A Note on the Pseudorandomness of Low-Degree Polynomials over the Integers

Aayush Jain* Alexis Korb† Paul Lou‡ Amit Sahai§

October 19, 2021

Abstract

We initiate the study of a problem called the Polynomial Independence Distinguishing Problem (PIDP). The problem is parameterized by a set of polynomials $\mathcal{Q} = (q_1, \ldots, q_m)$ where each $q_i : \mathbb{R}^n \to \mathbb{R}$ and an input distribution $\mathcal{D}$ over the reals. The goal of the problem is to distinguish a tuple of the form $\{q_i, q_i(\mathbf{x})\}_{i\in[m]}$ from $\{q_i, q_i(\mathbf{x}_i)\}_{i\in[m]}$ where $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_m$ are each sampled independently from the distribution $\mathcal{D}^n$. Refutation and search versions of this problem are conjectured to be hard in general for polynomial time algorithms (Feige, STOC 02) and are also subject to known theoretical lower bounds for various hierarchies (such as Sum-of-Squares and Sherali-Adams). Nevertheless, we show polynomial time distinguishers for the problem in several scenarios, including settings where such lower bounds apply to the search or refutation versions of the problem.

Our results apply to the setting when each polynomial is a constant degree multilinear polynomial. We show that this natural problem admits polynomial time distinguishing algorithms for the following scenarios:

- **Non-trivial Distinguishers.** We define a non-trivial distinguisher to be an algorithm that runs in time $n^{O(1)}$ and distinguishes between the two distributions with probability at least $n^{-O(1)}$. We show that such non-trivial distinguishers exist for large classes of *worst-case* families of polynomials, and essentially any non-trivial input distribution that is symmetric around zero, and isn't equivalent to a distribution over Boolean values.

  In particular, we show that when $m \geq n$ and the sets of indices corresponding to the variables present in each monomial exhibit a weak expansion property with expansion factor greater than $1/2$ for unions of at most 4 sets, then a non-trivial distinguisher exists.
- **Overwhelming Distinguishers.** Next we consider the problem of *amplifying* the success probability of the distinguisher, to guarantee that it succeeds with probability $1 - n^{-\omega(1)}$. We obtain such an overwhelming distinguisher for natural random classes of homogeneous multilinear constant degree d polynomials, denoted by $\mathcal{Q}_{n,d,p}$, and natural input distributions $\mathcal{D}$ such as discrete Gaussians or uniform distributions over bounded intervals. The polynomials are chosen by independently sampling each coefficient to be 0 with probability p and uniformly from $\mathcal{D}$ otherwise. For these polynomials, we show a surprisingly simple distinguisher that requires $p > n \log n / \binom{n}{d}$ and $m \geq \tilde{O}(n^2)$ samples, independent of the degree d. This is in contrast with the setting for refutation, where we have sum-of-squares lower bounds against constant degree sum-of-squares algorithms (Grigoriev, TCS 01; Schoenebeck, FOCS 08) for this parameter regime for degree $d > 6$.

*NTT. Email: aayushjain1728@gmail.com.
†UCLA. Email: alexiskorb@cs.ucla.edu.
‡UCLA. Email: pslou@cs.ucla.edu.
§UCLA. Email: sahai@cs.ucla.edu.

Contents

1 Introduction

In this work, we consider the following problem:

Definition 1.1 (Polynomial Independence Distinguishing Problem)**.** Let $\mathcal{Q} = \{q_1, ..., q_m\}$ denote a set of multivariate polynomials where $q_i : \mathbb{R}^n \to \mathbb{R}$ and $m = n^{O(1)}$. Let $\mathcal{D}$ be a distribution on $\mathbb{R}$, and let $\mathcal{D}^*$ be the distribution $\underbrace{\mathcal{D} \times \cdots \times \mathcal{D}}_{n \text{ times}}$ over $\mathbb{R}^n$ where $\mathbf{x} = (x_1, \ldots, x_n) \xleftarrow{R} \mathcal{D}^*$ means $x_1, \ldots, x_n$ are independently sampled from $\mathcal{D}$. The Polynomial Independence Distinguishing Problem with respect to $\mathcal{D}, \mathcal{Q}$ (or simply $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$) consists of distinguishing the following two distributions:

Distribution 1:	**Distribution 2:**
1. Sample $\mathbf{x} \xleftarrow{R} \mathcal{D}^*$	1. Sample $\mathbf{x}_1, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}^*$
2. Output $\{q_i, q_i(\mathbf{x})\}_{i \in [m]}$	2. Output $\{q_i, q_i(\mathbf{x_i})\}_{i \in [m]}$

Observe that the problem of recovering $\mathbf{x}$ from the output of Distribution 1 corresponds to solving the *search* version of a natural Constraint Satisfaction Problem (CSP for short). Similarly, the problem of certifying that no such $\mathbf{x}$ exists when in the scenario of Distribution 2 corresponds to the *refutation* version of the CSP. If it were possible to efficiently solve the search or refutation versions of our CSP above, then the distinguishing problem would immediately also be solved. The converse, however, is not true, and exploring this gap is the focus of this work.

Indeed, in many CSP problems, efficient search or refutation algorithms are not known to exist, and are even subject to theoretical lower bounds. For instance, there are abundant examples of CSPs where there are known Sum-of-Squares lower bounds [Gri01, Sch08, KMOW17]. In particular, the search and refutation versions of the Polynomial Independence Distinguishing Problem are subject to known Sum-of-Squares lower bounds for certain parameters [Jai19]. Nevertheless, in this work, we will show efficient distinguishers for those settings (and more).

Pseudorandomness over the Integers. The Polynomial Independence Distinguishing Problem is intimately tied with the notion of a pseudo-random generator (PRG). A PRG $\mathcal{G} : \mathcal{X}^n \to \mathcal{Y}^m$ with stretch $m > n$ takes as input $\mathbf{x} = (x_1, \ldots, x_n)$ where each x_i is a random sample from some distribution $\mathcal{D}_{in}$ with support over $\mathcal{X}$. The pseudorandomness property requires that the output $\mathcal{G}(\mathbf{x}) \in \mathcal{Y}^m$ is computationally indistinguishable from m independent copies of distribution $\mathcal{D}_{out}$ with support in $\mathcal{Y}$.

Traditionally, PRGs have been defined in the Boolean setting, where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, or in the setting of finite fields, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q$. A great deal of research has investigated these settings; much of this work has focused on investigating the possibility of the PRG $\mathcal{G}$ lying in a low complexity class such as low-locality [Gol00, AIK07, MST03, OW14, AL16, ABR12], block locality [LT17, LV17, BBKK18], low circuit-depth [AIK07], or low degree arithmetic circuits [KS99, KS98].

The goal of our work is to explore a new setting where $\mathcal{X} = \mathcal{Y} = \mathbb{Z}$. (By appropriate rescaling, this is equivalent to considering finite precision reals.)

More specifically, we consider the case where $\mathcal{D}_{in}$ and $\mathcal{D}_{out}$ are both distributions over the integers (or more broadly the reals) and $\mathcal{G}$ is a low degree multivariate polynomial over the integers. Furthermore, instead of aiming for a particular output distribution $\mathcal{D}_{out}$, one can simply require that the output of the generator is indistinguishable from the product of the marginals of the output components. One can therefore define a natural notion of a pseudorandom generator as follows (as defined by [ABKS17]).

Definition 1.2. (Pseudo-Independent Distribution Generator) A Pseudo-Independent Distribution Generator (or PIDG) is a tuple $(\mathcal{D}, \mathcal{F} = \{f_i\}_{i=1}^m))$ where m is called the stretch of the PIDG and

- $\mathcal{D}$ is an efficiently samplable distribution over $\mathbb{R}$.
- Each f_i for $i \in [m]$ is a polynomial-time computable multivariate function $f_i : \mathbb{R}^n \to \mathbb{R}$.

The security requirement is that for any probabilistic polynomial time adversary $\mathcal{A}$, the following holds:

$$\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}^*$$

$$\left| \Pr[\mathcal{A}(\mathcal{F}, \{f_i(\mathbf{x})\}_{i=1}^m) = 1] - \Pr[\mathcal{A}((\mathcal{F}, \{f_i(\mathbf{x}_i)\}_{i=1}^m))) = 1] \right| < n^{-\omega(1)}$$

We are interested in exploring the possibility of whether such PIDGs can exist, in settings that do not correspond to the well-studied Boolean case. Note that relaxing either the input domain to $\{0,1\}^n$ or letting the PIDG $\mathcal{G}$ be sufficiently complex trivialises the problem. If the input domain is allowed to be $\{0,1\}^n$, any such PIDG can be easily constructed using any standard Boolean PRG. Similarly, if $\mathcal{G}$ is allowed to be sufficiently complex then it is also trivial to construct a PIDG. The generator could treat the input as a string of bits and derive pseudorandom Boolean bits from the input bits, and then proceed again using a Boolean PRG.

This paper aims to initiate the study of limits on the existence of nontrivial PIDGs. In particular, we study the case where:

- **Input Distribution.** We require the input distribution to be a well-spread distribution over the integers (or reals) such as the standard discrete Gaussian distribution. Our results apply to different "spread" requirements, with several of our results applying to a quite minimal condition: that the distribution is symmetric, and at least three values in $\mathbb{Z}$ have noticeable probability mass.
- **Complexity of the PIDG.** The complexity class of the PIDG is the class of constant degree multilinear multivariate polynomials evaluated over the integers.

Connection to the Security of Indistinguishability Obfuscation. Indeed, the choice of input distribution and the complexity class above is motivated by recent progress [AJL+19, JLMS19, Agr19, JLS19, GJLS21] towards Indistinguishability Obfuscation ($i\mathcal{O}$) [BGI+01, GR10, GGH+13]. These works, relied on standard assumptions in addition to new assumptions which are very much related to the PIDP problem. The line culminated in the first construction of $i\mathcal{O}$ from well-studied assumptions [JLS21b, JLS21a]. Unfortunately, like the predecessors this work relies on bilinear maps and therefore, is not quantum secure.

Very recently there has been a lot of progress towards constructing plausibly post-quantum secure $i\mathcal{O}$ [WW21, GP21, BDGM20, DQV+21]. Underlying these works, there are new assumptions which employ random polynomial systems, which are plausibly susceptible to our observations.

1.1 Our Results

We consider two kinds of distinguishers - *non-trivial* and *overwhelming*. An algorithm $\mathcal{A}$ is a non-trivial distinguisher if it succeeds in distinguishing the two distribution of the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ with a noticable probability (in the input size). An overwhelming distinguisher is one where this probability is very close to 1. We define this formally below.

Definition 1.3. (Non-trivial PIDP Distinguisher) An algorithm $\mathcal{A}$ is a non-trivial PIDP distinguisher for the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem if

$$\Big|\Pr[\mathcal{A}(x_1) = 1] - \Pr[\mathcal{A}(x_2) = 1]\Big| \geq \frac{1}{n^{O(1)}}$$

where x_1 is sampled from Distribution 1 and x_2 is sampled from Distribution 2, as defined in Definition 1.1.

Definition 1.4. (Overwhelming PIDP Distinguisher) An algorithm $\mathcal{A}$ is an overwhelming PIDP distinguisher for the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem if

$$\Big|\Pr[\mathcal{A}(x_1) = 1] - \Pr[\mathcal{A}(x_2) = 1]\Big| \geq 1 - \frac{1}{n^{\omega(1)}}$$

where x_1 is sampled from Distribution 1 and x_2 is sampled from Distribution 2, as defined in Definition 1.1.

Results for Non-Trivial Distinguishers. We begin by building non-trivial distinguishers for large classes of input distributions and *worst-case* families of polynomials chosen by an adversary.

We require the input distribution to satisfy only a few basic structural properties. These input distributions are called weakly nice. A weakly nice distribution is a distribution that is intuitively well spread and symmetric around 0. We capture this by requiring all odd moments of the distribution to be 0, and in addition, requiring that for random variable X over $\mathcal{D}$ that $\big(\mathbb{E}[X^4]\big) \Big/ \big(\mathbb{E}[X^2]\big)^2 \geq 1 + \epsilon$ where $\epsilon > 0$ is some constant[1]. Refer to Definition 3.7 for a formal definition.

We obtain nontrivial distinguishers for the following classes of polynomials:

- We consider the set of constant degree multilinear polynomials where the monomials satisfy an expansion criteria $\mathcal{Q}_{Exp} \subseteq \mathbb{R}[x_1, \ldots, x_n]$. Namely, the expansion criteria, formally defined in Definition 5.3, captures the idea that the set of indices of variables in the monomials form an expanding set. Note that this is a key feature in low locality cryptographic Boolean PRGs [Gol00, KMOW17, ABR12, AL16, Gri01, Sch08], and CSPs with Sum-of-Squares Lower Bounds. Namely, we obtain:

  Theorem 1.1. *(Informal) Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$ where $\mathcal{Q}$ is an* Expander Based Polynomial Set *with coefficients bounded in absolute value by $n^{O(1)}$, and let $\mathcal{D}$ be a weakly-nice distribution with bounded support in $[-\beta, \beta]$ for $\beta = n^{O(1)}$. If $m > n$, then there exists a probabilistic polynomial algorithm can solve the $(\mathcal{D}, \mathcal{Q})$-*PIDP *with probability at least $\Omega(n^{-O(1)})$.*

- We also consider the set of constant degree multilinear polynomials with nonnegative coefficients $\mathcal{Q}_{n,nonneg} \subseteq \mathbb{R}[x_1, \ldots, x_n]$, obtaining:

  Theorem 1.2. *(Informal.) Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \in Q_{n,nonneg} \subset \mathbb{Z}[x_1, ..., x_n]$ with coefficients bounded in absolute value by $n^{O(1)}$, and let $\mathcal{D}$ be a weakly-nice distribution with bounded support in $[-\beta, \beta]$ for $\beta = n^{O(1)}$. If $m > n$, then there exists a probabilistic polynomial algorithm can solve the $(\mathcal{D}, \mathcal{Q})$-*PIDP *with probability at least $\Omega(n^{-O(1)})$.*

[1] Although, our results do apply to the case when $\epsilon = 1/n^{O(1)}$, we treat it as a constant for the sake of clarity of exposition.

We note that both our results correspond to *worst-case* properties that are checkable in polynomial time. In particular, the expansion condition that we refer to above only involves sets of size at most 4. Furthermore, the distinguisher also succeeds with non-trivial probability even if m is as small as 2, provided the conditions required by the algorithm are met.

Results for Overwhelming Distinguishers. We next consider the problem of *amplifying* the distinguishing advantage to yield overwhelming distinguishers for natural distributions of both inputs and polynomials.

We consider random families of polynomials, where each polynomial is sampled from some distribution $\mathcal{Q}_{n,d,p}$. The polynomials sampled from this distribution consist of homogeneous, multilinear degree d polynomials over the reals, where each coefficient is independently set to 0 with probability $1-p$, and otherwise sampled from some "nice" distribution. The distribution is *nice* if it satisfies certain conditions: The fourth moment is required to be sufficiently greater than the square of the second moment; it is required to take values within a bound that is roughly polylogarithmic in the second moment; and it must satisfy a weak anti-concentration property. We refer the reader to Definition 3.9 for a formal definition of a nice distribution. For the reader, it would be helpful to think of a (discrete) Gaussian distribution, or a uniform distribution over $[-n^c, n^c]$ for a constant $c > 0$ as examples of nice distributions.

The input distribution is also required to be nice. Then, our main result is:

Theorem 1.3. *Let d be any constant degree, and let $p > n \log n / \binom{N}{d}$. Let $\mathcal{D}_{nice}$ be a nice distribution as described above. If $m \geq n^2 \cdot (\log n)^{O(1)}$, then there exists a probabilistic polynomial time overwhelming distinguisher for the $(\mathcal{D}, \mathcal{Q}^m_{n,d,p}) - \mathsf{PIDP}$ problem.*

We stress that our overwhelming distinguisher applies in a context where strong sum-of-squares lower bounds apply to the search and refutation versions of our problem [Gri01, Sch08, KMOW17, HK22]. In particular, for $d > 6$, the value of m for which our attack applies is below the value of m for which sum-of-squares lower bounds apply.

2 Technical Overview

In this section, we give an intuitive technical guide to our results. Our objective, we recall, is to build efficient distinguishers for the Polynomial Independence Distinguishing Problem.

Correlations that arise over the integers, but not over Boolean values. The starting point for our work is that polynomials evaluated over natural distributions over the integers, instead of over uniform Boolean values, can lead to a detectable correlation between polynomials with shared variables. Consider the following example: Let $q_1, q_2 \in \mathbb{Z}[x_1, x_2]$ share the variable x_1 where

$$q_1(\mathbf{x}) = x_1$$

$$q_2(\mathbf{x}) = x_1 x_2$$

Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ where each X_i, Y_i are i.i.d. random variables with probability distribution $\mathcal{D}$. Now, if $\mathcal{D}$ were the uniform distribution over $\{-1, 1\}$, then the distributions $(q_1(X), q_2(X))$ and $(q_1(X), q_2(Y))$ are identical. However, if $\mathcal{D}$ is a non-Boolean distribution where $\mathbb{E}[X_1^2] \neq (\mathbb{E}[X_1])^2$, then

$$\mathbb{E}[q_1(X) q_2(X)] = \mathbb{E}[X_1^2] \, \mathbb{E}[X_2]$$

whereas

$$\mathbb{E}[q_1(X)q_2(Y)] = \mathbb{E}[X_1]\,\mathbb{E}[Y_1]\,\mathbb{E}[Y_2]$$

which differ as long as $\mathbb{E}[X_2] \neq 0$.

Unfortunately, if the distribution $\mathcal{D}$ has expectation 0, the above discrepancy will still yield the same overall expectation. As a result, we will instead consider the squared product distributions. For our simple example, this yields:

$$\mathbb{E}[q_1^2(X)q_2^2(X)] = \mathbb{E}[X_1^4]\,\mathbb{E}[X_2^2]$$

$$\mathbb{E}[q_1^2(X)q_2^2(Y)] = \mathbb{E}[X_1^2]\,\mathbb{E}[Y_1^2]\,\mathbb{E}[Y_2^2]$$

which differ as long as $\mathbb{E}[X_1^4] \neq (\mathbb{E}[X_1^2])^2$ and $\mathbb{E}[X_2^2] \neq 0$. Such conditions are reasonable for symmetric mean zero distributions over integers as we later show in Lemma 4.1. In fact, for any random variable Z, then $\mathbb{E}[Z^4] = \mathbb{E}[Z^2]^2$ if and only if $var[Z^2] = 0$. In other words, this will hold if and only if the input distribution either (1) is a point distribution, or (2) has support on $\{-k, k\}$ for some $k \in \mathbb{R}^+$, in which case it would be a scaled Boolean.

Polynomials. The $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem can be studied for any set of multivariate polynomials and input distributions over the reals. In this paper, we initiate this study by considering multilinear polynomials of constant degree over the reals. We leave it as an open question as to whether, and under what conditions, these results can be extended to arbitrary polynomials.

In all cases, we will consider m, the number of polynomials, to be larger than n, the number of variables. Otherwise, one can trivially build a set of m polynomials $\{q_i(\mathbf{x}) = x_i\}_{i\in[m]}$ for which $\{q_i, q_i(\mathbf{x})\}_{i\in[m]}$ and $\{q_i, q_i(\mathbf{x_i})\}_{i\in[m]}$ have identical distributions when $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n \xleftarrow{R} \mathcal{D}$ for some distribution $\mathcal{D}$ over the reals. We note that viewed as a pseudorandom number generator $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^m$ where $\mathcal{G}(\mathbf{x}) = \{q_i(\mathbf{x})\}$, this is just the identity function truncated to the first m values of the input.

