1,i_2,j_2\in[n]} a_{i_1,j_1}a_{i_2,j_2}\langle c_{i_1}, c_{i_2}\rangle\langle c_{j_1}, c_{j_2}\rangle \\ &= \sum_{\substack{i_1\neq j_1\\ i_2\neq j_2}} \sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)a_{i_1,j_1}a_{i_2,j_2}I\{h(i_1)=h(i_2), h(j_1)=h(j_2)\}.\end{aligned}$$

Let's take expectation over σ. We know that $\mathbb{E}_\sigma[\sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)] \neq 0$ iff $(i_1, j_1) = (i_2, j_2)$ or $(i_1, j_1) = (j_2, i_2)$.

Let T_1 denote the terms of the first kind, then we have $T_1 = \sum_{i_1,j_1} a_{i_1,j_1}^2 = |A|_F^2$. Let T_2 denote the terms of the second kind, then we have $T_2 = \sum_{i_1,j_1} a_{i_1,j_1}^2 I\{h(i_1)=h(j_1)\}$ and thus $E_L[T_2] = \sum_{i_1,j_1} a_{i_1,j_1}^2\frac{1}{m} = \frac{1}{m}|A|_F^2$.

Also

$$\langle B^tL, B^tL\rangle = \sum_{i_1,i_2\in[n]} b_{i_1}b_{i_2}\langle c_{i_1}, c_{i_2}\rangle = \sum_{i_1,i_2}\sigma(i_1)\sigma(i_2)b_{i_1}b_{i_2}I\{h(i_1) = h(i_2)\}$$

$$\mathbb{E}_\sigma\langle B^tL, B^tL\rangle = \sum_i b_i^2.$$

We now compute $\sum_{l\in[m]}(L^tAL)_{ll}^2$.

$$\begin{aligned}\sum_{l\in[m]}(L^tAL)_{ll}^2 &= \sum_{l=1}^{m}\Big(\sum_{i,j\in[n]} a_{ij}c_i^lc_j^l\Big)^2 = \sum_{i_1,i_2,j_1,j_2} a_{i_1j_1}a_{i_2j_2}\sum_{l=1}^{m} c_{i_1}^lc_{i_2}^lc_{j_1}^lc_{j_2}^l \\ &= \sum_{\substack{i_1\neq j_1\\ i_2\neq j_2}} \sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)a_{i_1,j_1}a_{i_2,j_2}I\{h(i_1)=h(i_2)=h(j_1)=h(j_2)\}\end{aligned}$$

Let's take expectation over σ. We know that $\mathbb{E}_\sigma[\sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)] \neq 0$ iff $(i_1, j_1) = (i_2, j_2)$ or $(i_1, j_1) = (j_2, i_2)$.

$$\mathbb{E}_\sigma\sum_{l\in[m]}(L^tAL)_{ll}^2 = 2\sum_{i_1,j_1} a_{i_1j_1}^2 I\{h(i_1) = h(j_1)\}$$

Thus,

$$\mathbb{E}_L \sum_{l\in[m]} (L^tAL)_{ll}^2 = \frac{2}{m}|A|_F^2. \qquad ◀$$

In the following Lemma we prove bounds on $Var_L[Var_y[pL]]$. This would help us show that $Var_y[pL] = \Theta(Var[p])$ whp.

▶ **Lemma 13.**

$$Var_L[Var_y[pL]] = \frac{O(1)}{m}.$$

Proof. From Lemma 12 we have

$$\begin{aligned} Var_y[pL] &= |L^tAL|_F^2 + |B^tL|_2^2 + \sum_{l=1}^{m}(L^tAL)_{ll}^2 \\ &= \sum_{i_1,i_2,j_1,j_2} a_{i_1j_1}a_{i_2j_2}\langle c_{i_1}, c_{i_2}\rangle\langle c_{j_1}, c_{j_2}\rangle + \sum_{r_1,r_2} b_{r_1}b_{r_2}\langle c_{r_1}, c_{r_2}\rangle + \sum_{l=1}^{m}(L^tAL)_{ll}^2 \end{aligned}$$

where

$$\sum_{l\in[m]} (L^tAL)_{ll}^2 = \sum_{\substack{i_1\neq j_1\\ i_2\neq j_2}} \sigma(i_1)\sigma(i_2)\sigma(j_1)\sigma(j_2)a_{i_1,j_1}a_{i_2,j_2}I\{h(i_1){=}h(i_2){=}h(j_1){=}h(j_2)\}$$

Thus we have

$$\begin{aligned} Var_y[pL] - \mathbb{E}_L[Var_y[pL]] = &\sum_{(i_1,j_1)\neq(i_2,j_2)} a_{i_1j_1}a_{i_2j_2}\langle c_{i_1}, c_{i_2}\rangle\langle c_{j_1}, c_{j_2}\rangle + \sum_{r_1\neq r_2} b_{r_1}b_{r_2}\langle c_{r_1}, c_{r_2}\rangle \\ &+ \sum_{l=1}^{m}(L^tAL)_{ll}^2 - \frac{3}{m}|A|_F^2. \end{aligned}$$

We skip showing the elaborate yet simple moment calculations but observe that when squared and averaged over L each term above will have atleast a $\frac{1}{m}$ term in it. Also the corresponding coefficients can be bounded using Cauchy Schwarz and noting that $|B|_2^2 \leq 1$ and $|A|_F^2 \leq 1$.

Thus

$$\mathbb{E}_L\Big(Var_y[pL] - \mathbb{E}_L[Var_y[pL]]\Big)^2 = O\Big(\frac{Var^2[p]}{m}\Big). \qquad ◀$$

Now we put together these two Lemmas to show that $Var_y[pL] = \Theta(Var[p])$ whp. We exclude the proof as it is a direct consequence of Chebyshev inequality using Lemma 12 and Lemma 13.

▶ **Lemma 14.**

$$|Var_y[pL] - Var[p]| \leq O\Big(\frac{Var[p]}{m^{1/3}}\Big) \ \textit{wp}\ \Big(1 - \frac{1}{m^{1/3}}\Big) \ \textit{over}\ L.$$

The following lemma would also be useful. Intuitively it means that L would not perturb an eigenvalue of A by a huge amount whp. In fact this would imply that all the eigenvalues of A would be in the *pseudospectrum* of L^tAL.

