

Sum of Squares Bounds for the Ordering Principle

Aaron Potechin

University of Chicago, IL, USA
potechin@uchicago.edu

Abstract

In this paper, we analyze the sum of squares hierarchy (SOS) on the ordering principle on n elements (which has $N = \Theta(n^2)$ variables). We prove that degree $O(\sqrt{n \log(n)})$ SOS can prove the ordering principle. We then show that this upper bound is essentially tight by proving that for any $\epsilon > 0$, SOS requires degree $\Omega(n^{\frac{1}{2}-\epsilon})$ to prove the ordering principle.

2012 ACM Subject Classification Theory of computation → Proof complexity

Keywords and phrases sum of squares hierarchy, proof complexity, ordering principle

Digital Object Identifier 10.4230/LIPIcs.CCC.2020.38

Funding This work was supported by the Knut and Alice Wallenberg Foundation, the European Research Council, and the Swedish Research Council.

Acknowledgements The author would like to thank Marc Vinyals for proposing this problem. The author would also like to thank Prahladh Harsha, Toniann Pitassi, and anonymous reviewers for helpful comments on this paper.

1 Introduction

In proof complexity, we study how easy or difficult it is to prove or refute various statements. Proof complexity is an extremely rich field, so we will not attempt to give an overview of proof complexity here (for a recent survey of proof complexity, see [15]). Instead, we will only describe the particular proof system and the particular statement we are considering, namely the sum of squares hierarchy (SOS), and the ordering principle.

SOS can be described in terms of sum of squares/Positivstellensatz proofs (which we write as SOS proofs for brevity). SOS proofs have the following nice properties:

1. SOS proofs are broadly applicable as they are complete for systems of polynomial equations over \mathbb{R} . In other words, for any system of polynomial equations over \mathbb{R} which is infeasible, there is an SOS proof that it is infeasible [16].
2. SOS proofs are surprisingly powerful. In particular, SOS captures both spectral methods and powerful inequalities such as Cauchy-Schwarz and variants of hypercontractivity [2], which means that much of our mathematical reasoning can be captured by SOS proofs.
3. In some sense, SOS proofs are simple. In particular, SOS proofs only use polynomial equalities and the fact that squares are non-negative over \mathbb{R} .

SOS has been extensively studied, so we will not give an overview of what is known about SOS here. To learn more about SOS, see the following survey on SOS [3] and the following recent seminars/courses on SOS [4, 9, 11, 12].

The ordering principle (which has $N = \Theta(n^2)$ variables) states that if we have elements a_1, \dots, a_n which have an ordering and no two elements are equal then some element a_i must be minimal. The ordering principle is a very interesting example in proof complexity because for several proof systems, it has a small size proof but any proof must have high width/degree.

The ordering principle was first considered by Krishnamurthy [10] who conjectured that it was hard for the resolution proof system. This conjecture was refuted by Stalmark [17], who showed that the ordering principle has a polynomial size resolution proof based on induction.



© Aaron Potechin;
licensed under Creative Commons License CC-BY
35th Computational Complexity Conference (CCC 2020).

Editor: Shubhangi Saraf; Article No. 38; pp. 38:1–38:37



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



However, any resolution proof of the ordering principle must have width $\Omega(n) = \Omega(\sqrt{N})$. While this is a trivial statement because it takes width n to even describe the ordering principle, Bonet and Galesi [6] showed that the ordering principle can be modified to give a statement which can be described with constant width but resolution still requires width $\Omega(n) = \Omega(\sqrt{N})$ to prove it. This showed that the width/size tradeoff of Ben-Sasson and Wigderson [5] (which was based on the degree/size tradeoff shown for polynomial calculus by Impagliazzo, Pudlák, and Sgall [8]) is essentially tight.

Later, $\Omega(n)$ degree lower bounds for the ordering principle were also shown for polynomial calculus [7] and for the Sherali-Adams hierarchy. However, non-trivial SOS degree bounds for the ordering principle were previously unknown. Thus, a natural question is, does SOS also require degree $\Omega(n)$ to prove the ordering principle or is there an SOS proof of the ordering principle which has smaller degree?

1.1 Our Results

In this paper, we show almost tight upper and lower SOS degree bounds for the ordering principle. In particular, we show the following theorems:

- **Theorem 1.** *Degree $O(\sqrt{n \log(n)})$ SOS can prove the ordering principle on n elements.*
- **Theorem 2.** *For any constant $\epsilon > 0$ there is a constant $C_\epsilon > 0$ such that for all $n \in \mathbb{N}$, degree $C_\epsilon n^{\frac{1}{2}-\epsilon}$ SOS cannot prove the ordering principle on n elements.*

Theorem 1 shows that looking at degree, SOS is more powerful than resolution, polynomial calculus, and the Sherali-Adams hierarchy for proving the ordering principle. This also implies that the ordering principle is not a tight example for the size/degree trade-off for SOS which was recently shown by Atserias and Hakoniemi [1]. In particular, Atserias and Hakoniemi show [1] that if the variables are Boolean and there is an SOS proof of size S then there must be an SOS proof of degree $O(\sqrt{N \log(S)} + k)$ where N is the number of variables and k is the maximum degree of an axiom (which is usually constant). According to this bound, a statement such as the ordering principle which has a polynomial size proof could still require degree $\tilde{\Omega}(\sqrt{N})$ to prove. However, Theorem 1 shows that the ordering principle has an SOS proof of degree $\tilde{O}(\sqrt[4]{N})$, so there is a considerable gap. On the other hand, Theorem 2 shows that Theorem 1 is essentially tight and thus the ordering principle does still give an example where there is a small SOS proof yet any SOS proof must have high degree.

► **Remark 3.** We should note that our encoding of the ordering principle is tailored to SOS and differs from previous encodings of the ordering principle. In particular, we use non-Boolean auxiliary variables, so the size/degree trade-off for SOS doesn't actually apply to our encoding. That said, our upper bound also applies to previous encodings of the ordering principle (for which the size/degree trade-off for SOS does apply) and with additional work, we can also prove a slightly worse SOS lower bound for an encoding of the ordering principle with only Boolean variables (for technical reasons which are likely an artefact of the proof, the number of variables is increased to $N = O(n^{\frac{5}{2}})$ so the final lower bound is $\Omega(N^{\frac{1}{5}-\epsilon})$). This is discussed in Section 2.2.1 and Appendix A.

1.2 Outline

The remainder of the paper is organized as follows. In Section 2 we give some preliminaries. In Section 3, we give natural pseudo-expectation values for the ordering principle. In Section 4, we prove our SOS upper bound. In Sections 5–9, we describe how to prove our SOS lower bound.

2 Preliminaries

In this section, we recall the definitions of SOS proofs and pseudo-expectation values and describe the encoding of the ordering principle which we will analyze.

2.1 Sum of Squares/Positivstellensatz Proofs and Pseudo-expectation Values

► **Definition 4.** *Given a system of polynomial equations $\{s_i = 0\}$ over \mathbb{R} , a degree d SOS proof of infeasibility is an equality of the form*

$$-1 = \sum_i f_i s_i + \sum_j g_j^2$$

where

1. $\forall i, \deg(f_i) + \deg(s_i) \leq d$
2. $\forall j, \deg(g_j) \leq \frac{d}{2}$.

► **Remark 5.** This is a proof of infeasibility because if the equations $\{s_i = 0\}$ were satisfied we would have that $\forall i, f_i s_i = 0$ and for all $\forall j, g_j^2 \geq 0$, so we cannot possibly have that $\sum_i f_i s_i + \sum_j g_j^2 < 0$.

In order to prove that there is no degree d SOS proof of infeasibility for a system of polynomial equations over \mathbb{R} , we use degree d pseudo-expectation values, which are defined as follows.

► **Definition 6.** *Given a system of polynomial equations $\{s_i = 0\}$ over \mathbb{R} , degree d pseudo expectation values are a linear map \tilde{E} from polynomials of degree at most d to \mathbb{R} such that*

1. $\tilde{E}[1] = 1$
2. $\forall f, i, \tilde{E}[f s_i] = 0$ whenever $\deg(f) + \deg(s_i) \leq d$
3. $\forall g, \tilde{E}[g^2] \geq 0$ whenever $\deg(g) \leq \frac{d}{2}$.

As shown by the following proposition, these conditions are a weakening of the constraint that we have expected values over an actual distribution of solutions. Thus, intuitively, pseudo-expectations look like they are the expected values (up to degree d) over a distribution of solutions, but they may not actually correspond to a distribution of solutions.

► **Proposition 7.** *If Ω is an actual distribution of solutions to the equations $\{s_i = 0\}$ over \mathbb{R} then*

1. $E_\Omega[1] = 1$
2. $\forall f, i, E_\Omega[f s_i] = 0$
3. $\forall g, E_\Omega[g^2] \geq 0$.

► **Proposition 8.** *For a given system of polynomial equations $\{s_i = 0\}$ over \mathbb{R} , there cannot be both degree d pseudo-expectation values and a degree d SOS proof of infeasibility.*

Proof. Assume we have both degree d pseudo-expectation values and a degree d SOS proof of infeasibility. Applying the pseudo-expectation values to the SOS proof gives the following contradiction:

$$-1 = \tilde{E}[-1] = \sum_i \tilde{E}[f_i s_i] + \sum_j \tilde{E}[g_j^2] \geq 0. \quad \blacktriangleleft$$

2.2 Equations for the ordering principle

For our SOS bounds, we analyze the following system of infeasible equations which corresponds to the negation of the ordering principle.

► **Definition 9** (Ordering principle equations). *The negation of the ordering principle says that it is possible to have distinct ordered elements $\{a_1, \dots, a_n\}$ such that no a_j is the minimum element. We encode this with the following equations:*

1. We have variables x_{ij} where we want that $x_{ij} = 1$ if $a_i < a_j$ and $x_{ij} = 0$ if $a_i > a_j$. We also have auxiliary variables $\{z_j : j \in [n]\}$.
 2. $\forall i \neq j, x_{ij}^2 = x_{ij}$ and $x_{ij} = 1 - x_{ji}$ (ordering)
 3. For all distinct i, j, k , $x_{ij}x_{jk}(1 - x_{ik}) = 0$ (transitivity)
 4. $\forall j, \sum_{i \neq j} x_{ij} = 1 + z_j^2$ (for all $j \in [n]$, a_j is not the minimum element of $\{a_1, \dots, a_n\}$)
- We call this system of equations the ordering principle equations.

► **Remark 10.** In this encoding of the negation of the ordering principle, we use the auxiliary variables $\{z_j : j \in [n]\}$ so that we can express the statement that $\forall j, \exists i \neq j : x_{ij} = 1$ as polynomial equalities of low degree.

2.2.1 Relationship to other encodings of the negation of the ordering principle

The equations in Definition 9 are tailored for SOS, so they are not the same as the encodings of the negation of ordering principle which were studied in prior works [6, 7]. We now discuss this difference and how it affects our results.

For resolution, the following axioms encode the negation of the ordering principle:

1. We have variables x_{ij} where we want that x_{ij} is true if $a_i < a_j$ and x_{ij} is false if $a_i > a_j$.
2. $\forall i \neq j, x_{ij} \vee x_{ji}$ and $\neg x_{ij} \vee \neg x_{ji}$ (ordering)
3. For all distinct i, j, k , $\neg x_{ij} \vee \neg x_{jk} \vee x_{ik} = 0$ (transitivity)
4. $\forall j, \bigvee_{i \neq j} x_{ij}$ (for all $j \in [n]$, a_j is not the minimum element of $\{a_1, \dots, a_n\}$)

Translating this into polynomial equations, this gives us the following equations for polynomial calculus:

1. We have variables x_{ij} where we want that $x_{ij} = 1$ if $a_i < a_j$ and $x_{ij} = 0$ if $a_i > a_j$.
2. $\forall i \neq j, x_{ij}^2 = x_{ij}$ and $x_{ij} = 1 - x_{ji}$ (ordering)
3. For all distinct i, j, k , $x_{ij}x_{jk}(1 - x_{ik}) = 0$ (transitivity)
4. $\forall j, \prod_{i \neq j} (1 - x_{ij}) = 0$ (for all $j \in [n]$, a_j is not the minimum element of $\{a_1, \dots, a_n\}$)

However, as noted in the introduction, a width/degree lower bound of $\Omega(n)$ is trivial for these encodings as the axioms already have width/degree n . To handle this, Bonet and Galesi [6] used auxiliary variables to break up the axioms into constant width axioms. Galesi and Lauria [7] instead considered the following ordering principle on graphs.

► **Definition 11** (Ordering principle on graphs). *Given a graph G with $V(G) = [n]$, the ordering principle on G says that if each vertex $i \in [n]$ has a value a_i and all of the values are distinct then there must be some vertex whose value is less than its neighbors' values.*

When we take the negation of the ordering principle on a graph G , this changes our axioms/equations as follows:

1. For resolution, instead of the axioms $\forall j, \bigvee_{i \neq j} x_{ij}$ we have the axioms $\forall j, \bigvee_{i:(i,j) \in E(G)} x_{ij}$.
2. For polynomial calculus, instead of the equations $\forall j, \prod_{i \neq j} (1 - x_{ij}) = 0$ we have the equations $\forall j, \prod_{i:(i,j) \in E(G)} (1 - x_{ij}) = 0$.

► **Remark 12.** If $G = K_n$ then the ordering principle on G is just the ordering principle Galesi and Lauria [7] showed that if G is an expander then polynomial calculus requires degree $\Omega(n)$ to refute these equations.

The ordering principle on graphs is a weaker statement than the ordering principle, so we would expect that its negation would be easier to refute. This is indeed the case. As shown by the following lemma, we can recover the equation $\sum_{i \neq j} x_{ij} = 1 + z_j^2$ from the equation $\prod_{i:(i,j) \in E(G)} (1 - x_{ij}) = 0$, except that z_j^2 is replaced by a sum of squares.

► **Lemma 13.** *Given Boolean variables $\{x_i : i \in [k]\}$ (i.e. $\forall i \in [k], x_i^2 = x_i$) and the equation $\prod_{i=1}^k (1 - x_i) = 0$, we can deduce that $(\sum_{i=1}^k x_i) - 1$ is a sum of squares.*

Proof. Observe that modulo the axioms $x_i^2 = x_i$,

$$\begin{aligned} \sum_{i=1}^k x_i - 1 &= - \left(\prod_{i=1}^k (1 - x_i) \right) + \sum_{J \subseteq [k]: J \neq \emptyset} (|J| - 1) \left(\prod_{i \in J} x_i \right) \left(\prod_{i \in [k] \setminus J} (1 - x_i) \right) \\ &= - \left(\prod_{i=1}^k (1 - x_i) \right) + \sum_{J \subseteq [k]: J \neq \emptyset} (|J| - 1) \left(\prod_{i \in J} x_i^2 \right) \left(\prod_{i \in [k] \setminus J} (1 - x_i)^2 \right). \end{aligned}$$

To see this, observe that for any non-empty $J \subseteq [k]$, if $x_i = 1$ for all $i \in J$ and $x_i = 0$ for all $i \notin J$ then the left and right sides are both $|J| - 1$. Similarly, if all of the x_i are 0 then the left and right sides are both -1 . ◀

This implies that our SOS upper bound holds for the graph ordering principle as well as the ordering principle. However, our SOS lower bound does not apply to the ordering principle on expander graphs. Part of the reason is that our SOS lower bound relies heavily on symmetry under permutations of $[n]$.