Results. We show how we leverage the simple starting observation above to achieve nontrivial distinguishers for a wide variety of worst-case polynomials and a very large class of input distributions. In the case of natural randomized families of polynomials and natural input distributions, we also show how to *amplify* the nontrivial correlations we identify in the case of our nontrivial distinguishers to obtain overwhelming distinguishers. We now elaborate.

2.1 Non-trivial Probability Distinguishers

We want to identify distributions $\mathcal{D}$ and classes of polynomials $\mathcal{C}$ such that for *any* set of $m > n$ polynomials $\mathcal{Q} \subseteq \mathbb{R}[x_1, \ldots, x_n]$ chosen from $\mathcal{C}$, there is an efficient algorithm that solves the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ with non-trivial probability.

Input Distributions. Our results apply to any bounded symmetric mean zero distribution over the reals with a wide enough spread. This is formalised by requiring $\mathbb{E}[Z^4]/(\mathbb{E}[Z^2])^2 \geq \gamma$ for some $\gamma > 1$ and $\mathbb{E}[Z^2] \geq \eta$ for some $\eta > 0$, where Z is a random variable with distribution $\mathcal{D}$. The property of having $\mathbb{E}[Z^4]/\,\mathbb{E}[Z^2]^2 \geq \gamma$ is called the $\gamma-$hyper expansion property of the distribution. For the technical overview, we will consider γ, η to be constants.

Leveraging Expectation Differences of the Squared Product Differences. Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$, let $\mathcal{D}$ be a distribution on $\mathbb{R}$, and let $\mathcal{D}^*$ sample an n-tuple of values each independently drawn from $\mathcal{D}$. Let X be a random variable on distribution $\mathcal{D}^*$. If $m > n$, then by the pigeonhole principle, there exist $i, j \in [m]$ such that q_i, q_j share a variable. We want to leverage the correlation between these two polynomials (or rather the correlation between the squares of these two polynomials). By definition of covariance,

$$cov[q_i^2(X), q_j^2(X)] = \mathbb{E}[q_i^2(X)q_j^2(X)] - \mathbb{E}[q_i^2(X)]\,\mathbb{E}[q_j^2(X)]$$

Therefore, if the covariance between q_i, q_j is large, then this expectation difference is also large. Note that in the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ problem, we either get samples of the form $\{q_i, z_i = q_i(\mathbf{x})\}_{i\in[m]}$ where $\mathbb{E}[Z_i^2 Z_j^2] = \mathbb{E}[q_i^2(X)q_j^2(X)]$ or samples of the form $\{q_i, z_i = q_i(\mathbf{x}_i)\}_{i\in[m]}$ where $\mathbb{E}[Z_i^2 Z_j^2] = \mathbb{E}[q_i^2(X)]\,\mathbb{E}[q_j^2(X)]$. Here the random variables $Z_i = q_i(X_i)$, where $X_i \xleftarrow{R} \mathcal{D}^*$, correspond to the samples z_i received. Thus, the covariance is equal to the difference in the expectations of the distribution of $Z_i^2 Z_j^2$ when getting evaluations on the same input, and the distribution of $Z_i^2 Z_j^2$ when getting evaluations on independent inputs. To build a distinguisher to solve the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$, we proceed in two steps.

1. **Expectation Distinguisher**: First, we build a general algorithm which, when given a single sample from one of two bounded non-negative distributions whose expectations differ by a non-negligible amount, can distinguish between the two distributions with non-negligible probability (Lemma 5.1). We will call this algorithm the Expectation Distinguisher.
2. **Covariance Guarantee**: Second, we show that for certain $\mathcal{Q}$ and $\mathcal{D}$, then $cov[q_i^2(X), q_j^2(X)] = \mathbb{E}[q_i^2(X)q_j^2(X)] - \mathbb{E}[q_i^2(X)]\,\mathbb{E}[q_j^2(X)]$ is non-negligible (Lemmas 5.2 and 5.3).

By combining these two steps, we get a distinguisher for the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$: We simply compute the product of the samples $z_i^2 z_j^2$ and send the product to the Expectation Distinguisher as input.

Expectation Distinguisher. As a basic tool for reasoning about the existence of nontrivial distinguishers, we prove the following general lemma which roughly says that if there exist two distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ with support in $[0, 1]$—which we can assume without loss of generality because we can shift and scale arbitrary bounded distributions—such that their expectations differ by some quantity q, then, we can show a distinguisher that runs in time $q^{-O(1)}$ and distinguishes these two distributions with probability $q^{O(1)}$. More generally, both the running time and the distinguishing probability is a function of the ratio of the absolute value of the difference in the expectation to the size of the support. More precisely,

Lemma 2.1. *Let p, q be two positive parameters. Let D_0 and D_1 be distributions with bounded support in $[0, p]$.*[2] *Let X_0 be a random variable distributed according to $\mathcal{D}_0$ and X_1 be a random variable distributed according to $\mathcal{D}_1$. If*

$$\Big|\,\mathbb{E}[X_0] - \mathbb{E}[X_1]\Big| > q$$

then the Expectation Distinguisher $\mathcal{A}$ (Algorithm 1) succeeds with probability

$$\Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| \geq \frac{q^2}{16p^2}$$

[2] More generally, the support is allowed to be $[-p/2, p/2]$ and then the result follows by appropriately shifting the two distributions by $p/2$.

Relying on this lemma, we prove a variety of results for PIDP. First, we describe how we prove this lemma. We construct a simple distinguisher that first creates an approximate histogram of the two distributions by randomly sampling from each of $\mathcal{D}_0$ and $\mathcal{D}_1$ a sufficient number of times. The distinguisher partitions the support into some ϵ-width intervals and estimates the probability of these distributions falling within these intervals. A Chernoff bound combined with a union bound ensures that these probability estimates are reasonably accurate.

Next, we show that there exists an interval where the following happens:

Lemma 2.2. *Let p, q be two positive parameters. Suppose D_0 and D_1 are distributions with bounded support in $[0, p]$, and X_0 be a random variable distributed according to $\mathcal{D}_0$ and X_1 be a random variable distributed according to $\mathcal{D}_1$. If*

$$\Big|\mathbb{E}[X_0] - \mathbb{E}[X_1]\Big| > q$$

Then, if $\{I_i\}_{i=1}^n$ is a partition of $[0, p]$ into equal-sized intervals and $n = \frac{2p}{q}$, then there exists an index i such that

$$\Big|\Pr[x \in I_i \mid x \xleftarrow{R} D_0] - \Pr[x \in I_i \mid x \xleftarrow{R} D_1]\Big| \geq \frac{q^2}{4p^2}.$$

The lower bound on the difference in probabilities follows by an averaging argument on the difference between the expectations.

A high-level perspective is that the algorithm uses the histogram to make its decisions for any input x by choosing the larger estimated probability. The existence of the interval guaranteed by Lemma 2.2, allows us to form a lower bound for the distinguishing probability through a careful argument involving the aforementioned partitioning and accuracy guarantees given by the Chernoff bound combined with a union bound.

Covariance Guarantee. We now hunt for families of polynomials where we can apply our Expectation Distinguisher to yield a nontrivial distinguisher. Let q_i, q_j be multilinear polynomials that share a variable x_k, and let $\mathcal{D}$ be a symmetric mean zero distribution with minimum spread as defined earlier. Let X be a random variable distributed according to the product distribution $\mathcal{D}^*$. We introduce some notation first. Let $x_1, ..., x_n$ be variables. For a set $S \in \mathcal{P}([n])$, define $x_S = \prod_{i \in S}(x_i)$. Then,

$$q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_S x_S$$
$$q_j(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} d_S x_S$$

where each $c_S, d_S \in \mathbb{R}$. Since expectation is linear, then

$$\mathbb{E}[q_i^2(X) q_j^2(X)] - \mathbb{E}[q_i^2(X)]\,\mathbb{E}[q_j^2(X)] = \sum_{S,T,U,V \in \mathcal{P}([n])} c_S c_T d_U d_V \left(\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V]\right)$$

Let us consider any single $(\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V])$. We will show that this value is always non-negative. Now, since $\mathcal{D}$ is symmetric, all odd moments of each X_i are zero. Consider the following two cases:

1. $X_S X_T X_U X_V$ is a square. In that case, if one of $X_S X_T$ or $X_U X_V$ is not a square, then observe that $\mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] = 0$ since all odd moments are 0. Therefore, the difference is non-negative, since the expectation of a square is always non-negative. Otherwise, $X_S X_T$ and $X_U X_V$ are squares, so the degree of all variables in these terms is 2. Also, the degree of any X_i for $i \in [n]$ occurring in $X_S X_T X_U X_V$ is even and is the sum of the degree of X_i in $X_S X_T$ and the degree of X_i in $X_U X_V$. Therefore, if Z is a random variable with distribution $\mathcal{D}$, then the difference in expectations is

$$\mathbb{E}[Z^4]^t \cdot \mathbb{E}[Z^2]^{u-2t} - \mathbb{E}[Z^2]^u = \left(\left(\frac{\mathbb{E}[Z^4]}{(\mathbb{E}[Z^2])^2}\right)^t - 1\right)\left(\mathbb{E}[Z^2]\right)^u$$

   for some $u > t \geq 0$. Since $\mathcal{D}$ has minimum spread, we have $\mathbb{E}[Z^4]/\,\mathbb{E}[Z^2] \geq \gamma$ for some $\gamma > 1$ and $\mathbb{E}[Z^2] \geq \eta$ for some $\eta > 0$, so this difference is non-negative. Note that whenever $t > 0$, then this difference is positive. This occurs at least once if q_i, q_j share a variable, as illustrated by the example at the start of this section.

2. $X_S X_T X_U X_V$ is not a square. Then, one of $X_S X_T$ or $X_U X_V$ is also not a square. So, the difference is 0 because all the odd moments are zero.

Although, each $(\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V]) \geq 0$, we may have $c_S c_T d_U d_V(\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V]) < 0$ depending on the coefficients. Thus, the total expectation difference may still be close to zero because these summation terms could cancel out. Applying certain conditions on the coefficients prevents this from occurring, ensuring that our expectation difference is large enough. We note immediately that if all coefficients are nonnegative, then all summation terms are nonnegative, so such a cancellation does not occur. However, we show another set of conditions also ensures this: Expander Based Coefficients.

Expander Based Coefficients. The following definitions will ensure that the coefficients of the summation terms where $\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] \neq 0$ are always nonnegative. As before, this implies that the summation terms of the expectation difference do not cancel each other out.

Definition 2.1 (n-Half-Expanding Set)**.** Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of sets. Then, $\mathcal{S}$ is a *n-half-expanding set* if for all $k \leq n$ and all distinct $a_1, a_2, ..., a_k \in [m]$

$$\left|\bigcup_{i=1}^{k} S_{a_i}\right| > \frac{1}{2}\sum_{i=1}^{k} |S_{a_i}|$$

Definition 2.2 (Expander Based Polynomial Set)**.** Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$ be a set of multilinear polynomials over the reals. Then, each $q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_{S,i} x_S$ for some coefficients $\{c_{S,i}\}_{S \in \mathcal{P}([n])} \in \mathbb{R}$. We say that $\mathcal{Q}$ is a Expander Based Polynomial Set if

- Each q_i is a polynomial of degree at most some constant d
- $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0 \textit{ for any } i \in [m]\}$ is a 4-half expanding set.
- $\mathcal{C}_S = \{c_{S,i}\}_{i \in [m]}$ contains at most one non-zero value. (i.e. All monomials appear at most once across all polynomials in $\mathcal{Q}$.)

Note that picking sufficiently sparse polynomials at random will yield an Expander Based Polynomial Set with good probability. Indeed, the random families of polynomials that yield sum-of-squares lower bounds for the search and refutation version of the natural CSP for our problem have this property [Jai19].

If q_i, q_j come from an Expander Based Polynomial Set $\mathcal{Q}$, then the following occurs: Consider the terms where $c_S, c_T, d_U, d_V \neq 0$. Then, since $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0 \textit{ for any } i \in [m]\}$ is a 4-half expanding set, then for distinct $S, T, U, V \in \mathcal{P}([n])$, we have $|S \cup T \cup U \cup V| > \frac{1}{2}(|S|+|T|+|U|+|V|)$. Therefore, some X_i occurs once in $X_S X_T X_U X_V$. Thus, $X_S X_T X_U X_V$ is not a square, which means that $\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] = 0$. Suppose then that S, T, U, V are not all distinct. Let one of S or T equal one of U or V. But since we assumed that $c_S, c_T, d_U, d_V \neq 0$, this means that c_A and d_A are both nonzero for some set A. But this contradicts the fact that all monomials appear at most once in all polynomials of $\mathcal{Q}$ since $\mathcal{Q}$ is an Expander Based Polynomial Set. Therefore, if S, T, U, V are not all distinct, we need either $S = T$ or $U = V$. Suppose without loss of generality, that $S = T$. Then, in order for $X_S X_S X_U X_V = X_S^2 X_U X_V$ to be a square (so that $\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] \neq 0$.), we need $U = V$ as well. Therefore, the actual coefficient that arises in the expectation calculation is $c_S c_T d_U d_V = c_S^2 d_U^2 \geq 0$ whenever $\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] \neq 0$. This implies that the summation terms of the expectation difference do not cancel each other out, and we obtain a non-trivial distinguisher.

2.2 Overwhelming Probability Distinguisher

We now describe how to amplify the correlations described above to yield our overwhelming probability distinguisher. In this setting, we are given polynomials $\{q_i\}_{i\in[m]}$ sampled from $\mathcal{Q}^m_{n,p,d}$ along with evaluations of the form $\{q_i(\mathbf{x}) = y_i\}_{i\in[m]}$ or $\{q_i(\mathbf{x}_i) = y'_i\}_{i\in[m]}$ where each $\mathbf{x}$ as well as $\{\mathbf{x}_i\}_{i\in[m]}$ are chosen at random from a distribution $\mathcal{D}^*$, as defined in Definition 1.1, where $\mathcal{D}$ is a nice distribution. For the purpose of this technical overview, the reader may assume that a nice distribution is simply a discrete Gaussian centered at zero with standard deviation $n^{O(1)}$.

Remark 2.1. Inputs to the generated polynomials are taken from $\mathcal{D}^*$ where the notation is as described in Definition 1.1. Throughout, we will treat x in small letters as an input variable to the polynomial, and X in capital letters as the corresponding random variable sampled from $\mathcal{D}^*$. Similarly, for any given random variables $X, \{X_i\}_{i=1}^m$ and given polynomials $\{q_i\}_{i=1}^m$ we will denote the random variable $q_i(X) = Y_i$ and $q_i(X_i) = Y'_i$ for all i.

Aside: Amplification in the case of Gaussian samples. If one observes $y_i = q_i(\mathbf{x}) = \sum_S c_S \Pi_{i\in S} x_i = \sum_S c_S \cdot x_S$, then a single sample should be distributed somewhat like a Gaussian distribution of mean 0 and appropriate standard deviation (this could be formalized for example using the Berry-Esseen theorem.). Thus, consider the following simplistic setting. Suppose we have been given either an instance of the form consisting of independently chosen Gaussian samples $\mathbf{z}' = (z'_1, ..., z'_m)$ or some arbitrarily correlated Gaussians $\mathbf{z} = (z_1, ..., z_m)$ and the goal is to identify the case. Consider the following ratio for Z_1, Z_2 random variables over the standard Gaussian.

$$\beta = \frac{\mathbb{E}_{Z_1}[Z_1^4]}{\mathbb{E}_{Z_1,Z_2}[Z_1^2 \cdot Z_2^2]}$$

If z_1, z_2 are sampled according to identical and independently distributed Gaussian distribution, then $\beta = \frac{\mathbb{E}_{Z_1}[Z_1^4]}{\mathbb{E}_{Z_1}[Z_1^2]^2}$. For a centered Gaussian variable Z_1, this quantity, which we will refer to as β_{diff} (diff for different) is exactly equal to 3 since for a centered Gaussian distribution the ratio of the fourth moment to the square of the second moment is 3. On the other hand, when Z_1 and Z_2 are

ρ correlated (i.e. $Z_2 = \rho \cdot Z_1 + \sqrt{1-\rho^2} Z^{\perp}$ where $Z^{\perp}$ is independently and identically distributed as Z_1), then, the ratio we get is denoted by β_{same}. Observe,

$$\beta_{\mathsf{same}} = \frac{3}{1 + 2 \cdot \rho^2}$$

Thus, as the correlation increases, this ratio (with maximum value 3) decreases until it attains a minimum value of 1 when $\rho \in \{+1, -1\}$. This example suggests that we consider the following idea:

Ratios for the PIDP problem. Define two ratios for Y_1, Y_2, Y_1', Y_2' random variables as defined in remark 2.1:

$$\alpha_{\mathsf{diff}} = \frac{\mathbb{E}_{Y_1'}[Y_1'^4]}{\mathbb{E}_{Y_1', Y_2'}[Y_1'^2 \cdot Y_2'^2]} \qquad \alpha_{\mathsf{same}} = \frac{\mathbb{E}_{Y_1}[Y_1^4]}{\mathbb{E}_{Y_1, Y_2}[Y_1^2 \cdot Y_2^2]}$$

One can compute $\alpha_{\mathsf{diff}} = \frac{\mathbb{E}[Y_1'^4]}{\mathbb{E}[Y_1'^2 \cdot Y_2'^2]}$ by expanding the random variables:

$$\begin{aligned}\alpha_{\mathsf{diff}} =& \frac{\mathbb{E}_{q_1, X_1}[q_1^4(X_1)]}{\mathbb{E}_{q_1, q_2, X_1, X_2}[q_1^2(X) \cdot q_2^2(X_2)]} \\ =& \frac{\mathbb{E}_{q_1, X_1}[q_1^4(X_1)]}{\mathbb{E}_{q_1, X_1}[q_1^2(X_1)] \cdot \mathbb{E}_{q_2, X_2}[q_2^2(X_2)]}\end{aligned}$$

Denote $q_1(X) = \sum_{|S|=d} c_S X_S$ and $q_2(Y) = \sum_{|S|=d} d_S Y_S$ where coefficients c_S and d_S are chosen independently from some nice distribution $\mathcal{D}$ with probability p and 0 otherwise. Assume $\mathbf{x}$ and $\mathbf{y}$ are chosen at random from $\mathcal{D}^*$. Let $\mathcal{D}$ be such that a random variable Z over $\mathcal{D}$ has $\frac{\mathbb{E}[Z^4]}{\mathbb{E}[Z^2]^2} = \gamma > 1$. A typical value of γ is some constant greater than 1. With this notation the numerator of α_{diff} can be computed as:

$$\begin{aligned}&\mathop{\mathbb{E}}_{q_1, X}[q_1^4(X)] \\ =& \mathop{\mathbb{E}}_{X} \mathop{\mathbb{E}}_{q_1} \left[\sum_{S_1} \sum_{S_2} \sum_{S_3} \sum_{S_4} c_{S_1} \cdot c_{S_2} \cdot c_{S_3} \cdot c_{S_4} \cdot X_{S_1} \cdot X_{S_2} \cdot X_{S_3} \cdot X_{S_4} \right] \\ =& \mathop{\mathbb{E}}_{X} \left[\sum_{S} p \cdot \gamma \cdot X_S^4 + 3 \cdot p^2 \cdot \sum_{S_1 \neq S_2} X_{S_1}^2 \cdot X_{S_2}^2 \right]\end{aligned}$$

Let $N = \binom{n}{d}$, then the numerator becomes,

$$N \cdot p \cdot \gamma \cdot \mathop{\mathbb{E}}_{X}[X_S^4] + 3 \cdot p^2 \cdot \sum_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1}^2 \cdot X_{S_2}^2]$$

Since, $\mathbb{E}_X[X_S^4] = \gamma^d$ and $\sum_{S_1 \neq S_2} \mathbb{E}_X[X_{S_1}^2 \cdot X_{S_2}^2] = N \cdot (N-1)\, \mathbb{E}_{S_1 \neq S_2}\, \mathbb{E}_X[X_{S_1} \cdot X_{S_2}]$, the numerator becomes,

$$N \cdot p \cdot \gamma \cdot \gamma^d + 3 \cdot p^2 \cdot N \cdot (N-1) \mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} \cdot X_{S_2}]$$

For $i \in [d-1]$, let g_i denote the probability that two randomly chosen sets $S_1 \neq S_2$ in $[n]$ of size d have i common elements.