▶ **Lemma 15.** *Let λ be an eigenvalue of A and let the unit vector v be the corresponding eigenvector. Then we have*

$$\mathbb{E}_L|(L^tAL - \lambda I_{m\times m})L^tv|_2^2 = O\Big(\frac{1}{m}\Big).$$

Proof. Substituting $Av = \lambda v$, we have

$$(L^tAL - \lambda I_{m\times m})L^tv = L^tALL^tv - L^tAv.$$

Expanding the product L^tALL^tv we have,

$$L^tALL^tv = \sum_{i,j,k\in[n]} a_{i,j}c_i\langle c_j, c_k\rangle v_k$$

$$L^tALL^tv - L^tAv = \sum_{\substack{i\\ j\neq k}} a_{i,j}c_i\langle c_j, c_k\rangle v_k$$

Thus

$$|L^tALL^tv - L^tAv|_2^2 = \sum_{\substack{i_1,i_2\\ j_1\neq k_1\\ j_2\neq k_2}} a_{i_1,j_1}a_{i_2,j_2}v_{k_1}v_{k_2}\langle c_{i_1}, c_{i_2}\rangle\langle c_{j_1}, c_{k_1}\rangle\langle c_{j_2}, c_{k_2}\rangle.$$

A term survives $\mathbb{E}_\sigma$ only if all the indices $\{i_1, i_2, j_1, j_2, k_1, k_2\}$ are paired appropriately. However when we take $\mathbb{E}_h$ since we have $j_1 \neq k_1, j_2 \neq k_2$ we would see atleast a $1/m$ in every term. Now the corresponding coefficient can be bounded using Cauchy Schwarz and noting that $|v|_2^2 = 1$ and $|A|_F^2 \leq 1$. Thus we have

$$\mathbb{E}_L|(L^tAL - \lambda I_{m\times m})L^tv|_2^2 = \frac{O(1)}{m}. \qquad ◀$$

3 The regular case

A polynomial is *regular* if a single variable can't influence its value by a lot. This comes into play when we try to employ the *Invariance principle.* Invariance principle shows that when the underlying variables are changed from Boolean to Gaussian the probability that the polynomial is positive will change by an amount proportional to the maximum influence of a variable over the polynomial. Thus let's assume *regularity* in this section so that we don't incur much error when we switch between Boolean and Gaussian inputs.

Proof under the assumption that polynomial is regular

In this section we assume that the degree 2 polynomial $p(x)$ is τ-regular and show that $|\underset{x\sim\{\pm1\}^n}{\mathbb{E}} sgn(p(x)) - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(pL(y))|$ is small whp over L.
We do this in three steps:

- *Replacement from Boolean to Gaussian for $p(x)$* We change the underlying input variables from Boolean to Gaussian. Since we assume the polynomial is regular, we do not incur much error when we do this via Invariance principle.
- *PRG error for Gaussian setting* Once we are in Gaussian setting we show that L is a *pseudorandom generator* for degree 2 polynomials for Gaussian inputs. The basic idea is that for PTFs in the Gaussian context one only needs to keep track of the top few eigenvalues and the total L^2 norm of rest of rest of the eigenvalues. We show that a Johnson-Lindenstrauss matrix preserves this top eigenvalue structure and the mass in the rest of the eigenvalues.

- *Replacement from Gaussian to Boolean for pL* Since p is regular we show that pL is regular whp too. This let's us go back from Gaussian to Boolean setting via the Invariance principle.

This is depicted in the following equation:

$$\begin{aligned}|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))| \leq\ & |\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x))| \\ & + |\mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y))| \\ & + |\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))|.\end{aligned}$$

The first term $|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x))| \leq O(\epsilon)$ using invariance principle from [16] since we assumed that $p(x)$ is τ-regular where $\tau = \epsilon^{O(1)}$.

Now we bound the other two terms in the following sections.

3.1 Gaussian PRG $|\mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y))|$

In this section we will show that in the Gaussian setting p cannot distinguish between x and Ly. The main idea is that to understand the average sign of a degree 2 polynomial you just need to keep track of the top few eigenvalues and the total mass in the rest of the eigenvalues. This is because either the latter eigenvalues are too small and thus the truncated part overall contributes very little mass to the total polynomial or these eigenvalues are small but do contribute a significant fraction of the total mass(we call this part the eigenregular part), then you could replace all of them by a single Gaussian with the same total mass via the CLT tools used in [3].

Thus let's think of the polynomial p as the top few eigenvalues and a lump mass of the rest of the eigenvalues. The Johnson-Lindenstrauss like matrix L we use preserves the top eigenvalue structure of the polynomial and also keeps the eigenregular part still *eigenregular*. It introduces some negligible dependence between the top eigenvalue part and the *eigenregular* part which we remove to begin with to keep them independent.

To begin with assume $p(x) = x^tAx + B^tx + C$ be a degree 2 multilinear polynomial with $|A|_F = 1$. Since A is a real symmetric matrix, let it be diagonalised as $A = V\Lambda V^t$, where V is an orthonormal matrix who columns are the eigenvectors of A. Let the eigenvalues of A be $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. Now let $k+1$ be the first index with $|\lambda_{k+1}| < \delta$ where we will choose $\delta = \epsilon^{O(1)}$ later on. Since $\sum_{i\in[n]} \lambda_i^2 = 1$, we know that $k \leq \frac{1}{\delta^2} = (\frac{1}{\epsilon})^{O(1)} \ll m$. Let $V_{\leq k}$ denote the first k eigenvectors of V and Λ_k denote the top $k \times k$ diagonal submatrix of Λ containing the top k eigenvalues of A.

▶ **Definition 16.** Define $A_1 = V_{\leq k}\Lambda_k V_{\leq k}^t$ to be the top eigenpart of A and $A_2 = V_{>k}\Lambda_{k+1}^n V_{>k}^t$ to be the lower eigenpart of A, we have $A = A_1 + A_2$.

Accordingly decompose $p(x) = q_1(x) + r_1(x)$ where

$$q_1(x) = x^tA_1x + B^tV_{\leq k}V_{\leq k}^tx + C,$$
$$r_1(x) = x^tA_2x + B^tV_{>k}V_{>k}^tx.$$

Note that $q_1(x)$ and $r_1(x)$ are independent of each other because the columns of $V_{\leq k}$ are orthogonal to the columns of $V_{>k}$. In the following lemma we replace $r_1(x)$ by just a single Gaussian that has the same mass and thus ignoring the total structure of $r_1(x)$. Let z be an one dimensional Gaussian independent of x.