There is one way in which the ordering principle equations are unsatisfactory for our purposes. We want to show that the size/degree tradeoffs for SOS [1] cannot be improved too much further. However, the auxiliary variables in the ordering principle equations are not Boolean and this tradeoff only applies when all of the variables are Boolean. To fix this, we show that we can modify the ordering principle equations so that we only have Boolean variables but our SOS upper and lower bounds still hold. For details, see Appendix A.

3 Pseudo-expectation values for the ordering principle

In this section, we give natural candidate pseudo-expectation values \tilde{E}_n for the ordering principle equations. In fact, \tilde{E}_n is essentially the only possible symmetric pseudo-expectation values. In particular, in section 4 we will show that if \tilde{E}_n fails at degree d then there is an SOS proof of degree at most $2d + 2$ that these equations are infeasible.

3.1 The candidate pseudo-expectation values \tilde{E}_n

As noted in Section 2, intuitively, pseudo-expectation values should look like the expected values over a distribution of solutions. Also, as shown by the following lemma, since the problem is symmetric under permutations of $[n]$, we can take \tilde{E} to be symmetric as well.

► **Lemma 14.** *If $\{s_i = 0\}$ is a system of polynomial equations which is symmetric under permutations of $[n]$ then if there are degree d pseudo-expectation values \tilde{E} then there are degree d pseudo-expectation values \tilde{E}' which are symmetric under permutations of $[n]$.*

Proof. Given degree d pseudo-expectation values \tilde{E} , take \tilde{E}' to be the linear map from polynomials of degree at most d to \mathbb{R} such that for all monomials p , $\tilde{E}'[p] = E_{\pi \in S_n}[\tilde{E}[\pi(p)]]$. Now observe that

1. $\tilde{E}'[1] = E_{\pi \in S_n}[\tilde{E}[1]] = 1$
2. $\forall f, i : \deg(f) + \deg(s_i) \leq d, \tilde{E}'[fs_i] = E_{\pi \in S_n}[\tilde{E}[\pi(f)\pi(s_i)]] = 0$ because the system of equations $\{s_i = 0\}$ is symmetric under permutations of $[n]$.
3. $\forall g : \deg(g) \leq \frac{d}{2}, \tilde{E}'[g^2] = E_{\pi \in S_n}[\tilde{E}[\pi(g)^2]] \geq 0$. ◀

Guided by this, we take the expected values over the uniform distribution over orderings of x_1, \dots, x_n . These orderings are not solutions to the equations because each ordering causes one equation $\sum_{i \neq j} x_{ij} = 1 + z_j^2$ to fail. However, a random ordering makes each individual equation $\sum_{i \neq j} x_{ij} = 1 + z_j^2$ true with high probability, so the intuition is that low degree SOS will not be able to detect that there is always one equation which fails.

► **Definition 15.** We define U_n to be the uniform distribution over orderings of x_1, \dots, x_n , i.e. for each permutation $\pi : [n] \rightarrow [n]$ we have that with probability $\frac{1}{n!}$, $x_{\pi(1)} < x_{\pi(2)} < \dots < x_{\pi(n)}$ and thus for all $i < j$, $x_{\pi(i)\pi(j)} = 1$ and $x_{\pi(j)\pi(i)} = 0$.

► **Definition 16.** Given a polynomial $f(\{x_{ij} : i \neq j\})$, we define

$$\tilde{E}_n[f(\{x_{ij} : i \neq j\})] = E_{U_n}[f].$$

► **Example 17.** $\forall i \neq j, \tilde{E}_n[x_{ij}] = \frac{1}{2}$ because there is a $\frac{1}{2}$ chance that i comes before j in a random ordering.

► **Example 18.** For all distinct i, j, k , $\tilde{E}_n[x_{ij}x_{jk}] = \frac{1}{6}$ because there is a $\frac{1}{6}$ chance that $i < j < k$ in a random ordering.

However, in order to fully define \tilde{E}_n we have to define $\tilde{E}_n[p]$ for monomials p involving the z variables. We can do this as follows.

► **Definition 19** (Candidate pseudo-expectation values).

1. For all monomials $p(\{x_{ij} : i, j \in [n], i \neq j\})$, we take $\tilde{E}_n[p] = E_{U_n}[p]$.
2. For all monomials $p(\{x_{ij} : i, j \in [n], i \neq j\})$, we take $\tilde{E}_n\left[\left(\prod_{j \in A} z_j\right)p\right] = 0$ whenever $A \subseteq [n]$ is non-empty because each z_j could be positive or negative.
3. For all monomials $p(\{x_{ij} : i, j \in [n], i \neq j\}, \{z_j : j \in [n]\})$ and all $j \in [n]$, we take $\tilde{E}_n[z_j^2 p] = \tilde{E}_n\left[\left(\sum_{i \neq j} x_{ij} - 1\right)p\right]$ because we have that for all j , $z_j^2 = \sum_{i \neq j} x_{ij} - 1$.

3.2 Checking if \tilde{E}_n are pseudo-expectation values

We now discuss what needs to be checked in order to determine whether our candidate pseudo-expectation values \tilde{E}_n are actually degree d pseudo-expectation values. To analyze \tilde{E}_n , it is convenient to create a new variable w_j which is equal to z_j^2 .

► **Definition 20.** Define $w_j = \sum_{i \neq j} x_{ij} - 1$.

Observe that viewing everything in terms of the variables $\{x_{ij}\}$ and $\{w_j\}$, \tilde{E}_n is the expected values over a distribution of solutions. This implies that the polynomial equalities obtained by multiplying one of the ordering principle equations in Definition 9 by a monomial will be satisfied at all degrees, not just up to degree d . However, each w_j is supposed to be a square but this is not actually the case for this distribution, so \tilde{E}_n may fail to give valid pseudo-expectation values. In fact, this is the only way in which \tilde{E}_n can fail to give valid pseudo-expectation values. We make this observation precise with the following lemma:

► **Lemma 21.** *If $\tilde{E}_n \left[\left(\prod_{j \in A} w_j \right) g_A(\{x_{ij} : i, j \in [n], i \neq j\})^2 \right] \geq 0$ whenever $A \subseteq [n]$ and $|A| + \deg(g_A) \leq \frac{d}{2}$ then \tilde{E}_n gives degree d pseudo-expectation values.*

Proof. We first check that $\tilde{E}_n[g^2] \geq 0$ whenever $\deg(g) \leq \frac{d}{2}$ as this is the more interesting part. Given a polynomial g of degree at most $\frac{d}{2}$, decompose g as

$$g = \sum_{A \subseteq [n]} \left(\prod_{j \in A} z_j \right) g_A(x_1, \dots, x_n)$$

where for all $A \subseteq [n]$, $|A| + \deg(g_A) \leq \frac{d}{2}$. Now observe that

$$\begin{aligned} \tilde{E}_n[g^2] &= \tilde{E}_n \left[\sum_{A, A' \subseteq [n]} \left(\prod_{j \in A} z_j \prod_{j \in A'} z_j \right) g_A(x_1, \dots, x_n) g_{A'}(x_1, \dots, x_n) \right] \\ &= \sum_{A \subseteq [n]} \tilde{E}_n \left[\left(\prod_{j \in A} w_j \right) g_A^2(x_1, \dots, x_n) \right] \geq 0. \end{aligned}$$

We now check that the polynomial equalities obtained by multiplying one of the ordering principle equations in Definition 9 by a monomial are satisfied. By the definition of \tilde{E}_n , we have that for all monomials p and all $j \in [n]$,

$$\tilde{E}_n \left[\left(\sum_{i \neq j} x_{ij} - 1 - z_j^2 \right) p \right] = \tilde{E}_n \left[\left(\sum_{i \neq j} x_{ij} - 1 \right) p \right] - \tilde{E}_n \left[\left(\sum_{i \neq j} x_{ij} - 1 \right) p \right] = 0$$

To prove the other polynomial equalities, we use induction on the total degree of the $\{z_j\}$ variables. For the base case, observe that

1. $\tilde{E}_n[1] = E_{U_n}[1] = 1$
2. For all monomials $p(\{x_{ij} : i, j \in [n], i \neq j\})$ and for all ordering or transitivity constraints $s_i = 0$, $\tilde{E}_n[ps_i] = E_{U_n}[ps_i] = 0$.

For the inductive step, if p is a monomial which is divisible by z_j^2 for some j then write $p = z_j^2 p'$. By the inductive hypothesis, for all ordering or transitivity constraints $s_i = 0$,

$$\tilde{E}_n[ps_i] = \tilde{E}_n[z_j^2 p' s_i] = \tilde{E}_n \left[\left(\sum_{i \neq j} x_{ij} - 1 \right) p' s_i \right] = 0.$$

Finally, if p is a monomial of the form $p = \left(\prod_{j \in A} z_j \right) p'(\{x_{ij} : i, j \in [n], i \neq j\})$ where $A \neq \emptyset$ then for all ordering or transitivity constraints $s_i = 0$,

$$\tilde{E}_n[ps_i] = \tilde{E}_n \left[\left(\prod_{j \in A} z_j \right) p' s_i \right] = 0. \quad \blacktriangleleft$$

4 $O(\sqrt{n} \log(n))$ Degree SOS Upper Bound

In this section, we prove theorem 1 by constructing a degree $O(\sqrt{n} \log(n))$ proof of the ordering principle. To construct this proof, we first find a polynomial g of degree $O(\sqrt{n} \log(n))$ such that $\tilde{E}_n[g^2] < 0$. We then show that there is an SOS proof (which in fact uses only polynomial equalities) that $E_{\pi \in S_n}[\pi(g)^2] = \tilde{E}_n[g^2] < 0$.

4.1 Failure of \tilde{E}_n

We now show that \tilde{E}_n fails to give valid pseudo-expectation values at degree $O(\sqrt{n}\log(n))$.

► **Theorem 22.** *For all $n \geq 4$ there exists a polynomial $g(w_1)$ of degree $\lceil \frac{1}{2}\sqrt{n}(\log_2(n) + 1) \rceil$ such that $\tilde{E}[(z_1 g(w_1))^2] = E_{U_n}[w_1 g^2(w_1)] < 0$.*

Proof. Observe that over the uniform distribution of orderings, w_1 is equally likely to be any integer in $[-1, n - 2]$. To make $E_{U_n}[w_1 g^2(w_1)]$ negative, we want $g(w_1)$ to have high magnitude at $w_1 = -1$ and small magnitude on $[1, n - 2]$. For this, we can use Chebyshev polynomials. From Wikipedia [18],

► **Definition 23.** *The m th Chebyshev polynomial can be expressed as*

1. $T_m(x) = \cos(m\cos^{-1}(x))$ if $|x| \leq 1$
2. $T_m(x) = \frac{1}{2} \left((x + \sqrt{x^2 - 1})^m + (x - \sqrt{x^2 - 1})^m \right)$ if $|x| \geq 1$.

► **Theorem 24.** *For all integers $m \geq 0$ and all $x \in [-1, 1]$, $|T_m(x)| \leq 1$.*

We now take $g(w_1) = T_m(-1 + \frac{2w_1}{n})$ where $m = \lceil \frac{1}{2}\sqrt{n}(\log_2(n) + 1) \rceil$ and analyze g .

► **Lemma 25.** *Taking $g(w_1) = T_m(-1 + \frac{2w_1}{n})$ where $m = \lceil \frac{1}{2}\sqrt{n}(\log_2(n) + 1) \rceil$,*

1. $|g(-1)| \geq n$
2. For all $w_1 \in [0, n - 2]$, $|g(w_1)| \leq 1$.

Proof. The second statement follows immediately from Theorem 24 as when $w_1 \in [0, n - 2]$, $-1 + \frac{2w_1}{n} \in [-1, 1]$ so $|g(w_1)| = |T_m(-1 + \frac{2w_1}{n})| \leq 1$. For the first statement, let $\Delta = \frac{2}{n}$ and observe that when $x = -1 - \Delta$,

1. $\sqrt{x^2 - 1} = \sqrt{(1 + \Delta)^2 - 1} \geq \sqrt{2\Delta}$. Thus, $|x - \sqrt{x^2 - 1}| \geq 1 + \sqrt{2\Delta}$.
2. $x + \sqrt{x^2 - 1} < 0$ and $x - \sqrt{x^2 - 1} < 0$ so

$$|T_m(x)| = \frac{1}{2} \left(\left| x + \sqrt{x^2 - 1} \right|^m + \left| x - \sqrt{x^2 - 1} \right|^m \right) \geq \frac{(1 + \sqrt{2\Delta})^m}{2}.$$

We now use the following proposition.

► **Proposition 26.** *For all $y \in [0, 1]$ and all $m \geq 0$, $(1 + y)^m \geq 2^{ym}$.*

Proof. Observe that $(1 + y)^m = \left((1 + y)^{\frac{1}{y}} \right)^{ym}$ and if $y \leq 1$ then $(1 + y)^{\frac{1}{y}} \geq 2$. ◀

Since $n \geq 4$, $\Delta = \frac{2}{n} \leq \frac{1}{2}$ and $\sqrt{2\Delta} \leq 1$. Applying Proposition 26 with $y = \sqrt{2\Delta}$ and recalling that $m = \lceil \frac{1}{2}\sqrt{n}(\log_2(n) + 1) \rceil$,

$$|g(-1)| = |T_m(-1 - \Delta)| \geq \frac{(1 + \sqrt{2\Delta})^m}{2} \geq \frac{2^{\sqrt{2\Delta}m}}{2} = \frac{2^{\frac{2m}{\sqrt{n}}}}{2} \geq 2^{\log_2(n)+1} 2 = n. \quad \blacktriangleleft$$

We can now complete the proof of Theorem 22. By Lemma 25,

$$E_{U_n}[w_1 g^2(w_1)] \leq \frac{1}{n-1} \left(-n^2 + \sum_{j=0}^{n-2} j \right) < 0. \quad \blacktriangleleft$$

4.2 Constructing an SOS proof of infeasibility

We now show that from the failure of \tilde{E}_n , we can construct an SOS proof that the ordering principle equations are infeasible.

► **Theorem 27.** *If there exists a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_n[g^2] < 0$ then there exists an SOS proof of degree at most $2d + 2$ that the ordering principle equations are infeasible.*

Proof. To prove this theorem, we will show that for any monomial $p(\{x_{i,j} : i, j \in [n], i \neq j\})$ of degree at most d , there is a proof of degree at most $2d + 2$ that $\frac{1}{n!} \sum_{\pi \in S_n} \pi(p) = \tilde{E}_n[p]$ which uses only polynomial equalities. To prove this, we observe that given arbitrary indices i_1, \dots, i_k , we can split things into cases based on the order of a_{i_1}, \dots, a_{i_k} .

► **Lemma 28.** *Given the ordering and transitivity axioms, for all $r \in \mathbb{N}$, tuples of distinct indices (i_1, \dots, i_{r+1}) , and permutations $\pi \in S_r$,*

$$\begin{aligned} \prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} &= x_{i_{r+1} i_{\pi(1)}} \prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \\ &+ \sum_{k=1}^{r-1} \left(\prod_{j=1}^{k-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) x_{i_{\pi(k)} i_r} x_{i_r i_{\pi(k+1)}} \left(\prod_{j=k+1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) \\ &+ \prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} x_{i_{\pi(r)} i_{r+1}} \end{aligned}$$

and there is a degree $r + 1$ proof of this fact which uses only polynomial equalities.

► **Remark 29.** The idea behind this lemma is that we have found the correct ordering for i_1, \dots, i_r and we are inserting i_{r+1} into the correct place.

