UNIVERSITY OF CALIFORNIA

Los Angeles

Limits on the Pseudorandomness of Low-Degree Polynomials over the Integers

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Computer Science

by

Alexis Lei Wan Korb

2020

ABSTRACT OF THE THESIS

Limits on the Pseudorandomness

of Low-Degree Polynomials over the Integers

by

Alexis Lei Wan Korb

Master of Science in Computer Science

University of California, Los Angeles, 2020

Professor Amit Sahai, Chair

We initiate the study of a problem called the Polynomial Independence Distinguishing Problem (PIDP). The problem is parameterized by a set of polynomials $\mathcal{Q} = (q_1, \ldots, q_m)$ of n variables 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.

The thesis of Alexis Lei Wan Korb is approved.

Rafail Ostrovsky

Alexander Sherstov

Amit Sahai, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This thesis is based on a joint work of the same title that I co-authored with Amit Sahai, Aayush Jain, and Paul Lou. The topic was first proposed by Amit Sahai and Aayush Jain due to its relation to their work on using low complexity pseudorandom generators to construct $i\mathcal{O}$. We all worked together to formally define the polynomial independence distinguishing problem and produce all results found in this paper. Apart from Appendix A, which was written by Amit Sahai and Aayush Jain, we all collaborated in writing and editing all sections of the original version of the manuscript. Though I have not added new results nor have I changed the overall organization from the original version, I have since edited the original manuscript for clarity and revised it into its current form here. The original manuscript is currently in preparation for submission.

We also gratefully thank Boaz Barak for several illuminating conversations about estimating features of inputs based on observations of random polynomial evaluations. We also thank both Pravesh Kothari and Rachel Lin for their encouragement of this work.

Research for this thesis was supported in part through funding from DARPA SAFEWARE and SIEVE awards, NTT Research, NSF Frontier Award 1413955, and NSF grant 1619348, BSF grant 2012378, a Xerox Faculty Research Award, a Google Faculty Research Award, an equipment grant from Intel, and an Okawa Foundation Research Grant. This material is based upon work supported by the Defense Advanced Research Projects Agency through the ARL under Contract W911NF-15-C-0205. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense, the National Science Foundation, NTT Research, or the U.S. Government.

1 Introduction

In this work, we consider the following problem:

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

Distribution 1:	**Distribution 2:**
1. Sample $\mathbf{x} \xleftarrow{R} \mathcal{D}_n^*$	1. Sample $\mathbf{x}_1, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}_n^*$
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). Similarly, the problem of certifying that no such $\mathbf{x}$ exists when given the output 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 to 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 collection of low degree multivariate polynomials 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 $(\mathcal{D}, \mathcal{F}, n, m)$ where m is called the stretch of the PIDG and

- $\mathcal{D}$ is an efficiently samplable distribution over $\mathbb{R}$.
- $\mathcal{F} = \{f_i\}_{i=1}^m$ where each f_i for $i \in [m]$ is a polynomial time 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:

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

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

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 $\mathcal{F}$ 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 $\mathcal{F}$ 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 using any standard Boolean PRG.

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

- **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, BHJ+19] towards a major problem in cryptography - Indistinguishability Obfuscation ($i\mathcal{O}$) [BGI+01, GR10, GGH+13]. Indistinguishability Obfuscation has had far-reaching consequences in cryptography and beyond (see, e.g., [BFM14, GGG+14, HSW13, SW14, KLW15, BPR15, CHN+16, GPS16, HJK+16]), including playing a pivotal role in establishing the hardness of Nash Equilibrium by creating provably hard instances for a PPAD complete problem called the End-of-Line EOL Problem [BPR15, GPS16, CHK+19]. Our results provide greater insight into the core objects that underlie constructions of $i\mathcal{O}$. See Appendix A for further discussion.

1.1 Our Results

We show that for certain classes of polynomials and input distributions, we can build distinguishers for the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$. Note that the existence of such distinguishers implies that these classes of polynomials and input distributions cannot form secure PIDGs. 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 distributions of the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$ 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 $(\mathcal{D}, \mathcal{Q}, n, m)$-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 $(\mathcal{D}, \mathcal{Q}, n, m)$-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. Such distributions, which we call weakly nice distributions, are distributions that are intuitively well spread and symmetric around 0. We formalize this by requiring all odd moments of the distribution $\mathcal{D}$ 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:

- **Expander-Based Polynomials**: We consider the set of constant degree multilinear polynomials where the monomials satisfy an expansion criteria. Namely, the expansion criteria, formally defined in Definition 5.3, captures the idea that the set of coefficients 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*

[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.

weakly-nice distribution with bounded support in $[-\beta, \beta]$ *for* $\beta = n^{O(1)}$. *If* $m > n$, *then there exists a probabilistic polynomial time algorithm can solve the* $(\mathcal{D}, \mathcal{Q}, n, m)$-PIDP *with probability at least* $\Omega(n^{-O(1)})$.

- **Polynomials with Non-negative Coefficients**: We also consider the set of constant degree multilinear polynomials with non-negative coefficients $\mathcal{Q}_{n,nonneg} \subseteq \mathbb{Z}[x_1, \ldots, x_n]$[2], obtaining:

  Theorem 1.2. *(Informal.) Let* $\mathcal{Q} = \{q_1, \ldots, q_m\} \in Q_{n,nonneg} \subseteq \mathbb{Z}[x_1, ..., x_n]$ *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* $(\mathcal{D}, \mathcal{Q}, n, m)$-PIDP *with probability at least* $\Omega(n^{-O(1)})$.

We note that both our results correspond to *worst-case* properties that are identifiable 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: 1) The fourth moment is required to be sufficiently greater than

[2]Our results also extend to polynomials over the reals provided that the values of the coefficients of the polynomials are at least $\Omega(n^{-O(1)})$.

the square of the second moment; 2) it is required to take values within a bound that is roughly polylogarithmic in the second moment; and 3) 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. *(Informal.) Let d be any constant degree, and let $p > n \log n / \binom{N}{d}$. Let $\mathcal{D}$ 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 $(\mathcal{D}, \mathcal{Q}_{n,d,p}, n, m) - \mathsf{PIDP}$ 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, Jai19]. 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. Recall that our objective 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 the observation that polynomials with shared variables may exhibit detectable correlations when evaluated over natural distributions over the integers instead of over uniform Boolean values. 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 is an i.i.d. random variable with probability distribution $\mathcal{D}$. Now, if $\mathcal{D}$ is 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, despite the above discrepancy, both cases will still yield the same overall expectation. Therefore, 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 $\mathbb{E}[X_2^2] \neq 0$. As we will later show in Lemma 4.1, such conditions are reasonable for symmetric mean zero distributions over integers. 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 is a scaled Boolean.

Polynomials. The $(\mathcal{D}, \mathcal{Q}, n, m)$-polynomial independence distinguishing problem can be studied for any set of multivariate polynomials and any input distribution 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, namely $\{q_i(\mathbf{x}) = x_i\}_{i\in[m]}$, 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 $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$ 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 that for a random variable

Z over our distribution $\mathcal{D}$, then $\mathbb{E}[Z^4]/(\mathbb{E}[Z^2])^2 \geq \gamma$ for some $\gamma > 1$ and $\mathbb{E}[Z^2] \geq \eta$ for some $\eta > 0$. 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 $\mathcal{D}_n^*$ sample an n-tuple of values each independently drawn from $\mathcal{D}$. Let X be a random variable on distribution $\mathcal{D}_n^*$. 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 and q_j is large, then this expectation difference is also large. Note that in the $(\mathcal{D}, \mathcal{Q}, n, m)-\mathsf{PIDP}$ problem, we either get samples of the form $\{q_i, y_i = q_i(\mathbf{x})\}_{i\in[m]}$ where $\mathbb{E}[Y_i^2Y_j^2] = \mathbb{E}[q_i^2(X)q_j^2(X)]$ or samples of the form $\{q_i, y_i = q_i(\mathbf{x}_i)\}_{i\in[m]}$ where $\mathbb{E}[Y_i^2Y_j^2] = \mathbb{E}[q_i^2(X)]\,\mathbb{E}[q_j^2(X)]$. Here, we use random variables Y_i to correspond to the samples y_i received. Thus, the covariance is equal to the difference in the expectation of the distribution of $Y_i^2Y_j^2$ when getting evaluations on the same input and the expectation of the distribution of $Y_i^2Y_j^2$ when getting evaluations on independent inputs. To build a distinguisher to solve the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$, 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 $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$: We simply compute the product of the samples $y_i^2 y_j^2$ 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. Roughly, this lemma 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 expectations 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 $[0, p]$.*[3] *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

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

[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 $p/2$.

To prove this lemma, we construct a simple distinguisher. The distinguisher first partitions the support of the two distributions into some ϵ-width intervals. Then, the distinguisher 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. This allows the distinguisher to estimate the probability for each interval and each distribution that the distribution falls within that interval. A Chernoff bound combined with a union bound ensures that these estimated interval probabilities do not differ too much from the actual probabilities.