▶ **Lemma 17.** *Given $\epsilon > 0$ let δ be a sufficiently large power of $\epsilon, \delta{=}\epsilon^{O(1)}$. If $p(x)$ can be written as a sum of two independent polynomials, that is $p(x) = q_1(x) + r_1(x)$ where $|\lambda_{max}(r_1)| < \delta$, then*

$$\Big|\underset{x\sim\mathcal{N}^n(0,1)}{\mathbb{E}} sgn(p(x)) - \underset{\substack{x\sim\mathcal{N}^n(0,1)\\ z\sim\mathcal{N}(0,1)}}{\mathbb{E}} sgn\Big(q_1(x) + \sqrt{Var[r_1]}z\Big)\Big| < O(\epsilon).$$

Proof. We consider two cases:

- Case I - Say r_1 has very small variance, that is $\sqrt{Var[r_1(x)]} < \frac{\delta}{\epsilon}$. Then we can use Lemma 5 to see that the replacement of $r_1(x)$ by $\sqrt{Var[r_1]}z$ will only incur an error of at most $O(\frac{\delta}{\epsilon})^{\frac{1}{3}}$. By an appropriate choice of $\delta = \epsilon^{O(1)}$ that we make later on this error will be $O(\epsilon)$.
- Case II - Say $\sqrt{Var[r_1(x)]} > \frac{\delta}{\epsilon}$, then note that every eigenvalue λ of $r_1(x)$ satisfies $|\lambda| < \epsilon\sqrt{Var[r_1]}$. Such a polynomial all of whose eigenvalues are small compared to its variance are called eigenregular polynomials and we could use Lemma 10 to replace $r_1(x)$ by $\sqrt{Var[r_1]}z$ and incur an error of at most $O(\epsilon)$. Note that we are using the independence of $q_1(x)$ and $r_1(x)$ in a *convolution* argument used to insert q_1 after applying the CLT.

Thus in either case the lemma holds after an appropriate choice of $\delta = \epsilon^{O(1)}$. ◀

To keep the presentation simple henceforth we assume that $L^tV_{\leq k}$ still has *orthonormal* columns, that is $V^t_{\leq k}LL^tV_{\leq k} = I_{k\times k}$. The exact computation proceeds by first using the *Gram Schmidt process* to orthonormalize $\{L^tv_1, \dots L^tv_k\}$. However this would not be very different from the exact analysis because L approximately preserves inner products and norms whp and we can union bound because k is a small constant depending on ϵ. In particular we have the following lemma.

▶ **Lemma 18.**

$$\mathbb{E}_L\Big|V^t_{\leq k}LL^tV_{\leq k}{-}I_{k\times k}\Big|^2_F = O\Big(\frac{k^2}{m}\Big).$$

Proof. This is a straightforward computation. Replacing $I_{k\times k}{=}V^t_{\leq k}V_{\leq k}$, we have $V^t_{\leq k}LL^tV_{\leq k}{-}I_{k\times k} = V^t_{\leq k}(LL^t{-}I_{n\times n})V_{\leq k}$. This gives,

$$\begin{aligned}\Big|V^t_{\leq k}(LL^t{-}I_{n\times n})V_{\leq k}\Big|^2_F &= \sum_{a,b\in[k]}\Big(\sum_{i_1\neq i_2} v_a^{i_1}v_b^{i_2}\langle c_{i_1}, c_{i_2}\rangle\Big)^2\\ &= \sum_{a,b\in[k]}\sum_{\substack{i_1\neq i_2\\ i_3\neq i_4}} v_a^{i_1}v_b^{i_2}v_a^{i_3}v_b^{i_4}\langle c_{i_1}, c_{i_2}\rangle\langle c_{i_3}, c_{i_4}\rangle.\end{aligned}$$

This gives

$$\mathbb{E}_L\Big|V^t_{\leq k}(LL^t{-}I_{n\times n})V_{\leq k}\Big|^2_F \leq O\Big(\frac{k^2}{m}\Big)$$ ◀

Let $y \sim \mathcal{N}^m(0,1)$ be a Gaussian independent of x, z. Since Gaussian distribution is invariant to rotations $V^t_{\leq k}x \sim \mathcal{N}^k(0,1)$ and $[V^t_{\leq k}L]y \sim \mathcal{N}^k(0,1)$ are identically distributed. Thus $q_1(x) = [x^tV_{\leq k}]\Lambda_k[V^t_{\leq k}x] + B^tV_{\leq k}[V^t_{\leq k}x] + C$ is *identically* distributed as $[y^tL^tV_{\leq k}]\Lambda_k[V^t_{\leq k}Ly] + B^tV_{\leq k}[V^t_{\leq k}Ly] + C$ which is exactly $q_1(Ly)$.

Thus we have,

$$\Big|\underset{x\sim\mathcal{N}^n(0,1)}{\mathbb{E}} sgn(p(x)) - \underset{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}}{\mathbb{E}} sgn\Big(q_1(Ly) + \sqrt{Var[r_1]}z\Big)\Big| < O(\epsilon).$$

Let's look at $p(Ly)$. We have

$$p(Ly) = y^t L^t A L y + B^t L y + C = y^t [L^t V] \Lambda [V^t L] y + B^t L y + C.$$

Let P denote the *projection matrix* onto the vector space spanned by $L^t v_1, \ldots, L^t v_k$. The *projection matrix* can be expressed by the $m \times m$ matrix $P \overset{def}{=} L^t V_{\leq k}(V_{\leq k}^t L L^t V_{\leq k})^{-1} V_{\leq k}^t L$. Note that $P^2 = P, P^t = P$. Since $V_{\leq k}^t L L^t V_{\leq k} = I_{k \times k}$, this simplifies to $P = L^t V_{\leq k} V_{\leq k}^t L$. Now as before we break $p(Ly)$ into two pieces $p(Ly) = q_2(y) + r_2(y)$, wherein

$$q_2(y) = y^t L^t A_1 L y + B^t L P y + C$$
$$r_2(y) = y^t L^t A_2 L y + B^t L [I - P] y$$

The goal is to do similar *CLT like analysis* but the problem is that $q_2(y)$ and $r_2(y)$ are not independent. We refine $r_2(y)$ to $r_3(y)$ to make it independent of $q_2(y)$ by separating the part of it that correlates with $q_2(y)$. That is, define

$$r_3(y) = y^t [I - P] L^t A_2 L [I - P] y + B^t L [I - P] y$$
$$s(y) = y^t P L^t A_2 L [I - P] y + y^t L^t A_2 L P y.$$

Observe that $r_3(y)$ is independent of $q_2(y)$. We have $p(Ly) = q_2(y) + r_3(y) + s(y)$. First let's get rid of $s(y)$ by showing that $Var[s]$ is small whp over L and invoking Lemma 5.