Proof. Using the ordering and transitivity axioms,

1. $\prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} = x_{i_{r+1} i_{\pi(1)}} \left(\prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) + x_{i_{\pi(1)} i_{r+1}} \left(\prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right)$
2. For all $k \in [r - 1]$,

$$\begin{aligned} &x_{i_{\pi(k)} i_{r+1}} \prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \\ &= (x_{i_{\pi(k)} i_{r+1}} x_{i_{r+1} i_{\pi(k+1)}} + x_{i_{\pi(k)} i_{r+1}} x_{i_{\pi(k+1)} i_{r+1}}) \left(\prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) \\ &= (x_{i_{\pi(k)} i_{r+1}} x_{i_{r+1} i_{\pi(k+1)}} + x_{i_{\pi(k+1)} i_{r+1}}) \left(\prod_{j=1}^{r-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) \end{aligned}$$

where the second equality follows because of the transitivity axiom

$$x_{i_{\pi(k)} i_{\pi(k+1)}} x_{i_{\pi(k+1)} i_{r+1}} (1 - x_{i_{\pi(k)} i_{r+1}}) = 0$$

which implies that $x_{i_{\pi(k)} i_{\pi(k+1)}} x_{i_{\pi(k+1)} i_{r+1}} x_{i_{\pi(k)} i_{r+1}} = x_{i_{\pi(k)} i_{\pi(k+1)}} x_{i_{\pi(k+1)} i_{r+1}}$.

3. For all $k \in [r - 1]$, we have the transitivity axiom

$$x_{i_{\pi(k)} i_{r+1}} x_{i_{r+1} i_{\pi(k+1)}} (1 - x_{i_{\pi(k)} i_{\pi(k+1)}}) = 0$$

38:10 Sum of Squares Bounds for the Ordering Principle

which implies that $x_{i_{\pi(k)}i_{r+1}}x_{i_{r+1}i_{\pi(k+1)}}x_{i_{\pi(k)}i_{\pi(k+1)}} = x_{i_{\pi(k)}i_{r+1}}x_{i_{r+1}i_{\pi(k+1)}}$. Thus,

$$\begin{aligned} & x_{i_{\pi(k)}i_{r+1}}x_{i_{r+1}i_{\pi(k+1)}} \prod_{j=1}^{r-1} x_{i_{\pi(j)}i_{\pi(j+1)}} \\ &= \left(\prod_{j=1}^{k-1} x_{i_{\pi(j)}i_{\pi(j+1)}} \right) x_{i_{\pi(k)}i_r}x_{i_r i_{\pi(k+1)}} \left(\prod_{j=k+1}^{r-1} x_{i_{\pi(j)}i_{\pi(j+1)}} \right). \end{aligned}$$

The result follows by combining all of these equalities. \blacktriangleleft

► **Corollary 30.** *Given the ordering and transitivity axioms, for all k and all sets of k distinct indices $\{i_1, \dots, i_k\}$,*

$$1 = \sum_{\pi \in S_k} \prod_{j=1}^{k-1} x_{i_{\pi(j)}i_{\pi(j+1)}}$$

and there is a degree $k+1$ proof of this fact which uses only polynomial equalities.

► **Corollary 31.** *For any monomial $p(\{x_{i,j} : i, j \in [n], i \neq j\})$ of degree d whose variables contain a total of k indices i_1, \dots, i_k , there is a proof of degree at most $d+k+1$ that $\frac{1}{n!} \sum_{\pi \in S_n} \pi(p) = \tilde{E}_n[p]$ which uses only polynomial equalities.*

Proof sketch. By Corollary 30,

$$p = \sum_{\pi \in S_k} \prod_{j=1}^{k-1} x_{i_{\pi(j)}i_{\pi(j+1)}} p$$

and there is a proof of this fact of degree at most $d+k+1$ which uses only polynomial equalities. Using the transitivity axioms, we can prove that

$$\sum_{\pi \in S_k} \prod_{j=1}^{k-1} x_{i_{\pi(j)}i_{\pi(j+1)}} p = \sum_{\pi \in S_k: p=1 \text{ when } x_{i_{\pi(1)}} < \dots < x_{i_{\pi(k)}}} \prod_{j=1}^{k-1} x_{i_{\pi(j)}i_{\pi(j+1)}}.$$

Using Corollary 30 again, this implies that there is a proof of degree at most $d+k+1$ which uses only polynomial equations that

$$\frac{1}{n!} \sum_{\pi \in S_n} \pi(p) = Pr_{\pi \in S_n} [p = 1 \text{ when } x_{\pi(1)} < \dots < x_{\pi(n)}] = E_{U_n}[p] = \tilde{E}_n[p]. \quad \blacktriangleleft$$

Now note that given a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_n[g^2] < 0$, g^2 is a polynomial of degree at most d in the variables $\{x_{i,j} : i, j \in [n], i \neq j\}$ and the variables of every monomial of g^2 contain a total of at most d indices. Thus, there is a proof of degree at most $2d+2$ that $\frac{1}{n!} \sum_{\pi \in S_n} \pi(g^2) = \tilde{E}_n[g^2] < 0$ which uses only polynomial equalities and this immediately gives us an SOS proof of degree at most $2d+2$ that the ordering principle equations are infeasible. \blacktriangleleft

Combining Theorem 22 and Theorem 27, we obtain an SOS proof of degree $O(\sqrt{n} \log(n))$ that the equations corresponding to the negation of the ordering principle are infeasible, which proves Theorem 1.

5 Lower Bound Overview

Proving the lower bound is surprisingly subtle. We proceed as follows.

► **Definition 32.** We define $\Omega_{n,d}$ to be the distribution on a variable u with support $[0, n-d] \cap \mathbb{Z}$ and the following probabilities:

$$\Pr[u = k] = \frac{\binom{n-k-1}{d-1}}{\binom{n}{d}}.$$

1. In Section 6, we show that to prove our sum of squares lower bound, it is sufficient to show that for all polynomials g_* of degree at most d , for some $d_2 \geq 2d$, $E_{\Omega_{n,d_2}}[(u-1)g_*(u)^2] \geq 0$. Equivalently,

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} k g_*^2(k+1) \geq g_*^2(0).$$

This reduces the problem to analyzing a distribution on one variable. For the precise statement of this result, see Theorem 33.

2. In Section 7, we observe that an approximation to the above statement is the statement that for some small $\Delta > 0$ (we will take $\Delta = \frac{2d_2}{n}$), taking $g_2(x) = g_*\left(\frac{x}{\Delta} + 1\right)$,

$$\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \geq \Delta^2 g_2^2(-\Delta).$$

We then prove this approximate statement. For the precise statement of this result, see Theorem 50.

3. In Section 8, we analyze the difference between $\Delta \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta)$ and $\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx$ and show that it is small. For the precise statement of this result, see Theorem 67.
4. In Section 9, we analyze the difference between $\Delta \sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} (k\Delta) g_2^2(k\Delta)$ and $\Delta \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta)$. For the precise statement of this result, see Theorem 70.

In Section 10, we put all of these pieces together to prove our SOS lower bound.

6 Reducing Checking \tilde{E} to Analyzing a Single-Variable Distribution

Recall that $\Omega_{n,d}$ is the distribution on a variable u with support $[0, n-d] \cap \mathbb{Z}$ and the following probabilities:

$$\Pr[u = k] = \frac{\binom{n-k-1}{d-1}}{\binom{n}{d}}.$$

In this section, we show that to check that our candidate pseudo-expectation values \tilde{E}_{2n} are valid, it is sufficient to analyze the distribution $\Omega_{n,d}$. In particular, we prove the following theorem:

► **Theorem 33.** For all $d, d_2, n \in \mathbb{N}$ such that $2d \leq d_2 \leq n$, if there is a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$ then there is a polynomial $g_* : \mathbb{R} \rightarrow \mathbb{R}$ of degree at most d such that $E_{\Omega_{n,d_2}}[(u-1)g_*(u)^2] < 0$. Equivalently,

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} k g_*^2(k+1) < g_*^2(0).$$

38:12 Sum of Squares Bounds for the Ordering Principle

To see why this statement is equivalent, observe that

$$\begin{aligned} E_{\Omega_{n,d_2}}[(u-1)g_*(u)^2] &= \sum_{k=0}^{n-d_2} \frac{\binom{n-k-1}{d_2-1}}{\binom{n}{d_2}} (k-1)g_*^2(k) \\ &= \left(\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n}{d_2}} kg_*^2(k+1) \right) - \frac{\binom{n-1}{d_2-1}}{\binom{n}{d_2}} g_*^2(0). \end{aligned}$$

Thus, $E_{\Omega_{n,d_2}}[(u-1)g_*(u)^2] < 0$ if and only if

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n}{d_2}} kg_*^2(k+1) < \frac{\binom{n-1}{d_2-1}}{\binom{n}{d_2}} g_*^2(0).$$

Multiplying both sides of this inequality by $\frac{\binom{n}{d_2}}{\binom{n-1}{d_2-1}}$, this is equivalent to

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} kg_*^2(k+1) < g_*^2(0).$$

In the remainder of this section, we prove this theorem by starting with the polynomial g and constructing the polynomial g_* .

6.1 Distinguished Indices of g

We first use symmetry to argue that we can take g to be symmetric under permutations of all but d distinguished indices. For this, we use Theorem 4.1 in [13], which is essentially implied by Corollary 2.6 of [14].

► **Definition 34.** *The index degree of a polynomial g is the maximum number of indices mentioned in any monomial of g .*

► **Example 35.** $g = x_{12}x_{13} + x_{45}^4$ has index degree 3 and degree 4.

► **Theorem 36.** *If \tilde{E} is a linear map from polynomials to \mathbb{R} which is symmetric with respect to permutations of $[1, n]$ then for any polynomial g , we can write*

$$\tilde{E}[g^2] = \sum_{I \subseteq [1, n], j: |I| \leq \text{indexdeg}(g)} \tilde{E}[g_{Ij}^2]$$

where for all I, j ,

1. g_{Ij} is symmetric with respect to permutations of $[1, n] \setminus I$
2. $\text{indexdeg}(g_{Ij}) \leq \text{indexdeg}(g)$ and $\text{deg}(g_{Ij}) \leq \text{deg}(g)$
3. $\forall i \in I, \sum_{\pi \in S_{[1, n] \setminus (I \setminus \{i\})}} \pi(g_{Ij}) = 0$.

► **Remark 37.** The statement that $\text{deg}(g_{Ij}) \leq \text{deg}(g)$ is not in Theorem 4.1 as stated in [13] but it follows from the proof.

By Theorem 36, if there is a polynomial g_0 of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g_0^2] < 0$ then there is a polynomial g of degree at most $\frac{d}{2}$ such that

1. $\tilde{E}_{2n}[g^2] < 0$
2. g is symmetric under permutations of $[2n] \setminus I$ for some $I \subseteq [2n]$ such that $|I| \leq \text{indexdeg}(g_0) \leq 2\text{deg}(g_0) \leq d$,

where $\text{indexdeg}(g_0) \leq 2\text{deg}(g_0)$ because all of our variables mention at most two indices.

6.2 Decomposing g Based on z_j Variables

We now show that we can choose g to be a polynomial of the form $g = \left(\prod_{j \in A} z_j\right) g_A$ where $A \subseteq [n]$ and g_A is a polynomial in the x_{ij} variables. To do this, just as in Section 3.2, we decompose g as $g = \sum_{A \subseteq [n]: |A| \leq \frac{d}{2}} \left(\prod_{j \in A} z_j\right) g_A$ where each g_A is a polynomial in the x_{ij} variables and observe that

$$\begin{aligned} \tilde{E}_{2n}[g^2] &= \tilde{E}_{2n} \left[\sum_{A, A' \subseteq [n]} \left(\prod_{j \in A} z_j \prod_{j \in A'} z_j \right) g_A g_{A'} \right] \\ &= \sum_{A \subseteq [2n]} \tilde{E}_{2n} \left[\left(\prod_{j \in A} z_j^2 \right) g_A^2 \right]. \end{aligned}$$

If $\tilde{E}_{2n}[g^2] < 0$ then there must be an $A \subseteq [2n]$ such that $\tilde{E}_{2n} \left[\left(\prod_{j \in A} z_j^2 \right) g_A^2 \right] < 0$. Thus, we can take $g = \left(\prod_{j \in A} z_j\right) g_A$. Note that $g = \left(\prod_{j \in A} z_j\right) g_A$ is symmetric under permutations of $[2n] \setminus I'$ where $I' = I \cup A$ and thus $|I'| \leq 2d$.

6.3 Choosing an Ordering on the Distinguished Indices and Changing Variables

We now further decompose $\tilde{E}_{2n}[g^2]$ by observing that for any set of indices $I'' = \{i_1, \dots, i_m\}$,

$$\tilde{E}_{2n}[g^2] = \tilde{E}_{2n} \left[\left(\sum_{\pi \in S_m} \prod_{j=1}^{m-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) g^2 \right].$$

Since $\tilde{E}_{2n}[g^2] < 0$, there must be a $\pi \in S_m$ such that $\tilde{E}_{2n} \left[\left(\sum_{\pi \in S_m} \prod_{j=1}^{m-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) g^2 \right] < 0$. Thus, we can restrict our attention to $\tilde{E}_{2n} \left[\left(\sum_{\pi \in S_m} \prod_{j=1}^{m-1} x_{i_{\pi(j)} i_{\pi(j+1)}} \right) g^2 \right]$ which effectively imposes the ordering $x_{i_{\pi(1)}} < \dots < x_{i_{\pi(m)}}$.

For technical reasons, we take I'' to be $I' = I \cup A$ plus some additional indices so that $|I''| = d_2$. Without loss of generality, we can assume that $I'' = [d_2]$ and π is the identity, giving the ordering $x_1 < x_2 < \dots < x_{d_2}$.

We now observe that under this ordering, for all $j \in [d_2]$,

$$z_j^2 = (j-2) + \sum_{i \in [2n] \setminus [d_2]} x_{ij}^2.$$

Thus, for all $j \in [2, d_2]$, z_j^2 is a sum of squares so $\left(\prod_{j \in A \setminus \{1\}} z_j^2\right) g_A^2$ is a sum of squares. This implies that $1 \in A$ as otherwise $\tilde{E}_{2n}[g^2] \geq 0$. Following similar logic as before, there is a polynomial $g_{\{1\}}$ in the x_{ij} variables of degree at most $\frac{d}{2} - 1$ such that

$$\tilde{E}_{2n} \left[\left(\prod_{i=1}^{d_2-1} x_{i(i+1)} \right) z_1^2 g_{\{1\}}^2 \right] < 0.$$

Now observe that by symmetry, under the ordering $x_1 < x_2 < \dots < x_{d_2}$, we can express $g_{\{1\}}$ in terms of the following new variables:

38:14 Sum of Squares Bounds for the Ordering Principle

► **Definition 38.** For $i \in [d_2] \cup \{0\}$, we define the variable u_i so that

1. $u_0 = \sum_{j \in [n] \setminus [d_2]} x_{j1}$ is the number of elements before a_1 .
2. For $i \in [d_2 - 1]$, $u_i = \sum_{j \in [n] \setminus [d_2]} x_{ij} x_{j(i+1)}$ is the number of elements between a_i and a_{i+1} .
3. $u_{d_2} = \sum_{j \in [n] \setminus [d_2]} x_{d_2 j}$ is the number of elements after a_{d_2} .

With these new variables,

$$\tilde{E}_{2n} \left[\left(\prod_{i=1}^{d_2-1} x_{i(i+1)} \right) z_1^2 g_{\{1\}}^2 \right] = \frac{1}{d_2!} E_{u_0, \dots, u_{d_2} \in \mathbb{N} \cup \{0\}: \sum_{j=0}^{d_2} u_j = 2n - d_2} [(u_0 - 1) g_{\{1\}}^2(u_0, \dots, u_{d_2})] < 0$$

where the $\frac{1}{d_2!}$ term appears because the probability of having the ordering $x_1 < x_2 < \dots < x_{d_2}$ is $\frac{1}{d_2!}$.