Then, the distinguisher uses the histogram to make its decisions for any input x by choosing the distribution with a larger estimated probability of producing a value in the same interval as x. To provide a lower bound on the distinguishing probability, we show that there exists an interval where the following occurs:

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 let X_0 be a random variable distributed according to $\mathcal{D}_0$ and X_1 be a random variable distributed according to $\mathcal{D}_1$. Suppose*

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

Then, if $\{I_i\}_{i=1}^n$ is a partition of $[0, p]$ into equal-sized intervals for $n = \frac{2p}{q}$, 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.

The existence of such an interval allows us to form a lower bound for the distinguishing probability through a careful argument involving the aforementioned partitioning and accuracy guarantees

given by a Chernoff bound combined with a union bound.

Covariance Guarantee. We now look 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 $\mathcal{D}_n^*$. 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

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

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

Recall that we want to form a lower bound on this expectation difference. Let us consider any single term $(\mathbb{E}[X_S X_T X_U X_V] - \mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V])$. First, we will show that this value is always non-negative. Now, since $\mathcal{D}$ is symmetric, all odd moments of each input variable X_i are zero. Consider the following three cases:

1. $X_S X_T X_U X_V$ is a square, but one of $X_S X_T$ or $X_U X_V$ is not a square. Observe that $\mathbb{E}[X_S X_T]\,\mathbb{E}[X_U X_V] = 0$ since the odd moments of the input variables are 0. Therefore, the difference is non-negative, since the expectation of a square is always non-negative.

2. $X_S X_T$ and $X_U X_V$ are both squares. Then, the degree of all variables in $X_S X_T$ and $X_U X_V$ 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. The magnitude of this difference is determined by γ, η, and the amount of overlap between the polynomials.

3. $X_S X_T X_U X_V$ 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 non-negative, then all summation terms are non-negative, so such a cancellation does not occur. However, we also show another set of conditions, which we call Expander Based Coefficients, that is sufficient to ensure this.

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 non-negative. As stated above, 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 an 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. Suppose without loss of generality that $S = U$. But since we assumed that $c_S, c_T, d_U, d_V \neq 0$, this means that c_S and $d_U = d_S$ are both nonzero. 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, which lets us 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 for certain parameter settings of the $(\mathcal{D}, \mathcal{Q}_{n,p,d}, n, m) - \mathsf{PIDP}$ where $\mathcal{D}$ is some nice input distribution and $\mathcal{Q}_{n,p,d}$ is the natural random family of polynomials described in Section 1.1 of the Introduction. In this setting, we are given polynomials $\{q_i\}_{i\in[m]}$ sampled from $\mathcal{Q}_{n,p,d}$ 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 distribution $\mathcal{D}_n^*$, the product distribution of $\mathcal{D}$, as defined in Definition 1.1. 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 $\mathcal{D}_n^*$ 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 $\mathcal{D}_n^*$. Let $X, \{X_i\}_{i=1}^m$ be random variables with distribution $\mathcal{D}_n^*$. We will let Y_i denote the random variable $q(X)$ which is a function of random variable X and the implicit random variables representing the coefficients of a polynomial q sampled from $\mathcal{Q}_{n,p,d}$. Similarly, we will let Y_i' denote the random variable $q(X_i)$ which is a function of random variable X_i and the implicit random variables representing the coefficients of q.

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 the ratio of the fourth moment to the square of the second moment of a centered Gaussian distribution 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

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

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

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

This follows because the odd moments of every coefficient variable are 0. Let $N = \binom{n}{d}$. Then, the numerator becomes

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

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

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

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:

$$g_i = \Pr_{S_1 \neq S_2}\left[|S_1 \cap S_2| = i\right].$$

This means that,

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

This means that the numerator is

$$\mathop{\mathbb{E}}_{q_1, X}[q_1^4(X)] = Np\gamma\gamma^d + 3p^2 N(N-1)\left((1 - g_1 - \ldots - g_{d-1}) + \gamma g_1 + \ldots + \gamma^{d-1} g_{d-1}\right)$$

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

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

Therefore,

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

Setting $t = Np$ as the average density of each polynomial (number of non-zero coefficients) and observing that $g_i \approx \theta(1/n^i)$ 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}

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

This will give us

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

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} which we compute 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, when given 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 analyze $\frac{1/m \sum_i v^4_i}{2/m \sum_{i\in[m/2]} v^2_{2i-1} \cdot v^2_{2i-1}}$. In general, analyzing the ratio of this form may not be an easy task as the expected ratio of a quantity is in general not the ratio of expectations. Thus, we analyze a slightly different objective. Define

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

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

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 lowercase 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 distribution D_1 is computationally indistinguishable from distribution D_2, 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 n, m be parameters. Let $\mathcal{Q} = \{q_1, ..., q_m\}$ denote a set of m multivariate polynomials $q_i : \mathbb{R}^n \to \mathbb{R}$. Let $\mathcal{D}$ be a distribution on $\mathbb{R}$, and let $\mathcal{D}_n^*$ be the distribution $\underbrace{\mathcal{D} \times \cdots \times \mathcal{D}}_{n \text{ times}}$ over $\mathbb{R}^n$ where $\mathbf{x} = (x_1, \ldots, x_n) \xleftarrow{R} \mathcal{D}_n^*$ means $x_1, \ldots, x_n$ are independently sampled from $\mathcal{D}$. The Polynomial Independence Distinguishing Problem with respect to $\mathcal{D}, \mathcal{Q}, n, m$ (or simply the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$) consists of distinguishing the following two distributions:

Distribution 1:	**Distribution 2:**
1. Sample $\mathbf{x} \xleftarrow{R} \mathcal{D}_n^*$	1. Sample $\mathbf{x}_1, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}_n^*$
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 overload notation and use $\mathcal{Q}$ to instead denote a distribution over some family of polynomials. In this case, the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$ consists of distinguishing the following two distributions:

Distribution 1*:	**Distribution 2*:**
0. Sample $q_1, \ldots, q_m \xleftarrow{R} \mathcal{Q}$.	0. Sample $q_1, \ldots, q_m \xleftarrow{R} \mathcal{Q}$
1. Sample $\mathbf{x} \xleftarrow{R} \mathcal{D}_n^*$	1. Sample $\mathbf{x}_1, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}_n^*$
2. Output $\{q_i, q_i(\mathbf{x})\}_{i\in[m]}$	2. Output $\{q_i, q_i(\mathbf{x}_i)\}_{i\in[m]}$

Remark 3.3. We say that an algorithm $\mathcal{A}$ solves the $(\mathcal{D}, \mathcal{Q}, n, m)-\mathsf{PIDP}$ with probability p if $\mathcal{A}$ can distinguish between Distribution 1 and Distribution 2 of the $(\mathcal{D}, \mathcal{Q}, n, m) - \mathsf{PIDP}$ 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 $(\mathcal{D}, \mathcal{F}, n, m)$ where m is called the stretch of the PIDG and

- $\mathcal{D}_n^*$ defined with respect to $\mathcal{D}$ as in Definition 3.4 above is a t-samplable distribution over $\mathbb{R}^n$ where $t = n^{O(1)}$.
- $\mathcal{F} = \{f_i\}_{i=1}^m$ where 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:

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

$$(\mathcal{F}, \{f_i(\mathbf{x})\}_{i=1}^m) \approx_c (\mathcal{F}, \{f_i(\mathbf{x}_i)\}_{i=1}^m))$$

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 $\mathcal{D}_n^*$.

Remark 3.4. If there exists a probabilistic polynomial time algorithm $\mathcal{A}$ that solves the $(\mathcal{D}, \mathcal{Q}, n, m)-$ PIDP with non-negligible probability, then $(\mathcal{D}, \mathcal{Q}, n, m)$ 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. We call a distribution $\mathcal{D}$ a hyper-expanding distribution if any random variable with distribution $\mathcal{D}$ is a hyper-expanding random variable.

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 non-negative 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. For some $\{c_i\}_{i=1}^{n}$ such that $0 \leq c_i \leq 4$ for all $i \in [n]$, then

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

If $X_S X_T X_U X_V$ contains a variable X_i of odd power (i.e. if any c_i is odd), then by Remark 3.6, $\mathbb{E}[X_S X_T X_U X_V] = 0$. Otherwise, each $c_i \in \{0, 2, 4\}$. Now, $c_i = 4$ if and only if X_i appears in each of X_S, X_U, X_V, X_T. Define

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

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

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

Define $b = |\{i \mid c_i = 2\}| = \frac{1}{2}(|S| + |T| + |U| + |V|) - 2a$. Therefore,

$$\mathbb{E}[X_S X_T X_U X_V] = \prod_{i=1}^{n} \mu_{c_i} = \mu_4^{|a|} \mu_2^{|b|}$$

□

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 $\Pr[X = k] = \max_{i=a}^{b}(\Pr[X = i])$

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

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

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,

$$var[X] = \sum_{i=-\beta}^{\beta} \left(\Pr[X = i](X - \mathbb{E}[X])^2\right)$$

$$\geq \Pr[X = y_1](y_1 - \mathbb{E}[X])^2 + (1 - \Pr[X = y_1])(y_2 - \mathbb{E}[X])^2$$

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

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

Which means that

$$var[X] \geq \Pr[|X| = \mathsf{mode}(|X|)](y_1 - \mathbb{E}[X])^2 + (1 - \Pr[|X| = \mathsf{mode}(|X|)])(y_2 - \mathbb{E}[X])^2$$

$$= (1 - \Pr[|X| \neq \mathsf{mode}(|X|)])(y_1 - \mathbb{E}[X])^2 + \Pr[|X| \neq \mathsf{mode}(|X|)](y_2 - \mathbb{E}[X])^2$$

To continue the proof, we will first prove the following claim.