▶ **Lemma 19.** *$Var[s] = O(\frac{1}{\sqrt{m}})$ wp $\left(1 - \frac{O(1)}{\sqrt{m}}\right)$ over L.*

Proof. It suffices to show that $|L^t A_2 L P|_F$ is small. Since P is a projection matrix we have,

$$|L^t A_2 L P|_F^2 = Tr[L^t A_2 L P L^t A_2 L] = |L^t A_2 L L^t V_{\leq k}|_F^2.$$

Since $A = A_1 + A_2$, we have

$$L^t A_2 L = L^t A L - L^t A_1 L = L^t A L - L^t V_{\leq k} \Lambda_k V_{\leq k}^t L.$$

Thus

$$L^t A_2 L L^t V_{\leq k} = (L^t A L) L^t V_{\leq k} - L^t V_{\leq k} \Lambda_k \overbrace{V_{\leq k}^t L L^t V_{\leq k}}^{I_{k \times k}}$$
$$= (L^t A L) L^t V_{\leq k} - L^t V_{\leq k} \Lambda_k,$$

Thus we have

$$|L^t A_2 L L^t V_{\leq k}|_F^2 = \sum_{l=1}^{k} |(L^t A L) L^t v_l - \lambda_l L^t v_l|_2^2.$$

Now we could use Lemma 15 to bound this. So we have,

$$\mathbb{E}_L |L^t A_2 L L^t V_{\leq k}|_F^2 = O\left(\frac{k}{m}\right).$$

Now the Lemma follows by Markov's inequality and noting that $Var[s] = O(|L^t A_2 L P|_F^2)$. ◀

Now we could apply Lemma 5 to remove s. That is,

$$\left|\underset{y \sim \mathcal{N}^m(0,1)}{\mathbb{E}} sgn(q_2(y) + r_3(y)) - \underset{y \sim \mathcal{N}^m(0,1)}{\mathbb{E}} sgn(pL(y))\right| \leq O\left(\frac{1}{m^{O(1)}}\right) \; wp \; \left(1 - \frac{O(1)}{\sqrt{m}}\right) \; over \; L$$

Now that $q_2(y)$ and $r_3(y)$ are independent, to go ahead with the CLT like analysis we first show that the largest eigenvalue of $r_3(y)$ is at most $\sqrt{\delta}$.

▶ **Lemma 20.** $\lambda_{max}[r_3(y)] \leq \sqrt{\delta}$ *whp*

Proof. We want to show that the eigenvalues of $[I{-}P](L^tA_2L)[I{-}P]$ are small. Its eigenvalues are interlaced into the eigenvalues of L^tA_2L because $[I{-}P]$ is a projection matrix. Thus it suffices to bound the eigenvalues of L^tA_2L, where $A_2 = V_{>k}\Lambda_{k+1}^{n}V_{>k}^{t}$. Note that A_2 is a symmetric matrix with spectrum $0^k, \lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_n$. To bound the eigenvalues of L^tA_2L we bound $Tr(L^tA_2L)^4 = |(L^tA_2L)^2|_F^2$. We have

$$\begin{aligned}
|(L^tA_2L)^2|_F^2 &= \sum_{j_1\cdots j_8\in[n]} A_{2_{j_1j_2}}A_{2_{j_3j_4}}A_{2_{j_5j_6}}A_{2_{j_7j_8}}\langle c_{j_1}, c_{j_5}\rangle\langle c_{j_2}, c_{j_3}\rangle\langle c_{j_4}, c_{j_8}\rangle\langle c_{j_6}, c_{j_7}\rangle \\
\mathbb{E}_L|(L^tA_2L)^2|_F^2 &= \sum_{j_1,j_2,j_4,j_6\in[n]} A_{2_{j_1j_2}}A_{2_{j_2j_4}}A_{2_{j_1j_6}}A_{2_{j_6j_4}} + \frac{O(1)}{m} \\
&= Tr(A_2^4) + \frac{O(1)}{m} \leq \delta^2 + \frac{O(1)}{m}.
\end{aligned}$$

This shows that the maximum absolute eigenvalue of $r_3(y)$ is at most $O(\sqrt{\delta})$ whp. This let's us either remove it as a low variance term or apply the CLT machinery on $r_3(y)$. ◀

Now that q_2 and r_3 are independent polynomials and since Lemma 20 gives $\lambda_{max}[r_3(y)] \leq \sqrt{\delta}$ we could use a slight variant of Lemma 17 to bound the following error:

$$|\mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_2(y) + \sqrt{Var[r_3]}z\Big) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(q_2(y){+}r_3(y))| < O(\epsilon).$$

We now bound the remaining term that finishes the telescoping for the Gaussian PRG part.

▶ **Lemma 21.**

$$\Big|\mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_1(Ly) + \sqrt{Var[r_1]}z\Big) - \mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_2(y) + \sqrt{Var[r_3]}z\Big)\Big| \leq O(\epsilon)whp$$

Proof. Since y and z are independent it suffices to show that $Var_y[q_1(Ly){-}q_2(y)]$ and $|Var[r_1]{-}Var[r_3]|$ are both small and invoke Lemma 5 to prove this Lemma.