This means that,

$$\mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} \cdot X_{S_2}] = (1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma \cdot g_1 + \ldots + \gamma^{d-1} \cdot g_{d-1}$$

This means that the numerator is,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{q_1, X}[q_1^4(X)] \\
=&N \cdot p \cdot \gamma \cdot \gamma^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma \cdot g_1 + \ldots + \gamma^{d-1} \cdot g_{d-1}\Big)
\end{aligned}$$

Now, consider the denominator, $\mathbb{E}_{q_1,q_2,X,Y}[q_1^2(X) \cdot q_2^2(Y)]$.

$$\begin{aligned}
\mathop{\mathbb{E}}_{q_1,q_2,X,Y}[q_1^2(X) \cdot q_2^2(Y)] =& \mathop{\mathbb{E}}_{q_1,q_2,X,Y}[\sum_{S_1,S_3} c_{S_1}^2 d_{S_3}^2 X_{S_1}^2 Y_{S_3}^2] \\
=& p^2 \cdot \mathop{\mathbb{E}}_{X,Y}[\sum_{S_1,S_3} X_{S_1}^2 Y_{S_3}^2] \\
=& N^2 \cdot p^2
\end{aligned}$$

This means that:

$$\alpha_{\text{diff}} = \frac{N \cdot p \cdot \gamma \cdot \gamma^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma \cdot g_1 + \ldots + \gamma^{d-1} \cdot g_{d-1}\big)}{N^2 \cdot p^2}$$

Setting $t = N \cdot p$ as the average density of each polynomial (number of non-zero coefficients) and setting $g_i \approx \theta(1/n^i)$ for $i \in [d-1]$, we get that:

$$\boxed{\alpha_{\text{diff}} = \frac{\gamma^{d+1}}{t} + 3 + \Omega(\frac{1}{n})}$$

Similarly, one can compute α_{same}

$$\alpha_{\mathsf{same}} = \frac{\mathbb{E}_{q_1,X}[q_1^4(X)]}{\mathbb{E}_{q_1,q_2,X}[q_1^2(X) \cdot q_2^2(X)]}$$

Lo and Behold,

$$\alpha_{\text{same}} = \frac{N \cdot p \cdot \gamma_2 \cdot \gamma^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma \cdot g_1 + \ldots + \gamma^{d-1} \cdot g_{d-1}\big)}{p^2 \cdot N \cdot \gamma^d + p^2 \cdot N \cdot (N-1) \cdot ((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma \cdot g_1 + \ldots + \gamma^{d-1} \cdot g_{d-1})}$$

Assuming $p < \gamma/3$,

$$\boxed{\alpha_{\text{same}} = 3 + \theta(\frac{\gamma_2 \cdot \gamma_1^d}{t})}$$

Thus, as expected $\alpha_{\text{diff}} > \alpha_{\text{same}}$. In fact, if $t \gg n$, then $\alpha_{\text{diff}} - \alpha_{\text{same}} > \Omega(1/n)$.

Amplifying from $\alpha_{\text{diff}} - \alpha_{\text{same}} > \Omega(1/n)$ to an overwhelming distinguisher. Above we observed that α_{same} and α_{diff} for the PIDP problem are apart by at least $1/n$. Can we somehow utilize this difference to construct an overwhelming distinguisher?

In order to do that, we construct empirical approximations of $\widehat{\alpha}_{\mathsf{same}}$ of α_{same} and $\widehat{\alpha}_{\mathsf{diff}}$ of α_{diff} computed as:

$$\widehat{\alpha}_{\mathsf{diff}} = \frac{1/m \sum_i y'^4_i}{2/m \sum_{i\in[m/2]} y'^2_{2i-1} \cdot y'^2_{2i-1}} \qquad \widehat{\alpha}_{\mathsf{same}} = \frac{1/m \sum_i y^4_i}{2/m \sum_{i\in[m/2]} y^2_{2i-1} \cdot y^2_{2i-1}}$$

If m is sufficiently large, then, $\widehat{\alpha}_{\mathsf{same}}$ will be close to α_{same} and $\widehat{\alpha}_{\mathsf{diff}}$ will be close to α_{diff} (at least in expectation). Thus, to prove this claim, what we do is that, given the samples $\{v_i\}_{i\in[m]}$ where $v_i = q_i(\mathbf{x})$ or $q_i(\mathbf{x}_i)$ for all $i \in [m]$, we compute the ratio:

$$\widehat{\alpha} = \frac{1/m \sum_i v^4_i}{2/m \sum_{i\in[m/2]} v^2_{2i-1} \cdot v^2_{2i-1}}$$

Then, we check if $\widehat{\alpha} - \frac{\alpha_{\text{same}}+\alpha_{\text{diff}}}{2} \stackrel{?}{>} 0$. If the check is true we declare independent, otherwise we declare same. Indeed, we show that the check identifies the distribution correctly if $m \geq n^2 \log^{O(1)}(n)$. Note that for showing this we need to analyse $\frac{1/m \sum_i v^4_i}{2/m \sum_{i\in[m/2]} v^2_{2i-1} \cdot v^2_{2i-1}}$. In general, analysing the ratio of this form may not be an easy task, as expected ratio of a quantity is in general not the ratio of expectations. Thus, we analyse a slightly different objective. Define $\alpha_{\text{th}} = \frac{\alpha_{\text{same}}+\alpha_{\text{diff}}}{2}$ and consider,

$$F = \sum_i v^4_i - 2 \cdot \alpha_{\text{th}} \sum_{i\in[m/2]} v^2_{2i-1} \cdot v^2_{2i-1}$$

In order to prove the result, we show two claims:

- If $v_1, ..., v_m$ is sampled using independent inputs then with probability $1 - n^{-\omega(1)}$, $F > 0$.
- If $v_1, ..., v_m$ is sampled using a single input then with probability $1 - n^{-\omega(1)}$, $F < 0$.

The analysis of this claim is somewhat involved, and includes careful algebraic manipulations and applications of concentration inequalities. Details can be found in Section 6.

3 Preliminaries

Let $\mathbb{N}, \mathbb{Z}$, and $\mathbb{R}$ denote the set of positive integers, integers, and real numbers respectively. For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, \ldots, n\}$. Let $\mathcal{P}(S)$ denote the power set of set S. We represent vectors using lower case bold-faced characters. For example, $\mathbf{v} \in \mathbb{R}^n$ indicates a vector over the reals of dimension n where $n \in \mathbb{N}$.

We use the usual Landau notations. A function $f(n)$ is said to be negligible if it is $n^{-\omega(1)}$ and we denote it by $f(n) = \mathsf{negl}(n)$. A probability $p(n)$ is said to be overwhelming if it is $1 - n^{-\omega(1)}$. For any distribution $\mathcal{D}$, we denote the process of sampling x at random from distribution D by $x \xleftarrow{R} \mathcal{D}$. We say that an algorithm or function $\mathcal{A}(x)$ is polynomial time if for all x, $\mathcal{A}$ is computable in time $t = O\left(|x|^{O(1)}\right)$.

Definition 3.1 (Computational Indistinguishability)**.** We say that D_1 is computationally indistinguishable from D_1, denoted $D_1 \approx_C D_2$, if no computationally-bounded adversary can distinguish between D_1 and D_2 except with advantage $\mathsf{negl}(\cdot)$. More formally, we write $D_1 \approx_C D_2$ if for any probabilistic polynomial time algorithm $\mathcal{A}$,

$$\left| \Pr_{x \xleftarrow{R} D_1} [\mathcal{A}(x) = 1] - \Pr_{x \xleftarrow{R} D_2} [\mathcal{A}(x) = 1] \right| \leq \mathsf{negl}(|x|)$$

where $\mathsf{negl}(\cdot)$ is a negligible function defined above and the probabilities are taken over the coins of $\mathcal{A}$ and the choice of x.

Remark 3.1. We will consider all real numbers used in our algorithms to be of some finite precision λ. When we talk about polynomial time algorithms with real inputs, we refer to algorithms that use a polynomial number of λ-precision operations.

Definition 3.2 (t-Samplable Distribution)**.** A probability distribution $\mathcal{D}$ is t-samplable if there is a probabilistic algorithm $\mathcal{A}$ that runs in time t such that $\mathcal{A}(0) = \mathcal{D}$.

For random variables X, Y, let $\mathbb{E}_X[f(X)]$ denote the expectation of $f(\cdot)$ over random variable X and let $\mathbb{E}_{X,Y}[f(X,Y)]$ denote $\mathbb{E}_X \mathbb{E}_Y[f(X,Y)]$.

Definition 3.3. Let X be a random variable. For any integer $i \geq 1$, we denote the ith moment of X as

$$\mu_i = \mathbb{E}[X^i]$$

In general, the random variable X we are referring to will be clear by context.

Theorem 3.1. *(Chernoff Bound) Suppose $X_1, \ldots, X_n$ are independent random variables taking values in $\{0,1\}$, and let $X = \sum_{i=1}^{n} X_i$ and $\mathbb{E}[X] = \mu$. Then a two-sided Chernoff bound for $\delta > 0$ is*

$$\Pr\left[|X - \mu| > \delta\mu\right] \leq 2 \cdot \exp\left(-\frac{\delta^2 \mu}{2+\delta}\right)$$

Theorem 3.2. *(Hoeffding Bound) Let $X_1, \ldots, X_n$ be independent bounded random variables with $X_i \in [a,b]$ for all i, where $-\infty < a \leq b < \infty$. Then*

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right] \leq \exp\Big(-\frac{2nt^2}{(b-a)^2}\Big)$$

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \leq -t\right] \leq \exp\Big(-\frac{2nt^2}{(b-a)^2}\Big)$$

for all $t \geq 0$.

3.1 Polynomial Independence Distinguishing Problem

Definition 3.4 (Polynomial Independence Distinguishing Problem)**.** Let $\mathcal{Q} = \{q_1, ..., q_m\}$ denote a set of multivariate polynomials where $q_i : \mathbb{R}^n \to \mathbb{R}$. Let $\mathcal{D}$ be a distribution on $\mathbb{R}$, and let $\mathcal{D}^*$ be the distribution $\underbrace{\mathcal{D} \times \cdots \times \mathcal{D}}_{n \text{ times}}$ over $\mathbb{R}^n$ where $\mathbf{x} = (x_1, \ldots, x_n) \xleftarrow{R} \mathcal{D}^*$ means $x_1, \ldots, x_n$ are independently sampled from $\mathcal{D}$. The Polynomial Independence Distinguishing Problem with respect to $\mathcal{D}, \mathcal{Q}$ (or simply $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$) consists of distinguishing the following two distributions:

Distribution 1:	**Distribution 2:**
1. Sample $\mathbf{x} \xleftarrow{R} \mathcal{D}^*$	1. Sample $\mathbf{x_1}, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}^*$
2. Output $\{q_i, q_i(\mathbf{x})\}_{i\in[m]}$	2. Output $\{q_i, q_i(\mathbf{x_i})\}_{i\in[m]}$

Remark 3.2. In the above definition, $\mathcal{Q}$ is a set of polynomials. However, we may refer to the $(\mathcal{D}, \mathcal{Q})$-Polynomial Independence Distinguishing Problem for some distribution $\mathcal{Q}$ over some family of polynomials. In this case, we mean that each polynomial q_i for $i \in [m]$ is sampled from the distribution $\mathcal{Q}$.

Remark 3.3. We say that an algorithm $\mathcal{A}$ solves the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ with probability p if $\mathcal{A}$ can distinguish between Distribution 1 and Distribution 2 of the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ with probability at least p.

3.2 Pseudo-Independent Distribution Generator

Definition 3.5. (Pseudo-Independent Distribution Generator) A Pseudo-Independent Distribution Generator (or PIDG) is a tuple $(\mathcal{D}, \mathcal{F} = \{f_i\}_{i=1}^m))$ where m is called the stretch of the PIDG and

- $\mathcal{D}^*$ is a t-samplable distribution over $\mathbb{R}^n$ where $t = n^{O(1)}$ and $\mathcal{D}^*$ is as defined in Definition 3.4 above.
- each f_i for $i \in [m]$ is a polynomial time multivariate function $f_i : \mathbb{R}^n \to \mathbb{R}$.

Further, we require the generator to satisfy the following security notion:

$$\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}^*$$
$$(\mathcal{F}, \{f_i(\mathbf{x})\}_{i=1}^m) \approx_c (\mathcal{F}, \{f_i(\mathbf{x}_i)\}_{i=1}^m))$$

In other words, a PIDG is a distribution along with a set of functions such that one cannot distinguish between evaluations of these functions on independent inputs and evaluations of these functions on the same input when the input(s) are sampled randomly from $\mathcal{D}^*$.

Remark 3.4. If there exists a probabilistic polynomial time algorithm $\mathcal{A}$ that solves the $(\mathcal{D}, \mathcal{Q}) - \mathsf{PIDP}$ with non-negligible probability, then $(\mathcal{D}, \mathcal{Q})$ is not a PIDG.

3.3 Distribution Definitions

Definition 3.6. A random variable X is called a (k, n, γ)-hyper-expanding random variable, if:

$$\frac{\mathbb{E}[X^n]}{\mathbb{E}[X^k]^{n/k}} \geq \gamma.$$

We will omit parameters n and k to denote $(2, 4, \gamma)$-hyper-expanding random variables and call them γ-hyper-expanding random variables. For example, a standard Gaussian random variable X is 3-hyper-expanding since

$$\frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]^2} = 3.$$

and a uniform random variable Y on $U_{[-\beta,\beta]}$ for any large enough β is $\frac{3}{2}$-hyper-expanding. Sometimes we will abuse the hyper-expanding notation to talk about hyper-expanding distributions, where we state that the same requirements hold for any random variable with distribution $\mathcal{D}$.

Definition 3.7. We say that a distribution $\mathcal{D}$ is (η, γ)-weakly-nice if

1. $\mathcal{D}$ is a symmetric distribution with mean 0
2. If X is a random variable over $\mathcal{D}$, then $\mathbb{E}[X^2] \geq \eta$ and $\frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]^2} \geq \gamma$.

Definition 3.8. We say that a distribution $\mathcal{D}$ is C bounded if

$$\Pr[x \xleftarrow{R} \mathcal{D}, |x| < C] = 1$$

Definition 3.9. We say that a distribution $\mathcal{D}$ is (γ, C, ϵ)-nice if

1. $\mathcal{D}$ is a symmetric distribution with mean 0
2. (Normalization.) If X is a random variable over $\mathcal{D}$, then $\mathbb{E}[X^2] = 1$ and $\mathbb{E}[X^4] = \gamma$.
3. $\mathcal{D}$ is C-bounded.
4. (Anti-concentration) $\Pr[x \xleftarrow{R} \mathcal{D}, |x| > \epsilon] > \Omega(1)$

Remark 3.5. If a distribution $\mathcal{D}$ is (γ, C, ϵ)-nice, then $\mathcal{D}$ is also $(1, \gamma)$-weakly-nice

We will be concerned with (η, γ)-weakly-nice distributions where $\eta, \gamma - 1$ are positive and large enough (to be quantified later). For bounded integer distributions, we can get a lower bound on these values provided that we don't have all (or almost all) of the weight of the distribution lie on k and $-k$ for some value $k \in \mathbb{Z}$.

3.4 Polynomial Notation and Expectations

Notation Let $x_1, ..., x_n$ be variables. For a set $S \in \mathcal{P}([n])$, define

$$x_S = \prod_{i \in S} x_i$$

Consider a multilinear polynomial $q \in \mathbb{R}[x_1, \ldots, x_n]$. Then, $q(\mathbf{x})$ is of the form

$$q(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_S x_S$$

where each $c_S \in \mathbb{R}$.

Fact If $\mathcal{D}$ is a symmetric distribution with mean 0, and X is a random variable with distribution D, then for all odd $i \in \mathbb{N}$, $u_i = \mathbb{E}[X^i] = 0$.

Remark 3.6. Let $\mathcal{D}$ be a symmetric distribution with mean 0. Let $X = (X_1, X_2, ..., X_n)$ where each X_i is an i.i.d. random variable with distribution $\mathcal{D}$. Let $f(\mathbf{x}) = \prod_{i=1}^n x_i^{a_i}$ where each a_i is a nonnegative integer. Then, if any a_i is odd, $\mathbb{E}[f(X)] = \prod_{i=1}^n \mathbb{E}[X_i^{a_i}] = 0$.

Lemma 3.1. *Let $\mathcal{D}$ be a symmetric distribution over $\mathbb{R}$ with mean 0. Let $X = (X_1, X_2, ..., X_n)$ where each X_i is an i.i.d. random variable with distribution $\mathcal{D}$. Let $S, T \in \mathcal{P}([n])$. Then,*

$$\mathbb{E}[X_S X_T] = \begin{cases} 0 & \text{if } S \neq T \\ \mu_2^{|S|} & \text{if } S = T \end{cases}$$

where μ_2 is the second moment of each X_i

Proof. If $S \neq T$, then since X_S and X_T contain each variable at most once, then $X_S X_T$ will contain some variable X_i of odd degree. By Remark 3.6, then $S \neq T$ implies $\mathbb{E}[X_S X_T] = 0$. If $S = T$, then $\mathbb{E}[X_S X_T] = \mathbb{E}[X_S^2] = \mathbb{E}[\prod_{i \in S} X_i^2] = \prod_{i \in S} \mathbb{E}[X_i^2] = \mu_2^{|S|}$. □

Lemma 3.2. *Let $\mathcal{D}$ be a symmetric distribution over $\mathbb{R}$ with mean 0. Let $X = (X_1, X_2, ..., X_n)$ where each X_i is an i.i.d. random variable with distribution $\mathcal{D}$. Let $S, T, U, V \in \mathcal{P}([n])$. Then,*

$$\mathbb{E}[X_S X_T X_U X_V] = \begin{cases} 0 & \text{if } X_S X_T X_U X_V \text{ contains a variable } X_i \text{ of odd power} \\ \mu_4^{|a|} \mu_2^{|b|} & \text{else} \end{cases}$$

where $a = |S \cap T \cap U \cap V|$, $b = \frac{1}{2}(|S| + |T| + |U| + |V|) - 2a$, and μ_2, μ_4 are the second and fourth moments respectively of each X_i.