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

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

$$xa^2 + (1 - x)b^2 \geq pa^2 + (1 - c)b^2$$

$$\geq \min(p, 1 - c)(a^2 + b^2)$$

$$\geq \frac{1}{2}\min(p, 1 - c)$$

By applying this claim to $a = |y_2 - \mathbb{E}[X]^2|$, $b = |y_1 - \mathbb{E}[X]^2|$, $\frac{1}{t} \leq x = \Pr[|X| \neq mode(|X|)] \leq$

$(1 - \frac{1}{\beta+1})$, 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 $\Pr[|X| = i] = \Pr[X^2 = i^2]$, and $\mathsf{mode}(|X|))^2 = \mathsf{mode}(X^2)$. Therefore, $\Pr[X^2 \neq \mathsf{mode}(X^2)] = \Pr[|X| \neq \mathsf{mode}(|X|)]$ so that

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

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,

$$var[X^2] = \sum_{i=0}^{\beta^2} \left(\Pr[X^2 = i](X^2 - \mathbb{E}[X^2])^2\right)$$

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

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 $\Pr[|X| \neq \mathsf{mode}(|X|)] \geq \frac{1}{t}$ 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 $[0, p]$.[4] Let X_0 be a random variable on $\mathcal{D}_0$ and X_1 be a random variable on $\mathcal{D}_1$. Suppose*

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

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

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

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 expectation 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 $a_{i-1} = \inf I_i$

[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 $p/2$.

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

We 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 $(\mathcal{D}, \mathcal{Q}, n, m) -$ PIDP with non-trivial probability. We then build such non-trivial distinguishers. 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 the specified classes. This implies that we cannot form any secure PIDGs with spread $m > n$ out of certain classes of polynomials and distributions. 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 $X = (X_1, \ldots, X_m)$, $Y = (Y_1, \ldots, Y_m)$, and 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 $(\mathcal{D}, \mathcal{Q}, n, m) -$ PIDP with noticeable probability.

5.1 An 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, can distinguish between the two distributions with probability proportional to the expectation difference and the bound.

This algorithm works by partitioning the support of the two distributions into sufficiently wide intervals. Then, the algorithm creates an approximate histogram of each of the two distributions by randomly sampling from each distribution a sufficient number of times. When given a value from some interval, the algorithm guesses that the value came from the distribution which, according to the approximate histograms, has a higher probability of landing in that interval.

Lemma 5.1. *Let p, q be two parameters. Let D_0 and D_1 be distributions with bounded support in $[0, p]$.*[5] *Let X_0 be a random variable on $\mathcal{D}_0$ and X_1 be a random variable on $\mathcal{D}_1$. If*

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

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

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

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 $\mathcal{D}_1$.

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

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 $\frac{q}{2}$
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 $r_{0,i} < r_{1,i}$ and 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.

Proof. To prove this, we will first use a Chernoff bound to show that the sample histograms of the distributions do not differ too much from the actual distributions. Then, we use Lemma 4.2 to claim that there exists some interval where the two distributions differ by a large enough amount that our algorithm will succeed with sufficient probability.

Partition $[0, p]$ into $n = \frac{2p}{q}$ disjoint intervals $\{I_i\}_{i\in[n]}$ each of width $\frac{q}{2}$, and let $a_i = \sup I_i$ 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.

Note 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:

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

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:

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

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:

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

□

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 $(\mathcal{D}, \mathcal{Q}, n, m)$-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 $\{q_i(\mathbf{x}_i)\}_{i=1}^m$ and $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \xleftarrow{R} \mathcal{D}$.

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

$$\left|\Pr\left[\mathcal{A}(\mathcal{Q}, \{q_i(\mathbf{x})\}_{i=1}^m) = 1\right] - \Pr\left[\mathcal{A}(\mathcal{Q}, \{q_i(\mathbf{x}_i)\}_{i=1}^m)\right]\right|$$

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

Therefore $\mathcal{A}$ is a probabilistic algorithm that solves the $(\mathcal{D}, \mathcal{Q}, n, m)$-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

Now, we will show that for a certain set of polynomials and distributions, we can find a probabilistic polynomial time algorithm that solves the $(\mathcal{D}, \mathcal{Q}, n, m)-\mathsf{PIDP}$ with non-negligible probability. 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 non-negative coefficients

Lemma 5.2. *Let n, m be parameters. 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

non-negative, 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 $\mu_2 \geq \eta > 0$. Therefore,

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

□

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 $(\mathcal{D}, \mathcal{Q}, n, m)$-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 $(\mathcal{D}, \mathcal{Q}, n, m)$ 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 $(\mathcal{D}, \mathcal{Q}, n, m)$ 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 Non-trivial 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 $(\mathcal{D}, \mathcal{Q}, n, m) -$ PIDP 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
- $\{S \in \mathcal{P}([n]) \mid c_{S,i} \neq 0 \text{ for some } i \in [m]\}$ 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 an* 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,

$$q_i(\mathbf{x}) = c_{S^*} x_{S^*} + p_i(\mathbf{x})$$

$$q_j(\mathbf{x}) = p_j(\mathbf{x})$$

where

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

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

Now,

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

On the other hand,

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

Therefore,

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

We will consider each term separately. First,

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

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,

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

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

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

For the next term, we have

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

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

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

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,

$$\mathop{\mathbb{E}}_{X}\left[X_{S^*}p_i(X)p_j^2(X)\right] - \mathop{\mathbb{E}}_{X}\left[X_{S^*}p_i(X)\right]\mathop{\mathbb{E}}_{X}\left[p_j^2(X)\right] = 0$$

For the last term,

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

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,

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

So we have that

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

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. Suppose without loss of generality that $S = U$. But since we assumed that $c_S, c_T, d_U, d_V \neq 0$, this means that c_S and $d_S = d_U$ are both nonzero. 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 $X_S X_T X_U X_V = X_S^2 X_U X_V$ 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 $U = V$.

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

As a result,

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

□

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 $(\mathcal{D}, \mathcal{Q}, n, m)$-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 $(\mathcal{D}, \mathcal{Q}, n, m)$ 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 $(\mathcal{D}, \mathcal{Q}, n, m)$ 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

We now show an efficient algorithm that solves the $(\mathcal{D}, \mathcal{Q}, n, m) -$ PIDP with overwhelming probability for natural random classes of homogeneous multilinear constant degree polynomials $\mathcal{Q}$ and natural input distributions $\mathcal{D}$. We note that in this section, we consider the distinguishing probability over polynomials randomly chosen from our class $\mathcal{Q}$ as opposed to over worst-case selections of polynomials. 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) $\Pr[x \xleftarrow{R} \mathcal{D}, |x| > \epsilon] > \Omega(1)$

Problem Setup. Let m be the number of polynomials, n be the number of variables, and d be the constant degree of each polynomial. Let $p, \gamma_I, \gamma_c, C_I, C_c, \epsilon_I, \epsilon_c$ be a set of parameters. We now describe the input and polynomial distributions as follows:

- **Input Variable Distribution** $\mathcal{D}_{\mathsf{Inp}}$: Let $\mathcal{D}_{\mathsf{Inp}}$ be a $(\gamma_I, C_I, \epsilon_I)-$nice distribution. If X_i is a random variable over $\mathcal{D}_{\mathsf{Inp}}$, observe that all odd moments of X_i are 0, $\mathbb{E}[X_i^2] = 1$, and $\mathbb{E}[X_i^4] = \gamma_I$.
- **Input Distribution** $\mathcal{D}^*_{\mathsf{Inp,n}}$: Let $\mathcal{D}^*_{\mathsf{Inp,n}}$ be the distribution $\underbrace{\mathcal{D}_{\mathsf{Inp}} \times \cdots \times \mathcal{D}_{\mathsf{Inp}}}_{n \text{ times}}$ where $\mathbf{x} = (x_1, \ldots, x_n) \xleftarrow{R} \mathcal{D}^*_{\mathsf{Inp,n}}$ means $x_1, \ldots, x_n$ are independently sampled from $\mathcal{D}_{\mathsf{Inp}}$. Inputs to the polynomials are sampled from $\mathcal{D}^*_{\mathsf{Inp,n}}$.
- **Coefficient Distribution** $\mathcal{D}_{\mathsf{Coeff,p}}$: Let $\mathcal{D}_{\mathsf{Coeff}}$ denote a $(\gamma_c, C_c, \epsilon_c)-$nice distribution. Then, for parameter $p \in [0, 1]$, let $\mathcal{D}_{\mathsf{Coeff,p}}$ be the distribution that outputs 0 with probability $1 - p$ and samples from $\mathcal{D}_{\mathsf{Coeff}}$ with probability p. If Z is a random variable over $\mathcal{D}_{\mathsf{Coeff,p}}$, observe that all odd moments of Z are 0, $\mathbb{E}[Z^2] = p$, and $\mathbb{E}[Z^4] = \gamma_c p$.
- **Polynomial Distribution** $\mathcal{Q}_{n,d,p}$: We define $\mathcal{Q}_{n,d,p}$ to be the distribution of polynomials

such that for $q(\mathbf{x})$ sampled from $\mathcal{Q}_{n,d,p}$, then

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

where each c_S is sampled independently from $\mathcal{D}_{\mathsf{Coeff,p}}$. We generate m polynomials by independently sampling each polynomial from $\mathcal{Q}_{n,d,p}$