We have

$$\begin{aligned}
q_2(y){-}q_1(Ly) &= B^t[LL^t{-}I]V_{\leq k}V_{\leq k}^tLy \\
Var\Big[q_2(y){-}q_1(Ly)\Big] &= \Big|B[LL^t{-}I]V_{\leq k}V_{\leq k}^tL\Big|_2^2
\end{aligned}$$

Since $L^tV_{\leq k}$ has orthonormal columns, this simplifies further to

$$\begin{aligned}
Var\Big[q_2(y){-}q_1(Ly)\Big] &= \Big|B^t[LL^t{-}I]V_{\leq k}\Big|_2^2 = \sum_{l=1}^{k}\Big[\sum_{j_1\neq j_2\in[n]}\langle c_{j_1}, c_{j_2}\rangle b_{j_1}v_l^{j_2}\Big]^2 \\
\textit{Thus} \quad \mathbb{E}_L Var\Big[q_2(y){-}q_1(Ly)\Big] &= O\Big(\frac{k|B|_2}{m}\Big).
\end{aligned}$$

To see that $Var[r_1] \approx Var[r_3]$, note that $Var[r_3] \approx Var[r_2]$ because $Var[s(y)]$ is small as shown above. Now to show that $Var[r_1] \approx Var[r_2]$ we need to show the following:

- $|A_2|_F \approx |L^tA_2L|_F$. This follows from Lemma 12.

- $|B^tV_{>k}V^t_{>k}|_2 \approx |B^tL[I-P]|_2$. To show this note that $|B^tV_{>k}V^t_{>k}|^2_2 = \sum_{t=k+1}^{n} \langle B^t, v_t\rangle^2$. Since $L^tV_{\leq k}$ has orthonormal columns, we have

$$|B^tL[I-P]|^2_2 = |B^tL|^2_2 - \sum_{l=1}^{k} \langle B^tL, L^tv_l\rangle^2.$$

Now this follows by noting that L^t approximately preserves the norms and inner products of vectors and that since k is a constant we can union bound. ◀

To summarize we telescoped the Gaussian PRG error as:

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y))| \leq \\
&\quad \Big|\mathop{\mathbb{E}}_{x\sim\mathcal{N}^n(0,1)} sgn(p(x)) - \mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_1(Ly) + \sqrt{Var[r_1]}z\Big)\Big| \\
&\quad + \Big|\mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_1(Ly) + \sqrt{Var[r_1]}z\Big) - \mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_2(y) + \sqrt{Var[r_3]}z\Big)\Big| \\
&\quad + |\mathop{\mathbb{E}}_{\substack{y\sim\mathcal{N}^m(0,1)\\ z\sim\mathcal{N}(0,1)}} sgn\Big(q_2(y) + \sqrt{Var[r_3]}z\Big) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(q_2(y) + r_3(y))| \\
&\quad + |\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(q_2(y) + r_3(y)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y))|
\end{aligned}$$

and showed that each of the terms is small whp over L. This completes the analysis of the Gaussian PRG error term.

Now we move back from Gaussian to Boolean setting to finish the analysis for regular polynomials.

3.2 Replacement for pL $|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))|$

We do this change in two parts:

- *Linearize pL to pL_{lin}* The application of invariance principle needs the polynomial to be multilinear but pL need not be multilinear even though p is. Thus we pre-process pL to convert to the multilinear polynomial pL_{lin}.
- *Replacement for pL_{lin}* Since p is multilinear, we show that pL_{lin} is regular whp and then apply the invariance principle.

Thus we split the replacement term for pL as an error between pL, pL_{lin} in Gaussian setting and a replacement error for pL_{lin}. This is depicted in the following equation.

$$\begin{aligned}
&|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))| \leq \\
&\qquad\qquad |\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y))| \\
&\qquad\qquad + |\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL_{lin}(y))|.
\end{aligned}$$

3.2.1 Linearize pL to pL_{lin} $|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y))|$

Note that p is a multilinear polynomial but pL need not be multilinear. For example L could map both x_i, x_j to y_l and thus the monomial x_ix_j to y_l^2. y_l^2 would be the constant 1 in the boolean case but would be a non linear term in the Gaussian case. However the invariance principle works only for multilinear polynomials. Thus we linearize pL as follows:

▶ **Definition 22.** pL_{lin} is the linearized version of pL - Every occurence of a term like y_l^2 is replaced by the constant 1.

Note that pL_{lin} satisfies the following properties:

(i) $pL_{lin}(y) = pL(y)$ when y is Boolean.

(ii) $\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y) - pL_{lin}(y)] = 0.$

We bound this linearization error using lemma 23 and lemma 5. Lemma 5 shows that the distributional distance between pL, pL_{lin} is small if $|pL{-}pL_{lin}|_2$ is small and Lemma 23 shows that this is the case.

▶ **Lemma 23.** *The non-linear part of pL has small variance with high probability over L,*

$$\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y){-}pL_{lin}(y)]^2 < \frac{Var[p]}{\sqrt{m}} \ wp\ \left(1 - \frac{O(1)}{\sqrt{m}}\right) \ over\ L.$$

Proof. From the basic definitions of p, L, pL_{lin} we have,

$$\begin{aligned} p &= \sum_{i,j} a_{ij}x_ix_j + \sum_k b_kx_k + C. \\ pL &= \sum_{i,j} a_{ij}\sigma(i)\sigma(j)y_{h(i)}y_{h(j)} + \sum_k b_k\sigma(k)y_{h(k)} + C. \\ pL - pL_{lin} &= \sum_{t\in[m]} [y_t^2 - 1] \sum_{\substack{i,j:\\ h(i)=h(j)=t}} a_{ij}\sigma(i)\sigma(j). \end{aligned}$$

We calculate the variance of $pL{-}pL_{lin}$ by noting that $\mathop{\mathbb{E}}_{y_t\sim\mathcal{N}^1(0,1)}[y_t^2 - 1]^2 = 2$,

$$\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y) - pL_{lin}(y)]^2 = 2\sum_{t\in[m]} \sum_{\substack{i,j,k,l:\\ h(i)=h(j)=h(k)=h(l)=t}} a_{ij}a_{kl}\sigma(i)\sigma(j)\sigma(k)\sigma(l).$$

We then calculate the expected variance over the sign σ. This makes a term 0 unless it is paired as $\{i,j\} = \{k,l\}$.

$$\mathbb{E}_\sigma \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y) - pL_{lin}(y)]^2 = 2\sum_{t\in[m]} \sum_{\substack{i,j:\\ h(i)=h(j)=t}} a_{ij}^2.$$

Now we calculate the expected value of this over the hash function h.

$$\mathbb{E}_h\mathbb{E}_\sigma \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y) - pL_{lin}(y)]^2 = 2\sum_t\sum_{i,j}\frac{1}{m^2}a_{ij}^2 = \frac{2}{m}||A||_F^2 \le \frac{2}{m}Var[p].$$