Proof. Now,

$$\mathbb{E}[X_S X_T X_U X_V] = \mathbb{E}\left[\prod_{i \in S} X_i \prod_{j \in T} X_j \prod_{k \in U} X_k \prod_{l \in V} X_l\right] = \mathbb{E}\left[\prod_{i=1}^{n} X_i^{a_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[X_i^{a_i}\right] = \prod_{i=1}^{n} \mu_{a_i}$$

for some $\{a_i\}_{i=1}^n$ such that $0 \leq a_i \leq 4$ for all $i \in [n]$. If $X_S X_T X_U X_V$ contains a variable X_i of odd power (i.e. if any a_i is odd), then by Remark 3.6, $\mathbb{E}[X_S X_T X_U X_V] = 0$. Otherwise, each $a_i \in \{0, 2, 4\}$. Now, $a_i = 4$ if and only if X_i appears in each of X_S, X_U, X_V, X_T. Define

$$a = |\{i \mid a_i = 4\}| = |S \cap T \cap U \cap V|$$

For any other variable X_i that appears in one of X_S, X_U, X_V, X_T, we must have that $a_i = 2$. Now,

$$\begin{aligned} \deg(X_S X_T X_U X_V) = |S| + |T| + |U| + |V| &= \sum_{i=1}^{n} a_i \\ &= 4|\{i \mid a_i = 4\}| + 2|\{i \mid a_i = 2\}| \\ &= 4a + 2|\{i \mid a_i = 2\}| \end{aligned}$$

Define $b = |\{i \mid a_i = 2\}| = \frac{1}{2}(|S| + |T| + |U| + |V|) - 2a$. Therefore, $\mathbb{E}[X_S X_T X_U X_V] = \prod_{i=1}^{n} \mu_{a_i} = \mu_4^{|a|} \mu_2^{|b|}$ □

4 Useful Lemmas

We show that for a bounded symmetric mean zero distribution $\mathcal{D}$ over the integers, then we only need a minimal notion of spread (namely that we have some noticeable probability mass on at least three points in $\mathbb{Z}$) to get a (η, γ)-weakly-nice distribution with reasonable lower bounds on $\eta, \gamma - 1$.

Definition 4.1. For a random variable X with integer support bounded by $[a, b]$, define $\mathsf{mode}(X)$ to be k such that $\Pr_X[X = k] = \max_{i=a}^{b}(\Pr_X[X = i])$

Lemma 4.1. *Let $\mathcal{D}$ be any distribution over $\mathbb{Z}$ with bounded support over $[-\beta, \beta]$. Let X be a random variable with distribution $\mathcal{D}$. Let $t > 0$. If $\Pr_X[|X| \neq \mathsf{mode}(|X|)] \geq \frac{1}{t}$, then*

$$\mu_2 \geq \mathbb{E}[X]^2 + \frac{1}{2 \cdot \max(\beta + 1, t)}$$

$$\frac{\mu_4}{\mu_2^2} \geq 1 + \frac{1}{2\mu_2^2 \cdot \max(\beta + 1, t)}$$

Proof. Since $\sum_{i=0}^{\beta} \Pr_X[|X| = i] = 1$ and $\Pr_X[|X| = \mathsf{mode}(|X|)] = \max_{i=0}^{\beta} \Pr_X[|X| = i]$, then $\Pr_X[|X| = \mathsf{mode}(|X|)] \geq \frac{1}{(\beta+1)}$. Therefore,

$$\frac{1}{t} \leq \Pr_X[|X| \neq \mathsf{mode}(|X|)] \leq 1 - \frac{1}{(\beta+1)}$$

By the definition of variance

$$\mu_2 = \mathbb{E}\left[X^2\right] = \mathbb{E}\left[X\right]^2 + var\left[X\right]$$

Let y_1 be the closest integer to $\mathbb{E}[X]$, and let y_2 be the next closest integer to $\mathbb{E}[X]$ with $y_1 \neq y_2$. Then, y_1 and y_2 are adjacent integers where

$$|y_1 - \mathbb{E}[X]| + |y_2 - \mathbb{E}[X]| = 1$$

Since y_1 and y_2 are the two closest integers to $\mathbb{E}[X]$, then for every integer $x \in \mathbb{Z}$ where $x \neq y_1$

$$(y_1 - \mathbb{E}[X])^2 \leq (y_2 - \mathbb{E}[X])^2 \leq (x - \mathbb{E}[X])^2$$

Therefore,

$$\begin{aligned} var[X] &= \sum_{i=-\beta}^{\beta} \left(\Pr_X[X = i](X - \mathbb{E}[X])^2 \right) \\ &\geq \Pr_X[X = y_1](y_1 - \mathbb{E}[X])^2 + (1 - \Pr_X[X = y_1])(y_2 - \mathbb{E}[X])^2 \end{aligned}$$

By definition of $\mathsf{mode}(|X|)$, then

$$\Pr_X[X = y_1] \leq \Pr_X[|X| = |y_1|] \leq \Pr_X[|X| = \mathsf{mode}(|X|)]$$

Which means that

$$\begin{aligned} var[X] &\geq \Pr_X[|X| = \mathsf{mode}(|X|)](y_1 - \mathbb{E}[X])^2 + (1 - \Pr_X[|X| = \mathsf{mode}(|X|)])(y_2 - \mathbb{E}[X])^2 \\ &= (1 - \Pr_X[|X| \neq \mathsf{mode}(|X|)])(y_1 - \mathbb{E}[X])^2 + \Pr_X[|X| \neq \mathsf{mode}(|X|)](y_2 - \mathbb{E}[X])^2 \end{aligned}$$

Claim 4.1. *If $a, b \geq 0$, $a + b \geq 1$, and $0 \leq p \leq t \leq c$, then $ta^2 + (1-t)b^2 \geq \frac{1}{2}\min(p, 1-c)$*

By the Cauchy Schwarz inequality, $(a+b)^2 = \langle (a,b),(1,1)\rangle^2 \leq \langle (a,b),(a,b)\rangle \cdot \langle (1,1),(1,1)\rangle = 2(a^2+b^2)$. So, $a + b \geq 1$ implies that $(a+b)^2 \geq 1$ which means $(a^2+b^2) \geq \frac{1}{2}$. Then,

$$\begin{aligned} ta^2 + (1-t)b^2 &\geq pa^2 + (1-c)b^2 \\ &\geq \min(p, 1-c)(a^2+b^2) \\ &\geq \frac{1}{2}\min(p, 1-c) \end{aligned}$$

By applying this claim to $a = |y_2 - \mathbb{E}[X]^2|$, $b = |y_1 - \mathbb{E}[X]^2|$, and $t = \Pr_X[|X| \neq mode(|X|)]$ then

$$var[X] \geq \frac{1}{2}\min\left(\frac{1}{\beta+1}, \frac{1}{t}\right) = \frac{1}{2 \cdot \max(\beta+1, t)}$$

$$\mu_2 \geq \mathbb{E}[X]^2 + \frac{1}{2 \cdot \max(\beta+1, t)}$$

Now, note that $\Pr_X[|X| = i] = \Pr_X[X^2 = i^2]$ and $\mathsf{mode}(|X|))^2 = \mathsf{mode}(X^2)$. Therefore, $\Pr_X[X^2 \neq \mathsf{mode}(X^2)] = \Pr_X[|X| \neq \mathsf{mode}(|X|)]$ so that

$$\frac{1}{t} \leq \Pr_X[X^2 \neq \mathsf{mode}(X^2)] \leq 1 - \frac{1}{(\beta+1)}$$

By the definition of variance

$$\mu_4 = \mathbb{E}\left[X^4\right] = \mathbb{E}\left[X^2\right]^2 + var\left[X^2\right] = \mu_2^2 + var\left[X^2\right]$$

Let y be the closest integer to $\mathbb{E}[X^2]$ where $y \neq \mathsf{mode}(X^2)$. Then, since y and $\mathsf{mode}(X^2)$ are nonequal integers

$$|y - \mathbb{E}[X^2]| + |\mathsf{mode}(X^2) - \mathbb{E}[X^2]| \geq 1$$

Now,

$$\begin{aligned} var[X^2] &= \sum_{i=0}^{\beta^2} \left(\Pr[X^2 = i](X^2 - \mathbb{E}[X^2])^2\right) \\ &\geq \Pr_X[X^2 \neq \mathsf{mode}(X^2)](y - \mathbb{E}[X^2])^2 + (1 - \Pr_X[X^2 \neq \mathsf{mode}(X^2)])(\mathsf{mode}(X^2) - \mathbb{E}[X^2])^2 \end{aligned}$$

By Claim 4.1

$$var[X^2] \geq \frac{1}{2}\min\left(\frac{1}{\beta+1}, \frac{1}{t}\right) = \frac{1}{2 \cdot \max(\beta+1, t)}$$

$$\mu_4 = \mu_2^2 + var[X^2] \geq \mu_2^2 + \frac{1}{2 \cdot \max(\beta+1, t)}$$

$$\frac{\mu_4}{\mu_2^2} \geq 1 + \frac{1}{2\mu_2^2 \cdot \max(\beta+1, t)}$$

□

Corollary 4.1. *Let $\mathcal{D}$ be any symmetric distribution over $\mathbb{Z}$ with mean 0 and bounded support over $[-\beta, \beta]$. Let X be a random variable with distribution $\mathcal{D}$. If $\Pr_X[|X| \neq \mathsf{mode}(|X|)] \geq \frac{1}{t}$ for some $t > 0$, then $\mathcal{D}$ is (η, γ)-weakly-nice where $\eta = (\min(\frac{1}{\beta}, \frac{1}{t}))^{O(1)}$ and $\gamma = 1 + (\min(\frac{1}{\beta}, \frac{1}{t}))^{O(1)}$.*

The following lemma proves that if the expectations of two distributions on bounded support $[0, 1]$ differ by some parameter q, then there exists a sufficiently large interval such that the difference between the probability that a sample from the first distributions lies in that interval and the probability that a sample from the second distribution lies in that interval is $O(q^{O(1)})$.

Lemma 4.2. *Let p, q be two parameters. Let D_0 and D_1 be distributions with bounded support in $[0, p]$.[3] Let X_0 be a random variable on $\mathcal{D}_0$ and X_1 be a random variable on $\mathcal{D}_1$. Suppose*

$$\left|\mathbb{E}[X_0] - \mathbb{E}[X_1]\right| \geq q.$$

If $[0, p]$ is partitioned into $n = \frac{2p}{q}$ intervals $\{I_i\}_{i=1}^n$ each of width $\frac{q}{2}$, then there exists an interval I_i such that

$$\left|\Pr[x \in I_i \mid x \xleftarrow{R} D_0] - \Pr[x \in I_i \mid x \xleftarrow{R} D_1]\right| \geq \frac{q^2}{4p^2}.$$

[3] More generally, the support is allowed to be $[-p/2, p/2]$ and then the result follows by appropriately shifting the two distributions by $p/2$.

Remark 4.1. Note that $\frac{p}{q} \geq 1$. Otherwise, $\frac{p}{q} < 1$ so $q > p$. But this means that the difference in expectations is bigger than the whole range of the support, which is a contradiction.

Proof. Without loss of generality, let $\mathbb{E}[X_0] \geq \mathbb{E}[X_1]$. Consider the following partition process. Partition $[0, p]$ into $n = \frac{p}{\epsilon}$ disjoint intervals I_i each of width ϵ where $a_i = \sup I_i$ and $a_{i-1} = \inf I_i$ for $i \in [n]$. Since $x \leq a_i$ for $x \in I_i$,

$$\mathbb{E}[X_0] \leq \sum_{i \in [n]} a_i \Pr[\mathcal{D}_0 \in I_i].$$

Similarly, a lower bound on $\mathbb{E}[\mathcal{D}_1]$ is given as follows:

$$\mathbb{E}[X_1] \geq \sum_{i \in [n]} a_{i-1} \Pr[\mathcal{D}_1 \in I_i].$$

Thus,

$$\begin{aligned} q \leq \mathbb{E}[X_0] - \mathbb{E}[X_1] &\leq \sum_{i \in [n]} a_i \Pr[\mathcal{D}_0 \in I_i] - \sum_{i \in [n]} a_{i-1} \Pr[\mathcal{D}_1 \in I_i] \\ &= \sum_{i \in [n]} a_i \left(\Pr[\mathcal{D}_0 \in I_i] - \Pr[\mathcal{D}_1 \in I_i]\right) + \epsilon \Pr[\mathcal{D}_1 \in I_i] \\ &= \epsilon + \sum_{i \in [n]} a_i \left(\Pr[\mathcal{D}_0 \in I_i] - \Pr[\mathcal{D}_1 \in I_i]\right) \end{aligned}$$

Therefore,

$$\sum_{i \in [n]} a_i \left(\Pr[\mathcal{D}_0 \in I_i] - \Pr[\mathcal{D}_1 \in I_i]\right) \geq q - \epsilon$$

By an averaging argument, there exists an index i^* such that

$$a_{i^*} \left[\Pr[\mathcal{D}_0 \in I_{i^*}] - \Pr[\mathcal{D}_1 \in I_{i^*}]\right] \geq \frac{1}{n} \cdot (q - \epsilon).$$

Note $a_{i^*} \leq p$ so by substitution we have:

$$\left| \Pr[\mathcal{D}_0 \in I_{i^*}] - \Pr[\mathcal{D}_1 \in I_{i^*}] \right| \geq \frac{q}{np} - \frac{1}{n^2}$$

Choosing $n = \frac{2p}{q}$ gives us

$$\left| \Pr[\mathcal{D}_0 \in I_{i^*}] - \Pr[\mathcal{D}_1 \in I_{i^*}] \right| \geq \frac{q^2}{4p^2}.$$

□

5 Non-trivial Probability Distinguishers

In this section, we show some of the limits of using polynomials to construct PIDGs. In particular, we show that a PIDG with large enough spread cannot be formed out of any set of polynomials and distributions taken from certain specific classes of polynomials and distributions. In this section, we consider selections of polynomials and distributions that lead to the smallest distinguishing

advantage; we want to distinguish between any choice of polynomials and distributions from these classes. In the next section, we will consider distinguishers when the polynomials are chosen randomly from some class of polynomials.

For these distinguishers, we consider the difference of $\mathbb{E}_{X,Y}[q_i^2(X)q_j^2(Y)]$ and $\mathbb{E}_X[q_i^2(X)q_j^2(X)]$ for polynomials q_i and q_j from some set $\mathcal{Q}$ where $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_m)$ where each X_i, Y_i is an i.i.d. random variable with probability distribution $\mathcal{D}$. When the polynomials are correlated in certain ways, then this difference will be noticeable and can be used to construct a weak probabilistic polynomial time distinguisher that can solve the $(\mathcal{Q}, \mathcal{D}) - \mathsf{PIDP}$ with noticeable probability.

5.1 An Expectation Distinguisher

Algorithm 1 (Expectation Distinguisher)**.**

Given: x from either distribution $\mathcal{D}_0$ or $\mathcal{D}_1$.

Goal: Output 0 if x was sampled from $\mathcal{D}_0$, and output 1 if x was sampled from $\mathcal{D}_1$.

Operation:

1. Let $t = 16000\frac{p^5}{q^5}$. Randomly sample t points from $\mathcal{D}_0$ and t points from $\mathcal{D}_1$. Let S_0 be the set of t points sampled from $\mathcal{D}_0$, and let S_1 be the set of t points sampled from $\mathcal{D}_1$.
2. Partition $[0, p]$ into $n = \frac{2p}{q}$ disjoint intervals $\{I_i\}_{i\in[n]}$ each of width $\frac{q}{2}$ where $a_i = \sup I_i$ and $a_{i-1} = \inf I_i$.
3. Count the number of samples in each interval and compute the sample probabilities, letting

$$S_{0,i} = \{s \in S_0 : s \in I_i\} \qquad r_{0,i} = \frac{|S_{0,i}|}{t}$$
$$S_{1,i} = \{s \in S_1 : s \in I_i\} \qquad r_{1,i} = \frac{|S_{1,i}|}{t}$$

   where $i \in [n]$.
4. Pick interval index i such that $x \in I_i$. If $r_{0,i} \geq r_{1,i}$, then output 0; else if $r_{0,i} < r_{1,i}$ then output 1.

Remark 5.1. If the samplers for $\mathcal{D}_0$ and $\mathcal{D}_1$ run in time at most k, then the Expectation Distinguisher $\mathcal{A}$ performs $(\frac{kp}{q})^{O(1)}$ operations over real numbers. The running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

Lemma 5.1. *Let p, q be two parameters. Let D_0 and D_1 be distributions with bounded support in $[0, p]$.*[4] *Let X_0 be a random variable on $\mathcal{D}_0$ and X_1 be a random variable on $\mathcal{D}_1$. If*

$$\Big|\, \mathbb{E}[X_0] - \mathbb{E}[X_1] \Big| \geq q$$

[4]More generally, the support is allowed to be $[-p/2, p/2]$ and then the result follows by appropriately shifting the two distributions by $p/2$.

then the Expectation Distinguisher $\mathcal{A}$ (Algorithm 1) succeeds with probability

$$\Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| \geq \frac{q^2}{16p^2}$$

Proof. Partition $[0, p]$ into $n = \frac{2p}{q}$ disjoint intervals $\{I_i\}_{i\in[n]}$ each of width $\frac{q}{2}$ where $a_i = \sup I_i$ and $a_{i-1} = \inf I_i$. Let $p_{0,i} = \Pr[\mathcal{D}_0 \in I_i]$ and $p_{1,i} = \Pr[\mathcal{D}_1 \in I_i]$. Define

$$\Delta_i = p_{0,i} - p_{1,i} = \Pr[\mathcal{D}_0 \in I_i] - \Pr[\mathcal{D}_1 \in I_i]$$
$$\delta_i = r_{0,i} - r_{1,i} = \frac{|S_{0,i}|}{t} - \frac{|S_{1,i}|}{t}$$

Note that δ_i is our approximation of Δ_i based on our t samples from each distribution.