The problem we are interested in is the $(\mathcal{D}_{\mathsf{lnp}}, \mathcal{Q}_{n,d,p}, n, m)-$Polynomial Independence Distinguishing Problem with respect to $\mathcal{D}_{\mathsf{lnp}}$ and $\mathcal{Q}_{n,d,p}$ as defined above with parameters $d, p, \gamma_I, \gamma_c, C_I, C_c, \epsilon_I, \epsilon_c$. We will now show that for certain values of these parameters, we can obtain an overwhelming distinguisher.

Theorem 6.1. *Let $n, m, d, p, \gamma_I, \gamma_c, C_I, C_c, \epsilon_I, \epsilon_c$ be parameters where $d \geq 2$ is an integer constant, $\gamma_I, \gamma_c, \epsilon_I = \theta(1)$, $\gamma_I > 1$, $p = \Omega(n \log n \cdot C_I^{4d}/\binom{n}{d})$, $p < \gamma_c/3$, and $m = \Omega(n^2 C_I^{8d} C_c^8 \log^{10} n)$. Then, Algorithm 3 is an overwhelming distinguisher for the $(\mathcal{D}_{\mathsf{lnp}}, \mathcal{Q}_{n,d,p}, n, m)-\mathsf{PIDP}$ problem with respect to $\mathcal{D}_{\mathsf{lnp}}$ and $\mathcal{Q}_{n,d,p}$ as defined above for these parameters.*

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

Given: Polynomials $\{q_i\}_{i=1}^m$ sampled from $\mathcal{Q}_{n,d,p}$, along with evaluations $\{y_i\}_{i\in[m]}$ where either

- $y_i = q_i(\mathbf{x})$ for a single $\mathbf{x}$ sampled according to $\mathcal{D}^*_{\mathsf{lnp,n}}$ (denoted by the event same),
- or $y_i = q_i(\mathbf{x}_i)$ for independent $\mathbf{x}_i$ sampled from $\mathcal{D}^*_{\mathsf{lnp,n}}$ (denoted by the event diff).

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 define α_{th} as:

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

Define α_{same} as:

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

Define α_{diff} as:

$$\alpha_{\text{diff}} = \frac{\mathbb{E}_{q_a, X_a}[q_a^4(X_a)]}{\mathbb{E}_{q_a, q_b, X_a, X_b}[q_a^2(X_a) \cdot q_b^2(X_b)]}$$

where the expectations are taken over random variables X, X_a, X_b with distribution $\mathcal{D}^*_{\mathsf{Inp},\mathsf{n}}$ and the random variables of the coefficients of q_a, q_b sampled from distribution $\mathcal{D}_{\mathsf{Coeff},\mathsf{p}}$.

We will prove correctness of the algorithm through a series of lemmas. Refer to section 2.2 in the Technical Overview for a proof overview and for further intuition on the proof. Then, we will analyze the algorithm's running time. But, first we define some notation.

Definition 6.3. Throughout this section, we will define $N = \binom{n}{d}$ and $t = Np$. For a homogeneous degree d polynomial, N denotes the number of possible monomials. If each monomial is present in the polynomial with probability p, then t denotes the expected number of monomials in the polynomial. In particular, t is the expected density of a polynomial q sampled from $\mathcal{Q}_{n,d,p}$.

Notation.

- The notation $\sum_S$ is shorthand for $\sum_{S \in \mathcal{P}[n]; |S|=d}$.
- The notation $\sum_{S_1 \neq S_2}$ is shorthand for $\sum_{S_1, S_2 \in \mathcal{P}([n]); |S_1|=|S_2|=d; S_1 \neq S_2}$. The notation $\mathbb{E}_{S_1 \neq S_2}$ and $\Pr_{S_1 \neq S_2}$ are shorthand for the expectation and probability respectively over two randomly chosen sets S_1, S_2 satisfying these constraints.
- Random variables X over $\mathcal{D}^*_{\mathsf{lnp},\mathsf{n}}$ are implicitly represented as tuples of random variables $X = (X_1, \ldots, X_n)$ where each variable X_i has distribution $\mathcal{D}_{\mathsf{lnp}}$. Recall that for a set $S \in \mathcal{P}([n])$, we define $X_S = \prod_{i \in S} X_i$.
- For polynomials of the form $q(\mathbf{x}) = \sum_S c_S x_S$, we will overload notation and use c_S to represent both a specific coefficient sampled from $\mathcal{D}_{\mathsf{Coeff},\mathsf{p}}$ and a random variable with distribution $\mathcal{D}_{\mathsf{Coeff},\mathsf{p}}$. Similarly, we will overload notation and use q to represent both a specific polynomial sampled from $\mathcal{Q}_{n,d,p}$ and as the implicit set of random variables $\{c_S\}$ representing the coefficients of q.
- For example, $\mathbb{E}_{X,q}[q(X)]$ denotes $\mathbb{E}_{X,\{c_S\}}[\sum_S c_S X_S] = \mathbb{E}_{\{X_i\},\{c_S\}}[\sum_S c_S \prod_{i \in S} X_i]$.

Our first goal is to show that α_{same} and α_{diff} differ by at least $\Omega(1/n)$. Our next several lemmas will accomplish this.

Lemma 6.1. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3 and, in particular, since $p < \gamma_c/3$ and $\gamma_I, \gamma_c = \theta(1)$, then*

$$\alpha_{\text{same}} = 3 + \theta(\frac{\gamma_c \gamma_I^d}{t}).$$

Proof. Recall that

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

Let us represent polynomials q_a and q_b as $q_a(\mathbf{x}) = \sum_S c_S x_S$ and $q_b(\mathbf{x}) = \sum_S d_S x_S$ where the coefficients are sampled from $\mathcal{D}_{\mathsf{Coeff,p}}$ and inputs are sampled from $\mathcal{D}_{\mathsf{Inp}}$. Now we compute the numerator.

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

The second equality follows because the odd moments of every coefficient variable are 0. Now, let $N = \binom{n}{d}$. Then, the numerator becomes

$$\mathop{\mathbb{E}}_{q_a,X}[q_a^4(X)] = Np\gamma_c \mathop{\mathbb{E}}_{X}[X_S^4] + 3p^2 \sum_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1}^2 X_{S_2}^2]$$

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

$$\mathop{\mathbb{E}}_{q_a,X}[q_a^4(X)] = Np\gamma_c\gamma_I^d + 3p^2 N(N-1) \mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} X_{S_2}]$$

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:

$$g_i = \Pr_{S_1 \neq S_2} \left[|S_1 \cap S_2| = i \right].$$

Since for each $j \in [n]$, $\mathbb{E}[X_j^2] = 1$ and $\mathbb{E}[X_j^4] = \gamma_I$, then

$$\begin{aligned}\mathop{\mathbb{E}}_{S_1 \neq S_2} \mathop{\mathbb{E}}_{X}[X_{S_1} X_{S_2}] &= \sum_{i=0}^{d-1} \gamma_I^i \Pr_{S_1 \neq S_2}[|S_1 \cap S_2| = i] \\ &= (1 - g_1 - \ldots - g_{d-1}) + \gamma_I g_1 + \ldots + \gamma_1^{d-1} g_{d-1}\end{aligned}$$

This means that the numerator is,

$$\mathop{\mathbb{E}}_{q_a, X}[q_a^4(X)] = Np\gamma_c\gamma_I^d + 3p^2N(N-1)\Big((1 - g_1 - \ldots - g_{d-1}) + \gamma_I g_1 + \ldots + \gamma_1^{d-1} g_{d-1}\Big)$$

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

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

From this, observe that

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

Since $p < \gamma_c/3$, the numerator of the additive term is $\theta(Np\gamma_c\gamma_I^d)$. Since $\gamma_I, \gamma_c = \theta(1)$, the denominator of the additive term is $\theta(p^2N^2)$. Thus, for $t = pN$, then $\alpha_{\text{same}} = 3 + \theta(\frac{\gamma_c\gamma_I^d}{t})$. □