The lemma now follows by the Markov inequality applied to $\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)}[pL(y){-}pL_{lin}(y)]^2$. ◀

Now we invoke Lemma 5 to finish the bound on the *linearization error.*

$$|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL(y)) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y))| \le \frac{O(1)}{m^{1/12}} \ wp\ \left(1 - \frac{O(1)}{\sqrt{m}}\right) \ over\ L.$$

3.2.2 Replacement for pL_{lin} $|\mathop{\mathbb{E}}\limits_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y)) - \mathop{\mathbb{E}}\limits_{y\sim\{\pm1\}^m} sgn(pL_{lin}(y))|$

Using Invariance principles from [16], Lemma 6 we have,

$$|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL_{lin}(y))| \le O\left[\frac{\sum_{r=1}^{m} Inf_r^2(pL_{lin})}{(Var[pL_{lin}])^2}\right]^{\frac{1}{9}}.$$

where

$$Inf_r(pL_{lin}) = 2\sum_{s\in[m]\setminus r}\left[\sum_{\substack{h(i)=r\\h(j)=s}} a_{ij}\sigma(i)\sigma(j)\right]^2 + \left[\sum_{h(l)=r} b_l\sigma(l)\right]^2$$

A simple but elaborate computation would show that

$$\mathbb{E}_L\left(\sum_{r=1}^{m} Inf_r^2(pL_{lin}) - \sum_{i=1}^{n} Inf_i^2(p)\right)^2 = O\left(\frac{Var^4[p]}{m}\right) = \frac{O(1)}{m}.$$

Thus using Chebyshev's inequality we have,

$$\sum_{r=1}^{m} Inf_r^2(pL_{lin}) \le \sum_{i=1}^{n} Inf_i^2(p) + \frac{Var^2[p]}{m^{\frac{1}{3}}} \;\; wp \;\left(1 - \frac{O(1)}{m^{\frac{1}{3}}}\right) \;\; over\ L.$$

Since p is τ-regular, we have

$$\sum_{j=1}^{n} Inf_j^2(p) \le \max_i Inf_i(p)\cdot\sum_{i=1}^{n} Inf_i(p) \le 2\tau Var^2[p].$$

Note that $Var[pL_{lin}] = \Theta(Var[p])$ wp $(1-\frac{1}{m^{O(1)}})$ over L by Lemma 12.
Thus we can bound the *replacement error* for pL_{lin} as follows:

$$|\mathop{\mathbb{E}}_{y\sim\mathcal{N}^m(0,1)} sgn(pL_{lin}(y)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL_{lin}(y))| \le O\left(2\tau + \frac{1}{\sqrt{m}}\right)^{\frac{1}{9}} \;\; wp \;\left(1 - \frac{1}{m^{O(1)}}\right) \;\; over\ L.$$

Note that we will be choosing $\tau = \epsilon^{O(1)}$ and $m = \frac{1}{\epsilon^{\Omega(1)}}$ which would also ensure that this error is $\le O(\epsilon)$.

4 Reduction to the regular case

If the polynomial p is *not regular* we fix few variables that have large influence to get to a polynomial that is either regular or constant. We note that under L the high influence variables would most likely have landed in separate bins and thus remain independent. The other variables that land in the same bin as one of these high influence variables do not contribute much to the size of the polynomial.

Proof that theorem holds for regular polynomials implies it holds for all polynomials

The idea of this reduction is a careful analysis of Lemma 8. This is the standard *Regularity Lemma* from [6].

We look at p as a decision tree using Lemma 8. If α denotes a path to the leaf in the decision tree then let p_α denote the restriction polynomial along the path. Let $S(\alpha)$ denote the set of variables that are set along the path α.

Note

p_α is only a function $n-|S(\alpha)|$ coordinates that are not set along the decision tree path α. For the ease of notation we look at it still as a function of n coordinates, wherein it just ignores the coordinates that are already fixed along α.

Averaging over all the decision tree paths we have,

$$\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) = \mathbb{E}_\alpha[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{[n]\setminus S(\alpha)}} sgn(p_\alpha(x))].$$

Thus let's split the change from x to Ly, $|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))|$ into two parts:

- *Change x to Ly only at decision tree leaves* Here we average the values along the decision paths based on the values set by α and only make the change x to Ly at the leaves. Note that analyzing this would be easier since the disagreement is only at the leaves and we know that the leaves are either regular or constant.
- *Changing Ly at leaves to Ly overall* Here we bound the error we incur by going from α setting the values along the decision path followed by Ly setting the values at the leaves to Ly setting the values overall. To analyse this we will be introducing a new distribution Ly' that only disagrees with Ly on very few variables.

This split is depicted in the following equation,

$$\begin{aligned}|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n} sgn(p(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))| \le \\ |\mathbb{E}_\alpha[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{[n]\setminus S(\alpha)}} sgn(p_\alpha(x))] - \mathbb{E}_\alpha[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))]| \\ +|\mathbb{E}_\alpha[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))] - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(pL(y))|\end{aligned}$$

4.1 Leaf change $|\mathbb{E}_\alpha[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{[n]\setminus S(\alpha)}} sgn(p_\alpha(x))] - \mathbb{E}_\alpha[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))]|$

Using Jensen's inequality, also known here as the triangle inequality we have

$$\begin{aligned}|\mathbb{E}_\alpha[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{S^c}} sgn(p_\alpha(x))] - \mathbb{E}_\alpha[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))]| \le \\ \mathbb{E}_\alpha|\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{S^c}} sgn(p_\alpha(x)) - \mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))|.\end{aligned}$$

By the regularity lemma we know that with probability atleast $(1-\theta)$ (where $\theta = \frac{1}{\sqrt{m}}$), the leaf p_α is either τ-regular (where $\tau = \epsilon^{O(1)}$) or p_α is almost constant (that is $Var[p_\alpha] \le \frac{1}{\sqrt{m}}$).

Now

- *Regular* If p_α is τ- regular, we could just bound the error by ϵ using our results from the previous section.
- *Constant* If p_α is almost constant, then wlog it is of the form $1+q$ where $Var[q] = \frac{1}{\sqrt{m}}$, thus $sgn(p_\alpha(x)) = sgn(p_\alpha(Ly))$ with probability $1-\frac{1}{\sqrt{m}}$.