6.4 Reducing to a Single Variable

We now complete the proof of Theorem 33 by constructing $g_*(u_0)$ and proving that it has the needed properties. To construct g_* , we take

$$g_*(u_0) = E_{u_1, \dots, u_{d_2} \in \mathbb{N} \cup \{0\}: \sum_{j=1}^{d_2} u_j = 2n - d_2 - u_0} [g_{\{1\}}(u_0, \dots, u_{d_2})^2].$$

We first need to check that $g_*(u_0)$ is indeed a polynomial of degree at most d in u_0 . This follows from the following lemma:

► **Lemma 39.** For all $d_2 \in \mathbb{N}$ and any polynomial $p(u_1, \dots, u_{d_2})$ of degree at most d ,

$$E_{u_1, \dots, u_{d_2} \in \mathbb{N} \cup \{0\}: \sum_{j=1}^{d_2} u_j = n'} p(u_1, \dots, u_{d_2})$$

is a polynomial of degree at most d in n' (for $n' \in \mathbb{N} \cup \{0\}$).

Proof sketch. We illustrate why this lemma is true by computing these expected values for a few monomials in the variables $\{u_1, \dots, u_{d_2}\}$. The ideas used in these computations can be generalized to any monomial. The idea is to consider placing $d_2 - 1$ dividing lines among n' labeled balls in a random order.

► **Example 40.** With 2 balls and 2 bins, the possibilities are as follows:

1. 12|: Balls 1 and 2 are in the first bin in the order 1, 2.
2. 21|: Balls 1 and 2 are in the first bin in the order 2, 1.
3. 1|2: Ball 1 is in the first bin and ball 2 is in the second bin.
4. 2|1: Ball 2 is in the first bin and ball 1 is in the second bin.
5. |12: Balls 1 and 2 are in the second bin in the order 1, 2.
6. |21: Balls 1 and 2 are in the second bin in the order 2, 1.

To analyze monomials in the variables $\{u_1, \dots, u_{d_2}\}$, we write $u_j = \sum_{i=1}^{n'} t_{ij}$ where $t_{ij} = 1$ if ball i is in bin j and $t_{ij} = 0$ otherwise.

1. By symmetry, the probability that a given ball is put into the first bin is $\frac{1}{d_2}$. Thus, $E[u_1] = n' E[t_{i1}] = \frac{n'}{d_2}$.
2. If we consider balls i and j where $i \neq j$, the probability that ball i is put into the first bin is $\frac{1}{d_2}$. If ball i is placed into the first bin, this effectively splits the first bin into two bins, the part before ball i and the part after ball i . For ball j , the probability that it is put into one of these parts is $\frac{2}{d_2+1}$ and the probability that it is put into the second bin is $\frac{1}{d_2+1}$. Thus, $E[t_{i1} t_{j1}] = \frac{2}{d_2(d_2+1)}$ and $E[t_{i1} t_{j2}] = \frac{1}{d_2(d_2+1)}$. This implies that $E[u_1 u_2] = n'(n' - 1) E[t_{i1} t_{j2}] = \frac{n'(n'-1)}{d_2(d_2+1)}$ and $E[u_1^2] = n' E[t_{i1}] + n'(n' - 1) E[t_{i1} t_{j1}] = \frac{n'}{d_2} + \frac{2n'(n'-1)}{d_2(d_2+1)}$.

Following similar ideas, we can analyze any monomial of degree at most d and show that its expected value is a polynomial in n' of degree at most d . ◀

To complete the proof of Theorem 33, we need one more technical lemma.

► **Lemma 41.** For all $n, k, d_2 \in \mathbb{N}$ such that $k \leq n - d_2$,

$$\frac{\binom{n-k-1}{d_2-1}}{\binom{n-1}{d_2-1}} \leq \left(\frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2.$$

Proof. Observe that for all k', n' such that $0 < k' \leq n'$, $\left(\frac{2n'-k'}{2n'} \right)^2 = 1 - \frac{k'}{n'} + \frac{k'^2}{4n'^2} > 1 - \frac{k'}{n'}$.

Now observe that

$$\begin{aligned} \left(\frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 &= \prod_{j=1}^{d_2-1} \left(\frac{2n-k-j}{2n-j} \right)^2 \geq \prod_{j=1}^{d_2-1} \left(\frac{2n-k-2j}{2n-2j} \right)^2 \\ &\geq \prod_{j=1}^{d_2-1} \frac{n-k-j}{n-j} = \frac{\binom{n-k-1}{d_2-1}}{\binom{n-1}{d_2-1}}. \end{aligned} \quad \blacktriangleleft$$

We now complete the proof of Theorem 33. Recall that Ω_{n,d_2} is the distribution on a variable u with support $[0, n - d_2] \cap \mathbb{Z}$ and the following probabilities

$$Pr[u = k] = \frac{\binom{n-k-1}{d_2-1}}{\binom{n}{d_2}}.$$

We have the following facts:

1. $E_{\Omega_{2n,d_2}}[(u-1)g_*(u)] = E_{u_0, \dots, u_{d_2} \in \mathbb{N} \cup \{0\}: \sum_{j=0}^{d_2} u_j = 2n-d_2} [(u_0-1)g_{\{1\}}^2(u_0, \dots, u_{d_2})] < 0$
2. For all $u_0 \in [0, 2n - d_2] \cap \mathbb{Z}$, $g_*(u) \geq 0$

► **Remark 42.** Intuitively, g_* should already be a sum of squares. However, we are not sure how to prove this, so we instead show that g_*^2 is sufficient for our purposes.

Since $E_{\Omega_{2n,d_2}}[(u-1)g_*(u)] < 0$,

$$\sum_{k=1}^{2n-d_2} \frac{\binom{2n-k-1}{d_2-1}}{\binom{2n}{d_2}} (k-1)g_*(k) < \frac{\binom{2n-1}{d_2-1}}{\binom{2n}{d_2}} g_*(0)$$

which implies that

$$\sum_{k=1}^{2n-d_2} \frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} (k-1) \frac{g_*(k)}{g_*(0)} < 1.$$

In turn, this implies that

$$\sum_{k=1}^{2n-d_2} \left(\frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 (k-1) \frac{g_*^2(k)}{g_*^2(0)} \leq \sum_{k=1}^{2n-d_2} \left(\frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 (k-1)^2 \frac{g_*^2(k)}{g_*^2(0)} < 1.$$

Using Lemma 41,

$$\sum_{k=1}^{n-d_2} \frac{\binom{n-k-1}{d_2-1}}{\binom{n-1}{d_2-1}} (k-1) \frac{g_*^2(k)}{g_*^2(0)} \leq \sum_{k=1}^{2n-d_2} \left(\frac{\binom{2n-k-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 (k-1) \frac{g_*^2(k)}{g_*^2(0)} < 1.$$

Multiplying both sides by $g_*^2(0)$,

$$\sum_{k=1}^{n-d_2} \frac{\binom{n-k-1}{d_2-1}}{\binom{n-1}{d_2-1}} (k-1)g_*^2(k) = \sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} kg_*^2(k+1) < g_*^2(0).$$

This implies that $E_{\Omega_{n,d_2}}[(u-1)g_*^2(u)] < 0$, as needed.

7 Approximate Analysis for Ω_{n,d_2}

To prove our SOS lower bound, we need to show that for any polynomial g_* of degree at most d , $E_{\Omega_{n,d_2}}[(u-1)g_*^2(u)] \geq 0$. Equivalently, we need to show that for any polynomial g_* of degree at most d ,

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} k g_*^2(k+1) \geq g_*^2(0).$$

7.1 Approximation by an Integral

The expression $\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} k g_*^2(k+1)$ is hard to analyze, so we approximate it by an integral. Observe that as long as $k \ll n$, $k d_2^2 \ll n^2$ and $k^2 d_2 \ll n$,

$$\frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} = \prod_{j=1}^{d_2-1} \left(1 - \frac{k}{n}\right) \frac{n-j-k-1}{1 - \frac{k}{n}} \approx \left(1 - \frac{k}{n}\right)^{d_2-1} \approx e^{-\frac{d_2 k}{n}}$$

as

$$1 - \frac{n-j-k-1}{1 - \frac{k}{n}} = \frac{(n-k)(n-j) - n(n-j-k-1)}{(n-k)(n-j)} = \frac{jk+n}{(n-k)(n-j)}$$

which is small. Taking $\Delta = \frac{d_2}{n}$ and $g_2(x) = g_*\left(\frac{x}{\Delta} + 1\right)$, approximately what we need to show is that for all polynomials g_2 of degree at most d ,

$$\frac{1}{\Delta} \sum_{j=0}^{\infty} (j\Delta) e^{-j\Delta} g_2^2(j\Delta) \geq g_2^2(-\Delta).$$

In turn, this statement is approximately the same as the statement that for all polynomials g_2 of degree at most d_2 ,

$$\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \geq \Delta^2 g_2^2(-\Delta).$$

In the remainder of this section, we prove this statement when $d d_2 \ll n$ by analyzing the distribution $\mu(x) = x e^{-x}$. In Sections 8 and 9, we will then analyze how to bound the difference between this statement and the statement which we actually need to prove.

► **Remark 43.** For technical reasons, we will actually take $\Delta = \frac{2d_2}{n}$ rather than $\Delta = \frac{d_2}{n}$. For details, see Section 9.

► **Remark 44.** We might think that the probability that x is much more than $\log(n)$ is very small and can be ignored. If so, then using Chebyshev polynomials would cause this statement to fail at degree $\tilde{O}\left(\sqrt{\frac{n}{d}}\right)$ which is much less than \sqrt{n} . However, this is not correct. Intuitively, since we are considering polynomials of degree up to d , we should consider the point where $x^d e^{-x}$ becomes negligible, which is when x is a sufficiently large constant times $d \log(d)$.

Based on this, we can only ignore u_0 which are a sufficiently large constant times $d \log(d) \frac{n}{d_2}$. Roughly speaking, we will want to ignore all $u_0 > \frac{n}{4}$, so we want d_2 to be at least $C d \log(d)$ for some sufficiently large constant C . For details, see Section 9.

7.2 Orthonormal Basis for $\mu(x) = xe^{-x}$

In order to analyze $\int_{x=0}^{\infty} g^2(x)xe^{-x}dx$, it is very useful to find the unique orthonormal basis $\{h_k : k \in \mathbb{N} \cup \{0\}\}$ for the distribution $\mu(x) = xe^{-x}$ such that h_k has degree k and the leading coefficient of h_k is positive. In this subsection, we find this orthonormal basis.

► **Definition 45.** Given two polynomials f and g , we define $f \cdot g = \int_{x=0}^{\infty} f(x)g(x)xe^{-x}dx$.

► **Definition 46.** We define h_k to be the degree k polynomial such that the leading coefficient of h_k is positive, $h_k \cdot h_k = 1$, and for all $j < k$, $h_j \cdot h_k = 0$.

► **Lemma 47.**

$$h_k(x) = \frac{1}{\sqrt{k!(k+1)!}} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} x^j$$

Proof.

► **Proposition 48.** $x^p \cdot x^q = (p+q+1)!$

Computing directly using Gram-Schmidt, the first few polynomials in the orthonormal basis are

1. $h_0 = 1$
2. $h_1 = \frac{1}{\sqrt{2}}(x - 2)$
3. $h_2 = \frac{1}{\sqrt{12}}(x^2 - 6x + 6)$
4. $h_3 = \frac{1}{\sqrt{144}}(x^3 - 12x^2 + 36x - 24)$
5. $h_4 = \frac{1}{\sqrt{2880}}(x^4 - 20x^3 + 120x^2 - 240x + 120)$.

To check the general pattern, we need to check that for all $i \in [0, k-1]$, $h_k \cdot x^i = 0$ and $h_k \cdot h_k = 1$. To see this, observe that for all $i \geq 0$,

$$\begin{aligned} h_k \cdot x^i &= \frac{1}{\sqrt{k!(k+1)!}} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} (i+j+1)! \\ &= \frac{1}{\sqrt{k!(k+1)!}} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \binom{i+j+1}{j+1} (k+1)!i! \end{aligned}$$

Now observe that for all k and all functions $f(j)$,

$$\sum_{j=0}^k (-1)^{k-j} \binom{k}{j} f(j) = (\Delta^k f)(0)$$

where $(\Delta f)(x) = f(x+1) - f(x)$.

► **Proposition 49.** If $f = x^i$ then $\Delta^k f = 0$ if $i < k$ and $\Delta^k f = k!$ if $i = k$.

Viewing $\binom{i+j+1}{j+1}$ as a polynomial in j ,

$$\binom{i+j+1}{j+1} = \frac{(i+j+1)!}{i!(j+1)!} \frac{j^i}{i!} + \text{lower order terms}$$

Putting everything together,

1. $h_k \cdot x^i = 0$ whenever $i \leq k$.
2. $h_k \cdot h_k = \frac{1}{\sqrt{k!(k+1)!}} (h_k \cdot x^k) = \frac{k!(k+1)!}{k!(k+1)!} (\Delta^k \binom{k+j+1}{j+1})(0) = 1$. ◀

7.3 Proof of the Approximate Statement

Now that we have the orthonormal basis for $\mu(x) = xe^{-x}$, we prove the approximate statement we need.

► **Theorem 50.** *For all $d \in \mathbb{N}$ and all $\Delta > 0$ such that $10(d+1)^2\Delta^2e^{2d\Delta} \leq 1$, for any polynomial g_2 of degree at most d ,*

$$\int_{x=0}^{\infty} g_2^2(x)xe^{-x}dx \geq 10\Delta^2g_2^2(-\Delta).$$

Proof. Given a polynomial g_2 of degree at most d , write $g_2 = \sum_{k=0}^d a_k h_k$. Since $\{h_k\}$ is an orthonormal basis for $\mu(x) = xe^{-x}$,

$$\int_{x=0}^{\infty} g_2^2(x)xe^{-x}dx = \sum_{k=0}^d a_k^2.$$

Using Cauchy Schwarz, we have that

$$\sum_{k=0}^d |a_k| \leq \sqrt{\left(\sum_{k=0}^d a_k^2\right) \left(\sum_{k=0}^d 1\right)} = \sqrt{(d+1)} \sqrt{\sum_{k=0}^d a_k^2}$$

which implies that $\sum_{k=0}^d a_k^2 \geq \frac{(\sum_{k=0}^d |a_k|)^2}{d+1}$.

In order to upper bound $|g_2(-\Delta)|$, we need to bound $h_k(x)$ near $x = 0$. For this, we use the following lemma:

► **Lemma 51.** *For all $k \in \mathbb{N}$ and all $x \in \mathbb{R}$,*

$$|h_k(x)| \leq \sqrt{k+1}e^{k|x|}.$$

Proof. Observe that

$$\begin{aligned} h_k(x) &\leq \frac{1}{\sqrt{k!(k+1)!}} \sum_{j=0}^k \binom{k}{j} \frac{(k+1)!}{(j+1)!} |x|^j \\ &\leq \sqrt{k+1} \sum_{j=0}^k \frac{(k|x|)^j}{j!(j+1)!} \leq \sqrt{k+1}e^{k|x|}. \end{aligned}$$

By Lemma 51,

$$|g_2(-\Delta)| \leq \sum_{k=0}^d |a_k| \sqrt{k+1}e^{k\Delta} \leq \sqrt{d+1}e^{d\Delta} \sum_{k=0}^d |a_k|.$$

Thus, $g_2^2(-\Delta) \leq (d+1)e^{2d\Delta} \left(\sum_{k=0}^d |a_k|\right)^2$.