Lemma 6.2. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3, and, in particular, since $d \geq 2$ is an integer constant and γ_I, γ_c are constants with $\gamma_I > 1$, then*

$$\alpha_{\text{diff}} = 3 + \frac{\gamma_c \cdot \gamma_I^d}{t} + \Omega(1/n)$$

Proof. Recall the definition α_{diff}:

$$\alpha_{\text{diff}} = \frac{\mathbb{E}_{q_a, X_a}[q_a^4(X_a)]}{\mathbb{E}_{q_a, q_b, X_a, X_b}[q_a^2(X_a) \cdot q_b^2(X_b)]}$$

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

$$\mathop{\mathbb{E}}_{q_a, X_a}[q_a^4(X_a)] = Np\gamma_c\gamma_I^d + 3p^2N(N-1)\left((1 - g_1 - \ldots - g_{d-1}) + \gamma_I g_1 + \ldots + \gamma_1^{d-1} g_{d-1}\right)$$

Now, let's compute the denominator. Let us represent polynomials q_a and q_b as $q_a(\mathbf{x}) = \sum_S c_S x_S$ and $q_b(\mathbf{x}) = \sum_S d_S x_S$ where the coefficients are sampled from $\mathcal{D}_{\mathsf{Coeff,p}}$ and inputs are sampled from $\mathcal{D}_{\mathsf{Inp}}$. Then,

$$\mathop{\mathbb{E}}_{q_a, q_b, X_a, X_b}[q_a^2(X_a) q_b^2(X_b)] = \mathop{\mathbb{E}}_{q_a, q_b, X_a, X_b}[\sum_{S_1, S_2, S_3, S_4} c_{S_1} c_{S_2} d_{S_3} d_{S_4} X_{a,S_1} X_{a,S_2} X_{b,S_3} X_{b,S_4}]$$

Now, since the odd moments of every coefficient variable are 0, and the second moment of the input

variables is 1, this becomes:

$$\begin{aligned}
\mathop{\mathbb{E}}_{q_a,q_b,X_a,X_b}[q_a^2(X_a)q_b^2(X_b)] &= \mathop{\mathbb{E}}_{q_a,q_b,X_a,X_b}[\sum_{S_1,S_3} c_{S_1}^2 d_{S_3}^2 X_{a,S_1}^2 X_{b,S_3}^2] \\
&= p^2 \mathop{\mathbb{E}}_{X_a,X_b}[\sum_{S_1,S_3} X_{a,S_1}^2 X_{b,S_3}^2] \\
&= N^2 p^2
\end{aligned}$$

This means that:

$$\begin{aligned}
\alpha_{\text{diff}} &= \frac{Np\gamma_c\gamma_I^d + 3p^2N(N-1)\left((1-g_1-\ldots-g_{d-1}) + \gamma_I g_1 + \ldots + \gamma_1^{d-1} g_{d-1}\right)}{N^2p^2} \\
&= \frac{Np\gamma_c\gamma_I^d + 3p^2N(N-1)\left(1 + (\gamma_I - 1)g_1 \ldots + (\gamma_I^{d-1} - 1)g_{d-1}\right)}{N^2p^2} \\
&= \frac{\gamma_c\gamma_I^d}{t} + 3\left(1 - \frac{1}{N}\right)\left(1 + (\gamma_I - 1)g_1 \ldots + (\gamma_I^{d-1} - 1)g_{d-1}\right)
\end{aligned}$$

Observe that $g_i = \Pr_{S_1 \neq S_2}[|S_1 \cap S_2| = i] = \theta(1/n^i)$ for $i \in [d]$. Hence, for $t = Np$ and since $\gamma_I > 1$, then

$$\alpha_{\text{diff}} \geq \frac{\gamma_c\gamma_I^d}{t} + 3\left(1 - \frac{1}{N}\right)\left(1 + \theta(\frac{1}{n})\right)$$

Since d is a constant integer greater than 1, then $N = \Omega(n^2)$ and

$$\alpha_{\text{diff}} = 3 + \frac{\gamma_c\gamma_I^d}{t} + \Omega\left(\frac{1}{n}\right)$$

□

Corollary 6.1. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3, then*

$$\alpha_{\text{th}} = 3 + \Omega(1/n)$$

Proof. This follows directly from Lemmas 6.1 and 6.2. □

Now we will show that our algorithm is correct with high probability by first showing correctness given the same distribution and then showing correctness given the diff distribution.

Lemma 6.3. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3, and, in particular, since* $\gamma_I, \gamma_c, \epsilon_I = \theta(1)$, $t = \Omega(n \log n \cdot C_I^{4d})$, *and* $m = \Omega(n^2 C_I^{8d} C_c^8 \log^{10} n)$, *then with probability* $1 - n^{-\omega(1)}$, *Algorithm 3 outputs* 0, *given a randomly chosen input from the* same *distribution.*

Proof. Suppose that we are given a randomly chosen input from the same distribution, that is we receive $\{q_i, q_i(\mathbf{x})\}_{i\in[m]}$ where each q_i is sampled from $\mathcal{Q}_{n,d,p}$ and $\mathbf{x}$ is randomly sampled from $\mathcal{D}^*_{\mathsf{lnp,n}}$. Let X be a random variable with distribution $\mathcal{D}^*_{\mathsf{lnp,n}}$. Define V_i for $i \in [m/2]$ to be the following random variable which is a function of random variable X and the implicit random variables representing the coefficients of q_{2i-1} and q_{2i}:

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

We now define random variable μ as

$$\mu = \mu_i = \mathop{\mathbb{E}}_{q_{2i-1}, q_{2i}}[V_i] = \mathop{\mathbb{E}}_{q_{2i-1}, q_{2i}}[q_{2i-1}^4(X) + q_{2i}^4(X) - 2\alpha_{\text{th}} \cdot q_{2i-1}^2(X) q_{2i}^2(X)].$$

Note that since the distributions of each q_i are i.i.d., then $\mu_i = \mu_j$ for any $i, j \in [m/2]$. Then,

observe that

$$\begin{aligned}
\Pr[\text{Algorithm 3 outputs } 0 \mid \mathsf{same}] &= \Pr_{X,q_1,\ldots,q_m}\left[\sum_i q_i^4(X) - 2\alpha_{\text{th}} \sum_{i\in[m/2]} q_{2i-1}^2(X) q_{2i}^2(X) < 0\right] \\
&= \Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} V_i < 0\right] \\
&= \Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (V_i - \mu) + m\mu/2 < 0\right]
\end{aligned}$$

Thus, in order to prove the lemma, it suffices to show that

$$\Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (V_i - \mu) + m\mu/2 < 0\right] \geq 1 - n^{-\omega(1)}. \tag{1}$$

To prove the above, we will show that the following two conditions hold:

1. $\Pr_X[\mu < 0] \geq 1 - n^{-\omega(1)}$
2. $\Pr_{X,q_1,\ldots q_m}\left[|\sum_{i\in[m/2]}(V_i - \mu)| < |m\mu/2|\right] \geq 1 - n^{-\omega(1)}$

Then, Equation 1 follows from these two conditions since

$$\begin{aligned}
\Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (V_i - \mu) + m\mu/2 < 0\right] &= \Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (V_i - \mu) < -m\mu/2\right] \\
&\geq \Pr_{X,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (V_i - \mu) < |m\mu/2|\right] \Pr_X[\mu < 0] \\
&\geq \Pr_{X,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]} (V_i - \mu)| < |m\mu/2|\right] \Pr_X[\mu < 0] \\
&\geq 1 - n^{-\omega(1)}
\end{aligned}$$

Claim 6.1. $\Pr_X[\mu < 0] \geq 1 - n^{-\omega(1)}$

Proof. Recall that if Z is a random variable over the coefficent distribution $\mathcal{D}_{\mathsf{Coeff},\mathsf{p}}$, then all odd moments of Z are 0 and $\mathbb{E}[Z^2] = p$, $\mathbb{E}[Z^4] = p\gamma_c$. Then, using a similar calculation as in the previous lemmata, we obtain that:

$$\begin{aligned}
\mu &= \mathop{\mathbb{E}}_{q_{2i-1},q_{2i}}[q_{2i-1}^4(X) + q_{2i}^4(X) - 2\alpha_{\text{th}} \cdot q_{2i-1}^2(X)q_{2i}^2(X)] \\
&= 2(\sum_S p\gamma_c X_S^4 + \sum_{S_1 \neq S_2} 3p^2 X_{S_1}^2 X_{S_2}^2) - 2\alpha_{\text{th}}(\sum_S p^2 X_S^4 + \sum_{S_1 \neq S_2} p^2 X_{S_1}^2 X_{S_2}^2) \\
&= 2p(\gamma_c - p\alpha_{\text{th}}) \sum_S X_S^4 + p^2(6 - 2\alpha_{\text{th}}) \sum_{S_1 \neq S_2} X_{S_1}^2 X_{S_2}^2
\end{aligned}$$

Observe, that since $\alpha_{\text{th}} > 3 + \Omega(1/n)$, then

$$\begin{aligned}
\mu &< 2p(\gamma_c - p\alpha_{\text{th}}) \sum_S X_S^4 - \Omega(1/n)p^2 \sum_{S_1 \neq S_2} X_{S_1}^2 X_{S_2}^2 \\
&< 2p\gamma_c \sum_S X_S^4 - \Omega(1/n)p^2 \sum_{S_1 \neq S_2} X_{S_1}^2 X_{S_2}^2
\end{aligned}$$