We remark that for $b, b' \in \{0, 1\}$ then

$$\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_b) = b'] = \sum_{i\in[n]} \Pr[\mathcal{A}(x) = b' | x \in I_i] \Pr[\mathcal{D}_b \in I_i] = \sum_{i\in[n]} p_{b,i} \Pr[\mathcal{A}(x) = b' | x \in I_i]$$

Therefore, we have

$$\begin{aligned}
&2\Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| \\
&= \Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| + \Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 1] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 1]\Big| \\
&\geq \Big|\Big(\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 1]\Big) - \Big(\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 1]\Big)\Big| \\
&= \Big|\sum_{i\in[n]} p_{0,i}\Big(\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] - \Pr[\mathcal{A}(x) = 1 \mid x \in I_i]\Big) \\
&\qquad - \sum_{i\in[n]} p_{1,i}\Big(\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] - \Pr[\mathcal{A}(x) = 1 \mid x \in I_i]\Big)\Big| \\
&= \Big|\sum_{i\in[n]} \Delta_i\Big(\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] - \Pr[\mathcal{A}(x) = 1 \mid x \in I_i]\Big)\Big|
\end{aligned}$$

Fix some $i \in [n]$. Suppose that $\Delta_i \geq 0$. Then $p_{0,i} \geq p_{1,i}$ and by construction of the algorithm:

$$\begin{aligned}
\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] &= \Pr[\delta_i > 0] \\
&= \frac{1}{2}\Pr[|\Delta_i - \delta_i| > \Delta_i] + \Pr[|\Delta_i - \delta_i| \leq \Delta_i] \\
&= \frac{1}{2} + \frac{1}{2}\Pr[|\Delta_i - \delta_i| \leq \Delta_i]
\end{aligned}$$

Define the random variable $X_{i,k}$ for $i \in [n]$, $k \in [t]$ representing whether the kth sample from $\mathcal{D}_0$ is in I_i and the random variable $Y_{i,k}$ for $i \in [n]$, $k \in [t]$ representing whether the kth sample from $\mathcal{D}_1$ is in I_i as:

$$X_{i,k} = \begin{cases} 1 & \text{if } k\text{th sample from } \mathcal{D}_0 \text{ is in } I_i \\ 0 & \text{o.w.} \end{cases} \qquad Y_{i,k} = \begin{cases} 1 & \text{if } k\text{th sample from } \mathcal{D}_1 \text{ is in } I_i \\ 0 & \text{o.w.} \end{cases}.$$

Then consider the sum of these random variables:

$$X_i = \sum_{k\in[t]} X_{i,k} \qquad \mathbb{E}[X_i] = tp_{0,i}$$
$$Y_i = \sum_{k\in[t]} Y_{i,k} \qquad \mathbb{E}[Y_i] = tp_{1,i}$$

where $X_{i,k}$ and $Y_{i,k}$ are i.i.d. Bernoulli random variable and X_i, Y_i are binomial random variables. Note that the distribution of δ_i is the same as the distribution of $\frac{X_i}{t} - \frac{Y_i}{t}$

Claim 5.1. *Assume that $\Delta_i \geq 0$. Then,*

$$\Pr\left[|\delta_i - \Delta_i| \leq \Delta_i\right] \geq 1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right).$$

Proof. Applying a two-sided Chernoff bound gives

$$\Pr\left[\left|\frac{X_i}{t} - p_{0,i}\right| > \delta p_{0,i}\right] = \Pr\left[|X_i - p_{0,i}t| > \delta p_{0,i} t\right] \leq 2\cdot\exp\left(-\frac{\delta^2 p_{0,i}}{2+\delta}\cdot t\right).$$

Set $\delta p_{0,i} = \theta_0$ to obtain

$$\Pr\left[\left|\frac{X_i}{t} - p_{0,i}\right| > \theta_0\right] \leq 2\exp\left(-\frac{\theta_0^2}{2+\theta_0}\cdot t\right).$$

By the same argument,

$$\Pr\left[\left|\frac{Y_i}{t} - p_{1,i}\right| > \theta_1\right] \leq 2\exp\left(-\frac{\theta_1^2}{2+\theta_1}\cdot t\right).$$

Fix $\theta = \frac{\Delta_i}{2}$. Then $\frac{\theta^2}{2+\theta} = \frac{\Delta_i^2}{8+2\Delta_i}$. Since $0 \leq \Delta_i \leq 1$, then $\exp(-\frac{\Delta_i^2 t}{8+2\Delta_i}) \leq \exp(-\frac{\Delta_i^2 t}{10})$. So by the union bound:

$$\Pr\left[\left(\left|\frac{X_i}{t} - p_{0,i}\right| \leq \frac{\Delta_i}{2}\right) \wedge \left(\left|\frac{Y_i}{t} - p_{1,i}\right| \leq \frac{\Delta_i}{2}\right)\right] \geq 1 - 4\exp\left(-\frac{\Delta_i^2 t}{8+2\Delta_i}\right) \geq 1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right).$$

Then it follows:

$$\Pr\left[\left|\left|\frac{X_i}{t} - p_{0,i}\right| - \left|\frac{Y_i}{t} - p_{1,i}\right|\right| \leq \frac{\Delta_i}{2}\right] \geq 1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right).$$

Since $\Delta_i \leq 1$,

$$\Pr\left[\left|\left(\frac{X_i}{t} - \frac{Y_i}{t}\right) - (p_{0,i} - p_{1,i})\right| \leq \Delta_i\right] \geq 1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)$$
$$\Rightarrow \Pr\left[|\delta_i - \Delta_i| \leq \Delta_i\right] \geq 1 - 4\exp(-\frac{\Delta_i^2 t}{10})$$

□

By the claim,

$$\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] \geq \frac{1}{2} + \frac{1}{2}\left(1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)\right)$$
$$\Pr[\mathcal{A}(x) = 1 \mid x \in I_i] \leq \frac{1}{2} - \frac{1}{2}\left(1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)\right).$$

Therefore,

$$\Delta_i\Big(\Pr[\mathcal{A}(x) = 0 \mid x \in I_i] - \Pr[\mathcal{A}(x) = 1 \mid x \in I_i]\Big) \geq |\Delta_i| \cdot \left(1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)\right)$$

By a symmetric argument, if $\Delta_i < 0$, then $p_{0,i} < p_{1,i}$ and

$$\Delta_i \cdot \Big(\Pr[\mathcal{A}(x) = 1 \mid x \in I_i] - \Pr[\mathcal{A}(x) = 0 \mid x \in I_i]\Big) \geq |\Delta_i| \cdot \left(1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)\right)$$

Since the inequality above holds for all values of Δ_i,

$$\begin{aligned} 2 \cdot \Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| &\geq \Big|\sum_{i\in[n]} |\Delta_i| \cdot \left(1 - 4\exp\left(-\frac{\Delta_i^2 t}{10}\right)\right)\Big| \\ &\geq \max_i |\Delta_i| \cdot \left(1 - 4\exp\left(-\frac{\max_i |\Delta_i|^2 t}{10}\right)\right) \end{aligned}$$

By Lemma 4.2, since $|\mathbb{E}[\mathcal{D}_0] - \mathbb{E}[\mathcal{D}_1]| \geq q$ and $[0, p]$ is partitioned into $n = \frac{2p}{q}$ intervals of equal width, there exists an interval indexed by j such that

$$\frac{q^2}{4p^2} \leq |\Delta_j| \leq \max_i |\Delta_i|.$$

Suppose that the algorithm makes $t = 16000\frac{p^5}{q^5}$ sampling calls for each of the distributions. Since $\frac{p}{q} \geq 1$ as noted in Lemma 4.2, the distinguishing advantage of the algorithm is given by:

$$\Big|\Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_0) = 0] - \Pr[\mathcal{A}(x \xleftarrow{R} \mathcal{D}_1) = 0]\Big| \geq \frac{1}{2} \cdot \left(\frac{q^2}{4p^2}\right) \cdot (1 - 4 \cdot \exp(-100 \cdot \frac{p}{q})) \geq \frac{q^2}{16p^2}$$

□

Corollary 5.1. *Let $\mathcal{Q} = \{q_i\}_{i=1}^m \subset \mathbb{R}[x_1, \ldots x_n]$ be a collection of multilinear polynomials over the reals of degree at most some constant d and coefficients bounded by $[-\nu, \nu]$. Let $\mathcal{D}$ be a samplable distribution over $\mathbb{R}$ with support bounded by $[-\beta, \beta]$. Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ where each X_i and each Y_i is an i.i.d. random variable with probability distribution $\mathcal{D}$. If a probabilistic algorithm can compute $i, j \in [m]$ such that $i \neq j$ and*

$$\left|\mathop{\mathbb{E}}_{X}[q_i^2(X)q_j^2(X)] - \mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)]\right| \geq t$$

then there exists a probabilistic algorithm $\mathcal{A}$ that solves the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem with probability at least

$$\frac{t^2}{16(dn^d\nu\beta^d)^8}$$

.

Proof. Since $\mathcal{Q}$ is of degree at most d, then $\mathcal{Q}$ has at most $\sum_{i=1}^{d}\binom{n}{d} \leq dn^d$ monomials. Since X, Y are bounded by $[-\beta, \beta]^n$ and the coefficients of $\mathcal{Q}$ are in $[-\nu, \nu]$, then for $x \in X$ or $y \in Y$, then $|q_i(\mathbf{x})|, |q_j(\mathbf{y})| \in [0, dn^d\nu\beta^d]$. Therefore, $q_i^2(x)q_j^2(x)$ and $q_i^2(x)q_j^2(y)$ are bounded by $[0, (dn^d\nu\beta^d)^4]$. Now, let $\mathcal{A}$ be the following adversary:

Algorithm 2 (Squared Expectation Distinguisher)**.**

Given: $(\mathcal{Q}, \mathcal{E})$ where $\mathcal{E}$ is either $\{q_i(\mathbf{x})\}_{i=1}^m$ or $\{q_i(\mathbf{x}_i)\}_{i=1}^m$ where $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}$

Operation:

1. Compute $i, j \in [m]$
2. Compute $\mathcal{E}_i^2\mathcal{E}_j^2$ which is either $q_i^2(\mathbf{x})q_j^2(\mathbf{x})$ or $q_i^2(\mathbf{x}_i)q_j^2(\mathbf{x}_j)$.
3. Let $\mathcal{B}$ be the Expectation Distinguisher (Algorithm 1) from Lemma 5.1. Let $\mathcal{D}_0$ be the distribution of $q_i^2(X)q_j^2(X)$ and let $\mathcal{D}_1$ be the distribution of $q_i^2(X)q_j^2(Y)$. Output $\mathcal{B}(\mathcal{D}_0, \mathcal{D}_1, \mathcal{E}_i^2\mathcal{E}_j^2)$.

Since $\mathcal{B}$ is a probabilistic algorithm, then $\mathcal{A}$ is also a probabilistic algorithm. Then, by Lemma 1 since $\mathcal{D}_0$ and $\mathcal{D}_1$ are bounded distributions over $[0, (dn^d\nu\beta^d)^4]$ then

$$\begin{aligned}
&|\Pr[\mathcal{A}(\mathcal{Q}, \{q_i(\mathbf{x})\}_{i=1}^m) = 1] - \Pr[\mathcal{A}(\mathcal{Q}, \{q_i(\mathbf{x}_i)\}_{i=1}^m)]| \\
&= \left|\Pr[\mathcal{B}(\mathcal{D}_0, \mathcal{D}_1, q_i^2(\mathbf{x})q_j^2(\mathbf{x})) = 1] - \Pr[\mathcal{B}(\mathcal{D}_0, \mathcal{D}_1, q_i^2(\mathbf{x}_i)q_j^2(\mathbf{x}_j)) = 1]\right| \geq \frac{t^2}{16(dn^d\nu\beta^d)^8}
\end{aligned}$$

Therefore $\mathcal{A}$ is a probabilistic algorithm that solves the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem with this advantage. □

Remark 5.2. Let the runtime of the sampler for $\mathcal{D}$ be $n^{O(1)}$, and let the algorithm to compute i, j make $n^{O(1)}$ operations over real numbers. Then if $m = n^{O(1)}$, by Remark 5.1, the Squared Expectation Algorithm (Algorithm 2) makes $\left(\frac{n\nu\beta}{t}\right)^{O(1)}$ operations over real numbers. The actual running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

5.2 Non-trivial Distinguisher for Polynomials with Non-negative Coefficients

First, we recall the definition of a (η, γ)-weakly-nice distribution:

Definition 5.1. We say that a distribution $\mathcal{D}$ is (η, γ)-weakly-nice if

1. $\mathcal{D}$ is a symmetric distribution with mean 0
2. If X is a random variable over $\mathcal{D}$, then $\mu_2 = \mathbb{E}[X^2] \geq \eta$ and $\frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]^2} \geq \gamma$.

Definition 5.2. Let $Q_{n,nonneg} \subset \mathbb{R}[x_1, \ldots, x_n]$ be the set of multilinear polynomials over the reals with degree at most some constant d and nonnegative coefficients

Lemma 5.2. *Let $q_1, \ldots, q_m \in Q_{n,nonneg}$, and let $\mathcal{D}$ be any (η, γ)-weakly-nice distribution with $\eta > 0$ and $\gamma > 1$. Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ where each X_i and each Y_i is an i.i.d.*

random variable with probability distribution $\mathcal{D}$. Then, if $m > n$ then a probabilistic algorithm can find $i, j \in [m]$ such that $i \neq j$, q_i, q_j share a variable x_k, and

$$\mathop{\mathbb{E}}_{X}[q_i^2(X)q_j^2(X)] - \mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] \geq (\gamma - 1)(\nu')^2(\nu'')^2\eta^{d'+d''}$$

for any ν', d' that are the coefficient and degree respectively of some monomial in q_i that contains variable x_k, and for any ν'', d'' that are the coefficient and degree respectively of some monomial in q_j that contains variable x_k.

Proof. By the pigeonhole principle, since $m > n$, there must exist $i, j \in [m]$ where $i \neq j$ such that q_i and q_j share a variable x_k. Furthermore, such i, j can be found by a probabilistic algorithm. We know that $q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_S x_S$ and $q_j(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} d_S x_S$ where each $c_S, d_S \in \mathbb{R}$. Consider any nonzero term $c_{S^*}x_{S^*}$ in q_i that contains x_k and any nonzero term $d_{T^*}x_{T^*}$ in q_j that contains x_k. Then, $S^*, T^* \in \mathcal{P}([n])$ such that $|S^* \cap T^*| \geq 1, |S^*| = d', |T^*| = d''$, $c_{S^*} = \nu' \neq 0$, and $d_{T^*} = \nu'' \neq 0$ for some d', d'', ν', ν''. Now,

$$\begin{aligned}
\mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] &= \mathop{\mathbb{E}}_{X}[q_i^2(X)] \mathop{\mathbb{E}}_{Y}[q_j^2(Y)] = \mathop{\mathbb{E}}_{X}[q_i^2(X)] \mathop{\mathbb{E}}_{X}[q_j^2(X)] \\
&= \mathop{\mathbb{E}}_{X}\left[\sum_{S,T \in \mathcal{P}([n])} c_S c_T X_S X_T\right] \mathop{\mathbb{E}}_{X}\left[\sum_{S,T \in \mathcal{P}([n])} d_S d_T X_S X_T\right] \\
&= \sum_{S,T \in \mathcal{P}([n])} c_S c_T \mathop{\mathbb{E}}_{X}[X_S X_T] \sum_{S,T \in \mathcal{P}([n])} d_S d_T \mathop{\mathbb{E}}_{X}[X_S X_T]
\end{aligned}$$

By Lemma 3.1, $\mathbb{E}_X[X_S X_T]$ equals 0 if $S \neq T$ and equals $\mu_2^{|S|}$ if $S = T$. Therefore,

$$\begin{aligned}
\mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] &= \sum_{S \in \mathcal{P}([n])} c_S^2 \mathop{\mathbb{E}}_{X}\left[X_S^2\right] \sum_{S \in \mathcal{P}([n])} d_S^2 \mathop{\mathbb{E}}_{X}\left[X_S^2\right] \\
&= \sum_{S \in \mathcal{P}([n])} c_S^2 \mu_2^{|S|} \sum_{S \in \mathcal{P}([n])} d_S^2 \mu_2^{|S|} \\
&= \sum_{S,T \in \mathcal{P}([n])} c_S^2 d_T^2 \mu_2^{|S|+|T|}
\end{aligned}$$

Now, in the other case, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{X}[q_i^2(X)q_j^2(X)] &= \mathop{\mathbb{E}}_{X}\left[\sum_{S,T,U,V \in \mathcal{P}([n])} c_S c_T d_U d_V X_S X_T X_U X_V\right] \\
&= \sum_{S,T,U,V \in \mathcal{P}([n])} c_S c_T d_U d_V \mathop{\mathbb{E}}_{X}[X_S X_T X_U X_V]
\end{aligned}$$

By Lemma 3.2, $\forall S, T, U, V \in \mathcal{P}([n])$, $\mathbb{E}_X[X_S X_T X_U X_V] \geq 0$. Since all coefficients of q_i and q_j are

nonnegative, then $c_S c_T d_U d_V \mathbb{E}_X [X_S X_T X_U X_V] \geq 0$ Therefore,

$$\begin{aligned}
\mathop{\mathbb{E}}_X[q_i^2(X)q_j^2(X)] &\geq \sum_{S,T\in\mathcal{P}([n])} c_S^2 d_T^2 \mathop{\mathbb{E}}_X \left[X_S^2 X_T^2\right] \\
&= \sum_{S,T\in\mathcal{P}([n])} c_S^2 d_T^2 \left(\frac{\mu_4}{\mu_2^2}\right)^{|S\cap T|} \mu_2^{|S|+|T|} \\
&\geq \sum_{S,T\in\mathcal{P}([n]);S\neq S^* or T\neq T^*} \left(c_S^2 d_T^2 \mu_2^{|S|+|T|}\right) + c_{S^*}^2 d_{T^*}^2 \left(\frac{\mu_4}{\mu_2^2}\right)^{|S^*\cap T^*|} \mu_2^{|S^*|+|T^*|} \\
&\geq \sum_{S,T\in\mathcal{P}([n]);S\neq S^* or T\neq T^*} \left(c_S^2 d_T^2 \mu_2^{|S|+|T|}\right) + c_{S^*}^2 d_{T^*}^2 \left(\frac{\mu_4}{\mu_2^2}\right) \mu_2^{|S^*|+|T^*|} \\
&= \sum_{S,T\in\mathcal{P}([n])} \left(c_S^2 d_T^2 \mu_2^{|S|+|T|}\right) + c_{S^*}^2 d_{T^*}^2 \left(\frac{\mu_4}{\mu_2^2} - 1\right) \mu_2^{|S^*|+|T^*|} \\
&= \mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] + c_{S^*}^2 d_{T^*}^2 \left(\frac{\mu_4}{\mu_2^2} - 1\right) \mu_2^{|S^*|+|T^*|}
\end{aligned}$$

Now, $|S^*| + |T^*| = d' + d''$, $c_{S^*}^2 d_{T^*}^2 = (\nu')^2(\nu'')^2 \neq 0$, $\frac{\mu_4}{\mu_2^2} \geq \gamma > 1$, and $\mu_2 \geq \eta > 0$

$$\mathop{\mathbb{E}}_X[q_i^2(X)q_j^2(X)] - \mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] \geq (\gamma - 1)(\nu')^2(\nu'')^2\eta^{d'+d''}$$

□

Remark 5.3. Since each polynomial $q_i \in \mathcal{Q}$ in the previous lemma is of degree at most some constant d, then q_i has $O(dn^d)$ monomials each of degree at most d. If $m = n^{O(1)}$ then finding $i \neq j$ such that q_i, q_j share a variable requires $n^{O(1)}$ operations over the reals. The running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

Theorem 5.1. *Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \in Q_{n,nonneg}$ with coefficients bounded by $[-\nu, \nu]$ and let $\mathcal{D}$ be a (η, γ)-weakly-nice distribution with $\eta > 0$, $\gamma > 1$ with bounded support in $[-\beta, \beta]$. If $m > n$, then a probabilistic algorithm can find $i, j \in [m]$ such that $i \neq j$ and q_i, q_j share a variable x_k and that solves the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem with probability at least*

$$\frac{(\gamma - 1)^2(\nu')^4(\nu'')^4\eta^{2d'+2d''}}{16(dn^d\nu\beta^d)^8}$$

for any ν', d' that are the coefficient and degree respectively of some monomial in q_i that contains variable x_k and for any ν'', d'' that are the coefficient and degree respectively of some monomial in q_j that contains variable x_k.

Proof. This follows directly from Corollary 5.1 and Lemma 5.2. □

Remark 5.4. Let the runtime of the sampler for $\mathcal{D}$ be $n^{O(1)}$ and let $m = n^{O(1)}$. By Remark 5.3, then the algorithm to compute i, j makes $n^{O(1)}$ operations over real numbers. Then, by Remark 5.1, the distinguisher in Theorem 5.1 makes $\left(\frac{n\nu\beta}{(\gamma-1)\nu'\nu''\eta}\right)^{O(1)}$ operations over real numbers. The actual running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

Corollary 5.2. *Any $(\mathcal{D}, \mathcal{Q})$ satisfying the conditions of Theorem 5.1 where $\gamma - 1, |\nu'|, |\nu''|, \eta$ are $\Omega(n^{-O(1)})$, and m, ν, β are $n^{O(1)}$ is not a* PIDG*.*

Corollary 5.3. *Suppose $\mathcal{D}$ and $\mathcal{Q}$ are over the integers $\mathbb{Z}$. Any $(\mathcal{D}, \mathcal{Q})$ satisfying the conditions of Theorem 5.1 where $\gamma - 1, \eta$ are $\Omega(n^{-O(1)})$, and m, ν, β are $n^{O(1)}$ is not a* PIDG*.*

5.3 Nontrivial Distinguisher for Expander Based Polynomials

Next, we will show that for a different set of polynomials and distributions, we can also find a probabilistic polynomial time algorithm that solves the $(\mathcal{D}, \mathcal{Q}) -$ PIDG with non-negligible probability.