Putting everything together, as long as $10(d+1)^2\Delta^2e^{2d\Delta} \leq 1$, $\int_{x=0}^{\infty} g_2^2(x)xe^{-x}dx \geq 10\Delta^2g_2^2(-\Delta)$, as needed. ◀

8 Handling Numerical Integration Error

In this section, we show how to bound the difference between $\Delta \sum_{j=0}^{\infty} (j\Delta)e^{-j\Delta}g_2^2(j\Delta)$ and $\int_{x=0}^{\infty} g_2^2(x)xe^{-x}dx$.

8.1 Bounding Numerical Integration Error via Higher Derivatives

In this subsection, we describe how the numerical integration error can be bounded using higher derivatives.

► **Lemma 52.** For any $\Delta > 0$ and any differentiable function $f : [0, \infty) \rightarrow \mathbb{R}$,

$$\left| \int_{x=0}^{\infty} f(x) dx - \Delta \sum_{j=0}^{\infty} f(j\Delta) \right| \leq \Delta \int_{x=0}^{\infty} |f'(x)| dx.$$

Proof. This result follows by summing the following proposition over all $j \in \mathcal{N} \cup \{0\}$ and using the fact that $|a + b| \leq |a| + |b|$.

► **Proposition 53.** For all $j \in \mathcal{N} \cup \{0\}$,

$$\left| \int_{x=j\Delta}^{(j+1)\Delta} (f(x) - f(j\Delta)) dx \right| \leq \Delta \int_{x=j\Delta}^{(j+1)\Delta} |f'(x)| dx.$$

Proof. Observe that for all $j \in \mathcal{N} \cup \{0\}$, for all $x \in [j\Delta, (j+1)\Delta]$,

$$|f(x) - f(j\Delta)| \leq \int_{x=j\Delta}^{(j+1)\Delta} |f'(x)| dx.$$

Thus,

$$\left| \int_{x=j\Delta}^{(j+1)\Delta} (f(x) - f(j\Delta)) dx \right| \leq \Delta \int_{x=j\Delta}^{(j+1)\Delta} |f'(x)| dx. \quad \blacktriangleleft$$

Using higher derivatives, we can get better bounds on the error.

► **Lemma 54.** For any $\Delta > 0$ and any twice differentiable function $f : [0, \infty) \rightarrow \mathbb{R}$,

$$\left| \int_{x=0}^{\infty} f(x) dx - \Delta \sum_{j=0}^{\infty} f(j\Delta) + \frac{\Delta}{2} f(0) \right| \leq \Delta^2 \int_{x=0}^{\infty} |f''(x)| dx.$$

Proof. This result follows from summing the following lemma over all $j \in \mathcal{N} \cup \{0\}$ and using the fact that $|a + b| \leq |a| + |b|$.

► **Lemma 55.** For all $j \in \mathcal{N} \cup \{0\}$,

$$\left| \int_{x=j\Delta}^{(j+1)\Delta} f(x) dx - \frac{\Delta}{2} (f(j\Delta) + f((j+1)\Delta)) \right| \leq \Delta^2 \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)| dx.$$

Proof. We prove this lemma using the following estimate of $f(x)$ for $x \in [j\Delta, (j+1)\Delta]$

► **Proposition 56.** For all $x \in [j\Delta, (j+1)\Delta]$,

$$|f(x) - f(j\Delta) - (x - j\Delta)f'(j\Delta)| \leq (x - j\Delta) \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)| dx.$$

38:20 Sum of Squares Bounds for the Ordering Principle

Proof. Observe that for all $j \in \mathcal{N} \cup \{0\}$, for all $x \in [j\Delta, (j+1)\Delta]$,

$$|f'(x) - f'(j\Delta)| \leq \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)|.$$

Taking the integral of this equation from $j\Delta$ to x and using the fact that $|a+b| \leq |a|+|b|$,

$$|f(x) - f(j\Delta) - (x - j\Delta)f'(j\Delta)| \leq (x - j\Delta) \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)|. \quad \blacktriangleleft$$

We now make the following observations:

1. By Proposition 56,

$$|f(j\Delta) + \Delta f'(j\Delta) - f((j+1)\Delta)| = |f((j+1)\Delta) - f(j\Delta) - \Delta f'(j\Delta)| \leq \Delta \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)| dx.$$

2. Taking the integral of Proposition 56 from $j\Delta$ to $(j+1)\Delta$ and using the fact that $|a+b| \leq |a|+|b|$,

$$\left| \int_{x=j\Delta}^{(j+1)\Delta} f(x) dx - \Delta f(j\Delta) - \frac{\Delta^2}{2} f'(j\Delta) \right| \leq \frac{\Delta^2}{2} \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)|$$

Adding $\frac{\Delta}{2}$ times the first equation to the second equation, we have that

$$\left| \int_{x=j\Delta}^{(j+1)\Delta} f(x) dx - \frac{\Delta}{2} (f(j\Delta) + f((j+1)\Delta)) \right| \leq \Delta^2 \int_{x=j\Delta}^{(j+1)\Delta} |f''(x)| dx$$

as needed. \(\blacktriangleleft\)

We now generalize this argument to $(t+1)$ th derivatives.

► **Definition 57.** For all $t \in \mathbb{N}$, we define M_t to be the $(t+1) \times (t+1)$ matrix with entries $(M_t)_{ab} = (b-1)^{(a-1)}$ where $(M_t)_{11} = 1$. Note that M_t is a Vandermonde matrix and is thus invertible.

► **Definition 58.** For all $t \in \mathbb{N}$, we define v_t to be the vector of length $t+1$ with entries $(v_t)_a = \frac{t^{a-1}}{a}$ and we define $c_t = M_t^{-1} v_t$.

► **Lemma 59.** For all $t \in \mathbb{N}$, for any $\Delta > 0$ and any function $f: [0, \infty) \rightarrow \mathbb{R}$ which can be differentiated $t+1$ times,

$$\begin{aligned} & \left| \frac{1}{t} \left(\sum_{b=0}^{t-1} \int_{x=b\Delta}^{\infty} f(x) dx \right) - \Delta \sum_{j=0}^{\infty} f(j\Delta) + \Delta \sum_{j=0}^{t-1} \left(\sum_{b=j+1}^t (c_t)_{b+1} \right) f(j\Delta) \right| \\ & \leq (t\Delta)^{t+1} \left(\frac{t}{(t+1)!} + \frac{\sum_{b=1}^t |(c_t)_{b+1}|}{t!} \right) \int_{x=0}^{\infty} |f^{(t+1)}(x)| dx. \end{aligned}$$

Proof. This result follows from summing the following lemma over all $j \in \mathcal{N} \cup \{0\}$:

► **Lemma 60.** For all $j \in \mathcal{N} \cup \{0\}$,

$$\begin{aligned} & \left| \frac{1}{t} \int_{x=j\Delta}^{(j+t)\Delta} f(x) dx - \Delta \sum_{b=0}^t (c_t)_{b+1} f((j+b)\Delta) \right| \\ & \leq (t\Delta)^{t+1} \left(\frac{1}{(t+1)!} + \frac{\sum_{b=1}^t |(c_t)_{b+1}|}{t(t!)} \right) \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx. \end{aligned}$$

Proof. We prove this lemma using the following estimate of $f(x)$ for $x \in [j\Delta, (j+t)\Delta]$.

► **Proposition 61.** For all $x \in [j\Delta, (j+t)\Delta]$,

$$\left| f(x) - \sum_{a=0}^t \frac{(x-j\Delta)^a}{a!} f^{(a)}(j\Delta) \right| \leq \frac{(x-j\Delta)^t}{t!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx.$$

Proof. Observe that for all $j \in \mathcal{N} \cup \{0\}$, for all $x \in [j\Delta, (j+t)\Delta]$,

$$|f^{(t)}(x) - f^{(t)}(j\Delta)| \leq \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)|.$$

Taking the integral of this equation from $j\Delta$ to x t times and using the fact that $|a+b| \leq |a| + |b|$,

$$\left| f(x) - \sum_{a=0}^t \frac{(x-j\Delta)^a}{a!} f^{(a)}(j\Delta) \right| \leq \frac{(x-j\Delta)^t}{t!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx. \quad \blacktriangleleft$$

We now make the following observations:

1. By Proposition 61, for all $b \in [t]$,

$$\left| \sum_{a=0}^t \frac{(b\Delta)^a}{a!} f^{(a)}(j\Delta) - f(j\Delta + b) \right| \leq \frac{(b\Delta)^t}{t!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)|.$$

2. Taking the integral of Proposition 61 from $j\Delta$ to $(j+t)\Delta$ and using the fact that $|a+b| \leq |a| + |b|$,

$$\left| \int_{x=j\Delta}^{(j+t)\Delta} f(x) dx - \sum_{a=0}^t \frac{(t\Delta)^{a+1}}{(a+1)!} f^{(a)}(j\Delta) \right| \leq \frac{(t\Delta)^{t+1}}{(t+1)!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx.$$

Adding $(c_t)_{b+1}$ times the first equation for each $b \in [t]$ to $\frac{1}{t}$ times the second equation, we have that

$$\begin{aligned} & \frac{1}{t} \int_{x=j\Delta}^{(j+t)\Delta} f(x) dx - \sum_{a=0}^t \Delta^{a+1} \left(\frac{t^a}{(a+1)!} - \sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} \right) f^{(a)}(j\Delta) - \sum_{b=1}^t \Delta (c_t)_{b+1} f(j\Delta + b) \\ & \leq \frac{(t\Delta)^{t+1}}{(t+1)!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx + \sum_{b=1}^t \frac{\Delta |(c_t)_{b+1}| (b\Delta)^t}{t!} \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx \\ & \leq (t\Delta)^{t+1} \left(\frac{1}{(t+1)!} + \frac{\sum_{b=1}^t |(c_t)_{b+1}|}{t(t!)} \right) \int_{x=j\Delta}^{(j+t)\Delta} |f^{(t+1)}(x)| dx. \end{aligned}$$

Thus, it is sufficient to show the following:

1. If $a = 0$ then $\frac{t^a}{(a+1)!} - \sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = (c_t)_1$.
2. If $a \in [t]$ then $\frac{t^a}{(a+1)!} - \sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = 0$.

38:22 Sum of Squares Bounds for the Ordering Principle

To see these statements, observe that

$$\sum_{b=0}^t b^a (c_t)_{b+1} = \sum_{b=1}^{t+1} (M_t)_{(a+1),b} (c_t)_b = (v_t)_{a+1} = \frac{t^a}{a+1}.$$

Thus, for all $a \in [t] \cup \{0\}$

$$\frac{t^a}{(a+1)!} - \sum_{b=0}^t \frac{b^a (c_t)_{b+1}}{a!} = 0.$$

When $a = 0$, $\sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = \sum_{b=0}^t \frac{b^a (c_t)_{b+1}}{a!} - (c_t)_1$ so $\frac{t^a}{(a+1)!} - \sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = (c_t)_1$.

When $a > 0$, $\sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = \sum_{b=0}^t \frac{b^a (c_t)_{b+1}}{a!}$ so $\frac{t^a}{(a+1)!} - \sum_{b=1}^t \frac{b^a (c_t)_{b+1}}{a!} = 0$. ◀

8.2 Bounds on h_k

In order to use our tools, we need bounds on the integrals of the functions h_k .

► **Lemma 62.** For all $j, j' \in \mathbb{N} \cup \{0\}$, $\int_{x=0}^{\infty} |h_j(x)h_{j'}(x)|xe^{-x} dx \leq 1$.

Proof. Observe that

$$\int_{x=0}^{\infty} |h_j(x)h_{j'}(x)|xe^{-x} dx \leq \int_{x=0}^{\infty} \frac{h_j^2(x) + h_{j'}^2(x)}{2} xe^{-x} dx = 1. \quad \blacktriangleleft$$

► **Lemma 63.** For all $j \in \mathbb{N} \cup \{0\}$, $\int_{x=0}^{\infty} h_j^2(x)e^{-x} dx \leq j + 8$.

Proof. The cases where $j = 0$ and $j = 1$ can be computed directly. For $j > 1$, observe that by Lemma 51,

$$\begin{aligned} \int_{x=0}^{\infty} h_j^2(x)e^{-x} dx &= \int_{x=0}^{\frac{1}{j}} h_j^2(x)e^{-x} dx + \int_{x=\frac{1}{j}}^{\infty} h_j^2(x)e^{-x} dx \\ &\leq \int_{x=0}^{\frac{1}{j}} (j+1)e^{2jx}e^{-x} dx + j \int_{x=\frac{1}{j}}^{\infty} h_j^2(x)xe^{-x} dx \\ &\leq \frac{(j+1)}{2j-1}(e^2 - 1) + j \leq j + 8. \quad \blacktriangleleft \end{aligned}$$

► **Corollary 64.** For all $j, j' \in \mathbb{N} \cup \{0\}$, $\int_{x=0}^{\infty} |h_j(x)h_{j'}(x)|e^{-x} dx \leq \sqrt{(j+8)(j'+8)}$.

Proof. Observe that by Lemma 63,

$$\int_{x=0}^{\infty} |h_j(x)h_{j'}(x)|e^{-x} dx \leq \int_{x=0}^{\infty} \frac{\sqrt{\frac{j'+8}{j+8}}h_j^2(x) + \sqrt{\frac{j+8}{j'+8}}h_{j'}^2(x)}{2} e^{-x} dx \leq \sqrt{(j+8)(j'+8)}. \quad \blacktriangleleft$$

8.3 Derivative of h_k

We also need to analyze what happens when we take the derivative of h_k . Calculating directly, the first few derivatives are:

1. $\frac{d(1)}{dx} = 0$
2. $\frac{d(x-2)}{dx} = 1$
3. $\frac{d(x^2-6x+6)}{dx} = 2x - 6 = 2(x-2) - 2$

$$4. \frac{d(x^3 - 12x^2 + 36x - 24)}{dx} = 3x^2 - 24x + 36 = 3(x^2 - 6x + 6) - 6(x - 2) + 6$$

5.

$$\begin{aligned} \frac{d(x^4 - 20x^3 + 120x^2 - 240x + 120)}{dx} &= 4x^3 - 60x^2 + 240x - 240 \\ &= 4(x^3 - 12x^2 + 36x - 24) - 12(x^2 - 6x + 6) + 24(x - 2) - 24. \end{aligned}$$

The general pattern is as follows:

► **Lemma 65.** $h'_k(x) = \sum_{k'=0}^{k-1} (-1)^{k-1-k'} \frac{k!}{k'!} \frac{\sqrt{k'!(k'+1)!}}{\sqrt{k!(k+1)!}} h_j(x)$

Proof. To prove this lemma, we need to show that the derivative of

$$\sqrt{k!(k+1)!} h_k = \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} x^j$$

is

$$\sum_{k'=0}^{k-1} (-1)^{k-1-k'} \frac{k!}{k'!} \left(\sum_{j=0}^{k'} (-1)^{k'-j} \binom{k'}{j} \frac{(k'+1)!}{(j+1)!} x^j \right) = \sum_{k'=0}^{k-1} (-1)^{k-1-k'} \frac{k!}{k'!} \left(\sqrt{k'!(k'+1)!} h_{k'} \right).$$

To prove this, we use the following proposition.