Since the input distribution is C_I bounded, then

$$\Pr_X\left[\mu < 2p\gamma_c N C_I^{4d} - \Omega(1/n)p^2 \sum_{S_1 \neq S_2} X_{S_1}^2 X_{S_2}^2\right] \geq 1 - n^{-\omega(1)}$$

Since the input distribution satisfies $\Pr[|\mathcal{D}_{\mathsf{Inp}}| > \epsilon_I] > \Omega(1)$, where $\epsilon_I = \Omega(1)$, then

$$\Pr_X\left[\sum_{S_1 \neq S_2} X_{S_1}^2 X_{S_2}^2 = \Omega(N^2)\right] \geq 1 - n^{-\omega(1)}$$

This means that if

$$p^2N^2/n >> \gamma_c pNC_I^{4d}$$

then

$$\Pr_X[\mu < 0] \geq 1 - n^{-\omega(1)}$$

Since $\gamma_c = O(1)$, this is ensured because $Np = t >> nC_I^{4d}$. □

To prove the second condition, we will first prove the following:

Claim 6.2. *With probability* $1 - e^{-\Omega(\log^2 n)}$ *over the coins of* q_i*, then for any* $\mathbf{x}$*,*

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

Proof. To prove this we will apply the Hoeffding bound. For a fixed $\mathbf{x}$, we define $q_i(\mathbf{x}) = \sum_S c_S x_S$ where each coefficient is chosen independently from $\mathcal{D}_{\mathsf{Coeff,p}}$. Recall that this means each coefficient is set to 0 with probability $1 - p$ and sampled from a distribution $\mathcal{D}_{\mathsf{Coeff}}$ with probability p. Now, we can instead consider sampling $q_i(\mathbf{x})$ by first sampling a set $\mathcal{T}$ representing all monomials with non-zero coefficients and then sampling coefficients c_S from $\mathcal{D}_{\mathsf{Coeff}}$ for each set $S \in \mathcal{T}$. If this set $\mathcal{T}$ is constructed by choosing each set S of size d with probability p, then we have an equivalent method of sampling $q_i(\mathbf{x})$. Thus,

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

where c_S is now chosen from $\mathcal{D}_{\mathsf{Coeff}}$ and $\mathcal{T}$ is randomly sampled as described above. Note that the expected number of elements inside set $\mathcal{T}$ is $t = Np$. Let k be the number of elements inside a set

$\mathcal{T}$. Since $\mathcal{D}_{\mathsf{Inp}}$ is C_I bounded and $\mathcal{D}_{\mathsf{Coeff}}$ is C_c bounded, then $|q_i(\mathbf{x})| \leq kC_I^d C_c$. we can now use the Hoeffding bound to prove

$$\Pr[|q_i(\mathbf{x})| < \sqrt{k}C_I^d C_c \log n] > 1 - e^{-\Omega(\log^2 n)}$$

Then, observe that by chernoff bound,

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

Thus, by the union bound and since $t > n$,

$$\begin{aligned}
\Pr\left[|q_i(\mathbf{x})| \leq O(C_I^d C_c \sqrt{t} \log n)\right] &\geq \Pr\left[(|q_i(\mathbf{x})| < \sqrt{k}C_I^d C_c \log n) \wedge (k \in [t/2, 3t/2])\right] \\
&\geq 1 - e^{-\Omega(t)} - e^{-\Omega(\log^2 n)} \\
&= 1 - e^{-\Omega(\log^2 n)}
\end{aligned}$$

□

Now, we prove the second condition.

Claim 6.3. $\Pr_{X,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(V_i - \mu)| < |m\mu/2|\right] \geq 1 - n^{-\omega(1)}$

Proof. We would like to use the Hoeffding bound. However, since the V_i's are not independent, we will first condition our variables on $X = \mathbf{x}$ for a specific value $\mathbf{x}$. We define random variable $V_{i,\mathbf{x}}$

to be V_i conditioned on $X = \mathbf{x}$. Similarly, we will define $\mu_{\mathbf{x}}$ to be μ conditioned on $X = \mathbf{x}$.

$$V_{i,\mathbf{x}} = q_{2i-1}^4(\mathbf{x}) + q_{2i}^4(\mathbf{x}) - 2\alpha_{\text{th}} \cdot q_{2i-1}^2(\mathbf{x}) q_{2i}^2(\mathbf{x})$$

$$\mu_{\mathbf{x}} = \mathbb{E}[V_{i,\mathbf{x}}]$$

Note that $V_{i,\mathbf{x}}$ is a function of the implicit random variables representing the coefficients of q_{2i-1} and q_{2i}, and that $\mu_{\mathbf{x}}$ is a value, not a random variable. Then, by Claim 6.2, for all $\mathbf{x}$,

$$\Pr_{q_{2i-1},q_{2i}}\left[|V_{i,\mathbf{x}}| \leq O(C_I^{4d} C_c^4 t^2 \log^4 n)\right] \geq 1 - n^{-\omega(1)}$$

Now we want to apply the Hoeffding bound to bound $\sum_{i\in[m/2]}(V_{i,\mathbf{x}} - \mu_{\mathbf{x}})$. However, the Hoeffding bound requires that each random variable $V_{i,\mathbf{x}}$ is bounded within an interval of $O(C_I^{4d} C_c^4 t^2 \log^4 n)$ with probability 1 over the coins of choosing the polynomials. But this happens only with probability $1 - n^{-\omega(1)}$ in our case. In order to deal with this issue, we define random variable $V'_{i,\mathbf{x}}$ as

$$V'_{i,\mathbf{x}} = \begin{cases} V_{i,\mathbf{x}} & \text{if } |V_{i,\mathbf{x}}| \leq O(C_I^{4d} C_c^4 t^2 \log^4 n) \\ 0 & \text{else} \end{cases}$$

and define

$$\mu'_{\mathbf{x}} = \mathbb{E}[V'_{i,\mathbf{x}}]$$

Observe that by Hoeffding's inequality, since $V'_{i,\mathbf{x}}$ is bounded in absolute value by $O(C_I^{4d} C_c^4 t^2 \log^4 n)$,

then

$$\Pr_{q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(V'_{i,\mathbf{x}}-\mu'_{\mathbf{x}})|\leq O(\sqrt{m/2}C_I^{4d}C_c^4t^2\log^5 n)\right]\geq 1-e^{-\Omega(\log^2 n)}$$

We will now relate this back to the variables we wish to bound. First, we will bound the difference between the expectations of $V_{i,\mathbf{x}}$ and $V'_{i,\mathbf{x}}$. Consider

$$\mathbb{E}[V_{i,\mathbf{x}}]=\mathbb{E}[V_{i,\mathbf{x}}\mid V_{i,\mathbf{x}}=V'_{i,\mathbf{x}}]\Pr[V_{i,\mathbf{x}}=V'_{i,\mathbf{x}}]+\mathbb{E}[V_{i,\mathbf{x}}\mid V_{i,\mathbf{x}}\neq V'_{i,\mathbf{x}}]\Pr[V_{i,\mathbf{x}}\neq V'_{i,\mathbf{x}}]$$

Note that due to the niceness of our coefficient and input distributions, each coefficient is bounded in absolute value by C_c and each input is bounded in absolute value by C_I. Thus each $q_i(\mathbf{x})$ is bounded in absolute value by $NC_I^dC_c$ and $V_{i,\mathbf{x}}$ is bounded in absolute value by $O(N^4C_I^{4d}C_c^4)$. Therefore, $\mathbb{E}[V_{i,\mathbf{x}}\mid V_{i,\mathbf{x}}\neq V'_{i,\mathbf{x}}]=O(N^4C_I^{4d}C_c^4)$. Since $\Pr[V_{i,\mathbf{x}}\neq V'_{i,\mathbf{x}}]=O(n^{-\omega(1)})$, then

$$\begin{aligned}\mathbb{E}[V_{i,\mathbf{x}}]&=\mathbb{E}[V_{i,\mathbf{x}}\mid V_{i,\mathbf{x}}=V'_{i,\mathbf{x}}]\Pr[V_{i,\mathbf{x}}=V'_{i,\mathbf{x}}]+O(n^{-\omega(1)})\\&=\mathbb{E}[V'_{i,\mathbf{x}}]+O(n^{-\omega(1)})\end{aligned}$$

This means that $|\mu_{\mathbf{x}}-\mu'_{\mathbf{x}}|=|\mathbb{E}[V_{i,\mathbf{x}}]-\mathbb{E}[V'_{i,\mathbf{x}}]|\leq O(n^{-\omega(1)})$. Now, consider

$$\sum_{i\in[m/2]}(V'_{i,\mathbf{x}}-\mu_{\mathbf{x}})=\sum_{i\in[m/2]}(V'_{i,\mathbf{x}}-\mu'_{\mathbf{x}})+\frac{m(\mu'_{\mathbf{x}}-\mu_{\mathbf{x}})}{2}$$