Definition 5.3 (n-Half-Expanding Set)**.** Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of sets. Then, $\mathcal{S}$ is a *n-half-expanding set* if for all $k \leq n$ and all distinct $a_1, a_2, ..., a_k \in [m]$

$$\left| \bigcup_{i=1}^{k} S_{a_i} \right| > \frac{1}{2} \sum_{i=1}^{k} |S_{a_i}|$$

Definition 5.4 (Expander Based Polynomial Set)**.** Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$ be a set of multilinear polynomials over the reals. Then, each $q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_{S,i} x_S$ for some coefficients $\{c_{S,i}\}_{S \in \mathcal{P}([n])} \in \mathbb{R}$. We say that $\mathcal{Q}$ is a Expander Based Polynomial Set if

- Each q_i is a polynomial of degree at most some constant d
- $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0$ *for some* $i \in [m]\}$ is a 4-half expanding set.
- $\mathcal{C}_S = \{c_{S,i}\}_{i \in [m]}$ contains at most one non-zero value. (i.e. All monomials appear at most once across all polynomials in $\mathcal{Q}$.)

Lemma 5.3. *Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$ be a* Expander Based Polynomial Set *and let $\mathcal{D}$ be any (η, γ)-weakly-nice distribution with $\eta > 0$ and $\gamma > 1$. Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ where each X_i and each Y_i is an i.i.d. random variable with probability distribution $\mathcal{D}$. Let d be the maximum degree of each polynomial q_i. Then, if $m > n$ then a probabilistic algorithm can find $i, j \in [m]$ such that $i \neq j$, q_i, q_j share a variable x_k, and*

$$\mathop{\mathbb{E}}_{X}[q_i^2(X) q_j^2(X)] - \mathop{\mathbb{E}}_{X,Y}[q_i^2(X) q_j^2(Y)] \geq (\gamma - 1)(\nu')^2 (\nu'')^2 \eta^{d' + d''}$$

for any ν', d' that are the coefficient and degree respectively of some monomial in q_i that contains variable x_k and for any ν'', d'' that are the coefficient and degree respectively of some monomial in q_j that contains variable x_k.

Proof. By the pigeonhole principle, since $m > n$, there must exist $i, j \in [m]$ where $i \neq j$ such that q_i and q_j share a variable x_k. Furthermore, such i, j can be found by a probabilistic algorithm. We know that $q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} c_S x_S$ and $q_j(\mathbf{x}) = \sum_{S \in \mathcal{P}([n])} d_S x_S$ where each $c_S, d_S \in \mathbb{R}$. Consider any nonzero monomial $c_{S^*} x_{S^*}$ in q_i that contains x_k and any nonzero monomial $d_{T^*} x_{T^*}$ in q_j that contains x_k. Then, $S^*, T^* \in \mathcal{P}([n])$ such that $|S^* \cap T^*| \geq 1, |S^*| = d', |T^*| = d''$, $c_{S^*} = \nu' \neq 0$, and $d_{T^*} = \nu'' \neq 0$ for some d', d'', ν', ν''. Since $\mathcal{Q}$ is a Expander Based Polynomial Set, then all monomials appear at most once in any polynomial. So, $d_{S^*} = 0$. Therefore,

$$q_1(\mathbf{x}) = c_{S^*} x_{S^*} + p_1(\mathbf{x})$$

$$q_2(\mathbf{x}) = p_2(\mathbf{x})$$

where

$$p_1(\mathbf{x}) = \sum_{S \in \mathcal{P}([n]); S \neq S^*} c_S x_S$$

$$p_2(\mathbf{x}) = \sum_{S \in \mathcal{P}([n]); S \neq S^*} d_S x_S.$$

Now,

$$\begin{aligned}
\mathop{\mathbb{E}}_{X,Y}[q_i^2(X) q_j^2(Y)] &= \mathop{\mathbb{E}}_{X}[q_i^2(X)] \mathop{\mathbb{E}}_{X}[q_j^2(X)] \\
&= \mathop{\mathbb{E}}_{X}\left[c_{S^*}^2 X_{S^*}^2 + 2c_{S^*} X_{S^*} p_1(X) + p_1^2(X)\right] \mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right] \\
&= c_{S^*}^2 \mathop{\mathbb{E}}_{X}\left[X_{S^*}^2\right] \mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right] + 2c_{S^*} \mathop{\mathbb{E}}_{X}\left[X_{S^*} p_1(X)\right] \mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right] + \mathop{\mathbb{E}}_{X}\left[p_1^2(X)\right] \mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right].
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathop{\mathbb{E}}_{X}[q_i^2(X) q_j^2(X)] &= \mathop{\mathbb{E}}_{X}\left[c_{S^*}^2 X_{S^*}^2 p_2^2 + 2c_{S^*} X_{S^*} p_1(X) p_2^2 + p_1^2(X) p_2^2\right] \\
&= c_{S^*}^2 \mathop{\mathbb{E}}_{X}\left[X_{S^*}^2 p_2^2(X)\right] + 2c_{S^*} \mathop{\mathbb{E}}_{X}\left[X_{S^*} p_1(X) p_2^2(X)\right] + \mathop{\mathbb{E}}_{X}\left[p_1^2(X) p_2^2(X)\right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X,Y}[q_i^2(X) q_j^2(Y)] - \mathop{\mathbb{E}}_{X}[q_i^2(X)] \mathop{\mathbb{E}}_{Y}[q_j^2(Y)] \\
&= c_{S^*}^2 \left(\mathop{\mathbb{E}}_{X}[X_{S^*}^2 p_2^2(X)] - \mathop{\mathbb{E}}_{X}[X_{S^*}^2] \mathop{\mathbb{E}}_{X}[p_2^2(X)]\right) + 2c_{S^*} \left(\mathop{\mathbb{E}}_{X}[X_{S^*} p_1(X) p_2^2(X)] - \mathop{\mathbb{E}}_{X}[X_{S^*} p_1(X)] \mathop{\mathbb{E}}_{X}[p_2^2(X)]\right) \\
&\qquad + \left(\mathop{\mathbb{E}}_{X}[p_1^2(X) p_2^2(X)] - \mathop{\mathbb{E}}_{X}[p_1^2(X)] \mathop{\mathbb{E}}_{X}[p_2^2(X)]\right)
\end{aligned}$$

We will consider each term separately. First,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X}\left[X_{S^*}^2 p_2^2(X)\right] - \mathop{\mathbb{E}}_{X}\left[X_{S^*}^2\right] \mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right] \\
&= \mathop{\mathbb{E}}_{X}\left[\sum_{S,T \in \mathcal{P}([n]); S,T \neq S^*} d_S d_T X_{S^*}^2 X_S X_T\right] - \mathop{\mathbb{E}}_{X}\left[\sum_{i \in S^*} X_i^2\right] \mathop{\mathbb{E}}_{X}\left[\sum_{S,T \in \mathcal{P}([n]); S,T \neq S^*} d_S d_T X_S X_T\right] \\
&= \sum_{S,T \in \mathcal{P}([n]); S,T \neq S^*} d_S d_T \mathop{\mathbb{E}}_{X}\left[X_{S^*}^2 X_S X_T\right] - \sum_{i \in S^*} \mathop{\mathbb{E}}_{X}\left[X_i^2\right] \sum_{S,T \in \mathcal{P}([n]); S,T \neq S^*} d_S d_T \mathop{\mathbb{E}}_{X}\left[X_S X_T\right]
\end{aligned}$$

By Lemma 3.1, $\mathbb{E}_X[X_S X_T]$ equals 0 if $S \neq T$ and equals $\mu_2^{|S|}$ if $S = T$. Furthermore, by Lemma 3.2, $\mathbb{E}_X[X_{S^*}^2 X_S X_T] \neq 0$ only if $X_{S^*}^2 X_S X_T$ does not contains a variable X_i of odd power. However,

since X_{S^*} is different from X_S, X_T, this only occurs when $S = T$. Therefore,

$$\begin{aligned}
&\mathop{\mathbb{E}}_X\left[X_{S^*}^2 p_2^2(X)\right] - \mathop{\mathbb{E}}_X\left[X_{S^*}^2\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] \\
&= \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2 \mathop{\mathbb{E}}_X\left[X_{S^*}^2 X_S^2\right] - \mu_2^{|S^*|}\sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2\mu_2^{|S|} \\
&= \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2 \mathop{\mathbb{E}}_X\left[X_{S^*}^2 X_S^2\right] - \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2\mu_2^{|S|+|S^*|} \\
&= \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2 \left(\frac{\mu_4}{\mu_2^2}\right)^{|S\cap S^*|}\mu_2^{|S|+|S^*|} - \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2\mu_2^{|S|+|S^*|} \\
&= \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S^2\mu_2^{|S|+|S^*|}\left(\left(\frac{\mu_4}{\mu_2^2}\right)^{|S\cap S^*|} - 1\right) \\
&\geq d_{T^*}^2\mu_2^{|S^*|+|T^*|}\left(\left(\frac{\mu_4}{\mu_2^2}\right)^{|S^*\cap T^*|} - 1\right)
\end{aligned}$$

Since $|S^*| + |T^*| = d' + d''$, $d_{T^*}^2 = (\nu'')^2 \neq 0$, $\frac{\mu_4}{\mu_2^2} \geq \gamma > 1$, and $\mu_2 \geq \eta > 0$.

$$\mathop{\mathbb{E}}_X\left[X_{S^*}^2 p_2^2(X)\right] - \mathop{\mathbb{E}}_X\left[X_{S^*}^2\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] \quad \geq (\gamma-1)(\nu'')^2\eta^{d'+d''}$$

For the next term, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)p_2^2(X)\right] - \mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] \\
&= \mathop{\mathbb{E}}_X\left[\sum_{S,T,U\in\mathcal{P}([n]);S,T,U\neq S^*} d_S X_{S^*}X_S X_T X_U\right] - \mathop{\mathbb{E}}_X\left[\sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S X_{S^*}X_S\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] \\
&= \sum_{S,T,U\in\mathcal{P}([n]);S,T,U\neq S^*} c_S d_T d_U \mathop{\mathbb{E}}_X\left[X_{S^*}X_S X_T X_U\right] - \sum_{S\in\mathcal{P}([n]);S\neq S^*} d_S \mathop{\mathbb{E}}_X\left[X_{S^*}X_S\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right]
\end{aligned}$$

Now by Lemma 3.1, then $\mathbb{E}_X[X_{S^*}X_S] = 0$ whenever $S^* \neq S$. So,

$$\mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)p_2^2(X)\right] - \mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] \quad = \sum_{S,T,U\in\mathcal{P}([n]);S,T,U\neq S^*} c_S d_T d_U \mathop{\mathbb{E}}_X\left[X_{S^*}X_S X_T X_U\right]$$

Now consider the terms where $c_S, d_T, d_U \neq 0$. Then, since $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0 \textit{ for some } i \in [m]\}$ is a 4-half expanding set and $c_{S^*} \neq 0$, then for distinct $S^*, T, U, V \in \mathcal{P}([n])$, then $|S^* \cup T \cup U \cup V| > \frac{1}{2}(|S^*| + |T| + |U| + |V|)$. Therefore, some X_i occurs once in $X_{S^*}X_S X_T X_U$. So, by Lemma 3.2, then $\mathbb{E}_X\left[X_{S^*}X_S X_T X_U\right] = 0$. Suppose then that S^*, T, U, V are not all distinct and that $S^* \neq T, U, V$. Without loss of generality, assume that $U = V$. Then, since $S^* \neq T$ and $S^* \neq U$, then $X_{S^*}X_T X_U^2$ must contain some X_i of odd power. So, by Lemma 3.2, then $\mathbb{E}_X\left[X_{S^*}X_S X_T X_U\right] = 0$. Therefore,

$$\mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)p_2^2(X)\right] - \mathop{\mathbb{E}}_X\left[X_{S^*}p_1(X)\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] = 0$$

For the last term,

$$\begin{aligned}
\mathop{\mathbb{E}}_X\left[p_1^2(X)\right]\mathop{\mathbb{E}}_X\left[p_2^2(X)\right] &= \mathop{\mathbb{E}}_X\left[\sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S c_T X_S X_T\right]\mathop{\mathbb{E}}_X\left[\sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} d_S d_T X_S X_T\right] \\
&= \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S c_T \mathop{\mathbb{E}}_X\left[X_S X_T\right] \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} d_S d_T \mathop{\mathbb{E}}_X\left[X_S X_T\right]
\end{aligned}$$

By Lemma 3.1, then $\mathbb{E}_X[X_S X_T]$ equals 0 whenever $S \neq T$ and equals $\mu_2^{|S|}$ whenever $S = T$. Therefore,

$$\mathop{\mathbb{E}}_{X}\left[p_1^2(X)\right]\mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right] = \sum_{S\in\mathcal{P}[n];S\neq S^*} c_S^2\mu_2^{|S|}\sum_{T\in\mathcal{P}[n];T\neq S^*} d_T^2\mu_2^{|T|} = \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|}$$

So we have that

$$\begin{aligned}&\mathop{\mathbb{E}}_{X}\left[p_1^2(X)p_2^2(X)\right] - \mathop{\mathbb{E}}_{X}\left[p_1^2(X)\right]\mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right]\\ &= \sum_{S,T,U,V\in\mathcal{P}[n];S,T,U,V\neq S^*} c_S c_T d_U d_V \mathop{\mathbb{E}}_{X}\left[X_S X_U X_D X_V\right] - \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|}\end{aligned}$$

Now consider the terms where $c_S, c_T, d_U, d_V \neq 0$. Then, since $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0 \text{ for some } i \in [m]\}$ is a 4-half expanding set, then for distinct $S, T, U, V \in \mathcal{P}([n])$, then $|S \cup T \cup U \cup V| > \frac{1}{2}(|S| + |T| + |U| + |V|)$. Therefore, some X_i occurs once in $X_S X_T X_U X_V$. So, by Lemma 3.2, then $\mathbb{E}_X\left[X_S X_T X_U X_V\right] = 0$. Suppose then that S, T, U, V are not all distinct. Let one of S or T equal one of U or V. But since we assumed that $c_S, c_T, d_U, d_V \neq 0$, this means that c_A and d_A are both nonzero for some set A. But this contradicts the fact that all monomials appear at most once in all polynomials of $\mathcal{Q}$ since $\mathcal{Q}$ is an Expander Based Polynomial Set. Therefore, if S, T, U, V are not all distinct, we need either $S = T$ or $U = V$. Suppose without loss of generality, that $S = T$. Then, in order for $X_S X_S X_U X_V = X_S^2 X_U X_V$ to not contain a variable X_i of odd power, we need $U = V$ as well. So, by Lemma 3.2, then $c_S c_T d_U d_V \mathbb{E}_X\left[X_S X_T X_U X_V\right] \neq 0$ if and only if $S = T$ and $U = V$.

$$\begin{aligned}&\mathop{\mathbb{E}}_{X}\left[p_1^2(X)p_2^2(X)\right] - \mathop{\mathbb{E}}_{X}\left[p_1^2(X)\right]\mathop{\mathbb{E}}_{X}\left[p_2^2(X)\right]\\ &= \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2 \mathop{\mathbb{E}}_{X}\left[X_S^2 X_T^2\right] - \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|}\\ &= \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\left(\frac{\mu_4}{\mu_2^2}\right)^{|S\cap T|}\mu_2^{|S|+|T|} - \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|}\\ &\geq \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|} - \sum_{S,T\in\mathcal{P}[n];S,T\neq S^*} c_S^2 d_T^2\mu_2^{|S|+|T|}\\ &= 0\end{aligned}$$

As a result,

$$\begin{aligned}&\mathop{\mathbb{E}}_{X,Y}[q_i^2(X)q_j^2(Y)] - \mathop{\mathbb{E}}_{X}[q_i^2(X)]\mathop{\mathbb{E}}_{Y}[q_j^2(Y)]\\ &= c_{S^*}^2\left(\mathop{\mathbb{E}}_{X}[X_{S^*}^2 p_2^2(X)] - \mathop{\mathbb{E}}_{X}[X_{S^*}^2]\mathop{\mathbb{E}}_{X}[p_2^2(X)]\right) + 2c_{S^*}\left(\mathop{\mathbb{E}}_{X}[X_{S^*}p_1(X)p_2^2(X)] - \mathop{\mathbb{E}}_{X}[X_{S^*}p_1(X)]\mathop{\mathbb{E}}_{X}[p_2^2(X)]\right)\\ &\qquad + \left(\mathop{\mathbb{E}}_{X}[p_1^2(X)p_2^2(X)] - \mathop{\mathbb{E}}_{X}[p_1^2(X)]\mathop{\mathbb{E}}_{X}[p_2^2(X)]\right)\\ &\geq c_{S^*}^2(\gamma-1)(\nu'')^2\eta^{d'+d''} + 0 + 0\\ &\geq (\gamma-1)(\nu')^2(\nu'')^2\eta^{d'+d''}\end{aligned}$$

□

Remark 5.5. Since each polynomial $q_i \in \mathcal{Q}$ in the previous lemma is of degree at most some constant d, then q_i has $O(dn^d)$ monomials each of degree at most d. Therefore, if $m = n^{O(1)}$, then finding $i \neq j$ such that q_i, q_j share a variable takes $n^{O(1)}$ operations over the reals.

Theorem 5.2. *Let $\mathcal{Q} = \{q_1, \ldots, q_m\} \subset \mathbb{R}[x_1, \ldots, x_n]$ where $\mathcal{Q}$ is a* Expander Based Polynomial Set *with coefficients bounded by $[-\nu, \nu]$, and let $\mathcal{D}$ be a (η, γ)-weakly-nice distribution with $\eta > 0$, $\gamma > 1$ and bounded support in $[-\beta, \beta]$. If $m > n$, then there exists a probabilistic algorithm $\mathcal{A}$ that can find $i, j \in [m]$ such that $i \neq j$ and q_i, q_j share a variable x_k, and that solves the $(\mathcal{D}, \mathcal{Q})$-polynomial independence distinguishing problem with probability at least*

$$\frac{(\gamma-1)^2(\nu')^4(\nu'')^4\eta^{2d'+2d''}}{16(dn^d\nu\beta^d)^8}$$

for any ν', d' that are the coefficient and degree respectively of some monomial in q_i that contains variable x_k, and for any ν'', d'' that are the coefficient and degree respectively of some monomial in q_j that contains variable x_k.

Proof. This follows directly from Corollary 5.1 and Lemma 5.3. □

Remark 5.6. Let the runtime of the sampler for $\mathcal{D}$ be $n^{O(1)}$ and let $m = n^{O(1)}$. By Remark 5.5, then the algorithm to compute i, j makes $n^{O(1)}$ operations over real numbers. Then, by Remark 5.1, the distinguisher in Theorem 5.2 makes $\left(\frac{n\nu\beta}{(\gamma-1)\nu'\nu''\eta}\right)^{O(1)}$ operations over real numbers. The actual running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

Corollary 5.4. *Any $(\mathcal{D}, \mathcal{Q})$ satisfying the conditions of Theorem 5.2 where $\gamma - 1, |\nu'|, |\nu''|, \eta$ are $n^{-O(1)}$, and m, ν, β are $n^{O(1)}$ is not a* PIDG.