► **Proposition 66.** For all n and all j ,

$$\sum_{k=j}^{n-1} \binom{k}{j} = \binom{n}{j+1}.$$

Proof. Observe that choosing $j+1$ objects out of n objects is equivalent to choosing the position $k+1$ of the last object and then choosing the remaining j objects from the first k objects. ◀

With this proposition in hand, we observe that

$$\begin{aligned} & \sum_{k'=0}^{k-1} (-1)^{k-1-k'} \frac{k!}{k'!} \left(\sum_{j=0}^{k'} (-1)^{k'-j} \binom{k'}{j} \frac{(k'+1)!}{(j+1)!} x^j \right) \\ &= \sum_{j=0}^{k-1} (-1)^{k-1-j} \frac{k!}{(j+1)!} \left(\sum_{k'=j}^{k-1} (k'+1) \binom{k'}{j} \right) x^j \\ &= \sum_{j=0}^{k-1} (-1)^{k-1-j} \frac{k!}{j!} \left(\sum_{k'=j}^{k-1} \binom{k'+1}{j+1} \right) x^j \\ &= \sum_{j=0}^{k-1} (-1)^{k-1-j} \frac{k!}{j!} \binom{k+1}{j+2} x^j \\ &= \sum_{j=1}^k (-1)^{k-j} \frac{k!}{j!} \binom{k+1}{j+1} (jx^{j-1}) = \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} (jx^{j-1}) \\ &= \frac{d \left(\sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} x^j \right)}{dx}. \end{aligned}$$

8.4 Bounding the Numerical Integration Error

In this subsection, we use our tools to bound the numerical integration error

$$\Delta \sum_{j=0}^{\infty} (j\Delta) e^{-j\Delta} g_2^2(j\Delta) - \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

► **Theorem 67.** *For all $t \in \mathbb{N}$, there exist constants $C_{t1}, C_{t2} > 0$ such that for all $\Delta > 0$, $d \in \mathbb{N}$, and polynomials g_2 of degree at most d ,*

$$\begin{aligned} & \left| \Delta \sum_{j=0}^{\infty} (j\Delta) e^{-j\Delta} g_2^2(j\Delta) - \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \right| \\ & \leq (C_{t1}(d\Delta)^2 e^{2td\Delta} + C_{t2}d(d\Delta)^{t+1}) \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx. \end{aligned}$$

Proof. To prove this, we bound $\int_{x=0}^{\infty} \left| \frac{d^{t+1}(g_2^2(x) x e^{-x})}{dx^{t+1}} \right| dx$.

► **Lemma 68.** *For all $t, d \in \mathbb{N}$ If $g_2 = \sum_{i=0}^d a_i h_i$ is a polynomial of degree at most d then*

$$\int_{x=0}^{\infty} \left| \frac{d^{t+1}(g_2(x)^2 x e^{-x})}{dx^{t+1}} \right| dx \leq (t+4)(d+8)(d+1)(3d)^t \left(\sum_{i=0}^d a_i^2 \right).$$

Proof. Observe that

$$\begin{aligned} \frac{d^{t+1}(g_2(x)^2 x e^{-x})}{dx^{t+1}} &= \sum_{j_1=0}^{t+1} \sum_{j_2=0}^{t+1-j_1} \frac{(t+1)!(-1)^{t+1-j_1-j_2}}{j_1!j_2!(t+1-j_1-j_2)!} \frac{d^{j_1}g_2(x)}{dx^{j_1}} \frac{d^{j_2}g_2(x)}{dx^{j_2}} x e^{-x} \\ &+ \sum_{j_1=0}^t \sum_{j_2=0}^{t-j_1} \frac{(t+1)!(-1)^{t-j_1-j_2}}{j_1!j_2!(t-j_1-j_2)!} \frac{d^{j_1}g_2(x)}{dx^{j_1}} \frac{d^{j_2}g_2(x)}{dx^{j_2}} e^{-x}. \end{aligned}$$

By Lemma 65, if $f = \sum_{i=0}^d b_i h_i$ is a polynomial of degree at most d then writing $\frac{df}{dx} = \sum_{i=0}^{d-1} b'_i h_i$, $\sum_{i=0}^{d-1} |b'_i| \leq d \sum_{i=0}^d |b_i|$. By Lemma 62 and Corollary 64, we have that for all $j, j' \in [0, d]$:

1. For all $j, j' \in \mathbb{N} \cup \{0\}$, $\int_{x=0}^{\infty} |h_j(x) h_{j'}(x)| x e^{-x} dx \leq 1$.
2. For all $j, j' \in \mathbb{N} \cup \{0\}$, $\int_{x=0}^{\infty} |h_j(x) h_{j'}(x)| e^{-x} dx \leq \sqrt{(j+8)(j'+8)}$.

Putting these facts together, if $g_2 = \sum_{i=0}^d a_i h_i$ then

$$\begin{aligned} \int_{x=0}^{\infty} \left| \frac{d^{t+1}(g_2^2(x) x e^{-x})}{dx^{t+1}} \right| dx &\leq (3^{t+1}d^{t+1} + (t+1)3^t d^t (d+8)) \left(\sum_{i=0}^d |a_i| \right)^2 \\ &\leq (t+4)(d+8)(d+1)(3d)^t \left(\sum_{i=0}^d a_i^2 \right). \quad \blacktriangleleft \end{aligned}$$

With this bound in hand, we now apply Lemma 59 with $f(x) = g_2^2(x) x e^{-x}$. For convenience, we recall the statement of Lemma 59 here:

For all $t \in \mathbb{N}$, for any $\Delta > 0$ and any function $f : [0, \infty) \rightarrow \mathbb{R}$ which can be differentiated $t + 1$ times,

$$\begin{aligned} & \left| \frac{1}{t} \left(\sum_{b=0}^{t-1} \int_{x=b\Delta}^{\infty} f(x) dx \right) - \Delta \sum_{j=0}^{\infty} f(j\Delta) + \Delta \sum_{j=0}^{t-1} \left(\sum_{b=j+1}^t (c_t)_{b+1} \right) f(j\Delta) \right| \\ & \leq (t\Delta)^{t+1} \left(\frac{t}{(t+1)!} + \frac{\sum_{b=1}^t |(c_t)_{b+1}|}{t!} \right) \int_{x=0}^{\infty} |f^{(t+1)}(x)| dx. \end{aligned}$$

To use this bound, we need to bound $f(x) = g_2^2(x)xe^{-x}$ when $x \in [0, t\Delta]$.

► **Lemma 69.** *If $g_2 = \sum_{i=0}^d a_i h_i$ then for all $x \in [0, t\Delta]$,*

$$f(x) = g_2(x)^2 x e^{-x} \leq t\Delta (d+1)^2 e^{2dt\Delta} \left(\sum_{i=0}^d a_i^2 \right).$$

Proof. By Lemma 51, for all $x \in \mathbb{R}$ and all $j \in \mathbb{N}$, $|h_j(x)| \leq \sqrt{j+1} e^{j|x|}$. Thus, for all $x \in [0, t\Delta]$,

$$|g_2(x)| = \left| \sum_{i=0}^d a_i h_i(x) \right| \leq \sqrt{d+1} e^{dt\Delta} \left(\sum_{i=0}^d |a_i| \right)$$

which implies that for all $x \in [0, t\Delta]$

$$g_2(x)^2 x e^{-x} \leq t\Delta (d+1)^2 e^{2dt\Delta} \left(\sum_{i=0}^d |a_i| \right)^2 \leq t\Delta (d+1)^2 e^{2dt\Delta} \left(\sum_{i=0}^d a_i^2 \right). \quad \blacktriangleleft$$

Using Lemma 69, we make the following observations:

1. $\left| \int_0^{\infty} f(x) dx - \frac{1}{t} \left(\sum_{b=0}^{t-1} \int_{x=b\Delta}^{\infty} f(x) dx \right) \right| \leq (t\Delta)^2 (d+1)^2 e^{2dt\Delta} \left(\sum_{i=0}^d a_i^2 \right)$
2. $\left| \Delta \sum_{j=0}^{t-1} \left(\sum_{b=j+1}^t (c_t)_{b+1} \right) f(j\Delta) \right| \leq (t\Delta)^2 (d+1)^2 e^{2dt\Delta} \left(\sum_{b=1}^t |(c_t)_{b+1}| \right) \left(\sum_{i=0}^d a_i^2 \right).$

Putting everything together, there exist constants C_{t1} and C_{t2} such that

$$\left| \Delta \sum_{j=0}^{\infty} (j\Delta) e^{-j\Delta} g_2^2(j\Delta) - \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \right| \leq (C_{t1} (d\Delta)^2 e^{2dt\Delta} + C_{t2} d (d\Delta)^{t+1}) \left(\sum_{i=0}^d a_i^2 \right).$$

Since $\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx = \sum_{i=0}^d a_i^2$, the result follows. \blacktriangleleft

9 Handling the Difference Between Distributions

In this section, we prove the following theorem:

► **Theorem 70.** *For all $d, d_2, n \in \mathbb{N}$ such that $(4d+2)\ln(d_2) + 2\ln(20) \leq d_2 \leq \frac{\sqrt{n}}{4}$, for all polynomials g_2 of degree at most d , taking $\Delta = \frac{2d_2}{n}$,*

$$\Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{\Delta}{2} \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) - \frac{1}{10} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

38:26 Sum of Squares Bounds for the Ordering Principle

Proof. To prove this theorem, we prove the following two statements:

1. $\Delta \sum_{k=0}^{\lceil \frac{n}{4} \rceil - 1} \binom{k\Delta}{2} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{\Delta}{4} \sum_{k=0}^{\lceil \frac{n}{4} \rceil - 1} (k\Delta) e^{-k\Delta} g_2^2(k\Delta)$
2. $\Delta \sum_{k=\lceil \frac{n}{4} \rceil}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) \leq \frac{1}{5} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$

Assuming these two statements, we have that

$$\begin{aligned} \Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) &\geq \Delta \sum_{k=0}^{\lceil \frac{n}{4} \rceil - 1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \\ &\geq \frac{\Delta}{2} \sum_{k=0}^{\lceil \frac{n}{4} \rceil - 1} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) \\ &= \frac{\Delta}{2} \left(\sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) - \sum_{k=\lceil \frac{n}{4} \rceil}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) \right) \\ &\geq \frac{\Delta}{2} \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) - \frac{1}{10} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx. \end{aligned}$$

We now prove these two statements. The first statement follows immediately from the following lemma:

► **Lemma 71.** For all $k, d_2, n \in \mathbb{N}$ such that $d_2 \leq \frac{\sqrt{n}}{4}$ and $k \leq \frac{n}{4}$, taking $\Delta = \frac{2d_2}{n}$,

$$\frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} \geq \frac{1}{2} e^{-k\Delta}.$$

Proof. Observe that

$$\begin{aligned} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} &= \prod_{j=1}^{d_2-1} \frac{n-j-k-1}{n-j} \\ &= \prod_{j=1}^{d_2-1} \left(e^{-\frac{2k}{n}} \cdot \frac{1-\frac{k}{n}}{e^{-\frac{2k}{n}}} \cdot \frac{n-j-k-1}{n-j} \right) \\ &\geq e^{-k\Delta} \left(\frac{1-\frac{k}{n}}{e^{-\frac{2k}{n}}} \right)^{d_2-1} \left(\prod_{j=1}^{d_2-1} \frac{n-j-k-1}{n-j} \right). \end{aligned}$$

Thus, to prove this result, it is sufficient to lower bound $\left(\frac{1-\frac{k}{n}}{e^{-\frac{2k}{n}}} \right)^{d_2-1}$ and $\prod_{j=1}^{d_2-1} \frac{n-j-k-1}{n-j}$.

► **Proposition 72.** For all $k \in \mathbb{N}$ such that $k \leq \frac{n}{4}$, $\frac{1-\frac{k}{n}}{e^{-\frac{2k}{n}}} \geq 1$.

Proof. Observe that for all $x \geq 0$, $e^{-x} \leq 1 - x + \frac{x^2}{2}$. Taking $x = \frac{2k}{n}$, if $k \leq \frac{n}{4}$ then

$$e^{-\frac{2k}{n}} \leq 1 - \frac{2k}{n} + \frac{2k^2}{n^2} \leq 1 - \frac{2k}{n} + \frac{k}{2n} \leq 1 - \frac{k}{n}$$

and thus $\frac{1-\frac{k}{n}}{e^{-\frac{2k}{n}}} \geq 1$. ◀

To bound $\prod_{j=1}^{d_2-1} \frac{n-j-k-1}{n-j}$, we prove the following lemma.

► **Lemma 73.** For all $j, k, n \in \mathbb{N}$ such that $j \leq \frac{n}{8}$ and $k \leq \frac{n}{4}$,

$$\frac{\frac{n-j-k-1}{n-j}}{1 - \frac{k}{n}} \geq \left(1 - \frac{2}{n}\right) \left(1 - \frac{2jk}{n^2}\right).$$

Proof. Observe that

$$\begin{aligned} 1 - \frac{\frac{n-j-k-1}{n-j}}{1 - \frac{k}{n}} &= \frac{(n-k)(n-j) - n(n-j-k-1)}{(n-k)(n-j)} = \frac{jk+n}{(n-k)(n-j)} \\ &\leq \frac{2jk + \frac{32n}{21}}{n^2} = \frac{2jk}{n^2} + \frac{2}{n} - \frac{11}{21n} \leq \frac{2jk}{n^2} + \frac{2}{n} - \frac{4jk}{n^3}. \end{aligned}$$

Rearranging this, we have that

$$\frac{\frac{n-j-k-1}{n-j}}{1 - \frac{k}{n}} \geq 1 - \frac{2jk}{n^2} - \frac{2}{n} + \frac{4jk}{n^3}$$

as needed. ◀

Combining this lemma with the following proposition, we have the following corollary.

► **Proposition 74.** For all $x \in [0, 1]$ and all $k \in \mathbb{N}$, $(1-x)^k \geq 1-kx$.

► **Corollary 75.** For all $d_2, k, n \in \mathbb{N}$ such that $d_2 \leq \frac{n}{8}$ and $k \leq \frac{n}{4}$,

$$\prod_{j=1}^{d_2-1} \frac{\frac{n-j-k-1}{n-j}}{1 - \frac{k}{n}} \geq \left(1 - \frac{2d_2}{n}\right) \left(1 - \frac{2d_2^2 k}{n^2}\right).$$

Since $d_2 \leq \frac{\sqrt{n}}{4} \leq \frac{n}{8}$ and $k \leq \frac{n}{4}$, combining Proposition 72 and Corollary 75 with the inequality

$$\frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} \geq e^{-k\Delta} \left(\frac{1 - \frac{k}{n}}{e^{-\frac{2k}{n}}}\right)^{d_2-1} \left(\prod_{j=1}^{d_2-1} \frac{\frac{n-j-k-1}{n-j}}{1 - \frac{k}{n}}\right)$$

we have that

$$\frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} \geq e^{-k\Delta} \left(1 - \frac{2d_2}{n}\right) \left(1 - \frac{2d_2^2 k}{n^2}\right) \geq \frac{1}{2} e^{-k\Delta}$$

which completes the proof of Lemma 71. ◀

We now prove the second statement needed to prove Theorem 70.

► **Lemma 76.** For all $d, d_2, n \in \mathbb{N}$ such that $d_2 \geq 4d \ln(d_2) + 2 \ln(10n)$, for any polynomial g_2 of degree at most d , taking $\Delta = \frac{2d_2}{n}$

$$\Delta \sum_{k=\lceil \frac{n}{4} \rceil}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(k\Delta) \leq 2n^2 \left(\frac{3}{4}\right)^{d_2-1} \left(\frac{4n}{d_2}\right)^d \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

Proof. To prove this, we upper bound $|h_k(x)|$ for large x .