Thus, since $|\mu_{\mathbf{x}}-\mu'_{\mathbf{x}}|\leq O(n^{-\omega(1)})$, then

$$\Pr_{q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(V'_{i,\mathbf{x}}-\mu_{\mathbf{x}})|\leq O(\sqrt{m/2}C_I^{4d}C_c^4t^2\log^5 n)\right]\geq 1-e^{-\Omega(\log^2 n)}$$

As $V_{i,\mathbf{x}} = V'_{i,\mathbf{x}}$ with probability $1 - n^{-\omega(1)}$, then by a union bound,

$$\Pr_{q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(V_{i,\mathbf{x}} - \mu_{\mathbf{x}})| \leq |\sum_{i\in[m/2]} V'_{i,\mathbf{x}} - \mu_{\mathbf{x}}|\right] \geq 1 - n^{-\omega(1)}$$

Since the above probabilities are true for all $\mathbf{x}$, it holds that

$$\Pr_{X,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(V_i - \mu)| \leq O(\sqrt{m/2}C_I^{4d}C_c^4 t^2 \log^5 n)\right] \geq 1 - n^{-\omega(1)}$$

We will now conclude a lower bound on $|\mu|$. Recall that in the proof of Claim 6.1, we showed

$$\mu = 2p(\gamma_c - p\alpha_{\text{th}})\sum_S X_S^4 + p^2(6 - 2\alpha_{\text{th}})\sum_{S_1\neq S_2} X_{S_1}^2 X_{S_2}^2$$

Since $\Pr_X[\mu < 0] \geq 1 - n^{-\omega(1)}$ by Claim 6.1, then

$$\Pr_X\left[|\mu| = 2p(p\alpha_{\text{th}} - \gamma_c)\sum_S X_S^4 + p^2(2\alpha_{\text{th}} - 6)\sum_{S_1\neq S_2} X_{S_1}^2 X_{S_2}^2\right] \geq 1 - n^{-\omega(1)}$$

Since $\alpha_{\text{th}} = 3 + \Omega(1/n)$, then

$$\Pr_X\left[|\mu| \geq -2p\gamma_c\sum_S X_S^4 + (p^2/n)\sum_{S_1\neq S_2} X_{S_1}^2 X_{S_2}^2\right] \geq 1 - n^{-\omega(1)}$$

Since the input distribution $\mathcal{D}_{\mathsf{lnp}}$ is $(\gamma_I, C_I, \epsilon_I)$ nice and $\epsilon_I = \theta(1)$, then

$$\Pr_X\left[|\mu| \geq -2p\gamma_c N C_I^{4d} + \Omega((p^2/n)N(N-1))\right] \geq 1 - n^{-\omega(1)}$$

Since $\gamma_c = \theta(1)$ and $Np = t = \omega(nC_I^{4d})$, then

$$\Pr_X \left[|\mu| = \Omega(t^2/n) \right] \geq 1 - n^{-\omega(1)}$$

This means that

$$\Pr_{X, q_1, \ldots, q_m} \left[|\sum_{i \in [m/2]} (V_i - \mu)| < |m\mu/2| \right] \geq 1 - n^{-\omega(1)}$$

as long as

$$\sqrt{m/2} C_I^{4d} C_c^4 t^2 \log^5 n << \frac{mt^2}{2n}$$

which is true since

$$m > 2n^2 C_I^{8d} C_c^8 \log^{10} n$$

□

□

Lemma 6.4. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3, and, in particular, since* $\gamma_I, \gamma_c, \epsilon_I = \theta(1)$, $t = \Omega(n \log n \cdot C_I^{4d})$, *and* $m = \Omega(n^2 C_I^{8d} C_c^8 \log^{10} n)$, *then with probability* $1 - n^{-\omega(1)}$, *Algorithm 3 outputs* 1, *given a randomly chosen input from the* diff *distribution.*

Proof. Suppose that we are given a randomly chosen input from the diff distribution, that is we receive $\{q_i, q_i(\mathbf{x}_i)\}_{i \in [m]}$ where each q_i is sampled from $\mathcal{Q}_{n,d,p}$ and each x_i is sampled from $\mathcal{D}^*_{\mathsf{Inp},\mathsf{n}}$. Let $X_1, \ldots, X_m$ be random variables with distribution $\mathcal{D}^*_{\mathsf{Inp},\mathsf{n}}$.[6] Define U_i for $i \in [m/2]$ to be the

[6]For this proof, we will switch from our usual custom of using X_i to denote a random variable with distribution

following random variable which is a function of random variables X_{2i-1}, X_{2i} and the implicit random variables representing the coefficients of q_{2i-1} and q_{2i}:

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

We now define μ as

$$\mu = \mu_i = \mathbb{E}[U_i].$$

Note that since the distributions of each q_i are i.i.d., then $\mu_i = \mu_j$ for any $i, j \in [m/2]$. Similarly to before, observe that

$$\begin{aligned}
\Pr[\text{Algorithm 3 outputs } 0 \mid \mathsf{diff}] &= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_i q_i^4(X) - 2\alpha_{\text{th}} \sum_{i\in[m/2]} q_{2i-1}^2(X) q_{2i}^2(X) \geq 0\right] \\
&= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} U_i \geq 0\right] \\
&= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (U_i - \mu) + m\mu/2 \geq 0\right]
\end{aligned}$$

Thus, in order to prove the lemma, it suffices to show that

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]} (U_i - \mu) + m\mu/2 \geq 0\right] \geq 1 - n^{-\omega(1)}. \tag{2}$$

Using a similar argument as in the previous lemma, to prove the above, it suffices to show that the following two conditions hold:

1. $\mu > 0$

$\mathcal{D}_{\mathsf{Inp}}$ and instead use X_i to denote a random variables with distribution $\mathcal{D}^*_{\mathsf{Inp},\mathsf{n}}$.

2. $\Pr_{X_1,\ldots,X_m,q_1,\ldots q_m}\left[|\sum_{i\in[m/2]}(U_i-\mu)| < |m\mu/2|\right] \geq 1-n^{-\omega(1)}$

Equation 2 follows from these two conditions since

$$\begin{aligned}
\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]}(U_i-\mu)+m\mu/2\geq 0\right] &= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[\sum_{i\in[m/2]}(U_i-\mu)\geq -m\mu/2\right] \\
&\geq \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[-1\cdot|\sum_{i\in[m/2]}(U_i-\mu)|\geq -m\mu/2\right] \\
&= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(U_i-\mu)|\leq m\mu/2\right] \\
&= \Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(U_i-\mu)|\leq |m\mu/2|\right] \\
&\geq 1-n^{-\omega(1)}
\end{aligned}$$

Claim 6.4. $\mu=\Omega(t^2/n)>0$

Proof. First, observe that by definition of α_{diff} then

$$\alpha_{\text{diff}}=\frac{\mathbb{E}_{q_a,X_a}[q_a^4(X_a)]}{\mathbb{E}_{q_a,q_b,X_a,X_b}[q_a^2(X_a)\cdot q_b^2(X_b)]}$$

which implies

$$\mathop{\mathbb{E}}_{q_a,X_a}[q_a^4(X_a)]-\alpha_{\text{diff}}\cdot\mathop{\mathbb{E}}_{q_a,q_b,X_a,X_b}[q_a^2(X_a)\cdot q_b^2(X_b)]=0$$

Thus,

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

Therefore,

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

Observe that q_{2i-1} and q_{2i} have expected density $t = Np$. Then, since the odd moments of each input and coefficient variable are zero, and the second moment of each input and coefficient variable is 1, then

$$\mathop{\mathbb{E}}_{q_{2i},q_{2i-1},X_{2i-1},X_{2i}}[q_{2i-1}^2(X_{2i-1}) q_{2i}^2(X_{2i})] = N^2 p^2 = t^2$$

Therefore, since $\alpha_{\text{diff}} - \alpha_{\text{same}} = \Omega(1/n)$ by Lemmas 6.3 and 6.4, then

$$\mu = \Omega(t^2/n) > 0$$

□

Claim 6.5. $\Pr_{X_1,\ldots,X_m,q_1,\ldots q_m}\left[|\sum_{i\in[m/2]}(U_i-\mu)|<|m\mu/2|\right]\geq 1-n^{-\omega(1)}$

Proof. Since $\mu>0$, this will follow if we show that

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots q_m}\left[|\sum_{i\in[m/2]}(U_i-\mu)|<\frac{m\mu}{2}\right]\geq 1-n^{-\omega(1)}$$

Now, for all $\mathbf{x}_i$,

$$\Pr_{q_i}\left[|q_i(\mathbf{x}_i)|=O(C_I^d C_c\sqrt{t}\log n)\right]\geq 1-e^{\Omega(-\log^2 n)}$$

This can be proven with a proof identical to that of Claim 6.2. This means that

$$\Pr_{X_{2i-1},X_{2i},q_{2i-1},q_{2i}}\left[|U_i|=O(C_I^{4d}C_c^4t^2\log^4 n)\right]\geq 1-e^{\Omega(-\log^2 n)}$$