In either case we have

$$\begin{aligned}&|\mathbb{E}_\alpha[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{[n]\setminus S(\alpha)}} sgn(p_\alpha(x))] - \mathbb{E}_\alpha[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^m} sgn(p_\alpha(Ly))]| \\ &\le (1-\theta)\epsilon + 2\theta \;\; wp \left(1 - \frac{1}{m^{O(1)}}\right) \; over \; L \\ &= \epsilon + \frac{1}{m^{O(1)}} \;\; wp \left(1 - \frac{1}{m^{O(1)}}\right) \; over \; L \; since \; \theta = \frac{1}{\sqrt{m}}.\end{aligned}$$

4.2 Changing Ly overall $|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly))] - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(pL(y))|$

To bound this term we introduce a new intermediate distribution Ly'. As long as L doesn't hash collide two high influence decision path α variables, Ly and Ly' are *identically* distributed. In fact they agree on all variables except on those variables that hash collide with the decision path variables. On these variables Ly' just assigns the decision path variable's value to every other variable that lands in its bin.

▶ **Definition 24.** We define Ly' as follows:

$$(Ly')_i = \begin{cases} x_i & i \in S(\alpha) \\ (Ly)_i & i \in LS(\alpha)^c \\ x_{h^{-1}(h(i))} & i \in LS(\alpha) \setminus S(\alpha). \end{cases}$$

where $LS = \{i \in [n] : h(i) \in h(S)\}$ and for $j \in [m], h^{-1}(j)$ is the smallest $i \in [n]$ such that $h(i) = j$. This essentially fixes all the bits that hash collide with the decision path variables.

Now we split the error in changing Ly on leaves to Ly overall further into two steps.

- *Changing Ly on leaves to Ly'* Here we observe that Ly and Ly' agree everywhere except the variables that hash collide with the decision path variables. The depth of the tree is small $D = \log m$, thus each variable collides with a decision path variable with very small probability $\frac{\log m}{m}$. Thus this difference amounts to a negligible fraction of the variance of the polynomial.
- *Changing from Ly' to Ly overall* Here we only need to bound the probability that two decision path variables along α dont hash collide. As long as thing doesn't happen Ly and Ly' are identically distributed.

This is depicted in the following equation:

$$\begin{aligned}|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly))] - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(pL(y))| \leq \\ |\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly))] - \mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))]| \\ +|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))] - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(pL(y))|\end{aligned}$$

4.2.1 Changing from Ly' to Ly overall $|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))] - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(pL(y))|$

Here we only need to bound the probability that there is no hash collision on $S(\alpha)$. This is because if h does not have a hash collision on $S(\alpha)$, then Ly and Ly' are identically distributed. That is,

$$\begin{aligned}\left|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^{[m]\setminus h(S(\alpha))}}{\mathbb{E}} sgn(p_\alpha(Ly'))] - \mathbb{E}_{y_{h[S(\alpha)]}}[\underset{y\sim\{\pm1\}^{[m]\setminus h[S(\alpha)]}}{\mathbb{E}} sgn(pL(y))]\right| \\ \leq 2\Pr_\alpha\left(|h[S(\alpha)]| \neq |S(\alpha)|\right).\end{aligned}$$

Note that the above equation is for a *fixed hash map* h. The probability is over the choice of random paths α of the decision tree but for this fixed h. It is the probability that a *random decision tree path* sees a collision wrt this fixed hash map h.

We show that for most hash maps (whp over L) this term is very small.

▶ **Lemma 25.**

$$\mathbb{E}_h\Big[\Pr_\alpha\Big(|h[S(\alpha)]| \neq |S(\alpha)|\Big)\Big] \leq \frac{D^2}{m}.$$

Proof. Interchange the expectations and fix a depth D decision tree path and observe that the probability that a random h has a collision along this path is at most $\frac{\binom{D}{2}}{m}$. ◀

Hence by Markov's inequality we have,

$$\Pr_\alpha\Big(|h[S(\alpha)]| \neq |S(\alpha)|\Big) \leq \frac{D^2}{\sqrt{m}} \text{ } \textit{wp} \text{ } \Big(1 - \frac{O(1)}{\sqrt{m}}\Big) \text{ } \textit{over } L.$$

Note that D here is $\mathrm{poly}(\log m)$.

4.2.2 Changing Ly on leaves to Ly' $|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly))] - \mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))]|$

Note that Ly and Ly' agree everywhere except the variables that hash collide with the decision path variables. Thus the part of the polynomial that disagrees wrt Ly and Ly' only contributes little mass to the total polynomial. However in order to use this fact to bound this term we would need to move back to Gaussian setting so that we could use Anticoncentration like ideas.

Jensen's inequality, also known here as the triangle inequality gives

$$\begin{aligned}|\mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly))] - \mathbb{E}_\alpha[\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))]| \\ \leq \mathbb{E}_\alpha\Big[|\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))|\Big].\end{aligned}$$

Now we move back to Gaussian setting by doing a replacement on both $p_\alpha(Ly), p_\alpha(Ly')$.

$$\begin{aligned}\mathbb{E}_\alpha\Big[|\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))|\Big] \leq \\ \mathbb{E}_\alpha\Big[|\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly))|\Big] \\ +\mathbb{E}_\alpha\Big[|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly'))|\Big] \\ +\mathbb{E}_\alpha\Big[|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly')) - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))|\Big].\end{aligned}$$

Replacement Errors:

Note that with probability $(1-\theta)$(where $\theta = \frac{1}{\sqrt{m}}$) the decision tree path α is such that $p_\alpha(\cdot)$ is either regular or almost constant. We use similar analysis as done in the previous section to bound both the replacement error terms as follows:

$$\mathbb{E}_\alpha\Big[|\underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly))|\Big] \leq O\Big(\epsilon + \frac{1}{m^{O(1)}}\Big) \text{ } \textit{wp} \text{ } \Big(1 - \frac{O(1)}{\sqrt{m}}\Big) \text{ } \textit{over } L$$

and

$$\mathbb{E}_\alpha\Big[|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly')) - \underset{y\sim\{\pm1\}^m}{\mathbb{E}} sgn(p_\alpha(Ly'))|\Big] \leq O\Big(\epsilon + \frac{1}{m^{O(1)}}\Big) \text{ } \textit{wp} \text{ } \Big(1 - \frac{O(1)}{\sqrt{m}}\Big) \text{ } \textit{over } L$$

Change Ly on leaves to Ly' in Gaussian setting

We just need to bound $\mathbb{E}_\alpha\Big[\big|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly'))\big|\Big]$.