► **Lemma 77.** For all $k \in \mathbb{N}$ and all $x \geq 1$, $|h_k(x)| \leq (2x)^k$.

38:28 Sum of Squares Bounds for the Ordering Principle

Proof. Recall that

$$h_k(x) = \frac{1}{\sqrt{k!(k+1)!}} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \frac{(k+1)!}{(j+1)!} x^j.$$

Now observe that

$$\sum_{j=0}^k \binom{k}{j} \frac{1}{(j+1)!} \leq \sum_{j=0}^k \frac{k^j}{j!2^j} = e^{\frac{k}{2}}.$$

Thus, for all $k \in \mathbb{N}$ and all $x \geq 1$, $|h_k(x)| \leq \sqrt{k+1} e^{\frac{k}{2}} x^k$.

If $k \geq 5$ then $\sqrt{k+1} e^{\frac{k}{2}} \leq 2^k$ and we are done. For $k \in [1, 4]$ we check the polynomials directly.

1. $|h_1(x)| = \left| \frac{1}{\sqrt{2}}(x-2) \right| \leq \max \left\{ \frac{x}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} < 2x$
2. $|h_2(x)| = \left| \frac{1}{\sqrt{12}}(x^2 - 6x + 6) \right| \leq \max \left\{ \frac{x^2}{\sqrt{12}}, \frac{5x}{\sqrt{12}} \right\} < 4x^2$
3. $|h_3(x)| = \left| \frac{1}{\sqrt{144}}(x^3 - 12x^2 + 36x - 24) \right| \leq \max \left\{ \frac{25x^3}{\sqrt{144}}, \frac{12x^2}{\sqrt{144}} \right\} < 8x^3$
4. $|h_4(x)| = \left| \frac{1}{\sqrt{2880}}(x^4 - 20x^3 + 120x^2 - 240x + 120) \right| \leq \max \left\{ \frac{101x^4}{\sqrt{2880}}, \frac{139x^3}{\sqrt{2880}} \right\} < 16x^4. \quad \blacktriangleleft$

► **Corollary 78.** For all $d \in \mathbb{N}$, if g_2 is a polynomial of degree at most d then for all $y \geq 1$,

$$g_2^2(y) \leq 2(2y)^{2d} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

Proof. Writing $g_2 = \sum_{i=0}^d a_i h_i$, we have that $\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx = \sum_{i=0}^d a_i^2$ and

$$\begin{aligned} g_2^2(y) &\leq \sum_{i=0}^d \sum_{i'=0}^d a_i a_{i'} (2y)^{i+i'} \\ &\leq \sum_{i=0}^d \sum_{i'=0}^d \left(\frac{1}{2^{d-i'+1}} a_i^2 + \frac{1}{2^{d-i+1}} a_{i'}^2 \right) (2y)^{2d} \\ &\leq \left(\sum_{i=0}^d a_i^2 + \sum_{i'=0}^d a_{i'}^2 \right) (2y)^{2d} = 2(2y)^{2d} \sum_{i=0}^d a_i^2. \quad \blacktriangleleft \end{aligned}$$

With this bound, we can now prove Lemma 76. By Corollary 78,

$$g_2^2(y) \leq 2(2y)^{2d} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx$$

Applying this with $y = k\Delta$, since $d_2 \geq (4d + 2)\ln(d_2) + 2\ln(20)$,

$$\begin{aligned}
\Delta \sum_{k=\lceil \frac{n}{4} \rceil}^{\infty} (k\Delta)e^{-k\Delta} g_2^2(k\Delta) &\leq \Delta \sum_{k=\lceil \frac{n}{4} \rceil}^{\infty} (2k\Delta)e^{-k\Delta} (2k\Delta)^{2d} \int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \\
&\leq \left(\int_{x=(\lceil \frac{n}{4} \rceil - 1)\Delta}^{\infty} (2x)^{2d+1} e^{-x} dx \right) \int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \\
&\leq 2^{2d+1} \sum_{j=0}^{2d+1} \frac{(2d+1)!}{(2d+1-j)!} \left((\lceil \frac{n}{4} \rceil - 1)\Delta \right)^{2d+1-j} e^{-(\lceil \frac{n}{4} \rceil - 1)\Delta} \left(\int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \right) \\
&\leq 2^{2d+2} \left((\lceil \frac{n}{4} \rceil - 1)\Delta \right)^{2d+1} e^{-(\lceil \frac{n}{4} \rceil - 1)\Delta} \left(\int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \right) \\
&\leq 2e^\Delta d_2^{2d+1} e^{-\frac{d_2}{2}} \left(\int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \right) \\
&\leq 4e^{(2d+1)\ln(d_2) - \frac{d_2}{2}} \left(\int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx \right) \\
&\leq \frac{1}{5} \int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx.
\end{aligned}$$

where the second inequality holds because $(2x)^{2d+1}e^{-x}$ is a decreasing function whenever $x \geq 2d + 1$. ◀

10 Putting Everything Together

In this section, we put everything together to prove our SOS lower bound. ◀

10.1 Lower bounding our sum with an integral

We first combine Theorems 67 and 70 to lower bound our sum with an integral.

► **Theorem 79.** For all $d, d_2, t, n \in \mathbb{N}$, taking $\Delta = \frac{2d_2}{n}$, if the following conditions hold:

1. $(4d + 2)\ln(d_2) + 2\ln(20) \leq d_2 \leq \frac{\sqrt{n}}{4}$.
2. Letting C_{t1} and C_{t2} be the constants given by Theorem 67,

$$C_{t1}(d\Delta)^2 e^{2td\Delta} + C_{t2}d(d\Delta)^{t+1} \leq \frac{1}{2}$$

then for any polynomial g_2 of degree at most d ,

$$\Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{3}{20} \int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx.$$

Proof. By Theorem 70, for all $d, d_2, n \in \mathbb{N}$ such that $4d\ln(d_2) + 2\ln(10n) \leq d_2 \leq \frac{\sqrt{n}}{4}$, for all polynomials g_2 of degree at most d , taking $\Delta = \frac{2d_2}{n}$,

$$\Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{\Delta}{2} \sum_{k=0}^{\infty} (k\Delta)e^{-k\Delta} g_2^2(k\Delta) - \frac{1}{10} \int_{x=0}^{\infty} g_2^2(x)xe^{-x} dx.$$

38:30 Sum of Squares Bounds for the Ordering Principle

By Theorem 67, for all $\Delta > 0$, $d \in \mathbb{N}$, and polynomials g_2 of degree at most d ,

$$\begin{aligned} & \left| \Delta \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(j\Delta) - \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \right| \\ & \leq (C_{t_1}(d\Delta)^2 e^{2td\Delta} + C_{t_2} d (d\Delta)^{t+1}) \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \\ & \leq \frac{1}{2} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx. \end{aligned}$$

Thus,

$$\Delta \sum_{k=0}^{\infty} (k\Delta) e^{-k\Delta} g_2^2(j\Delta) \geq \frac{1}{2} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

Combining these statements, $\Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx$, as needed. \blacktriangleleft

10.2 Proof of the SOS lower bound

We now prove our SOS lower bound.

► **Theorem 80.** For all $d, d_2, t, n \in \mathbb{N}$, taking $\Delta = \frac{2d_2}{n}$, if the following conditions hold:

1. $(4d+2)\ln(d_2) + 2\ln(20) \leq d_2 \leq \frac{\sqrt{n}}{4}$.
2. $10(d+1)^2 \Delta^2 e^{2d\Delta} \leq 1$.
3. Letting C_{t_1} and C_{t_2} be the constants given by Theorem 67,

$$C_{t_1}(d\Delta)^2 e^{2td\Delta} + C_{t_2} d (d\Delta)^{t+1} \leq \frac{1}{2}$$

then there is no polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$.

Proof. We recall the following results.

1. By Theorem 33, since $2d \leq d_2 \leq n$, if there is a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$ then there is a polynomial $g_* : \mathbb{R} \rightarrow \mathbb{R}$ of degree at most d such that $E_{\Omega_{n,d_2}}[(u-1)g_*(u)^2] < 0$. Equivalently,

$$\sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} k g_*^2(k+1) < g_*^2(0)$$

Taking $g_2(x) = g_*\left(\frac{x}{\Delta} + 1\right)$,

$$\Delta \sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} (k\Delta) g_2^2(k\Delta) < \Delta^2 g_2^2(-\Delta).$$

2. By Theorem 79, under the given conditions,

$$\Delta \sum_{k=0}^{n-d_2-1} (k\Delta) \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} g_2^2(k\Delta) \geq \frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

3. By Theorem 50, since $10(d+1)^2 \Delta^2 e^{2d\Delta} \leq 1$, for any polynomial g_2 of degree at most d ,

$$\int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx \geq 10\Delta^2 g_2^2(-\Delta).$$

Putting everything together, if there is a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$ then there is a polynomial $g_2 : \mathbb{R} \rightarrow \mathbb{R}$ of degree at most d such that

$$\frac{3}{2}\Delta^2 g_2^2(-\Delta) \leq \Delta \sum_{k=0}^{n-d_2-1} \frac{\binom{n-k-2}{d_2-1}}{\binom{n-1}{d_2-1}} (k\Delta) g_2^2(k\Delta) \leq \frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx < \Delta^2 g_2^2(-\Delta)$$

which is impossible. Thus, there is no polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$. ◀

► **Corollary 81.** *For all $\epsilon > 0$, there exists a constant C_ϵ such that for all $n \in \mathbb{N}$, degree $C_\epsilon n^{\frac{1}{2}-\epsilon}$ sum of squares cannot prove the ordering principle on n elements.*

11 Conclusion

In this paper, we analyzed the performance of SOS for proving the ordering principle, showing that SOS requires degree roughly \sqrt{n} to prove the ordering principle on n elements. This shows that in terms of degree, SOS is more powerful than resolution, polynomial calculus, and the Sherali-Adams hierarchy, but SOS still requires high degree to prove the ordering principle. While this mostly resolves the question of how powerful SOS is for proving the ordering principle, there are several open questions remaining including the following:

1. Can we find a tight example for the size/degree trade-off for SOS which was recently shown by Atserias and Hakoniemi [1]?
2. Can we prove SOS lower bounds for the graph ordering principle on expanders?

References

- 1 Albert Atserias and Tuomas Hakoniemi. Size-degree trade-offs for sums-of-squares and positivstellensatz proofs. In *Computational Complexity Conference*, 2019.
- 2 Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, page 307–326, New York, NY, USA, 2012. Association for Computing Machinery. doi:10.1145/2213977.2214006.
- 3 Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:59, 2014.
- 4 Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares, 2016. URL: <https://www.sumofsquares.org/public/index.html>.
- 5 Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow – resolution made simple. *J. ACM*, 48(2):149–169, March 2001. doi:10.1145/375827.375835.
- 6 Maria Luisa Bonet and Nicola Galesi. Optimality of size-width tradeoffs for resolution. *Comput. Complex.*, 10(4):261–276, May 2002. doi:10.1007/s000370100000.
- 7 Nicola Galesi and Massimo Lauria. Optimality of size-degree tradeoffs for polynomial calculus. *ACM Trans. Comput. Logic*, 12(1), November 2010. doi:10.1145/1838552.1838556.
- 8 Russell Impagliazzo, Pavel Pudlák, and Jirí Sgall. Lower bounds for the polynomial calculus and the gröbner basis algorithm. *computational complexity*, 8:127–144, 1999.
- 9 Pravesh Kothari and David Steurer. Sum-of-squares: proofs, beliefs, and algorithms, 2016. URL: <https://sos16.dsteurer.org/>.
- 10 Balakrishnan Krishnamurthy. Short proofs for tricky formulas. *Acta Inf.*, 22(3):253–275, August 1985.
- 11 Massimo Lauria. Dd3013 sum of squares and integer programming relaxations, 2014. URL: <http://www.csc.kth.se/~lauria/sos14/>.

- 12 Aaron Potechin. Cmsc 39600 1 (autumn 2018) topics in theoretical computer science: The sum of squares hierarchy, 2018. URL: <https://canvas.uchicago.edu/courses/17604>.
- 13 Aaron Potechin. Sum of squares lower bounds from symmetry and a good story. In *ITCS*, 2019.
- 14 Annie Raymond, James Saunderson, Mohit Singh, and Rekha R. Thomas. Symmetric sums of squares over k-subset hypercubes. *Math. Program.*, 167(2):315–354, February 2018. doi:10.1007/s10107-017-1127-6.
- 15 Alexander Razborov. Guest column: Proof complexity and beyond. *SIGACT News*, 47(2):66–86, June 2016. doi:10.1145/2951860.2951875.
- 16 Gilbert Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207:87–97, 1974.
- 17 Gunnar Stålmarck. Short resolution proofs for a sequence of tricky formulas. *Acta Inf.*, 33(3):277–280, May 1996. doi:10.1007/s002360050044.
- 18 Wikipedia. Chebyshev polynomials, Accessed May 30, 2020. URL: https://en.wikipedia.org/wiki/Chebyshev_polynomials.

A Analyzing the Ordering Principle with Boolean Variables

In this appendix, we describe how to modify the ordering principle equations so that they only have Boolean variables. We then describe how to modify the pseudo-expectation values and the SOS lower bound proof for these equations.

A.1 Equations for the ordering principle with Boolean auxiliary variables

To encode the negation of the ordering principle using only Boolean variables, we simply replace each z_j^2 with a sum of squares of Boolean auxiliary variables. This gives us the following equations for the negation of the ordering principle:

1. We have variables x_{ij} where we want that $x_{ij} = 1$ if $a_i < a_j$ and $x_{ij} = 0$ if $a_i > a_j$. We also have auxiliary variables $\{z_{jk} : j \in [n], k \in [m]\}$ where $m \geq n - 2$.
2. $\forall i \neq j, x_{ij}^2 = x_{ij}$ and $\forall j \in [n], \forall k \in [m], z_{jk}^2 = z_{jk}$ (variables are Boolean)
3. $\forall i \neq j, x_{ij} = 1 - x_{ji}$ (ordering)
4. For all distinct $i, j, k, x_{ij}x_{jk}(1 - x_{ik}) = 0$ (transitivity)
5. $\forall j, \sum_{i \neq j} x_{ij} = 1 + \sum_{k=1}^m z_{jk}^2$ (for all $j \in [n]$, a_j is not the minimum element of $\{a_1, \dots, a_n\}$)

A.2 Pseudo-expectation values with Boolean auxiliary variables

In order to give pseudo-expectation values for these equations, we need to give pseudo-expectation values for polynomials involving the auxiliary variables. The idea for this is as follows. Letting $w_j = \left(\sum_{i \neq j} x_{ij}\right) - 1$, we want that w_j of the auxiliary variables $\{z_{jk} : k \in [m]\}$ are 1. If $w_j \in [0, m] \cap \mathbb{Z}$, if we choose which of these auxiliary variables are 1 at random,

1. $Pr(z_{j1} = 1) = \frac{w_j}{m}$,
2. $Pr(z_{j1} = 1, z_{j2} = 1) = \frac{w_j(w_j-1)}{m(m-1)}$
3. More generally, for all $K \subseteq [m]$, $Pr(\forall k \in K, z_{jk} = 1) = \frac{\prod_{a=0}^{|K|-1} (w_j-a)}{\prod_{a=0}^{|K|-1} (m-a)}$.

Note that these expressions are still defined for other w_j including $w_j = -1$ (though in this case they aren't actual probabilities over a distribution of solutions). Based on this, we have the following candidate pseudo-expectation values:

► **Definition 82** (Candidate pseudo-expectation values with Boolean auxiliary variables).