We now want to apply the Hoeffding bound to bound $|\sum_{i\in[m/2]}(U_i-\mu)|$. We will proceed in the same manner as in the previous lemma. As before, we only have $|U_i|=O(C_I^{4d}C_c^4t^2\log^4 n)$ with probability $1-e^{\Omega(-\log^2 n)}$ as opposed to probability 1. To deal with this, we define random variable U_i' as

$$U_i'=\begin{cases}U_i & \text{if } |U_i|\leq O(C_I^{4d}C_c^4t^2\log^4 n)\\ 0 & \text{else}\end{cases}$$

and define

$$\mu'=\mu_i'=\mathbb{E}[U_i']$$

Observe that by Hoeffding's inequality, since U_i' is bounded in absolute value by $O(C_I^{4d}C_c^4t^2\log^4 n)$,

then

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(U_i'-\mu')| \leq O(\sqrt{m/2}C_I^{4d}C_c^4t^2\log^5 n)\right] \geq 1-e^{-\Omega(\log^2 n)}$$

We will now relate this back to the variables we wish to bound. First, we will bound the difference between the expectations of U_i and U_i'. Consider

$$\mathbb{E}[U_i] = \mathbb{E}[U_i \mid U_i = U_i']\Pr[U_i = U_i'] + \mathbb{E}[U_i \mid U_i \neq U_i']\Pr[U_i \neq U_i']$$

Note that due to the niceness of our coefficient and input distributions, each coefficient is bounded in absolute value by C_c and each input is bounded in absolute value by C_I. Thus each $q_i(\mathbf{x}_i)$ is bounded in absolute value by $NC_I^dC_c$ and U_i is bounded in absolute value by $O(N^4C_I^{4d}C_c^4)$. Therefore, $\mathbb{E}[U_i \mid U_i \neq U_i'] = O(N^4C_I^{4d}C_c^4)$. Since $\Pr[U_i \neq U_i'] = O(n^{-\omega(1)})$, then

$$\begin{aligned}\mathbb{E}[U_i] &= \mathbb{E}[U_i \mid U_i = U_i']\Pr[U_i = U_i'] + O(n^{-\omega(1)})\\ &= \mathbb{E}[U_i'] + O(n^{-\omega(1)})\end{aligned}$$

This means that $|\mu-\mu'| = |\mathbb{E}[U_i] - \mathbb{E}[U_i']| \leq O(n^{-\omega(1)})$. Now, consider

$$\sum_{i\in[m/2]}(U_i'-\mu) = \sum_{i\in[m/2]}(U_i'-\mu') + \frac{m(\mu'-\mu)}{2}$$

Thus, since $|\mu-\mu'| \leq O(n^{-\omega(1)})$, then

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]}(U_i'-\mu)| \leq O(\sqrt{m/2}C_I^{4d}C_c^4t^2\log^5 n)\right] \geq 1-e^{-\Omega(\log^2 n)}$$

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

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]} U_i - \mu| \leq |\sum_{i\in[m/2]} U_i' - \mu|\right] \geq 1 - n^{-\omega(1)}$$

Therefore,

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]} (U_i - \mu)| \leq O(\sqrt{m/2}C_I^{4d}C_c^4 t^2 \log^5 n)\right] \geq 1 - n^{-\omega(1)}$$

Now, by Claim 6.4, then

$$\mu = \Omega(t^2/n) > 0$$

This means that

$$\Pr_{X_1,\ldots,X_m,q_1,\ldots,q_m}\left[|\sum_{i\in[m/2]} (U_i - \mu)| < m\mu/2\right] \geq 1 - n^{-\omega(1)}$$

as long as

$$\sqrt{m/2}C_I^{4d}C_c^4 t^2 \log^5 n << \frac{mt^2}{2n}$$

which is true since

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

□

□

Corollary 6.2. *For the parameters and terms defined in Theorem 6.1 and Algorithm 3, then with probability* $1 - n^{-\omega(1)}$*, Algorithm 3 outputs* 0 *when given a randomly chosen input from the* same *distribution and outputs* 1 *when given a randomly chosen input from the* diff *distribution.*

Proof. This follows directly from Lemmas 6.3 and 6.4. □

Running Time. Algorithm 3 first computes ratio α_{th} which can be computed exactly using the formulae described in Lemmas 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.

Thus, from the correctness and running time results above, we prove Theorem 6.1.

A On PIDGs, $i\mathcal{O}$, and Pseudo-Flawed Smudging Generators

Lin and Matt [LM18] propose the notion of a pseudo-flawed smudging generator as a tool for building $i\mathcal{O}$, and they propose using the candidates from the work of Ananth, Jain, and Sahai [AJS18] to instantiate this object (see also [JLMS19]). While the definition of this object is quite complex, Lin and Matt suggest by way of example (see [LM18], p. 26), that if the candidates of [AJS18] satisfied the notion of a PIDG, and a little more, then this would yield a pseudo-flawed smudging generator. However, the polynomial families suggested by [AJS18, JLMS19] in fact satisfy the conditions we require for our non-trivial distinguishers to exist.

Intuitively, this attack arises because pseudo-flawed smudging generators require that polynomials over the integers achieve computational indistinguishability with respect to a distribution that satisfies a statistical "flawed smudging" property. The most natural such distributions would

be product distributions, and attacking the assumption with respect to a product distribution corresponds to solving the PIDP.

An interesting open question is whether there are non-product distributions that also satisfy the flawed smudging property, thereby potentially allowing the existence of pseudo-flawed smudging generators despite our attacks.

B 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.

[AJS18] Prabhanjan Ananth, Aayush Jain, and Amit Sahai. Indistinguishability obfuscation without multilinear maps: io from lwe, bilinear maps, and weak pseudorandomness. *IACR Cryptology ePrint Archive*, 2018:615, 2018.

[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.

[BFM14] Christina Brzuska, Pooya Farshim, and Arno Mittelbach. Indistinguishability obfuscation and UCEs: The case of computationally unpredictable sources. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO 2014, Part I*, volume 8616 of *LNCS*, pages 188–205. Springer, Heidelberg, August 2014.

[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.

[BHJ+19] Boaz Barak, Samuel B. Hopkins, Aayush Jain, Pravesh Kothari, and Amit Sahai. Sum-of-squares meets program obfuscation, revisited. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part I*, volume 11476 of *LNCS*, pages 226–250. Springer, Heidelberg, May 2019.

[BPR15] Nir Bitansky, Omer Paneth, and Alon Rosen. On the cryptographic hardness of finding a Nash equilibrium. In Venkatesan Guruswami, editor, *56th FOCS*, pages 1480–1498. IEEE Computer Society Press, October 2015.

[CHK+19] Arka Rai Choudhuri, Pavel Hubácek, Chethan Kamath, Krzysztof Pietrzak, Alon Rosen, and Guy N. Rothblum. Finding a nash equilibrium is no easier than breaking Fiat-Shamir. In Moses Charikar and Edith Cohen, editors, *51st ACM STOC*, pages 1103–1114. ACM Press, June 2019.

[CHN+16] Aloni Cohen, Justin Holmgren, Ryo Nishimaki, Vinod Vaikuntanathan, and Daniel Wichs. Watermarking cryptographic capabilities. In Daniel Wichs and Yishay Mansour, editors, *48th ACM STOC*, pages 1115–1127. ACM Press, June 2016.

[GGG+14] Shafi Goldwasser, S. Dov Gordon, Vipul Goyal, Abhishek Jain, Jonathan Katz, Feng-Hao Liu, Amit Sahai, Elaine Shi, and Hong-Sheng Zhou. Multi-input functional encryption. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 578–602. Springer, Heidelberg, May 2014.

[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.

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

[GPS16] Sanjam Garg, Omkant Pandey, and Akshayaram Srinivasan. Revisiting the cryptographic hardness of finding a nash equilibrium. In Matthew Robshaw and Jonathan Katz, editors, *CRYPTO 2016, Part II*, volume 9815 of *LNCS*, pages 579–604. Springer, Heidelberg, August 2016.

[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.

[HJK+16] Dennis Hofheinz, Tibor Jager, Dakshita Khurana, Amit Sahai, Brent Waters, and Mark Zhandry. How to generate and use universal samplers. In Jung Hee Cheon and Tsuyoshi Takagi, editors, *ASIACRYPT 2016, Part II*, volume 10032 of *LNCS*, pages 715–744. Springer, Heidelberg, December 2016.

[HSW13] Susan Hohenberger, Amit Sahai, and Brent Waters. Full domain hash from (leveled) multilinear maps and identity-based aggregate signatures. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part I*, volume 8042 of *LNCS*, pages 494–512. Springer, Heidelberg, August 2013.

[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`.

[KLW15] Venkata Koppula, Allison Bishop Lewko, and Brent Waters. Indistinguishability obfuscation for turing machines with unbounded memory. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *47th ACM STOC*, pages 419–428. ACM Press, June 2015.

[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.

[LM18] Huijia Lin and Christian Matt. Pseudo flawed-smudging generators and their application to indistinguishability obfuscation. *IACR Cryptology ePrint Archive*, 2018:646, 2018.

[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.

[SW14] Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In David B. Shmoys, editor, *46th ACM STOC*, pages 475–484. ACM Press, May / June 2014.