We show that $\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}}[p_\alpha(Ly)-p_\alpha(Ly')]^2$ is small in Lemma 26 and then invoke Lemma 5 to bound $|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly'))|$.

We expect $\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}}[p_\alpha(Ly) - p_\alpha(Ly')]^2$ to be small because $p_\alpha(Ly), p_\alpha(Ly')$ agree on all terms except those that contain a variable that hash collides with the decision tree path. Since this is a low probability event ($\frac{D}{m} = \frac{(\log m)^{O(1)}}{m}$), we expect the overall mass in this difference $p_\alpha(Ly) - p_\alpha(Ly')$ to be very small.

▶ **Lemma 26.**

$$\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}}[p_\alpha(Ly) - p_\alpha(Ly')]^2 \leq \frac{1}{\sqrt{m}} \;wp\; \left(1 - \frac{O(1)}{\sqrt{m}}\right) \;over\; L.$$

Proof. Let $W = LS(\alpha) \setminus S(\alpha)$. Expanding out the terms we have

$$\begin{aligned} p_\alpha(Ly) - p_\alpha(Ly') = 2 &\sum_{j_1\in S, j_2\in W} a_{j_1j_2}x_{j_1}[\sigma(j_2)y_{h(j_2)} - x_{h^{-1}[h(j_2)]}] \\ &+ \sum_{l\in W} b_l[\sigma(l)y_{h(l)} - x_{h^{-1}(h(l))}] \\ &+ 2 \sum_{j_1\in LS^c, j_2\in W} \sigma(j_1)a_{j_1j_2}y_{h(j_1)}[\sigma(j_2)y_{h(j_2)} - x_{h^{-1}[h(j_2)]}] \\ &+ \sum_{j_1,j_2\in W} a_{j_1j_2}[\sigma(j_1)\sigma(j_2)y_{h(j_1)}y_{h(j_2)} - x_{h^{-1}[h(j_1)]}x_{h^{-1}[h(j_2)]}]. \end{aligned}$$

Note that j_2 needs to be in W in all of the terms above. This is a very low probability event. In fact $Pr_h(j_2 \in W) = \frac{D}{m}$ where $D = |S(\alpha)|$ is the depth of the decision tree and is chosen to be $(\log m)^{O(1)}$.

$$\mathbb{E}_L \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}}[p_\alpha(Ly) - p_\alpha(Ly')]^2 = O\Big(\frac{DVar[p_\alpha]}{m}\Big) = O\Big(\frac{(\log m)^{O(1)}}{m}\Big).$$

Thus

$$\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}}[p_\alpha(Ly) - p_\alpha(Ly')]^2 < \frac{(\log m)^{O(1)}}{\sqrt{m}} \;wp\; \left(1 - \frac{O(1)}{\sqrt{m}}\right) \;over\; L.$$ ◀

Now we invoke Lemma 5 to finish bounding this term,

$$\begin{aligned} \mathbb{E}_\alpha\Big[\big|\underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly)) - \underset{y\sim\mathcal{N}^m(0,1)}{\mathbb{E}} sgn(p_\alpha(Ly'))\big|\Big] &< \left(\frac{(\log m)^{O(1)}}{\sqrt{m}}\right)^{1/6} \\ &wp\; \left(1 - \frac{O(1)}{\sqrt{m}}\right) \;over\; L \end{aligned}$$

References

1 J. Wright A. Carbery. Distributional and l^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, 2001.

2 Richard Beigel. The polynomial method in circuit complexity. In *Proceedings of the Eigth Annual Structure in Complexity Theory Conference, San Diego, CA, USA, May 18-21, 1993*, pages 82–95. IEEE Computer Society, 1993. doi:10.1109/SCT.1993.336538.

3 Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. Deterministic approximate counting for degree-\$2\$ polynomial threshold functions. *CoRR*, abs/1311.7105, 2013. arXiv:1311.7105.

4 Anindya De and Rocco A. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. *CoRR*, abs/1311.7178, 2013. arXiv:1311.7178.

5 I. Diakonikolas, P. Gopalan, R. Jaiswal, R.A. Servedio, and E. Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 2010.

6 Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. *CoRR*, abs/0909.4727, 2009. arXiv:0909.4727.

7 Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 903–922. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.60.

8 Jelani Nelson Ilias Diakonikolas, Daniel M. Kane. Bounded independence fools degree-2 threshold functions. *Foundations of Computer Science (FOCS)*, 2010.

9 Daniel M. Kane. k-independent gaussians fool polynomial threshold functions. *Conference on Computational Complexity (CCC)*, 2011.

10 Daniel M. Kane. A small prg for polynomial threshold functions of gaussians. *Symposium on the Foundations Of Computer Science (FOCS)*, 2011.

11 Daniel M. Kane. A structure theorem for poorly anticoncentrated gaussian chaoses and applications to the study of polynomial threshold functions. *CORR*, 2012. URL: http://arxiv.org/abs/1204.0543.

12 Daniel M. Kane. A pseudorandom generator for polynomial threshold functions of gaussians with subpolynomial seed length. *Conference on Computational Complexity (CCC)*, 2014.

13 Daniel M. Kane. A polylogarithmic PRG for degree 2 threshold functions in the gaussian setting. In David Zuckerman, editor, *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, volume 33 of *LIPIcs*, pages 567–581. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015. doi:10.4230/LIPIcs.CCC.2015.567.

14 Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{\tilde{o}(n^{1/3})}$. In Jeffrey Scott Vitter, Paul G. Spirakis, and Mihalis Yannakakis, editors, *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pages 258–265. ACM, 2001. doi:10.1145/380752.380809.

15 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *CoRR*, abs/0910.4122, 2009. arXiv:0910.4122.

16 E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *ArXiv Mathematics e-prints*, 2005. arXiv:math/0503503.

17 Alexander A. Sherstov. Separating ac^0 from depth-2 majority circuits. In David S. Johnson and Uriel Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 294–301. ACM, 2007. doi:10.1145/1250790.1250834.

18 Valentine Kabanets and Zhenjian Lu. Nisan-wigderson pseudorandom generators for circuits with polynomial threshold gates. *ECCC*, https://eccc.weizmann.ac.il/report/2018/012/, 2018. URL: https://eccc.weizmann.ac.il/report/2018/012/.