1. For all polynomials $p(\{x_{ij} : i, j \in [n], i \neq j\})$, we take $\tilde{E}_n[p] = E_{U_n}[p]$.
2. For all $j \in [n]$, for all $K \subseteq [m]$ and all polynomials p which do not contain any of the auxiliary variables $\{z_{jk} : k \in [m]\}$, we take

$$\tilde{E} \left[\left(\prod_{k \in K} z_{jk} \right) p \right] = \frac{\tilde{E} \left[\left(\prod_{a=0}^{|K|-1} (w_j - a) \right) p \right]}{\prod_{a=0}^{|K|-1} (m - a)}.$$

A.3 Reducing to one variable with Boolean auxiliary variables

Unfortunately, our lower bound for the ordering principle equations in Definition 9 does not directly imply a lower bound for the ordering principle equations with Boolean auxiliary variables. That said, we can still reduce the problem to one variable by using the same techniques we used to prove Theorem 33. The resulting theorem is similar but not quite the same as Theorem 33.

► **Theorem 83.** For all $d, d_2, n, m \in \mathbb{N}$ such that $2d \leq d_2 \leq n$ and $m \geq 15nd$, if there is a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$ then there is a polynomial $g_* : \mathbb{R} \rightarrow \mathbb{R}$ of degree at most d and a $j \in [d]$ such that

$$\sum_{u=1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right) g_*^2(u) < 1.2g_*^2(0).$$

Proof sketch. Having Boolean auxiliary variables affects each part of the proof of Theorem 33 as follows:

1. Since the equations and pseudo-expectation values are still symmetric under permutations of $[2n]$, the argument in Section 6.1 that we can reduce to the case when g is symmetric under permutations of $[2n] \setminus I$ for some subset $I \subseteq [2n]$ where $|I| \leq d$ still applies.
2. In Section 6.2, we decomposed $\tilde{E}_{2n}[g^2]$ as

$$\tilde{E}_{2n}[g^2] = \sum_{A \subseteq [2n]} \tilde{E}_{2n} \left[\left(\prod_{j \in A} z_j^2 \right) g_A^2 \right].$$

Here we can do a similar decomposition but it is somewhat more complicated.

► **Definition 84.** Given a $j \in [n]$ and a nonempty $K \subseteq [m]$, define

$$y_{jK} = \prod_{k \in K} z_{jk} - \frac{\prod_{a=0}^{|K|-1} (w_j - a)}{\prod_{a=0}^{|K|-1} (m - a)}.$$

► **Proposition 85.** For any $j \in [2n]$, any nonempty $K \subseteq [m]$, and any polynomial p which does not depend on the auxiliary variables $\{z_{jk} : k \in [m]\}$, $\tilde{E}_{2n}[y_{jK}p] = 0$

With this proposition in mind, we decompose g as $g = \sum_{A \subseteq [2n]} g_A$ where

$$g_A = \sum_{\{K_j : j \in A\}} \left(\prod_{j \in A} y_{jK_j} \right) p_{\{K_j : j \in A\}} \text{ (where each } K_j \text{ is a nonempty subset of } [m]\text{)}$$

where each K_j is a nonempty subset of $[m]$. With this decomposition, we have that

$$\tilde{E}_{2n}[g^2] = \sum_{A \subseteq [2n]} \tilde{E}_{2n}[g_A^2].$$

38:34 Sum of Squares Bounds for the Ordering Principle

Note that unlike before, here we have the auxiliary variables be part of g_A . That said, this allows us to assume that there are no auxiliary variables z_{jk} where $j \notin A$ and that everything is symmetric under permutations of $[2n] \setminus I'$ where $|I'| \leq 2d$.

3. In Section 6.3, we restricted ourselves to a single ordering for the distinguished indices and expressed everything in terms of the new variables u_0, \dots, u_{d_2} . We can still do this with Boolean auxiliary variables, but this no longer removes all of the auxiliary variables. What we get is a polynomial $g_{\{1\}}(u_0, \dots, u_{d_2}, \{z_{j'k} : j' \in [d_2]\})$ of degree at most $\frac{d}{2}$ such that

$$\begin{aligned} & \tilde{E}_{2n} \left[\left(\prod_{i=1}^{d_2-1} x_{i(i+1)} \right) g_{\{1\}}^2 \right] \\ &= \frac{1}{d_2!} E_{u_0, \dots, u_{d_2} \in \mathbb{N} \cup \{0\} : \sum_{j=0}^{d_2} u_j = 2n - d_2} \left[\tilde{E}'_{u_0, \dots, u_{d_2}} [g_{\{1\}}^2(u_0, \dots, u_{d_2}, \{z_{j'k} : j' \in [d_2]\})] \right] < 0 \end{aligned}$$

where $\tilde{E}'_{u_0, \dots, u_{d_2}}$ gives the pseudo-expectation values of the auxiliary variables for given values of u_0, \dots, u_{d_2} .

4. In Section 6.4, we took

$$g_*(u_0) = E_{u_1, \dots, u_{d_2} \in \mathbb{N} \cup \{0\} : \sum_{j=1}^{d_2} u_j = 2n - d_2 - u_0} [g_{\{1\}}(u_0, \dots, u_{d_2})^2].$$

Before we can do this here, we need to remove the auxiliary variables $\{z_{1k} : k \in [m]\}$. We can do this as follows:

- a. Observe that looking at the auxiliary variables $\{z_{1k} : k \in [m]\}$, $\tilde{E}'_{u_0, \dots, u_{d_2}}$ (and thus \tilde{E}_{2n}) is symmetric under permutations of $[m]$. Using Theorem 36, we can assume that $g_{\{1\}}$ is symmetric (as far as the auxiliary variables $\{z_{1k} : k \in [m]\}$ are concerned) under permutations of $[m] \setminus K$ for some $K \subseteq [m]$ where $|K| \leq d$.
- b. Breaking things into cases based on the values of the auxiliary variables $\{z_{1k} : k \in K\}$, we can assume that

$$\begin{aligned} & g_{\{1\}}(u_0, \dots, u_{d_2}, \{z_{j'k} : j' \in [d_2]\}) = \\ & \left(\prod_{k \in K_1} z_{1k} \right) \left(\prod_{k \in K_2} (1 - z_{1k}) \right) p_{\{1\}}(u_0, \dots, u_{d_2}, \{z_{j'k} : j' \in [2, d_2]\}) \end{aligned}$$

for some $K_1, K_2 \subseteq [m]$ such that $K_1 \cap K_2 = \emptyset$ and $|K_1 \cup K_2| \leq d$.

We now take

$$g_* = E_{u_1, \dots, u_{d_2} \in \mathbb{N} \cup \{0\} : \sum_{j=1}^{d_2} u_j = 2n - d_2 - u_0} \left[\tilde{E}'_{u_0, \dots, u_{d_2}} [p_{\{1\}}^2(u_0, \dots, u_{d_2}, \{z_{j'k} : j' \in [2, d_2]\})] \right]$$

and we have that

- a. $g_*(u_0)$ is a polynomial of degree at most d in u_0 .
- b. For all $u_0 \in [0, 2n - d_2] \cap \mathbb{Z}$, $g_*(u_0) \geq 0$.
- c. $E_{\Omega_{2n, d_2}} \left[\left(\prod_{a=1}^{|K_1|} (u - a) \right) \left(\prod_{a=1}^{|K_2|} (m + 2 - u - a) \right) g_*(u) \right] < 0$.

Equivalently, taking $j = |K_1|$ and $j_2 = |K_2|$,

$$\sum_{u=0}^{2n-d_2} \frac{\binom{2n-u-1}{d_2-1}}{\binom{2n}{d_2}} \left(\prod_{a=1}^j (u - a) \right) \left(\prod_{a=1}^{j_2} (m + 2 - u - a) \right) g_*(u) < 0.$$

Manipulating this gives

$$\sum_{u=1}^{2n-d_2} \frac{\binom{2n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u - a)}{j!} \right) \left(\prod_{a=1}^{j_2} \frac{m + 2 - u - a}{m + 2 - a} \right) \left(\frac{g_*(u)}{g_*(0)} \right) < 1,$$

which implies that

$$\sum_{u=1}^{2n-d_2} \left(\frac{\binom{2n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right)^2 \left(\prod_{a=1}^{j_2} \frac{m+2-u-a}{m+2-a} \right)^2 \left(\frac{g_*(u)}{g_*(0)} \right)^2 < 1.$$

We now make the following observations:

a. By Lemma 41, for all $u \in \mathbb{N}$ such that $u \leq n - d_2$, $\left(\frac{\binom{2n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \right)^2 \geq \frac{\binom{n-u-1}{d_2-1}}{\binom{n-1}{d_2-1}}$.

b. For all $u \in \mathbb{N}$, $\left(\frac{\prod_{a=1}^j (u-a)}{j!} \right)^2 \geq \frac{\prod_{a=1}^j (u-a)}{j!}$.

c. Since $m \geq 15nd$, for all $u \in \mathbb{N}$ such that $u \leq n - d_2$,

$$\left(\prod_{a=1}^{j_2} \frac{m+2-u-a}{m+2-a} \right)^2 \geq \left(\frac{14nd-n}{14nd} \right)^{2d} \geq \frac{12}{14}.$$

Putting everything together,

$$\sum_{u=1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right) g_*^2(u) < 1.2g_*^2(0)$$

as needed. ◀

A.4 SOS lower bound with Boolean auxiliary variables

When we have Boolean auxiliary variables, our SOS lower bound is modified as follows:

► **Theorem 86.** For all $d, d_2, t, n, m \in \mathbb{N}$ such that $m \geq 15nd$, if the following conditions hold for all $j \in [d]$:

1. $(4d+2)\ln(d_2) + 2\ln(20) \leq d_2 \leq \frac{\sqrt{n'}}{4}$
2. $\frac{(n-1)!(n'-d_2)!}{(n'-1)!(n-d_2)!\binom{2j-1}{j}} \Delta^2 (d+1)^2 e^{2d(2j-1)\Delta} \leq \frac{1}{10}$
3. Letting C_{t1} and C_{t2} be the constants given by Theorem 67,

$$C_{t1}(d\Delta)^2 e^{2td\Delta} + C_{t2}d(d\Delta)^{t+1} \leq \frac{1}{2}$$

where $n' = n - 2d + 2$ and $\Delta = \frac{2d_2}{n'}$ then there is no polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$.

► **Remark 87.** We believe the condition on m is an artefact of the proof and that we should have essentially the same lower bound as long as $m \geq n - 2$, though proving this would require modifying the analysis further.

Proof. Assume there is a polynomial g of degree at most $\frac{d}{2}$ such that $\tilde{E}_{2n}[g^2] < 0$. By Theorem 83, since $2d \leq d_2 \leq n$, there is a polynomial $g_* : \mathbb{R} \rightarrow \mathbb{R}$ of degree at most d and a $j \in [d]$ such that

$$\sum_{u=1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{2n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right) g_*^2(u) < 1.2g_*^2(0).$$

We transform this left side of this equation into the same form as the left hand side of Theorem 79 using the following lemma.

38:36 Sum of Squares Bounds for the Ordering Principle

► **Lemma 88.** For all $j, u \in \mathbb{N}$, $\frac{1}{j!} \left(\prod_{a=1}^j (u-a) \right) \geq \binom{2j-1}{j} (u-2j+1)$.

Proof. We prove this lemma by induction. When $u \leq 2j-1$, the result is trivial. When $u = 2j$,

$$\frac{1}{j!} \left(\prod_{a=1}^j (u-a) \right) = \binom{2j-1}{j} = \binom{2j-1}{j} (u-2j+1).$$

Now assume the result is true for $u = k$ where $k \geq 2j$ and consider the case when $u = k+1$. By the inductive hypothesis,

$$\begin{aligned} \frac{1}{j!} \left(\prod_{a=1}^j ((k+1)-a) \right) &= \frac{k}{k-j} \left(\frac{1}{j!} \prod_{a=1}^j (k-a) \right) \geq \frac{k}{k-j} \left(\binom{2j-1}{j-1} (k-2j+1) \right) \\ &\geq \frac{k-2j+2}{k-2j+1} \left(\binom{2j-1}{j-1} (k-2j+1) \right) = \binom{2j-1}{j} (k-2j+2). \blacktriangleleft \end{aligned}$$

Applying Lemma 88, we have that

$$\begin{aligned} 1.2g_*^2(0) &> \sum_{u=1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right) g_*^2(u) \geq \sum_{u=2j-1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{n-1}{d_2-1}} \left(\frac{\prod_{a=1}^j (u-a)}{j!} \right) g_*^2(u) \\ &\geq \sum_{u=2j-1}^{n-d_2} \frac{\binom{n-u-1}{d_2-1}}{\binom{n-1}{d_2-1}} \binom{2j-1}{j} (u-2j+1) g_*^2(u). \end{aligned}$$

Taking $k = u - 2j + 1$, $n' = n - 2j + 2$, $\Delta = \frac{2d_2}{n'}$, and $g_2(x) = g_* \left(\frac{x}{\Delta} + 2j - 1 \right)$,

$$\frac{1.2(n-1)!(n'-d_2)!}{(n'-1)!(n-d_2)! \binom{2j-1}{j}} \Delta^2 g_2^2(-(2j-1)\Delta) > \Delta \sum_{k=0}^{n'-d_2-1} \frac{\binom{n'-k-2}{d_2-1}}{\binom{n'-1}{d_2-1}} (k\Delta) g_2^2(k\Delta)$$

By Theorem 79, under the given conditions,

$$\Delta \sum_{k=0}^{n'-d_2-1} \frac{\binom{n'-k-2}{d_2-1}}{\binom{n'-1}{d_2-1}} (k\Delta) g_2^2(k\Delta) \geq \frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx.$$

Thus,

$$\frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx < \frac{1.2(n-1)!(n'-d_2)!}{(n'-1)!(n-d_2)! \binom{2j-1}{j}} \Delta^2 g_2^2(-(2j-1)\Delta).$$

Decomposing g_2 as $g_2 = \sum_{i=0}^d c_i h_i$, observe that

1. $\frac{3}{20} \int_{x=0}^{\infty} g_2^2(x) x e^{-x} dx = \frac{3}{20} \left(\sum_{i=0}^d c_i^2 \right)$.
2. By Cauchy-Schwarz,

$$g_2^2(-(2j-1)\Delta) = \left(\sum_{i=0}^d c_i h_i(-(2j-1)\Delta) \right)^2 \leq \left(\sum_{i=0}^d c_i^2 \right) \left(\sum_{i=0}^d h_i^2(-(2j-1)\Delta) \right).$$

By Lemma 51, for all $i \in \mathbb{N}$ and all $x \in \mathbb{R}$, $|h_i(x)| \leq \sqrt{i+1} e^{i|x|}$. Thus,

$$g_2^2(-(2j-1)\Delta) \leq (d+1)^2 e^{2d(2j-1)\Delta}.$$

Putting these pieces together,

$$\frac{3}{20} < \frac{1.2(n-1)!(n'-d_2)!}{(n'-1)!(n-d_2)! \binom{2j-1}{j}} \Delta^2 (d+1)^2 e^{2d(2j-1)\Delta}.$$

However, $\frac{(n-1)!(n'-d_2)!}{(n'-1)!(n-d_2)! \binom{2j-1}{j}} \Delta^2 (d+1)^2 e^{2d(2j-1)\Delta} \leq \frac{1}{10}$ so this gives $\frac{3}{20} = .15 < .12$, which is a contradiction. ◀