Corollary 5.5. *Suppose $\mathcal{D}$ and $\mathcal{Q}$ are over the integers $\mathbb{Z}$. Any $(\mathcal{D}, \mathcal{Q})$ satisfying the conditions of Theorem 5.2 where $\gamma - 1, \eta$ are $n^{-O(1)}$, and m, ν, β are $n^{O(1)}$ is not a* PIDG.

6 Overwhelming Probability Distinguisher

First, we recall the definitions of C-bounded and nice distributions.

Definition 6.1. We say that a distribution $\mathcal{D}$ is C-bounded if

$$\Pr[x \xleftarrow{R} \mathcal{D}, |x| < C] = 1.$$

Remark 6.1. Note that our results also apply if the probability specified above is greater than $1 - n^{-\omega(1)}$ where n is the number of inputs. This follows from a simple union bound.

Definition 6.2. We say that a distribution $\mathcal{D}$ is (γ, C, ϵ)-nice if

1. $\mathcal{D}$ is a symmetric distribution with mean 0
2. (Normalization.) If X is a random variable over $\mathcal{D}$, then $\mathbb{E}[X^2] = 1$ and $\mathbb{E}[X^4] = \gamma$.
3. $\mathcal{D}$ is C-bounded.
4. (Anti-concentration) $\Pr[x \xleftarrow{R} \mathcal{D}, |x| > \epsilon] > \Omega(1)$

Definition 6.3. (Inputs and Coefficient Distributions.)

- We now define the input distribution $\mathcal{D}_{\mathsf{Inp}}$ used when sampling inputs $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. $\mathcal{D}_{\mathsf{Inp}}$ is a $(\gamma_1, C_1, \epsilon_1)-$nice distribution. The input vectors are sampled by sampling each coordinate independently from $\mathcal{D}_{\mathsf{Inp}}$.

- The coefficient distribution $\mathcal{D}_{\mathsf{Coeff}}$ is defined as follows. Let $\mathcal{D}$ denote a $(\gamma_2, C_2, \epsilon_2)-$nice distribution. Then, for some $p \in [0,1]$, let $\mathcal{D}_{\mathsf{Coeff}}$ be the distribution that outputs 0 with probability p and samples from $\mathcal{D}$ with probability $1-p$. Observe that $\mathbb{E}[X^2] = p$ and $\mathbb{E}[X^4] = \gamma_2 \cdot p$ for X a random variable over $\mathcal{D}$. The ratio for Z a random variable over $\mathcal{D}_{\mathsf{Coeff}}$,

$$\frac{\mathbb{E}[Z^4]}{\mathbb{E}[Z^2]^2} = \gamma_2/p$$

  Also observe that all odd moments of Z are 0.

Problem Setup Let n denote the number of variables/inputs for any given polynomial. Let m be the number of polynomial evaluations. Let d be the constant degree of every polynomial. Let $\gamma_1, \gamma_2, C_1, C_2, \epsilon_1, \epsilon_2$ be a set of parameters. We now describe the process of sampling polynomials as follows:

- Each polynomial is generated as

$$q_i(\mathbf{x}) = \sum_{S \in \mathcal{P}([n]), |S|=d} c_S x_S$$

  where $i \in [m]$, $\mathcal{P}(\cdot)$ denotes the power set, and each c_S is sampled randomly from $\mathcal{D}_{\mathsf{Coeff}}$ for a given probability parameter p. Note that this means that roughly the density of each polynomial is $t = \binom{n}{d}p$.

- Inputs to the generated polynomials are vectors where each coordinate is sampled from $\mathcal{D}_{\mathsf{Inp}}$ as mentioned above. Throughout, we will treat x_i in small letters as an input variable to the polynomial and X_i in capital letters as a tuple of random variables, each of which has distribution $\mathcal{D}_{\mathsf{Inp}}$.

The problem we are interested in is the $(\mathcal{D}_{\mathsf{Inp}}, \mathcal{Q} = (q_1, \ldots, q_m))-$Polynomial Independence Distinguishing Problem.

Consider Algorithm 3. We now prove correctness of the algorithm and then analyze its running time.

Theorem 6.1. *Assume $\gamma_1, \gamma_2, \epsilon_1 = \theta(1)$, $p = \Omega(n \log n \cdot C_1^d / \binom{n}{d})$, $p < \gamma_2/3$, $m = \Omega(n^2 \cdot C_1^{8d} \cdot C_2^8 \cdot \log^{10} n)$ then, Algorithm 3 is an overwhelming distinguisher for the* PIDP *problem with respect to the input and polynomial distributions specified above.*

Algorithm 3 (Strong Distinguishing Algorithm)**.**

Given: Polynomials $\{q_i\}_{i=1}^m$ where $q_1, ..., q_m : \mathbb{R}^n \to \mathbb{R}$, along with evaluations $\{y_i\}_{i \in [m]}$ for some $\mathbf{e}$ sampled from $\mathcal{D}_{\mathsf{Inp}}$ as described above where either $y_i = q_i(\mathbf{e})$ for a fixed $\mathbf{e}$ (denoted by the event same)for all $i \in [m]$ or $y_i = q_i(\mathbf{e}_i)$ (denoted by the event diff) for $\mathbf{e}_i$ sampled from $\mathcal{D}_{\mathsf{Inp}}$ as described above.

Goal: Output 0 if same holds and 1 otherwise.

Operation:

1. Let α_{th} be as defined below.
2. Compute $F(\alpha_{th}, y_1, \ldots, y_m) = \sum_i y_i^4 - 2 \cdot \alpha_{\text{th}} \sum_{i \in [m/2]} y_{2i-1}^2 \cdot y_{2i}^2$

3. If $F(\alpha_{\text{th}}, y_1 \ldots, y_m) \geq 0$ output 1 otherwise output 0.

We now define α_{th} as:

$$\alpha_{\text{th}} = \frac{\alpha_{\text{same}} + \alpha_{\text{diff}}}{2}$$

Define α_{same} as:

$$\alpha_{\text{same}} = \frac{\mathbb{E}_{q_1,X}[q_1^4(X)]}{\mathbb{E}_{q_1,q_2,X}[q_1^2(X) \cdot q_2^2(X)]}$$

Define α_{diff} as:

$$\alpha_{\text{diff}} = \frac{\mathbb{E}_{q_1,X_1}[q_1^4(X_1)]}{\mathbb{E}_{q_1,q_2,X_1,X_2}[q_1^2(X_1) \cdot q_2^2(X_2)]}$$

where the expectations are taken over the input distribution $\mathcal{D}_{\mathsf{Inp}}$ and coefficient distribution $\mathcal{D}_{\mathsf{Coeff}}$.

Lemma 6.1. *Assuming $p < \gamma_2/3$, $\alpha_{\text{same}} = 3 + \theta(\frac{\gamma_2 \cdot \gamma_1^d}{t})$*

Proof. Recall that,

$$\alpha_{\text{same}} = \frac{\mathbb{E}_{q_1,X}[q_1^4(X)]}{\mathbb{E}_{q_1,q_2,X}[q_1^2(X) \cdot q_2^2(X)]}$$

Let $q_1(\mathbf{x}) = \sum_S c_S \cdot x_S$ and $q_2(\mathbf{x}) = \sum_S d_S \cdot x_S$. Here the coefficients are sampled from $\mathcal{D}_{\mathsf{Coeff}}$ and inputs are sampled from $\mathcal{D}_{\mathsf{Inp}}$. Now we compute the numerator.

$$\begin{aligned}
\mathop{\mathbb{E}}_{q_1,X}[q_1^4(X)] &= \mathop{\mathbb{E}}_{X} \mathop{\mathbb{E}}_{q_1} \left[\sum_{S_1} \sum_{S_2} \sum_{S_3} \sum_{S_4} c_{S_1} \cdot c_{S_2} \cdot c_{S_3} \cdot c_{S_4} \cdot X_{S_1} \cdot X_{S_2} \cdot X_{S_3} \cdot X_{S_4} \right] \\
&= \mathop{\mathbb{E}}_{X} \left[\sum_{S_1} \sum_{S_2} \sum_{S_3} \sum_{S_4} \mathop{\mathbb{E}}_{q_1} \left[c_{S_1} \cdot c_{S_2} \cdot c_{S_3} \cdot c_{S_4} \cdot X_{S_1} \cdot X_{S_2} \cdot X_{S_3} \cdot X_{S_4} \right] \right] \\
&= \mathop{\mathbb{E}}_{X} \left[\sum_{S} p \cdot \gamma_2 \cdot X_S^4 + 3 \cdot p^2 \cdot \sum_{S_1 \neq S_2} X_{S_1}^2 \cdot X_{S_2}^2 \right]
\end{aligned}$$

The last equality follows because the odd moments of each coefficient are 0. Let $N = \binom{n}{d}$, then the numerator becomes,

$$= N \cdot p \cdot \gamma_2 \cdot \mathop{\mathbb{E}}_{X}[X_S^4] + 3 \cdot p^2 \cdot \sum_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1}^2 \cdot X_{S_2}^2]$$

Since, $\mathbb{E}_X[X_S^4] = \gamma_1^d$ and $\sum_{S_1 \neq S_2} \mathbb{E}_X[X_{S_1}^2 \cdot X_{S_2}^2] = N \cdot (N-1)\, \mathbb{E}_{S_1 \neq S_2}\, \mathbb{E}_X[X_{S_1} \cdot X_{S_2}]$, the numerator becomes,

$$= N \cdot p \cdot \gamma_2 \cdot \gamma_1^d + 3 \cdot p^2 \cdot N \cdot (N-1) \mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} \cdot X_{S_2}]$$

For $i \in [d-1]$, let g_i denote the probability that two randomly chosen sets $S_1 \neq S_2$ in $[n]$ of size d have i common elements.

This means that,

$$\mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} \cdot X_{S_2}] = (1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}$$

This means that the numerator is,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{q_1, X}[q_1^4(X)] \\
=&N \cdot p \cdot \gamma_2 \cdot \gamma_1^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)
\end{aligned}$$

Now, consider the denominator, $\mathbb{E}_{q_1, q_2, X}[q_1^2(X) \cdot q_2^2(X)]$. By a similar calculation above, we can show that:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{q_1, q_2, X}[q_1^2(X) \cdot q_2^2(X)] \\
=&\mathop{\mathbb{E}}_{X}[p^2 \cdot \sum_{S_1, S_2} X_{S_1}^2 \cdot X_{S_2}^2] \\
=&\mathop{\mathbb{E}}_{X}[p^2 \cdot \sum_{S} X_S^4 + \sum_{S_1 \neq S_2} X_{S_1}^2 \cdot X_{S_2}^2] \\
=& \, p^2 \cdot N \cdot \gamma_1^d + p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)
\end{aligned}$$

From this, observe that

$$\begin{aligned}
\alpha_{\text{same}} =& \\
=&\frac{N \cdot p \cdot \gamma_2 \cdot \gamma_1^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)}{p^2 \cdot N \cdot \gamma_1^d + p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)}
\end{aligned}$$

Let $t = p \cdot N$, then observe that:

$$\begin{aligned}
\alpha_{\text{same}} =& \\
=&3 + \frac{N \cdot p \cdot \gamma_2 \cdot \gamma_1^d - 3 \cdot p^2 \cdot N \cdot \gamma_1^d}{p^2 \cdot N \cdot \gamma_1^d + p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)}
\end{aligned}$$

Assuming $p < \gamma_2/3$, numerator of additive term is $\theta(N \cdot p \cdot \gamma_2 \cdot \gamma_1^d)$. Since $\gamma_1, \gamma_2 > 1$, denominator is $\theta(p^2 \cdot N^2)$. Thus, $\alpha_{\text{same}} = 3 + \theta(\frac{\gamma_2 \cdot \gamma_1^d}{t})$.

□

Lemma 6.2. *Assuming $d > 1$ and γ_1, γ_2 are constants $\alpha_{\text{diff}} = 3 + \frac{\gamma_2 \cdot \gamma_1^d}{t} + \Omega(1/n)$*

Proof. Recall the definition α_{diff},

$$\alpha_{\text{diff}} = \frac{\mathbb{E}_{q_1, \mathbf{X}_1}[q_1^4(X_1)]}{\mathbb{E}_{q_1, q_2, X_1, X_2}[q_1^2(X_1) \cdot q_2^2(X_2)]}$$

The numerator is identical to the calculation done for α_{same}. Hence, the numerator is:

$$\mathop{\mathbb{E}}_{q_1,X_1}[q_1^4(X_1)] =$$
$$= N \cdot p \cdot \gamma_2 \cdot \gamma_1^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)$$

Now, let's compute the denominator:

$$\mathop{\mathbb{E}}_{q_1,q_2,X_1,X_2}[q_1^2(X_1) \cdot q_2^2(X_2)] =$$
$$= \mathop{\mathbb{E}}_{q_1,q_2,X_1,X_2}[\sum_{S_1,S_2,S_3,S_4} c_{S_1} c_{S_2} d_{S_3} d_{S_4} X_{1,S_1} X_{1,S_2} X_{2,S_3} X_{2,S_4}]$$

where we write $q_1(\mathbf{x}) = \sum_S c_S x_S$ and $q_2(\mathbf{y}) = \sum_S d_S y_S$. Now, since the odd moments of coefficients are 0, this becomes:

$$\mathop{\mathbb{E}}_{q_1,q_2,X_1,X_2}[q_1^2(X_1) \cdot q_2^2(X_2)] =$$
$$= \mathop{\mathbb{E}}_{q_1,q_2,X_1,X_2}[\sum_{S_1,S_3} c_{S_1}^2 d_{S_3}^2 X_{1,S_1}^2 X_{2,S_3}^2]$$
$$= p^2 \cdot \mathop{\mathbb{E}}_{X_1,X_2}[\sum_{S_1,S_3} X_{1,S_1}^2 X_{2,S_3}^2]$$
$$= N^2 \cdot p^2$$

This means that:

$$\alpha_{\text{diff}} = \frac{N \cdot p \cdot \gamma_2 \cdot \gamma_1^d + 3 \cdot p^2 \cdot N \cdot (N-1) \cdot \Big((1 - g_1 - \ldots - g_{d-1}) \cdot 1 + \gamma_1 \cdot g_1 + \ldots + \gamma_1^{d-1} \cdot g_{d-1}\Big)}{N^2 \cdot p^2}$$

When d is a constant integer greater than 1, observe that $g_i = \theta(1/n^i)$ for $i \in [d]$. Hence,

$$\alpha_{\text{diff}} \geq \frac{\gamma_2 \gamma_1^d}{t} + 3 \cdot (1 - \frac{1}{N}) \cdot (1 + \theta(\frac{1}{n}))$$

Since $N \geq n^2$, the claim holds. □

From the above two lemmata it holds that:

Corollary 6.1. *Assuming $d \geq 2$, γ_1 and γ_2 are constants greater than 1, $\alpha_{\text{th}} = 3 + \Omega(1/n)$*

Lemma 6.3. *Assume $\gamma_1, \gamma_2, \epsilon_1 = \theta(1)$, $t = \Omega(n \log n \cdot C_1^d)$, $m = \Omega(n^2 \cdot C_1^{8d} \cdot C_2^8 \cdot \log^{10} n)$ then, with probability $1 - n^{-\omega(1)}$, Algorithm 3 outputs 0, given randomly chosen input from the* same *distribution.*

Proof. Define V_i for $i \in [m/2]$ to be the random variable denoting:

$$V_i = q_{2i-1}^4(X) + q_{2i}^4(X) - 2\alpha_{\text{th}} q_{2i-1}^2(X) \cdot q_{2i}^2(X)$$

For a given vector $\mathbf{x}$, let us calculate $\mathbb{E}_{q_{2i-1},q_{2i}}[V_i] = \mu_{\mathbf{x}}$. Using a similar calculation as done in the previous lemmata, we obtain that:

$$\mu_{\mathbf{x}} = p \cdot (2 \cdot \gamma_2 - 2\alpha_{\text{th}} \cdot p) \sum_S x_S^2 + (6 - 2 \cdot \alpha_{\text{th}}) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2$$

Now, in order to prove the claim, we need to show that with probability $1 - n^{-\omega(1)}$ over the choice of the polynomials and input :

$$\sum_{i \in [m/2]} (V_i - \mu_{\mathbf{x}}) + m \cdot \mu_{\mathbf{x}}/2 < 0$$

In order to prove the lemma, we prove the following two conditions hold with probability $1 - n^{-\omega(1)}$,

$$\mu_{\mathbf{x}} < 0 \tag{1}$$

Secondly,

$$\left| \sum_{i \in [m/2]} (V_i - \mu_{\mathbf{x}}) \right| < |m\mu_{\mathbf{x}}/2| \tag{2}$$

We now show Equation 1 holds with probability at least $1 - n^{\omega(1)}$. Observe,

$$\mu_{\mathbf{x}} = p \cdot (2 \cdot \gamma_2 - 2\alpha_{\text{th}} \cdot p) \sum_S x_S^2 + (6 - 2 \cdot \alpha_{\text{th}}) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2$$

As $\alpha_{\text{th}} > 3 + \Omega(1/n)$,

$$\mu_{\mathbf{x}} < p \cdot (2 \cdot \gamma_2 - 2\alpha_{\text{th}} \cdot p) \sum_S x_S^2 - \Omega(1/n) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2$$

We show that with probability $1 - n^{-\omega(1)}$,

$$\mu_{\mathbf{x}} < 2 \cdot p \cdot \gamma_2 \sum_S x_S^2 - \Omega(1/n) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2 < 0$$

Since input distribution is C_1 bounded, this can be proven if we show that with probability $1 - n^{-\omega(1)}$,

$$\mu_{\mathbf{x}} < 2 \cdot p \cdot \gamma_2 \cdot N \cdot C_1^{2d} - \Omega(1/n) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2 < 0$$

Since the input distribution satisfies $\Pr[|\mathcal{D}_{\mathsf{Inp}}| > \epsilon_1] > \Omega(1)$, where $\epsilon_1 = \Omega(1)$, then with probability $1 - n^{-\omega(1)}$

$$\sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2 = \Omega(N^2)$$

This means, that with probability $1 - n^{-\omega(1)}$, $\mu_{\mathbf{x}} < 0$, if:

$$p^2 \cdot N^2/n >> \gamma_2 \cdot p \cdot N \cdot C_1^d$$

As $\gamma_2 = O(1)$, this can be ensured if $t >> n \cdot C_1^d$.

Now we prove the second claim. With probability $1 - n^{-\omega(1)}$,

$$\left| \sum_{i \in [m/2]} (V_i - \mu_{\mathbf{x}}) \right| < |m\mu_{\mathbf{x}}/2| \tag{3}$$

Due to C_1 boundedness it holds that $|x_i| < C_1$ for $i \in [n]$.

Claim 6.1. *With probability* $1 - e^{-\Omega(\log^2 n)}$ *over the coins of* q_i*,*

$$|q_i(\mathbf{x}) = \sum_S c_S x_S| \leq O(C_1^d \cdot C_2 \cdot \sqrt{t} \cdot \log n)$$

Proof. To prove this we apply hoeffding bound. Note that for a fixed $\mathbf{x}$, $q_i(\mathbf{x}) = \sum_S c_S x_S$. Here the coefficeints are chosen independently from $\mathcal{D}_{\mathsf{Coeff}}$. The coefficients are chosen to be 0 with probability $1 - p$ and from a distribution $\mathcal{D}$ with probability p. We replace it by chosing a set $\mathcal{S}$ containing all monomials that have non zero coefficients that are sampled from $\mathcal{D}$. This set is constructed by choosing each set S of size d with probability p. The expected number of elements inside this set is $t = N \cdot p$. Let k be the number of elements inside this set $\mathcal{S}$. Then,

$$q_i(\mathbf{x}) = \sum_{S \in \mathcal{S}} c_S \cdot x_S$$

where c_S is now chosen from $\mathcal{D}$. Since $\mathcal{D}_{\mathsf{lnp}}$ is C_1 bounded and $\mathcal{D}$ is C_2 bounded, we can now use hoeffding bound to bound with probability the absolute value $|q_i(\mathbf{x})|$ to show:

$$\Pr[|q_i(\mathbf{x})| < \sqrt{k} \cdot C_1^d \cdot C_2 \cdot \log n] > 1 - e^{-\Omega(\log^2 n)} \tag{4}$$

Then, observe that by chernoff bound,

$$\Pr[|k - t| < t/2] > 1 - e^{-\Omega(t)}$$

Thus the required probability is computed as follows. Let A_1 be the event stated in the claim. A_2 be the event that when the set $\mathcal{S}$ is selected for the non zero coefficients, the condition in Equation 4 is satisfied. Let A_3 be event that size of $\mathcal{S}$ is within $[t/2, 3t/2]$

$$\Pr[A_1] \geq \Pr[A_2 \wedge A_3]$$

Thus by a union bound, the probability is at least $1 - e^{-\Omega(t)} - e^{-\Omega(\log^2 n)}$. The claim follows by observing that $t > n$.

□

Now, consider:

$$\sum_{i \in [m/2]} V_i - \mu_x$$

With probability $1 - n^{-\omega(1)}$ over polynomials q_i, each V_i is bounded in absolute value by $O(C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^4 n)$.

Now we apply Hoeffding bound again to bound

$$\sum_{i \in [m/2]} V_i - \mu_x$$

However, Hoeffding bound requires that each random variable V_i to be bounded with an interval of $O(C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^4 n)$ with probability 1 over the coins of choosing the polynomials. However, this happens only with probability $1 - n^{-\omega(1)}$ in our case. In order to deal with this issue, note that due to niceness of the input distribution, each coefficient is bounded by C_2 and inputs are bounded in absolute value by C_1. Thus each $q_i(\mathbf{x})$ are bounded by $N \cdot C_1^d \cdot C_2$ in absolute value. Define, V_i' to be the random variable denoting V_i if the underlying polynomials sampled force the absoulte value to be smaller than $O(C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^4 n)$, and 0 otherwise. Now consider,

$$\sum_{i \in [m/2]} V_i - \mu_x$$

Let E_i be the event that $V_i = V_i'$. Observe that, $\mathbb{E}[V_i] = \mu_{\mathbf{x}}$ and $|\mathbb{E}[V_i' - V_i]| = n^{-\omega(1)}$. Note that $\mathbb{E}[V_i] = \mu_{\mathbf{x}} = \mathbb{E}[V_i/E_i] \cdot \Pr[E_i] + \mathbb{E}[V_i/\overline{E}_i] \cdot \Pr[\overline{E}_i]$. Note that $\Pr[\overline{E}_i] = O(n^{-\omega(1)})$. Since $\mathbb{E}[V_i/\overline{E_i}] = O(N^4 \cdot C_1^{4d} \cdot C_2^4)$, $\mu_{\mathbf{x}} = \mathbb{E}[V_i/E_i] \cdot \Pr[E_i] + O(n^{-\omega(1)})$. Also $\mathbb{E}[V_i/E_i] = \mathbb{E}[V_i']$. This means that $|\mathbb{E}[V_i] - \mathbb{E}[V_i']| \leq O(n^{-\omega(1)})$. Denote by $\mu_{\mathbf{x}}' = \mathbb{E}[V_i']$. Consider,

$$\sum_{i \in [m/2]} V_i' - \mu_x = \sum_{i \in [m/2]} V_i' - \mu_x' + \frac{m(\mu_{\mathbf{x}}' - \mu_{\mathbf{x}})}{2}$$

Observe that by Hoeffding's inequality,

$$\left| \sum_{i \in [m/2]} V_i' - \mu_x' \right| \leq O(\sqrt{m/2} \cdot C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^5 n)$$

with probability at least $1 - e^{-\Omega(\log^2 n)}$. Thus,

$$\left| \sum_{i \in [m/2]} V_i' - \mu_x \right| \leq O(\sqrt{m/2} \cdot C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^5 n)$$

as $|\mu_{\mathbf{x}} - \mu_{\mathbf{x}}'| \leq n^{-\omega(1)}$.

As $V_i = V_i'$ with probability $1 - n^{-\omega(1)}$, by a union bound, with probability $1 - n^{-\omega(1)}$,

$$\left| \sum_{i \in [m/2]} V_i - \mu_x \right| \leq \left| \sum_{i \in [m/2]} V_i' - \mu_x \right|$$

Let's now conclude a lower bound on $|\mu_{\mathbf{x}}|$. Observe

$$|\mu_{\mathbf{x}}| = |p \cdot (2 \cdot \gamma_2 - 2\alpha_{\text{th}} \cdot p) \sum_S x_S^2 + (6 - 2 \cdot \alpha_{\text{th}}) p^2 \sum_{S_1 \neq S_2} x_{S_1}^2 \cdot x_{S_2}^2|$$

Since the distribution on inputs $\mathcal{D}_{\mathsf{Inp}}$ is $(\gamma_1, C_1, \epsilon_1)$ nice where $\epsilon_1, \gamma_1 = \theta(1)$ and $\alpha_{\text{th}} = 3+\Omega(1/n)$, with probability $1 - n^{-\omega(1)}$,

$$|\mu_{\mathbf{x}}| \geq p^2 \cdot N \cdot (N-1)a_\epsilon/n - (2\gamma_2 + 2\alpha_{\text{th}} \cdot p)p \cdot N \cdot C_1^d$$

for some constant a_ϵ. Thus, this is $\Omega(t^2/n)$ if t is $\omega(C_1^d)$.

This means that with probability at least $1 - n^{-\omega(1)}$ over the choice of $\mathbf{x}$ and polynomials $q_1, .., q_m$, the claim holds true if:

$$\sqrt{m/2} \cdot C_1^{4d} \cdot C_2^4 \cdot t^2 \cdot \log^5 n << \frac{m \cdot t^2}{2n}$$

This is true if:

$$m > 2 \cdot n^2 \cdot C_1^{8d} \cdot C_2^8 \cdot \log^{10} n$$

□

Lemma 6.4. *Assume $\gamma_1, \gamma_2, \epsilon_1 = \theta(1)$, $t = \Omega(n \log n \cdot C_1^d)$, $m = \Omega(n^2 C_1^{8d} \cdot C_2^8 \cdot \log^{10} n)$ then, with probability $1-n^{-\omega(1)}$, Algorithm 3 outputs 1, given randomly chosen input from the* diff *distribution.*

Proof. Define U_i for $i \in [m/2]$ to be the random variable denoting:

$$U_i = q_{2i-1}^4(X_{2i-1}) + q_{2i}^4(X_{2i}) - 2\alpha_{\text{th}} q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})$$

First observe that

$$\begin{aligned}
\mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[U_i] = & \mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[q_{2i-1}^4(X_{2i-1}) + q_{2i}^4(X_{2i}) \\
& \qquad - 2\alpha_{\text{th}} q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})] \\
= & \mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[q_{2i-1}^4(X_{2i-1}) + q_{2i}^4(X_{2i}) \\
& \qquad - (\alpha_{\text{same}} + \alpha_{\text{diff}}) q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})] \\
= & \mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[q_{2i-1}^4(X_{2i-1}) + q_{2i}^4(X_{2i}) \\
& \qquad - 2\alpha_{\text{diff}} q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i}) \\
& \qquad + (\alpha_{\text{diff}} - \alpha_{\text{same}}) q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})] \\
= & \mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[(\alpha_{\text{diff}} - \alpha_{\text{same}}) q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})] \\
= & \Omega(p^2 \cdot N^2/n)
\end{aligned}$$

The above calculation uses the fact that:

$$\mathop{\mathbb{E}}_{q_{2i}, q_{2i-1}, X_{2i-1}, X_{2i}}[q_{2i-1}^4(X_{2i-1}) + q_{2i}^4(X_{2i}) - 2\alpha_{\text{diff}} q_{2i-1}^2(X_{2i-1}) \cdot q_{2i}^2(X_{2i})] = 0$$

and $\alpha_{\text{diff}} - \alpha_{\text{same}} = \Omega(1/n)$. Denote $\mathbb{E}[U_i] = \mu$. As before, we will be done if we prove that with probability $1 - n^{-\omega(1)}$,

$$\sum_{i\in[m/2]}(U_i-\mu)+\frac{m\cdot\mu}{2}>0$$

This will follow if we show that with probability $1-n^{-\omega(1)}$:

$$\left|\sum_{i\in[m/2]}(U_i-\mu)\right|<\frac{m\cdot\mu}{2}$$

Note that this is enough as $\mu>0$. First observe that with probability $1-e^{\Omega(-\log^2 n)}$,

$$|q_i(\mathbf{x}_i)|=O(\sqrt{t}\cdot C_1^d\cdot C_2\log n)$$

This is proven similar to claim 6.1 This uses chernoff bound to bound density of each polynomial within $[t/2,3t/2]$ monomials and then a hoeffding bound relying on C_1 and C_2 boundedness of $\mathcal{D}_{\mathsf{Inp}}$ and $\mathcal{D}_{\mathsf{Coeff}}$. This means that with probability probability $1-e^{\Omega(-\log^2 n)}$,

$$|U_i-\mu|=O(t^2\cdot C_1^{4d}\cdot C_2^4\log^4 n)$$

We now apply hoeffding bound. Denote $Z_i=U_i-\mu$. Note that with probability $1-e^{\Omega(-\log^2 n)}$,

$$|Z_i|=O(t^2\cdot C_1^{4d}\cdot C_2^4\log^4 n)$$

but with probability 1,

$$|Z_i|=O(N^4\cdot C_1^{4d}\cdot C_2^4)$$

Let E_i be the event that

$$|Z_i|=O(t^2\cdot C_1^{4d}\cdot C_2^4\log^4 n)$$

and define Z_i' to be equal to Z_i if E_i occurs and 0 otherwise. Note that $\mathbb{E}[Z_i]=0$ and,

$$\begin{aligned}\mathbb{E}[Z_i]&=0\\&=\mathbb{E}[Z_i/E_i]\Pr[E_i]+\mathbb{E}[Z_i/\overline{E}_i]\Pr[\overline{E}_i]\\&=\mathbb{E}[Z_i']\Pr[E_i]+\mathbb{E}[Z_i/\overline{E}_i]\Pr[\overline{E}_i]\\&=\mu'\cdot(1-e^{-\Omega(\log^2 n)})+\mathbb{E}[Z_i/\overline{E}_i]\cdot e^{-\Omega(\log^2 n)}\end{aligned}$$

This means that $|\mu'|=|\mathbb{E}[Z_i']|=O(e^{-\log^2 n})$ as $\mathbb{E}[Z_i/\overline{E}_i]<|Z_i|=O(N^4\cdot C_1^{4d}\cdot C_2^4)$ with probability 1.

Now consider the probability

$$\left|\sum_{i\in[m/2]}Z_i\right|<\frac{m\cdot\mu}{2}$$

Let's denote this event by E^*. Thus,

$$\Pr[E^*] = \Pr[E^* / \wedge_{i\in[m/2]} E_i] \cdot \Pr[\wedge_{i\in[m/2]} E_i] + \Pr[E^* / \vee_{i\in[m/2]} \overline{E}_i] \cdot \Pr[\vee_{i\in[m/2]} \overline{E}_i]$$

Observe that $\Pr[\wedge_{i\in[m/2]} E_i] \geq 1 - m \cdot e^{-\Omega(-\log^2 n)}$ using a union bound. Now let's analyse $\Pr[E^* / \wedge_{i\in[m/2]} E_i]$. This probability is the same as the probability

$$\left| \sum_{i\in[m/2]} Z'_i \right| < \frac{m \cdot \mu}{2}$$

By using hoeffding bound, with probability $1 - e^{-\Omega(\log^2 n)}$,

$$\left| \sum_{i\in[m/2]} Z'_i \right| < |m\mu'/2| + O(\sqrt{m/2} \cdot t^2 \cdot C_1^{4d} \cdot C_2^4 \cdot \log^5 n)$$

By substituting $|\mu'| = O(n^{-\omega(1)})$, if $m = \Omega(n^2 \cdot C_1^{8d} \cdot C_2^8 \cdot \log^{10} n)$, the claim holds.

□

Running Time. The algorithm 3 first computes ratio α_{th} which can be computed exactly using the formulae described in lemma 6.1 and 6.2. This step consists of $O(d^{O(1)})$ operations. Then, the algorithm computes a simple objective function which consists of $O(m)$ real operations. The running time scales multiplicatively as the number of real operations times the cost of manipulating ℓ bit numbers where ℓ is the precision of the input to the algorithm.

7 Acknowledgements

We gratefully thank Boaz Barak, Pravesh Kothari, and Rachel Lin for several illuminating conversations about estimating features of inputs based on observations of random polynomial evaluations.

8 References

[ABKS17] Prabhanjan Ananth, Zvika Brakerski, Dakshita Khurana, and Amit Sahai. Constructing indistinguishability obfuscation using preprocessing-friendly pseudoindependence generators. Unpublished Work, 2017.

[ABR12] Benny Applebaum, Andrej Bogdanov, and Alon Rosen. A dichotomy for local small-bias generators. In Ronald Cramer, editor, *TCC 2012*, volume 7194 of *LNCS*, pages 600–617. Springer, Heidelberg, March 2012.

[Agr19] Shweta Agrawal. Indistinguishability obfuscation without multilinear maps: New methods for bootstrapping and instantiation. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part I*, volume 11476 of *LNCS*, pages 191–225. Springer, Heidelberg, May 2019.

[AIK07] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography with constant input locality. In Alfred Menezes, editor, *CRYPTO 2007*, volume 4622 of *LNCS*, pages 92–110. Springer, Heidelberg, August 2007.

[AJL+19] Prabhanjan Ananth, Aayush Jain, Huijia Lin, Christian Matt, and Amit Sahai. Indistinguishability obfuscation without multilinear maps: New paradigms via low degree weak pseudorandomness and security amplification. In Alexandra Boldyreva and Daniele Micciancio, editors, *CRYPTO 2019, Part III*, volume 11694 of *LNCS*, pages 284–332. Springer, Heidelberg, August 2019.

[AL16] Benny Applebaum and Shachar Lovett. Algebraic attacks against random local functions and their countermeasures. In Daniel Wichs and Yishay Mansour, editors, *48th ACM STOC*, pages 1087–1100. ACM Press, June 2016.

[BBKK18] Boaz Barak, Zvika Brakerski, Ilan Komargodski, and Pravesh K. Kothari. Limits on low-degree pseudorandom generators (or: Sum-of-squares meets program obfuscation). In Jesper Buus Nielsen and Vincent Rijmen, editors, *EUROCRYPT 2018, Part II*, volume 10821 of *LNCS*, pages 649–679. Springer, Heidelberg, April / May 2018.

[BDGM20] Zvika Brakerski, Nico Döttling, Sanjam Garg, and Giulio Malavolta. Factoring and pairings are not necessary for io: Circular-secure LWE suffices. *IACR Cryptol. ePrint Arch.*, page 1024, 2020.

[BGI+01] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In Joe Kilian, editor, *CRYPTO 2001*, volume 2139 of *LNCS*, pages 1–18. Springer, Heidelberg, August 2001.

[DQV+21] Lalita Devadas, Willy Quach, Vinod Vaikuntanathan, Hoeteck Wee, and Daniel Wichs. Succinct lwe sampling, random polynomials, and obfuscation. *Cryptology ePrint Archive*, 2021.

[GGH+13] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *54th FOCS*, pages 40–49. IEEE Computer Society Press, October 2013.

[GJLS21] Romain Gay, Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from simple-to-state hard problems: New assumptions, new techniques, and simplification. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part III*, volume 12698 of *Lecture Notes in Computer Science*, pages 97–126. Springer, 2021.

[Gol00] Oded Goldreich. Candidate one-way functions based on expander graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 7(90), 2000.

[GP21] Romain Gay and Rafael Pass. Indistinguishability obfuscation from circular security. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 736–749, 2021.

[GR10] Shafi Goldwasser and Guy N. Rothblum. Securing computation against continuous leakage. In Tal Rabin, editor, *CRYPTO 2010*, volume 6223 of *LNCS*, pages 59–79. Springer, Heidelberg, August 2010.

[Gri01] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[HK22] Tim Hsieh and Pravesh Kothari. Algorithmic thresholds for refuting random polynomial systems. In *SODA*, 2022.

[Jai19] Aayush Jain. Public talk: Evidence for resilient generators. New Roads to Cryptopia, CRYPTO, 2019. `https://crypto.iacr.org/2019/affevents/nrc/page.html`.

[JLMS19] Aayush Jain, Huijia Lin, Christian Matt, and Amit Sahai. How to leverage hardness of constant-degree expanding polynomials overa $\mathbb{R}$ to build $i\mathcal{O}$. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part I*, volume 11476 of *LNCS*, pages 251–281. Springer, Heidelberg, May 2019.

[JLS19] Aayush Jain, Huijia Lin, and Amit Sahai. Simplifying constructions and assumptions for $i\mathcal{O}$. Cryptology ePrint Archive, Report 2019/1252, 2019. `https://eprint.iacr.org/2019/1252`.

[JLS21a] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from lpn over f_p, dlin and prgs in nc^0. Cryptology ePrint Archive, Report 2021/1334, 2021. `https://eprint.iacr.org/2019/1334`.

[JLS21b] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from well-founded assumptions. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 60–73. ACM, 2021.

[KMOW17] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *49th ACM STOC*, pages 132–145. ACM Press, June 2017.

[KS98] Aviad Kipnis and Adi Shamir. Cryptanalysis of the oil & vinegar signature scheme. In Hugo Krawczyk, editor, *CRYPTO'98*, volume 1462 of *LNCS*, pages 257–266. Springer, Heidelberg, August 1998.

[KS99] Aviad Kipnis and Adi Shamir. Cryptanalysis of the HFE public key cryptosystem by relinearization. In Michael J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 19–30. Springer, Heidelberg, August 1999.

[LT17] Huijia Lin and Stefano Tessaro. Indistinguishability obfuscation from trilinear maps and block-wise local PRGs. In Jonathan Katz and Hovav Shacham, editors, *CRYPTO 2017, Part I*, volume 10401 of *LNCS*, pages 630–660. Springer, Heidelberg, August 2017.

[LV17] Alex Lombardi and Vinod Vaikuntanathan. Limits on the locality of pseudorandom generators and applications to indistinguishability obfuscation. In Yael Kalai and Leonid Reyzin, editors, *TCC 2017, Part I*, volume 10677 of *LNCS*, pages 119–137. Springer, Heidelberg, November 2017.

[MST03] Elchanan Mossel, Amir Shpilka, and Luca Trevisan. On e-biased generators in NC0. In *44th FOCS*, pages 136–145. IEEE Computer Society Press, October 2003.

[OW14] Ryan O'Donnell and David Witmer. Goldreich's PRG: evidence for near-optimal polynomial stretch. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 1–12, 2014.

[Sch08] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-CSPs. In *49th FOCS*, pages 593–602. IEEE Computer Society Press, October 2008.

[WW21] Hoeteck Wee and Daniel Wichs. Candidate obfuscation via oblivious lwe sampling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 127–156. Springer, 2021.